The lexeme in descriptive and theoretical morphology

Edited by

Olivier Bonami
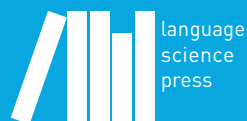
Gilles Boyé

Georgette Dal

Hélène Giraudo

Fiammetta Namer

language
science
press

Empirically Oriented Theoretical Morphology and Syntax

In this series:

1. Lichte, Timm. Syntax und Valenz: Zur Modellierung kohärenter und elliptischer Strukturen mit Baumadjunktionsgrammatiken

2. Bîlbîie, Gabriela. Grammaire des constructions elliptiques: Une étude comparative des phrases sans verbe en roumain et en français

3. Bowern, Claire, Laurence Horn & Raffaella Zanuttini (eds.). On looking into words (and beyond): Structures, Relations, Analyses

4. Bonami, Olivier, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer. The lexeme in descriptive and theoretical morphology.

The lexeme in descriptive and theoretical morphology

Edited by

Olivier Bonami

Gilles Boyé

Georgette Dal

Hélène Giraudo

Fiammetta Namer

Olivier Bonami , Gilles Boyé , Georgette Dal , Hélène Giraudo & Fiammetta Namer (eds.). 2018. *The lexeme in descriptive and theoretical morphology* (Empirically Oriented Theoretical Morphology and Syntax 4). Berlin: Language Science Press.

Freie Universität Berlin

Contents

Contents

Contents

Introduction

Olivier Bonami
Université Paris Diderot

Gilles Boyé
Université Bordeaux-Montaigne

Georgette Dal
Université de Lille

Hélène Giraudo
CLLE, Université de Toulouse, CNRS, Toulouse

Fiammetta Namer
Université de Lorraine

1 Introducing the lexeme

It is customary (see for instance Aronoff 1994: 4) to associate the notion of a lexeme with Peter H. Matthews (1965, 1972, 1974, 1991).[1] Matthews (1972: 160-161) contrasts three uses of the term *word* that may be differentiated as follows.

- The term *word* may denote a certain type of syntactic constituent. In this sense, the term unambiguously designates a kind of Saussurean sign, possibly complex: it associates a phonological representation with a meaning.

---

[1] Matthews (1972: 160) himself notes that the use of the word *lexeme* in this sense originates in Lyons (1963), and that his understanding of the lexeme is very close to that of the *semanteme* in Bally (1944: 287). See also Trnka (1949: 28). On the other hand, the use of *lexeme* in the tradition starting with Matthews has little to do with Martinet's *lexème* (e.g. Martinet 1960), which designates what in the English-speaking world would be called a morpheme with lexical meaning, or a root.

Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer

- The term *word* is used to denote the phonological sequence that is the shape of a word in the first sense. Matthews coins the term *wordform* to designate this and avoid ambiguity.[2]

- The term *word* often denotes that lexical object dictionaries talk about: an item characterized by a stable lexical meaning and a set of syntactic properties, but that abstracts away from inflection. This unit is what Matthews calls the *lexeme*.

One may illustrate these definitions by saying that the French lexeme VIEUX 'old' is associated with four words filling the four cells in its paradigm: M.SG *vieux*, F.SG *vieille*, M.PL *vieux*, and F.PL *vieilles*. To these four words correspond only three wordforms, since the M.SG and M.PL are phonologically identical. This characterization of the lexeme is deliberately silent on phonology: the lexeme is defined in terms of the syntactic and semantic cohesion of a family of words, ignoring phonology. Literature from the 1990s was not so prudent, and presented the lexeme as an underspecified sign. The following quote is representative of the dominant view:

> Each lexeme can be viewed as a set of properties, which will in some sense be present in all occurrences of the lexeme. These crucially include some semantic properties, some phonological properties [...], and some syntactic properties. (Zwicky 1992: 333)

Such a definition is obviously not adequate if one wants to be able to take into account the full spectrum of stem allomorphy, including suppletion. In some cases, there is no phonological property that is shared by all forms of the lexeme; e.g. there is nothing common between the 3sg forms of the French lexeme ALLER 'go' in the imperfect (*allait*), present (*va*) and future (*ira*). This example shows that lexemes are ineffable: one can't utter a lexeme, but only one of its forms. It also highlights the importance of cleanly distinguishing lexemes from their CITATION FORM.[3] The French grammatical tradition happens to use infinitives as citation forms, and the infinitive of ALLER happens to use the *al-* stem. From this, no conclusion can be drawn as to *al-* being a more reflective of the fundamental phonological identity of that lexeme: if French grammarians had kept the Latin tradition of using the present 1SG as a citation form, we would call the lexeme VAIS, and the *v-* stem would seem crucial.

Because the definition of a lexeme derives from that of an inflectional paradigm (lexemes abstract away from inflection), using the notion commits one to a particular view of morphology. It presupposes the existence of a split between inflectional and derivational

---

[2]Lyons (1968) and some more recent authors use *phonological word* instead of *wordform*. This is problematic, "phonological word" being standardly used to denote a particular type of prosodic constituent, which may or may not be coextensive with a wordform. Matthews is explicit on the difference between wordforms and phonological words, both in Matthews (1972: 2, 96, 161) and in the second edition of his textbook (Matthews 1991: 42, 216). Unfortunately, the first edition was somewhat confusing on this particular issue (Matthews 1974: 32-33, 35). Adding to the confusion, Mel'čuk (1993) and Fradin (2003) use the French term *mot-forme* (litterally, "word-form") to denote what Matthews, and after him the whole English-speaking literature, simply calls *word*.

[3]The unfortunate use of the term *lemma* in many discussions in psycholinguistics and Natural Language Processing rests on such a confusion between lexeme and citation form.

morphology (Matthews 1965: 140, note 4; Anderson 1982; Perlmutter 1988). Delineating the sets of words instantiating the same lexeme, such as the one shown in (1a), requires one to distinguish it from a set of words that merely belong to the same morphological family, as the one in (1b).

(1)   a. { *vieux* 'old' M.SG, *vieille* 'old' F.SG, *vieux* 'old' M.PL, *vieilles* 'old' F.PL }
      b. { *vieux* 'old' M.SG, *vieillard* 'old man' SG, *vieillesse* 'old age' SG }

As characterised above, the lexeme is a descriptive category. As such it is compatible with diverse models of morphology, as long as they implement a notion of structured paradigms and split morphology. In practice, however, the notion of a lexeme is mainly used within theoretical frameworks that adopt a constructive view of morphology (Blevins 2006) and use the lexeme as the pivot of the theory, linking inflection and derivation. Following Fradin (2003), we may call this family of frameworks LEXEMIC MORPHOLOGY, and assume that they rely on the series of key hypotheses in (2). The wording is deliberately noncommittal as to how inflection is to be modeled, since proponents of lexemic morphology have assumed either *Item and Process* or *Word and Paradigm* approaches (Hockett 1954).

(2)   a. Atoms of morphological description are SIMPLE LEXEMES.
      b. LEXEME FORMATION RULES predict the possibility of COMPLEX LEXEMES from either a single pre-established lexeme (DERIVATION) or a pair of pre-established lexemes (COMPOSITION).
      c. Inflectional morphology deduces, for each lexeme, the set of words constituting its inflected forms.

It is noteworthy that such a conception of morphology predates the coining of the term *lexeme*. It is very clearly outlined by Kuryłowicz (1945–1949), where *theme* plays a role analogous to *lexeme* as used by lexemic morphology:

> When we say that *lupulus* is derived from *lupus*, or, more precisely, that the theme *lup-ul-* is derived from the theme *lup-*, this means that the *paradigm* of *lupulus* is derived from the *paradigm* of *lupus*.
>
> […]
>
> The derivation process for *lupulus* takes the following concrete form:

$$
\begin{array}{llcll}
lupus & \textit{-i, -o, -um, -orum, -is, -os} & \text{ou} & lup\text{-} & (\textit{-us, -i, -o}, \text{etc.}) \\
\downarrow & & & \downarrow & \\
lupulus & \textit{-i, -o, -um, -orum, -is, -os} & & lupul\text{-} & (\textit{-us, -i, -o}, \text{etc.})
\end{array}
$$

(Kuryłowicz 1945–1949: p. 123; my translation)

Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer

2 Morpheme, lexeme, and the recent history of morphology

The notion of a *morpheme* is without doubt the most popular theoretical innovation of 20th century morphology.[4] Although questions about its usefulness were raised from the 1950s, most notably by Hockett (1954, 1967), Robins (1959), Chomsky (1965) and Matthews (1965), morphemic analysis firmly occupied the center of the stage until the 1990s. Accordingly, the notion of a lexeme barely figured in discussions of morphology. For example, although he adopts a word-based (vs. morpheme-based) approach of morphology, Aronoff (1976) claims in his preface that he has "avoided the term *lexeme* [instead of *word*] for personal reasons" and used "the term *morpheme* in the American structuralist sense, which means that a morpheme must have phonological substance and cannot be simply a unit of meaning".

In the 1980s, most generative morphologists (Lieber 1981, Williams 1981, Selkirk 1982) explicitly reject word-based models and assume that the traditional morpheme is a legitimate unit of analysis (Lieber 2015b). Aronoff (2007) claims that the classical lexicalist hypothesis (Chomsky 1970) holds instead that the central basic meaningful constituents of language are not morphemes but lexemes. However, even among supporters of the lexicalist hypothesis, things are not so clear. Some of them, such as Halle (1973), explicitly adopt a so-called *Item-and-Arrangement* (IA) model while others, such as Jackendoff (1975), adopt a so-called *Item-and-Process* (IP) model. Hockett (1954) coined these two terms *IA* and *IP* to refer to two different views of mapping between phonological form and morphosyntactic and semantic information. In IA models, complex words are viewed as arrangements of lexical and derivational morphemes; in IP models, they are viewed as the result of an operation, called a Word Formation Rule (Aronoff 1976), applying to a root paired with a set of morphosyntactic features and possibly modifying its phonological form. In such models, a complex word is not a concatenation of morphemes but is considered as a single piece. IA models clearly reject lexemes as a pertinent unity. IP models are not so consensual and hesitate between morpheme-based and word (or lexeme)-based theory, and some of them continue to involve morphemes. Corbin's position illustrates this hesitation. While adopting the lexicalist hypothesis, Corbin (1987) never uses the term *lexeme*: she claims "une morphologie du morphème (…) ou plus exactement une morphologie du morphème-mot" (p. 183) and treats affixes as morphemes (p. 285).

Indeed, "this conflict between morpheme-based and lexeme-based theories has haunted generative grammar ever since" (Lieber 2015a).

The work collected in this volume is representative of the growingly dominant view that the lexeme is an unavoidable component of useful morphological descriptions as well as theorizing. The high number of French scholars represented in the volume re-

---

[4]Although the term *morpheme* was coined by Baudouin de Courtenay in 1895 with a meaning close to the contemporary one, its widespread usage with that meaning can be traced back to Bloomfield (1933) and his immediate readers. See Anderson (2015) and Blevins (2016) for relevant discussion of the history of the morpheme.

flects the importance that the notion of a lexeme has played for that community for the past twenty years, mostly under the impulsion of Bernard Fradin (1993, 2003), and the group of researchers involved in the CNRS cooperation network *Groupe de Recherche Description et modélisation en morphologie* he coordinated between 2000 and 2007. We are happy to dedicate this volume to him.

3 Presentation of the volume

While the notion of lexeme is in widespread use in contemporary descriptive and theoretical morphology, many questions remain unresolved. Among others: what is exactly a lexeme: a theoretical description or an object manipulated by rules? Is the difference between lexemes and word-forms as clear as in Matthews' definition? Are lexemes and Lexeme Formation Rules (LFR) always sufficient to explain the formation of lexicon? Do LFR always apply to lexemes?

The twenty papers collected in this volume address the previous questions and some others. They are organized in four sections:

3.1 Lexemes in standard descriptive and theoretical lexeme-based morphology

Three papers centrally deal with this first theme.

In his atypical but stimulating contribution based on his own intellectual biography, Aronoff traces the emergence of lexeme in descriptive and theoretical morphology since the 1960's in Generative Grammar.

In his paper, Boyé focuses on French cardinals and their place in Word and Paradigm models. He argues that, like simple French cardinals, complex cardinals are lexemes, and that their phonological idiosyncrasies can better be modeled in a morpholexical system than in syntax.

Rainer studies the linguistic history of two keywords of economics and politics, viz. CAPITALIST and CAPITALISM, in which semantic change, calques and word formation – suffixation, conversion, suffix substitution – interacted in a complex manner. He argues that, within a morpheme-based model, it would not be possible to account for this history, which, consequently, supports the hypothesis of a lexeme-based conception of the word.

3.2 Lexeme Formation Rules

Lexeme Formation Rules (LFRs) are the main theme of four contributions.

Amiot & Tribout deal with the category of outputs of French suffixation(s) in *-iste*: are they basically adjectives, nouns, lexically underspecified or do we need two different suffixations to account for data-observation? Their proposal is the last one. They consider that, categorically and semantically, the French morphological system contains two suffixations: one of them forms basically professional nouns, the other basically adjective

meaning "in relation to (a practice, an ideology, an activity, a behavior)". They argue that, because such properties can apply to humans, these adjective can easily converted in nouns.

In her contribution, Dal addresses the status of French adverbs in -*ment*. Although they are usually considered derivational, she shows that this status is highly questionable. For her, neither inputs nor outputs respect undoubtedly constraints imposed by a LFR and her conclusion is that they can be regarded as word-forms belonging to the paradigm of adjectives.

Villoing & Deglas focus two morphological patterns in Creole languages based on nouns to form verbs: suffixation N-é and parasynthetic verbs dé-N-é. The hypothesis is that these two patterns emerged following the reanalysis of converted and prefixed French verbs.

Strictly speaking, clipping of deverbal nouns is not a standard LFR. However, the treatment proposed in Štichauer's paper, which applies Fradin & Kerleroux's (2003) Hypothesis of a Maximal (Semantic) Specification, conforms to standard conception of LFRs: in case of polysemous lexemes, clipping applies to specific semantic features of lexeme-bases, and outputs inherit these features, without being synonymous to the full parental form.

3.3 Troubles with lexemes

Six of the contributions centrally address the issue of the definition of lexeme and its use in morphological theories.

Bonami & Crysman's contribution reevaluates the role of the lexeme in recent Head-Driven Phrase Structure Grammar (HPSG) integrating a truly realisational theory of inflection within the HPSG frameword (Bonami & Crysmann 2016). After having distinguished two notions of an abstract lexical object: lexemes, which are characterized in terms of their syntax and semantics, and flexemes (Fradin 2003: 159; Fradin & Kerleroux 2003), which are characterized in terms of their inflectional paradigm, they show how the two notions interact to capture various inflectional phenomena, most prominently heteroclisis and overabundance.

Cruz & Stump deal with essence predicates in San Juan Quiahije Chatino: do they fall in the domain of morphology or in the domain of syntax? Their conclusion is that, even though their structure comprises a predicate base and a nominal component, their inflectional morphology differs from that of simple lexemes.

In his paper on traces of feminine agreement within complex words in Norwegian and Istro-Romanian, Enger tries to overcome troubles with lexemes. He combines a modified version of the Agreement Hierarchy (Corbett 1979) and grammaticalisation to explain what he considers as intra-morphological meaning.

Kihm examines the realization of the copula in Haitian Creole, suggesting that the absence of an overt copula in some contexts should be modeled by postulating an empty stem alternant. He outlines a formal account based on Crysmann & Bonami's (2016) Information-based Morphology, but extending that framework to the analysis of periphrastic inflection.

Spencer questioned whether lexemes are abstract representations of properties unifying a set of inflected word-forms or objects manipulated by rules. Using the architecture of his model of lexical relatedness Generalized Paradigm Function Morphology (GPFM) (Spencer 2013), he proposes an answer to verb-to-adjective transpositions (participles), which can be seen as lexemes-within-lexemes according to their double status of word-forms in relation to verbs, and lexemes in relation to their adjective properties. His proposal is that a lexeme is not a theoretical observation but is best regarded as a maximally underspecified object, bearing all and only those properties which are not predictable from default specification.

Flexemes are also the central issue of Thornton's paper. After reviewing the development of this notion since Fradin (2003) and Fradin & Kerleroux (2003), she focuses on the concept of overabundance in inflectional paradigms and presents data illustrating cases in which a single lexeme maps to two distinct flexemes.

3.4 Troubles with Lexeme Formation Rules

LFRs are questioned in seven papers.

In their study on reduplication in Mandarin Chinese where difference between lexemes and word-forms is less apparent than in languages with clear inflection, Basciano & Melloni claim that the domain of application of reduplication is below the level of the word, or below X° in the standard X-bar approach: for them, in Mandarin Chinese, base units do not have a lexical category and should be vague enough to make them compatible with nominal, verbal and adjectival meanings.

Hathout & Namer explore limits of LFRs to explain and predict the formation of the lexicon. They confront parasynthetics lexemes, in other words complex lexemes that apparently result from simultaneous application of a prefixation and a suffixation, with different hypothesis. This recurrent theme leads them to propose the system ParaDis (for: Paradigms and Discrepancies). ParaDis is a model particularly useful to analyze, explain and predict noncanonical formations (Corbett 2010). It is lexeme-based and combines independency of the three dimensions of LFRs (Fradin 2003) and constraints on outputs founded on derivational families and derivational series (Hathout 2011, Blevins 2016).

Giraudo validates this double view of complex words articulating syntagmatic and paradigmatic dimensions, from a psycholinguistic perspective. She identifies two levels in processing of complex lexemes: the first decomposes complex lexemes into pieces called "morcemes"; the second deals with the internal structure of words according to LFRs and contains lexemes. Her model poses family clustering as an organizational principle of the mental lexicon. She argues that, during language acquisition, growing of family size consecutively continually strengthens links between complex lexemes.

Montermini is devoted to variation of derivational exponents. Adapting the frame developed in Plénat & Roché (2014) and Roché & Plénat (2014, 2016), he argues that this variation obeys to the same constraints as those which explain forms of complex lexemes.

Plag, Andreou & Kawaletz tackle a recurrent and central problem with LFRs: polysemy. They rely frame semantics (Barsalou 1992a,b; Löbner 2013), an approach to lexical

semantics based on elaborate structured representations modelling mental representations of concepts. They hypothesize that the semantics of a derivational process can be described as its potential to perform certain operations on the frames of the bases to which they apply.

Schwarze deals also with the semantic outputs of LFRs. His hypothesis is they are semantically underspecified. The model he proposes is multilayered: it comprises four layers of representation: phonology, constituent structure, functional feature structure and lexical semantics. The meaning of complex words is treated in the framework of two-level semantics. It is assumed that LFRs derive underspecified semantic forms, parting from which the actual meanings are construed by recourse to conceptual structure. Three morphological processes are studied: French *é-* prefixation, Italian denominal verbs of removal, and French noun-to-verb conversion.

Strnadová addresses the issue of apparent rivalry between French denominal adjectives and prepositional phrases in *de*+N where N is the lexeme-base of the adjective (or in relation to it). She discusses some motivations explaining the choice between the former and the latter strategy, and shows that they usually do not have the same distribution and, therefore, are not interchangeable.

Acknowledgements

References

Anderson, Stephen R. 1982. Where's Morphology? *Linguistic Inquiry* 13(4). 571–612.

Anderson, Stephen R. 2015. The morpheme: Its nature and use. In Matthew Baerman (ed.), *The Oxford handbook of inflection*, 11–33. Oxford: Oxford University Press.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: MIT Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.

Aronoff, Mark. 2007. In the beginning was the word. *Language* 83(4). 803–830.

Bally, Charles. 1944. *Linguistique générale et linguistique française*. Paris: PUF.

Barsalou, Lawrence W. 1992a. *Cognitive psychology: An overview for cognitive scientists*. Hillsdale: Erlbaum.

Barsalou, Lawrence W. 1992b. Frames, concepts, and conceptual fields. In Adrienne Lehrer (ed.), *Frames, fields, and contrasts*, 21–74. Hillsdale: Erlbaum.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.

Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.

Bloomfield, Leonard. 1933. *Language*. Londres: George Allen & Unwin Ltd.

Bonami, Olivier & Berthold Crysmann. 2016. The role of morphology in constraint-based lexicalist grammars. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge handbook of morphology*. Cambridge: Cambridge University Press.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: The MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Waltham: Blaisdell.

Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15. 202–224.

Corbett, Greville G. 2010. Canonical derivational morphology. *Word Structure* 3(2). 141–155.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Max Niemeyer Verlag.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374.

Fradin, Bernard. 1993. *Organisation de l'information lexicale et interface morphologie/syntaxe dans le domaine verbal*. Paris 8 dissertation.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* (4). 3–16.

Hathout, Nabil. 2011. Une approche topologique de la construction des mots: propositions théoriques et application à la préfixation en *anti-*. In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (eds.), *Des unités morphologiques au lexique*, 251–318. Paris: Hermès / Lavoisier.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–234.

Hockett, Charles F. 1967. The Yawelmani basic verb. *Language* 43. 208–222.

Jackendoff, Ray S. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3). 639–671.

Kuryłowicz, Jerzy. 1945–1949. La nature des procès dits "analogiques". *Acta Linguistica* 5. 121–138.

Lieber, Rochelle. 1981. Morphological Conversion Within a Restrictive Theory of the Lexicon. In Michael Moortgat, Harry van der Hulst & Teun Hoekstra (eds.), *The scope of lexical rules*, 161–200. Dordrecht: Foris Publications.

Lieber, Rochelle. 2015a. *Introducing morphology*. Cambridge: Cambridge University Press.

Lieber, Rochelle. 2015b. The semantics of transposition. *Morphology* 25(4). 353–369.

Löbner, Sebastian. 2013. *Understanding semantics*. 2nd, revised edition. London: Arnold.

Lyons, John. 1963. *Structural semantics: An analysis of part of the vocabulary of Plato*. Oxford: Publications of the Philological Society.

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Martinet, André. 1960. *Eléments de linguistique générale*. Paris: Armand Colin.

Matthews, P. H. 1965. The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics* 1(2). 139–171.

Matthews, P. H. 1972. *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Matthews, P. H. 1974. *Morphology*. Cambridge: Cambridge University Press.

Matthews, P. H. 1991. *Morphology*. 2nd edition. Cambridge: Cambridge University Press.

Mel'čuk, Igor. 1993. *Cours de morphologie générale. Première partie : Le mot*. Vol. 1. Montréal /Paris: Les Presses de l'Université de Montréal/CNRS Éditions.

Perlmutter, David M. 1988. The split morphology hypothesis: Evidence from Yiddish. In Michael Hammond & M. Noonan (eds.), *Theoretical morphology: Approaches in modern linguistics*, 79–100. San Diego: Academic Press.

Plénat, Marc & Michel Roché. 2014. La suffixation dénominale en *-at* et la loi des (sous-)séries. In Florence Villoing, Sophie David & Sarah Leroy (eds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 47–74. Nanterre: Presses Universitaires de Paris Ouest.

Robins, R. H. 1959. In defense of WP. *Transactions of the Philological Society*. 116–144.

Roché, Michel & Marc Plénat. 2014. Le jeu des contraintes dans la sélection du thème présuffixal. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberger, Judith Meinschaefer & Sophie Prévost (eds.), *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1863–1878. Paris: Institut de Linguistique Française.

Roché, Michel & Marc Plénat. 2016. De l'harmonie dans la construction des mots français. In Franck Neveu, Gabriel Bergounioux, Marie-Hélène Côté, Jean-Marc Fournier, Linda Hriba & Sophie Prévost (eds.), *Actes du 5e Congrès Mondial de Linguis- tique Française. Tours, 4-8 juillet 2016*. Paris: Institut de Linguistique Française.

Selkirk, Elisabeth. 1982. *The syntax of words*. Cambridge: MIT Press.

Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press.

Trnka, Bohumil. 1949. Rapport, question III: Peut-on poser une définition universellement valable des domaines respectifs de la morphologie et de la syntaxe? In Michel Lejeune (ed.), *Actes du sixième congrès international des linguistes*. Paris: Klincksieck.

Williams, Edwin. 1981. On the notions "lexically related" and "head of a word". *Linguistic Inquiry* 12. 245–274.

Zwicky, Arnold M. 1992. Some choices in the theory of morphology. In Robert D. Levine (ed.), *Formal grammar. Theory and implementation*, 327–371. Oxford: Oxford University Press.

Part I

Lexemes in standard descriptive and theoretical lexeme-based morphology

Chapter 1

Morphology and words: A memoir

Mark Aronoff
Stony Brook University

Lexicographers agree with Saussure that the basic units of language are not morphemes but words, or more precisely lexemes. Here I describe my early journey from the former to the latter, driven by a love of words, a belief that every word has its own properties, and a lack of enthusiasm for either phonology or syntax, the only areas available to me as a student. The greatest influences on this development were Chomsky's *Remarks on Nominalization*, in which it was shown that not all morphologically complex words are compositional, and research on English word-formation that grew out of the European philological tradition, especially the work of Hans Marchand. The combination leads to a panchronic analysis of word-formation that remains incompatible with modern linguistic theories.

Since the end of the nineteenth century, most academic linguistic theories have described the internal structure of words in terms of the concept of the *morpheme*, a term first coined and defined by Baudouin de Courtenay (1895/1972, p. 153):

> that part of a word which is endowed with psychological autonomy and is for the very same reason not further divisible. It consequently subsumes such concepts as the root (radix), all possible affixes, (suffixes, prefixes), endings which are exponents of syntactic relationships, and the like.

This is not the traditional view of lexicographers or lexicologists or, surprising to many, Saussure, as Anderson (2015) has reminded us. Since people have written down lexicons, these lexicons have been lists of words. The earliest known ordered word list is Egyptian and dates from about 1500 BCE (Haring 2015). In the last half century, linguists have distinguished different sorts of words. Those that constitute dictionary entries are usually called *lexemes*. Since the theme of this volume is the lexeme, I thought that it might be useful to describe my own academic journey from morphemes to lexemes. Certainly, when I began this journey, the morpheme, both the term and the notion, seemed so modern, so scientific, while the word was out of fashion and undefined. Morphemes were, after all, atomic units in a way that words could never be, and if linguistics were to have any hope of being a science, it needed atomic units.

I grew up with morphemes. The structuralist phoneme may have fallen victim to the generative weapons of the 1960s, but no one questioned the validity of morphemes at

MIT. They were needed to construct the beautiful syntactic war machines that drove all before them, beginning with the analysis of English verbs in *Syntactic Structures*, which featured such stunners as the morpheme S, which "is singular for verbs and plural for nouns ('comes', 'boys')" and ∅, "the morpheme which is singular for nouns and plural for verbs, ('boy', 'come')" (Chomsky 1957: 29, fn. 3).

Aside from brief mentions here and there in *Syntactic Structures* and the cogent but little noted discussion at the end of Chomsky's other masterwork, *Aspects* (Chomsky 1965), by the time I arrived at MIT as a graduate student in 1970 there was no talk of morphology; the place was all about phonology and syntax. These two engines, which everyone was hard at work constructing, would undoubtedly handle everything in language worth thinking about. My problem was that I very quickly discovered that I had little taste for either of the choices, phonology or syntax. It was like having a taste for neither poppy seed bagels nor sesame seed bagels, and having no other variety available at the best bagel bakery in the world, but still wanting a bagel. This had never happened to me before, and not just with bagels. Maybe I should go to another store, but I liked the atmosphere in this one a lot and, like the St. Viateur bagel shop, famous to this day (www.stviateurbagel.com), it was acknowledged to be the best in the world.

What I did love was words. I had purchased a copy of the two-volume compact edition of the *Oxford English Dictionary* (OED) as soon as I could scrape together the money to buy one, even though reading the microform-formatted pages of the dictionary required a magnifying glass. I also owned a copy of *Webster's III*. I kept these dictionaries at home, not at my desk in the department. Dictionaries and the words they contained were my dark secret. Why should I tell anyone I owned them? These dictionaries served no purpose in our education, where the meanings of individual words were seldom of much use, though we did talk a lot about the word classes that were relevant to syntax: *raising verbs, psych verbs, ditransitive verbs*. The only dictionary we ever used in our courses was *Walker's Rhyming Dictionary*, a reverse-alphabetical dictionary of English, first published in 1775. Its main value, as Walker had noted in his original preface, was "the information, as to the structure of our language, that might be derived from the juxtaposition of words of similar terminations." Chomsky & Halle had mined it extensively in their research for *The Sound Pattern of English* and it was to prove invaluable in my work on English suffixes, though I did not know it at first.

The 1960's had seen the brief flowering of ordinary language philosophy, whose proponents, beginning with the very late Wittgenstein (1953), were most interested in how individual everyday words were used, in opposition to the logical project of Wittgenstein's early work. Despite the popularity of such works as Austin (1962) and Searle (1969), ordinary language philosophy never went very far, at least in part because its proponents never developed more than anecdotal methods of mining the idiosyncratic subtleties of usage of individual words. But there was no contradicting the view that every word is a mysterious object with its own singular properties, a fact that most of my colleagues willfully ignored, in their search for the beautiful generality of rules. The question for me was and remains how to balance the two, words and rules.

Morris Halle had given a course on morphology in the spring of 1972, in preparation for his presentation at the International Congress of Linguists in the summer. Noam Chomsky had published a paper on derived nominal two years before, in 1970, which, though it was directed at syntacticians, provided a different kind of legitimation for the study of the individual words that my beloved dictionaries held. Maybe I could find something there, I said to myself with faint hope, though the approach that Halle had outlined did not open a clear path for me and I knew that I was not a syntactician, so Chomsky's framework did not appear at first to provide much hope, despite his attention to words.

Beginning in early 1972, I spent close to a year reading everything I could lay my hands on that had anything to do with morphology. I started with Bloomfield and the classic American Structuralist works of the 1950s that had been collected in Martin Joos's (1958) *Readings in Linguistics*, almost all of which dealt with inflection. Though I learned a lot, I couldn't find much of anything in that literature to connect with the sort of work that was going on in the department or in generative linguistics more broadly at the time.

In the end, I did find something to study in morphology, though not in generative linguistics. I have come back to this topic, English word formation, again and again ever since, but only now am I beginning to gain some real grasp of how it works. The seeds of my understanding were sown in my earliest work on the topic but they lay dormant for decades, until they fell on fertile ground, far outside conventional linguistic tradition. And though again I did not come to understand it for decades, word-formation was also a fine fit for the Boasian approach that I had learned to love in my first undergraduate linguistics training, in which the most interesting generalizations are often emergent, rather than following from a theory. Also, the nature of the system in morphology, and especially word-formation, is much better suited to someone of my intellectual predilections. This is an area of research in which regular patterns can best be understood in their interplay with irregular phenomena. I enjoy this kind of play.

Word-formation and morphology in general had had an odd history within the short history of generative linguistics before 1972, generously twenty years. One of the best-known early generative works was about word-formation, Robert Lees's immensely successful *Grammar of English Nominalizations* (1960). This book, though, despite its title, dealt mostly with compounds and not nominalizations, using purely syntactic mechanisms to derive compounds from sentences, seemingly modeled on the method of *Syntactic Structures*.[1] Lees's book directly inspired very little research on word-formation in its wake, though the idea of trying to derive words from syntactic structures has surfaced regularly ever since (Marchand 1969, Hale & Keyser 1993, Pesetsky 1995).

Chomsky's 1970 "Remarks on nominalization" (henceforth Remarks) echoed Lees's book in title only. It was in fact its complete opposite in spirit, method and conclusions, although Chomsky never said so. After all, he owed Lees a great personal debt. Lees had played a large role in making Chomsky famous with his (1957) review in *Language* of Chomsky (1957). Remarks injected for the first time into generative circles the observa-

---

[1]Lees's book went through five printings between 1960 and 1968, extraordinary for a technical monograph that was first published as a supplement to a journal and then reissued by a university research center.

tion that some linguist units, in this case derived words, are semantically idiosyncratic and not derivable in syntax (unless one is willing to give up on the bedrock principle of semantic compositionality). Word-formation, it turns out, is centered on the interplay between the idiosyncrasies of individual words that Chomsky noted and the regular sorts of phenomena that are enshrined in the rules of grammar.

My first excursion into original morphological research took place in the fall and winter of 1972–73, a time when I was entirely adrift. I had begun to read widely and desperately on morphology early in 1972, hoping it might save me from myself, but had not yet lit on any phenomenon that held the faintest glimmer of real promise. This is the lifelong agony of an academic: the struggle to find something that is both new and of sufficient current interest for others to give it more than a passing glance. For some reason, I embarked on a study of Latinate verbs in English and their derivative nouns and adjectives, verbs like *permit* and *repel*, and their derivatives: *permission* and *permissive*; *repulsion* and *repulsive,* which contained a Latin prefix followed by a Latin root that did not occur independently in English. All the verbs had been borrowed into English and I can't recall for the life of me what led me to study this peculiar class of words.

What I first noticed about these verbs and their derivatives was that the individual roots very nicely determined the forms of the nouns and adjectives from the verb by affixation. Each individual root such as *pel* generally set the form of the following noun suffix (always *-ion* after *pel*). Also, a given root often also had an idiosyncratic form (here *puls-*) before both the noun and adjective suffix: com***puls***ion, com***puls***ive; ex***puls***ion, ex***puls***ive; and so on for all verbs containing this Latinate root. With a very small number of exceptions, the pattern of root and suffix forms was entirely systematic for any given root but idiosyncratic to it, and therefore predictable for many hundreds of English verbs, nouns, and adjectives. The whole system was also obviously entirely morphological. And best of all, no one had noticed it before. I had discovered something new in morphology and I quickly outlined my findings in by far the longest paper that I had ever written, almost fifty pages, filled with typos, which I completed in April 1973.

The central results of this first work were entirely empirically driven. I have prized empirical findings above all other aspects of research ever since, because these findings don't change with the theoretical wind. The generalizations I found are as true today as they were in 1973. In this emphasis on factual generalization I differ from most of my linguist colleagues. Of the empirical discoveries that I have made over the years, I am proudest of three: this one, the morphome, and the morphological stem.

It wasn't long before I realized that Latinate roots presented a fundamental problem for standard structural linguistic theories of morphology. All of these theories were – and many still are –based on the still unproven assumption that Baudouin de Courtenay had first made explicit almost a century before in linguistics, that all complex linguistic units could be broken down exhaustively into indivisible meaningful units, which were reassembled compositionally (in a completely rule-bound manner) to make up utterances.[2] The problem was that, although these Latinate roots could not be said to have

---

[2]The idea that morphology and syntax are both compositional was simply assumed from the beginning, though it should be noted that Baudouin's work predates Frege's discussion of compositionality.

constant meaning, or in some cases any meaning at all that could be generalized over all their occurrences, they had constant morphological properties. The English verbs *ad**mit**, com**mit**, e**mit**, o**mit**, per**mit**, re**mit**, sub**mit**, trans**mit***, and so on, do not share any common meaning. What they do share are the morphological peculiarities of the root *mit*. The classical Latin verb *mittere* meant 'send' and the prefixed Latin verbs to which the English verbs are traceable may have had something to do with this meaning in the deep historical past of Latin, but even in classical times the prefixed verbs had begun to diverge semantically from their base and from each other. What ties them so closely together in English is only the structural fact that, without exception, they share the alternant *miss* before the noun suffix *-ion* and the adjective suffix *-ive*, and that the form of the noun suffix that they take is similarly always *-ion*, and not *-ation* or *-ition*.

The verb root *mit/miss* has very consistent, unmistakable, and idiosyncratic morphological properties in English today. Unless we choose to disregard them, these properties must be part of the morphology of the language. But the root has no meaning, so it can't be a morpheme in the standard sense. How can we make sense of this apparent paradox?

The answer is found in the empirical observation that formed the core of Chomsky's Remarks: derived words are not always semantically compositional. This observation, which Chomsky called the *lexicalist hypothesis*, is the single greatest legacy of Remarks. It is far from original; only its audience is new. Jespersen, for example, writing about compound words, had pointed out many times over several decades that the relations between the members of a compound are so various as to defy any semantically predictive analysis. Jespersen concluded that the possible relations between the two members of a compound are innumerable:

> Compounds express a relation between two objects or notions, but say nothing of the way in which the relation is to be understood. That must be inferred from the context or otherwise. Theoretically, this leaves room for a large number of different interpretations of one and the same compound […] On account of all this it is difficult to find a satisfactory classification of all the logical relations that may be encountered in compounds. In many case the relation is hard to define accurately […] The analysis of the possible sense-relations can never be exhaustive. (Jespersen 1954: 137-138)

The purpose of Remarks had been tactical. As Harris (1993) recounts in detail, at the time of writing the article, Chomsky was locked in fierce combat with a resurgent group of younger colleagues, the generative semanticists, who sought to ground all of syntax in semantics. Syntax at the time was assumed to encompass word-formation, though in truth almost no work had been done on word-formation besides Lees (1960). Reminding everyone in the room that at least some word-formation was not compositional, a purely empirical observation, cut the legs out from under generative semantics in a single stroke from which the movement never recovered. More importantly, although Chomsky never mentioned it and may not have realized it, the demonstration that some complex words are not semantically compositional also destroyed Baudouin's traditional morpheme and lent support to Saussure's sign theory of words. The non-compositional complex words

at the core of Remarks lie within the class of what Jespersen (1954) called *naked words*: uninflected words. Complex naked words are formed by derivational morphology and compounding. Inflected forms, by contrast, are always compositional, because they realize cells in the morphosyntactic paradigm of the naked word. Their properties are accidental, in the traditional grammatical sense of the term, not essential.

What I had learned from Remarks about compositionality within words, combined with my discoveries about meaningless Latinate roots, led me to realize that word-formation needed to be studied in a way that was free from Baudouin's axiom, an axiom that had held sway for over a century: that complex words can be broken down exhaustively into meaningful morphemes. Although I was entirely unaware of the consequence at the time, and remained unaware of it for decades, this discovery freed me to do linguistics in the way I loved to, not deductively as I had been taught to do at MIT, following some current theory where it led, and not inductively, but by working towards what the great Barbara McClintock had called "a feeling for the organism" (Keller 1983). My first two years at MIT had taught me that the theory and deduction game held little charm for me. Perhaps that's because I wasn't very good at it. Working on my own terms made me feel better about myself than I had for the entire preceding two years. I could stop worrying whether I was as smart as all those other people. It turned out I didn't have to be smart. Common sense was at least as valuable, and much rarer in those circles.

English had been an exotic object of inquiry for American linguistics from the start. The first American Structuralists were anthropological field workers who confined themselves deliberately to the native languages of North America. Only in his very last years did Edward Sapir turn to English. Bloomfield discussed English in his *Language* (1933), presumably to engage a broad readership, but in his technical writing he too dealt mostly with languages of North America on which he did original fieldwork. Bloomfield's successors, notably Trager & Lee Smith (1951) did important work on English, but they were in a decided minority.

Generative grammar was different. The vast bulk of research in the first two decades, beginning with Chomsky et al. (1956), had been on English. This English bias was especially true of generative syntax, whose success was due in no small part to the analyst being able to come up with novel sentences on the fly that the grammar could label as either grammatical or ungrammatical. Only a native English speaker could have come up with the most important sentence in the history of linguistics, Chomsky's *colorless green ideas sleep furiously*.[3] Even in generative phonology, whose earliest works, Chomsky (1951) on Modern Hebrew and Halle (1959) on Russian had dealt with other languages, the high-water mark of this tradition was an analysis of English, *The Sound Pattern of English*. It was therefore not entirely unexpected that I should turn my attention to English word formation. Even my earliest excursion into morphology had dealt with English, albeit Latin roots that had been borrowed into English. It would be a decade before I looked seriously at word-formation in other languages (Aronoff & Sridhar 1984).

American linguists had not written much about word-formation in the preceding quarter century. The great Structuralists from Bloomfield to Hockett had done seminal work

---

[3]All the data in the most important American structuralist work on syntax before *Syntactic Structures*, Wells (1947), is from English, except for one small example from Japanese.

on morphology. Much of it was collected in Martin Joos's (1958) *Readings in Linguistics*, which I read carefully, along with the chapters on morphology in Bloomfield's *Language* (1933). But the Structuralists had dealt almost exclusively with inflection. I could find almost nothing on uninflected words. There was Lees's (1960) monograph, but his approach was not useful in a post-Remarks environment, and besides, he mostly dealt with compounds.

The most notable exception of the previous decade had been Karl Zimmer's monograph on English negative prefixes (Zimmer 1964). This book opened up an entirely new world for me, the tradition of English linguistics. This world had existed for a century and more, parallel to the one I inhabited but completely unknown to us, and it was one in which the study of word-formation had always occupied an important place.

English linguistics had emerged in departments of English language and literature, where in the 1970s it still retained the connections to philology that most of the rest of the field had left behind in the 19th century. To this day, it is much more rooted in texts than other kinds of linguistics, because of its closeness to literature. Much of English linguistics was historically oriented, but in a very different way from the comparative historical linguistics that lay at the root of modern structural linguistics. Its focus was on the linguistic history of a single language, the record of English since its emergence as a distinct written language around 800 CE. The connection to philology lay in this shared basis of written texts, though philologists were much more literarily oriented. People who read Beowulf and Chaucer and Shakespeare had to know something about the language these people were writing in and English linguistics served this purpose.

Every undergraduate English major—and there were many more in those days—had to take a course on the history of the English language. For the same reasons, English linguistics had sister disciplines in the other major standard European languages and language families: French, German, Italian, Spanish, Romance, Scandinavian, etc. As I learned much later, the OED was the greatest monument of this tradition of English linguistics, but much of the best work had been done on the European continent, especially in German departments of *Anglistik*. The best-known exponent of this tradition was a Dane, Otto Jespersen.

Hans Marchand reviewed Zimmer's monograph in *Language* in 1966. Marchand had fled from Germany to Istanbul in 1934 as a Catholic political refugee with the help of his mentor, the Jewish Romance philologist Leo Spitzer. He gradually turned towards the study of language rather than literature, remaining in Istanbul until 1953. Marchand returned to Germany in 1957, after a stint in the United States, to teach *Anglistik* at the University of Tuebingen. His book, *The Categories and Types of Present-Day English Word-Formation*, published in 1960 and greatly revised in 1969, has remained the authoritative description of English word-formation since its first publication. Remarkably, Marchand had written most of the book while in internal exile in Turkey in an Anatolian village from 1944 to 1945, under threat of repatriation to Germany, which had drafted him into the military in absentia in 1944. He had sought unsuccessfully for years to publish this early version while still in Turkey.

Mark Aronoff

Marchand and Zimmer follow very similar approaches, quite different from that of American structural linguistics. They ask what a given derivational affix meant (what Zimmer calls its "semantic content"), what it applied to, and what it produced. The prefix *un-* that most occupies Zimmer's mind, for example, is negative in meaning and derives adjectives from adjectives.[4] This is all very traditional and in line with the treatment of derivational affixes in the OED, which contained entries for derivational affixes from the beginning, though not for inflectional affixes. The adjectival negative prefix *un-* has a very extensive entry in OED, with many observations similar to those of Marchand and Zimmer, and hundreds of examples (my favorite being *unpolicemanly*). The OED even notes the morphological environments in which a given derivational affix is particularly productive, which was of special importance to Zimmer and to my own work. For *un-*, the OED notes that it is especially common with adjectives ending in *-able*: "In the modern period the examples become too numerous for illustration; in addition to those entered as main words, those given below will serve as specimens of the freedom with which new formations are created."

This traditional approach to word-formation provided an intuitively satisfying solution to the problem of the morpheme that my work on Latinate roots had uncovered. If derivation is not a matter of combining morphemes but of attaching affixes to words, then we don't need all the morpheme components of words to be meaningful and we don't need the internal semantics of words to be compositionally derived from these components. All we need is for words to be meaningful. We don't need to worry about morphemes at all, only words and what the derivational affixes do with them.

This traditional approach circumvented the problem of meaningless morphemes for a simple reason: it predated the notion of the morpheme. The earliest citation in OED by far for any sense of the word *derivation* equates it with *formation.* It comes from Palsgrave's 1530 English-language grammar of French, *L'esclarcissement de la langue françoyse*, the first known grammar of French ever written in any language: "1530 J. Palsgrave *Lesclarcissement* 68 Derivatyon or formation, that is to saye, substantyves somtyme be fourmed of other substantyves." This has become my favorite citation of the words derivation and (word) formation and, though I did not know it at first, it encompasses the claim that words are formed from words; my observation that words are formed from words merely updates Palsgrave's remark. This claim is the essence of the traditional treatment of word-formation and it is the motto that I adopted, elevating the observation to a principle.[5]

In my dissertation and subsequent monograph, I took complete credit for the axiom that morphology was word-based. Even decades later, when I clarified the terminology and called it *lexeme-based morphology*, I did not provide any direct attribution to the tradition of English word-formation studies. My only defense is that neither Marchand nor Zimmer ever stated what for them was simply an unspoken assumption. All I did was to make this assumption clear as an axiom. I can therefore at least take credit for the realization that this was a useful axiom on which to base the analysis of word-formation.

---

[4] *Un-* also attaches to verbs and has the sense of undoing the action of the verb. Whether these two are one and the same affix has been much discussed (Horn 1984).

[5] The idea that words are formed from words may ultimately be traceable to the Greek and Latin grammatical traditions, which were entirely word-based, even at the level of inflection (Robins 1959).

Notation meant everything in those days. Chomsky & Halle (1968) had even gone so far as to extol the explanatory power of parentheses. My most important task was therefore to create a simple notation in which traditional OED-style generalizations about word-formation could be stated in a way that generative linguists might understand. This was the word-formation rule (WFR). It bore close resemblance in form to the rewrite rules that were standard in generative grammar. A WFR took a word from one of the three major lexical categories (Noun, Verb, or Adjective) and mapped it onto a lexical category (the same or another), usually adding an affix, and making another word. The rule of *un-* prefixation, for example, could be written as $[X]_A \rightarrow [\text{un-}[X]_A]_A$ or it could be written simply as the output $[\text{un-}[X]_A]_A$. This notation was transparent and made generative linguists, myself included, think that this way of dealing with word-formation could be easily assimilated into their way of thinking. The acronym WFR added a nice touch. The title of the published version of my dissertation, *Word Formation in Generative Grammar* (Aronoff 1976) was suggested by S. Jay Keyser, the editor of the series of which this would be the inaugural monograph. It only served to strengthen the impression that I had integrated the study of word-formation into generative grammar. The monograph was a great success, thanks in no small part to its title, and most accounts treat the book as central to the treatment of morphology and word-formation within generative grammar.

Nothing could be further from the truth. The title of the monograph was deeply deceptive and in agreeing to it I was also deceiving myself. Word formation rules, as conceived of and discussed in that monograph, are incompatible with generative grammar or with any grammar-based linguistic framework, because, like the tradition they encode, these rules cross the synchronic-diachronic boundary that is central to all post-Saussurean structural linguistics. I have only recently come to appreciate this fact. I certainly believed at the time that I was doing generative grammar, as have most of the book's readers since. What is true is that I was a member of a social community self-organized around generative grammar. I did my work on word-formation within that community and it was accepted as legitimate almost entirely on those social grounds.

In his great posthumous work, Saussure 1916/1959 set up a distinction that has been accepted throughout the field ever since, between *synchronic* and *diachronic* linguistics. Synchronic linguistics deals with a single state of a language—the present—while diachronic linguistics deals with successive states—history. Generative grammar seeks to provide a theory of what is a possible synchronic grammar of a language, the basic idea being that the grammar generates the language (Chomsky 1957). The theory is also supposed to mirror the innate capacity that a child brings to the task of constructing a grammar for the input that the child receives (Chomsky 1965). But traditional research on word-formation, which preceded Saussure in its origins, is neither synchronic nor diachronic: it is about how new derived words accumulate in a language **over time**. That is why Marchand gave his *magnum opus* the subtitle "A Synchronic-Diachronic Approach" and why Jespersen called his monumental six-volume life's work *A Modern English Grammar on Historical Principles*, both titles in direct contradiction of the Saussurean split, both by scholars working within the tradition of English linguistics. In

truth, Marchand's approach was neither synchronic nor diachronic, in spite of its fashionable title, because the study of word formation lends itself to neither synchrony nor diachrony: the word formation system of the language at any given moment can only be understood through the historical accumulation of the lexicon. The study of word-formation is concerned at its core with how words are created, how they are formed, and how they are added to the language. Unlike sentences, words, once formed, accumulate, and this accumulated storehouse has an effect on new words. Words accumulate both in the mental lexicon of an individual speaker and in the collective lexicon of the larger linguistic community.

This brings us back to Chomsky's lexicalist hypothesis. To understand this hypothesis, we need to clarify two distinct senses of the word *lexical* (Aronoff 1988). One is Bloomfield's lexicon, the list of what DiSciullo and Di Sciullo & Williams (1987) later so nicely called the "unruly." The other encompasses the word-formation rules themselves and maybe all morphology including inflection too. The term *lexical component* is usually meant to include both the rules of morphology and the lexicon. Chomsky's original lexicalist hypothesis says no more than that the lexical component is responsible for forming and storing some of the complex words of the language, in addition to the simple monomorphemic words that have always been thought of as arbitrary signs stored in the lexicon. His major criterion for distinguishing lexically from 'transformationally' derived words is semantic predictability or compositionality (lexically derived words are not compositional) though most later lexicalist theorists used others as well (Aronoff 1994, Pesetsky 1995).

Halle's (1973) lexicon, which he described as "a special filter through which the words have to pass after they have been generated by the word formation rules" (p. 5), is a Bloomfieldian list of words, separate from the morphological rules. Halle suggested that "the list of morphemes together with the rules of word-formation define the set of *potential* words of the language. It is the filter and the information that is contained therein which turn this larger set into the smaller subset of *actual* words" (p. 6). This way of looking at the relation between word-formation and the lexicon appears to permit us to include word-formation in a synchronic grammar: the morphemes and the abstract rules of word-formation will be part of the grammar, not the lexicon, while the actual results of the application of the rules to the morphemes, which can be quite messy and idiosyncratic, as Chomsky had already emphasized, will be housed outside the grammar in the Bloomfieldian lexicon. Words will be formed by rules in the grammar, just as sentences are, though perhaps by a distinct lexical component, along the lines of the theory of Remarks. On this story, though, once words are formed they are stored in the lexicon and should accordingly have no further interaction with the grammar or the rules.

Over the years, this general strategy of strictly separating the rules from the unruly in order to better assimilate word-formation to syntax, what Marantz much later called the *single engine hypothesis* (Marantz 2005) has faced a number of problems, all of which are traceable to the fact that the strategy allows for no interaction between the rules (and the morphemes they operate on) and the set of words formed by the rules, which are stored in the lexicon. The insulation of the rules from the lexicon makes it impossible to

ask many interesting questions with even more interesting answers. I will discuss briefly here only the two most important ones, morphological productivity and blocking.

Unlike most rules of syntax, rules of word-formation vary widely in their productivity. A standard example is the trio of suffixes -*ness*, -*ity*, and -*th*, all of which form nouns from adjectives in English. of the three, -*th* is the least productive; only a handful of words end in this suffix. The only one I can identify as having been added to the language in the last couple of centuries is *illth*, which was coined on purpose by John Ruskin in 1862 to denote the opposite of *wealth*. The word is almost never used today, except in close proximity to *wealth* or *health*. Speakers of English know that new or infrequent words in -*th* have an odd flavor about them. The OED remarks about the word *coolth*, for example, that it is "Now chiefly literary, arch[aic], or humorous."

The suffix -*ity* is more productive, but limited in the morphology of what it can attach to. The OED lists approximately 2400 nouns in current use ending in the letter sequence <ity>, most of which contain the suffix, compared with about 3600 ending in the letters <ness>. But a closer look reveals that <ity> is much more likely to appear after a select set of suffixes. With -*ic* it is preferred by a ratio of almost 7/1 over -*ness*. This preference is reflected in speakers' judgments and in the relative frequency of members of individual pairs. The word *automaticity* feels much more natural than *automaticness* and a simple Google search shows 109,000 "hits" for *automaticity* but only 242 for *automaticness*. Even for very rare words, the same pattern emerges. While *oceanicity*, a word I have never heard of, gets only 762 hits, its counterpart, *oceanicness*, gets only 5!

Once we leave the few affixes that -*ity* is attracted to, though, -*ness* is ascendant. *Greenness* outnumbers *greenity* 1000/1. Google even thinks that you have made a mistake when you search for *greenity and* asks: "Did you mean: greenify?" A similar pattern of results is found for all the other color words. In the same vein, we can find examples of humorous uses of words like *sillity* or *slowity* in the Urban Dictionary, but not in many other places on the Web.

There are numerous ways of distinguishing the productivity of these three suffixes, but productivity is clearly related to the number of words that are already present in the language: the more you have, the more you get. Productivity depends on the accumulation of words. It is a dance between the lexicon and the grammar. If we try to make a strict separation between the two, we will never understand how the dance works. Both Marchand and Zimmer knew about the nuances of productivity. Marchand closes his review of Zimmer's book with the following somewhat backhanded compliment: "Zimmer's investigation is a valuable contribution not to the study of semantic universals, which it planned to be, but to the problem of productivity in word-formation" (Marchand 1966: 142).

The other problem that productivity poses for modern linguistics is that it is variable. Mainstream formal linguistics, with its roots in the triumphal 19th century neo-grammarian slogan that sound change laws have no exceptions (Paul 1880) has never dealt well with variation. If anything, formal linguists continue to be blind to the fact that variation is a part of language (I-language). One response to variability is simply to deny that a phenomenon like productivity exists. Another is to admit that it exists,

but to deny that the phenomenon is variable, claiming instead that it is all or none. That is what Marchand does. Referring to Harris (1951: 225), Marchand notes disapprovingly that "a descriptivist like Zellig S. Harris maintained that 'the methods of descriptive linguistics cannot treat of the degree of productivity of elements'" (Marchand 1966: 141) . But he himself only dichotomizes word-formation rules into those that are productive and those that are, in his words, restricted:

> Zimmer's merit is to have seen an important problem in word-formation, that of productivity. . . . Zimmer's study . . . calls our attention to the fact that what seems to be the same type of combination, viz. derivation by means of a negative prefix, is in reality split up into two groups, one of restricted productivity (instanced by *unkind*) and another, deverbal group (instanced by *unread*) which is of more or less unrestricted productivity (Marchand 1966: 141).

Even here, Marchand is not talking about one productive rule vs. a different unproductive rule, but rather a single rule, which is more productive in one environment (with past participles and *-able* derivatives, both of which have a passive reading) and less productive in another (with underived adjectives like *kind*). As Zimmer demonstrates, there is not in fact a dichotomy, but rather a cline in productivity that depends on both environments and rules. In the half century since, the nondiscrete nature of productivity has been demonstrated time and again, most definitively in Bauer (2001).

Productivity is a question of fecundity, how many words there can be and how easily they can be created. A pattern is highly productive if there can be many new words in that pattern. It is unproductive if there can be only a few new words. When we say that the English nominal suffix *-ness* is highly productive we mean that the pattern can form many nouns from adjectives; when we say that the suffix *-th*, which also derives nouns from adjectives, is unproductive, we mean that it cannot. And because words are formed from words, there is a direct relation between how easy it is to form words in a pattern and how many already exist in that pattern, in either the mind of a speaker or the language of a community. As we have just seen, there are many *-ness* nouns in English. The OED lists over 4000 nouns ending in the letters <ness>, the great majority of them containing the suffix. There are no more than a handful of *-th* nouns derived from adjectives. If how many words there can be of a given type depends on a combination of how many words there are already of this type and how many there are for the type to feed on, then words differ sharply from sentences. For starters, it makes little sense to even ask how many sentences there are of a given type. Sentences are not stored, they are produced and then vanish.

Blocking is the second phenomenon that demonstrates how the formation of individual words depends intimately on the words we already know. For four decades, since the moment that I first stumbled on this phenomenon, it has been clear to me that blocking is a real empirical phenomenon and that it is just what I first defined it to be: "the nonoccurrence of one form due to the simple existence of another" (Aronoff 1976: 43). A few pages later, I made an explicit connection to synonymy: "Blocking is basically a constraint against listing synonyms in a given stem" (Aronoff 1976: 55). And on the

same page I wrote: "To exclude having two words with the same meaning is to exclude synonymy, and that is ill-advised." A few pages later, I referred to "the blocking rule." Clearly, I had no idea precisely what blocking was, beyond an empirical phenomenon. Only now, though, do I understand why my empirical observation might be true: the avoidance of synonymy in general and blocking in particular are the result of competition, a topic I have spent the last half decade investigating.

The tradition of word-based morphology dates to the first grammarians, although it was eclipsed for much of the twentieth century by the rise of synchronic linguistics. In Cambridge, Massachusetts one didn't learn much about what was happening in Cambridge, England, but soon after leaving for Stony Brook I learned that word-based morphology had been revived in England in the decade or so before my own research, notably by R. H. Robins (1959) and Peter Matthews (1965, 1972). This line of research, especially in derivational morphology, has grown in the decades since, notably in France, led by Danielle Corbin (1987), Françoise Kerleroux (1996), and Bernard Fradin (2003). Together, they created a new thriving research community, of which I am proud to be a member.

References

Anderson, Stephen R. 2015. The morpheme: Its nature and use. In Matthew Baerman (ed.), *The Oxford handbook of inflection*, 11–33. Oxford: Oxford University Press.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: MIT Press.

Aronoff, Mark. 1988. Two senses of *lexical*. In *Proceedings of the fifth eastern states conference on linguistics*, 1–11.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.

Aronoff, Mark & S. N. Sridhar. 1984. Agglutination and composition in Kannada verb morphology. In *Proceedings of the 20th meeting of the Chicago linguistics society: Papers from the parasession on lexical semantics*, 3–20.

Austin, J. L. 1962. *How to do things with words*. Oxford: Clarendon Press.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Chomsky, Noam. 1951. *Morphophonemics of Modern Hebrew*. University of Pennsylvania MA thesis. Published in 1979 by Garland Publishing, New York.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: The MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Waltham: Blaisdell.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chomsky, Noam, Morris Halle & Fred Lukoff. 1956. On accent and juncture in English. In Morris Halle, Horace Lunt, Hugh McLean & Cornelis van Schooneveld (eds.), *For Roman Jakobson. essays on the occasion of his sixtieth birthday*. The Hague: Mouton.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Max Niemeyer Verlag.

de Courtenay, Baudouin. 1895. An attempt at a theory of phonetic alternations. In Edward Stankiewicz (ed.), *A Baudouin de Courtenay anthology (1972)*. Bloomington: Indiana University Press.

Di Sciullo, Anna Maria & Edwin Williams. 1987. *On the definition of Word*. Cambridge: MIT Press.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Hale, Kenneth & Samuel Jay Keyser. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in honor of Sylvain Bromberger*, 53–109. Cambridge: The MIT Press.

Halle, Morris. 1959. *The sound pattern of Russian*. The Hague: Mouton.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* (4). 3–16.

Harris, Randy. 1993. *The linguistics wars*. Oxford: Oxford University Press.

Harris, Zellig. 1951. *Methods in structural linguistics*. Chicago: The University of Chicago Press.

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Deborah Schiffrin (ed.), *Meaning, form, and use in context: Linguistic applications*, 11–42. Washington, D. C: Georgetown University Press.

Jespersen, Otto. 1954. *A modern English grammar on historical principles*. London: Allan & Unwin.

Joos, Martin (ed.). 1958. *Readings in linguistics*. American Council of Learned Societies.

Keller, Evelyn Fox. 1983. *A feeling for the organism*. New York: Henry Holt & Company.

Kerleroux, Françoise. 1996. *La coupure invisible : études de syntaxe et de morphologie*. Lille: Presses Universitaires du Septentrion.

Lees, Robert B. 1960. The grammar of English nominalizations. *International Journal of American Linguistics* 26, Part 2.

Marantz, Alec. 2005. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review* 22. 426–445.

Marchand, Hans. 1966. Review of zimmer 1964. *Language* 42. 134–142.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation*. München: Beck.

Matthews, P. H. 1965. The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics* 1(2). 139–171.

Matthews, P. H. 1972. *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.

Pesetsky, David. 1995. *Zero syntax: Experiencers and cascades*. Cambridge: MIT Press.

Robins, R. H. 1959. In defense of WP. *Transactions of the Philological Society*. 116–144.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Payot.

Searle, John R. 1969. *Speech acts.* Cambridge: Cambridge University Press.

Trager, George L. & Henry Lee Smith. 1951. *An outline of English structure.* Washington, D. C.: American Council of Learned Societies.

Wells, Rulon S. 1947. Immediate constituents. *Language* 23(1). 81–117.

Wittgenstein, Ludwig. 1953. *Philosophical investigations.* Trans. by Elizabeth Anscombe. Third edition. Oxford: Basil Blackwell.

Zimmer, Karl. 1964. Affixal negation in English and other languages: An investigation of restricted productivity. Supplement to *Word* 20.2, Monograph 5.

Chapter 2

Lexemes, categories and paradigms: What about cardinals?

Gilles Boyé

Université Bordeaux-Montaigne & UMR5263 (CNRS)

In Word and Paradigm frameworks such as Network Morphology (Corbett & Fraser 1993) and Paradigm Function Morphology (Stump 2001), categories and lexemes are taken as granted and usually associated with an inflectional paradigm relevant for all the lexemes in a given category. In Section 2, we explore the status of French cardinals as lexemes based on the characteristic properties defined by Fradin (2003): i) abstraction over form-variation, ii) autonomous forms, iii) stable meaning, iv) belonging to a major category, v) open-ended set of units that can serve as input and/or output of morphology. We start with the simple cardinals and argue, following Saulnier (2008)'s discussion, that French cardinals fit all the lexemic criteria but (iv), belonging to a major category, and should be considered full lexemes even though they constitute a sub-category of determiner, a minor category in Fradin's terms. In Section 3, moving from simple cardinals to complex ones, we show that the idiosyncratic morphophonological properties of French cardinals plead for a morphological analysis rather than a syntactic one, giving an analysis of their construction as multi-layered compounds. In Section 4, we describe the inflectional paradigms of French cardinals as dependent on their rightmost element using the Right Edge mechanism introduced by Miller (1992) and Tseng (2003) for other phenomena in French. In the conclusion, we show that some complex cardinals have to be analyzed as multi-layered morphological compounds due to their morphophonological idiosyncrasies but this does not entail that all complex cardinal should be. The fact that syntactic combinations of French cardinals do not respect lexical integrity indicates that to some extent, complex cardinals are in the shared custody of morphology and syntax.

1 Introduction

In this paper, following the lead of Saulnier (2008, 2010), we explore the status of French cardinals and their place in Word and Paradigm frameworks, within theories of morphology focusing on lexemes as their fundamental unit. In general, this topic poses interesting problems for linguistic theories:

- Are they lexemes? To what category do they belong: determiners, nouns, adjectives?

- Are they built by syntax or in the lexicon?

- Is there an inflectional paradigm for cardinals? If so, where does it come from?

In Section 2, we explore the categorial status of simple cardinals. In Section 3, we argue that complex cardinals are lexemes, like simple cardinals, even though they constitute a subcategory of determiners.[1] We outline a syntagmatic analysis to create complex cardinals in morphology as compounds. In the last section, we propose an analysis of the inflectional paradigm of cardinals based on the Right Edge mechanism introduced by Miller (1992) and Tseng (2003) for other phenomena in French.

2 French cardinals: Lexemes?

In this section, we examine the lexical status of French cardinals.[2]

Following Fradin (2003: 102), we distinguish two types of atomic units in the lexicon: lexemes and grammemes. Lexemes are typically nouns, verbs, adjectives, adverbs, while grammemes are grammatical units such as prepositions, determiners, conjunctions. Fradin identifies the following characteristic properties of lexemes:

(1)   a. It is an abstract unit to which word-forms are related; this unit captures the variations across word-forms.
   b. It possesses a phonological representation which gives it prosodic autonomy.
   c. Its meaning is stable and unique.
   d. It belongs to a category and can have an argument structure.
   e. It belongs to an open-ended set and can serve as output and input of derivational morphology.

Whatever the analysis of French complex cardinals such as *vingt-et-un* '21', simple cardinals like *vingt* or *un* are underived and therefore have to be listed in the lexicon. In what follows, we argue that simple cardinals in French pattern with lexemes rather than grammemes.

In French, the simple cardinals are the elements listed in (2) that serve as cardinals and as building blocks for complex cardinals.[3]

(2)   *un* '1', *deux* '2', *trois* '3', *quatre* '4', *cinq* '5', *six* '6', *sept* '7', *huit* '8', *neuf* '9', *dix* '10', *onze* '11', *douze* '12', *treize* '13', *quatorze* '14', *quinze* '15', *seize* '16',

---

[1] This does not mean that all determiners are lexemes but rather that cardinals have to be treated as an exception.

[2] For complex cardinals, see Section 3.

[3] The elements *million* and *milliard* are not simple cardinals in French; their respective values are realized as *un million* ('one million') and *un milliard* ('one billion'). They semantically belong to the quantity noun series in *-aine* (see Table 2, p. 23)

> *vingt* '20', *trente* '30', *quarante* '40', *cinquante* '50', *soixante* '60',
> *cent* '100', *mille* '1,000'

Simple cardinals have the properties (1b–c). They can be used as single word answers, meaning they have an autonomous phonological representation. They have straightforward semantics, denoting counting values.

2.1 Form variation abstraction

As for property (1a), while *un* '1' is the only simple cardinal varying in gender (м: [œ̃] *un*, ꜰ: [yn] *une*), many simple cardinals are subject to *liaison* (linking), a morphosyntactic phenomenon whereby French words can change in form depending on the phonological properties of the following word. For example, in (3), the adjective ʙᴏɴ agrees in gender and number with the following noun, in both cases masculine and singular. But in a liaison context such as prenominally, the form bɔ̃ appears in (3a) in front of a word starting with a consonant (not a liaison trigger: ⊖) and the form bɔn appears in (3b) in front of a vowel-initial word (a liaison trigger: ⊕). Outside liaison context (⊘), adjectives assume the same form as in liaison context without trigger (⊘=⊖).[4]

(3) a. un bon collègue
         œ̃ bɔ̃⊖ kolɛg
         'a good colleague'

     b. un bon ami
         œ̃ bɔn⊕ ami
         'a good friend'

     c. bon à manger
         bɔ̃⊘ a mɑ̃ʒe
         'ready to eat'

Unlike adjectives, cardinals can have three different forms for the three contexts above.[5] For example, *six* '6' has different realizations (si, siz, sis) for the three contexts:

(4) a. in liaison context without a liaison trigger ⟹ si ⊖
         six souris
         si⊖ suʁi
         'six mice'

     b. in liaison context with a liaison trigger ⟹ siz ⊕
         six   écureuils
         siz⊕ ekyʁœj
         'six squirrels'

     c. not in liaison context ⟹ sis ⊘
         six à attraper
         sis⊘ a atʁape
         'six to catch'

---

[4]For more details about the morpho-syntactic aspects of liaison see Bonami et al. (2004).
[5]See Plénat (2008), Plénat & Plénat (2011) and the citations therein for a detailed description.

Not all cardinals have different forms in all three contexts. Table 1 gives the five different patterns of syncretism found with the simple cardinals. Type A cardinals are not sensitive to liaison and thus display only one form; in type B the ⊖ and the ⊘ are identical and the ⊕ has an additional consonant at the end, while in type C all three forms are distinct. In type D, ⊖ is overabundant with a long form and a short form, and the long form is also used in the two other contexts. Type E is a variant of type B where instead of having an additional consonant for ⊕, the final fricative alternates between voiceless f and its voiced counterpart v.[6]

Table 1: Type of simple cardinal variation according to liaison

Type	Example	⊖	⊕	⊘	Cardinals
A	4	katʁ	katʁ	katʁ	4, 7, 11, 12, 13, 14, 15, 16, 30, 40, 50, 60, 1000
B	2	dø	døz	dø	1, 2, 3, 20, 100
C	6	si	siz	sis	6, 10
D	5	sɛ̃/sɛ̃k	sɛ̃k	sɛ̃k	5, 8
E	9	nœf	nœv	nœf	9

The simple cardinals in (2) have an associated form paradigm for liaison, which fit Fradin's property (1a). This property is part of the conceptual definition of lexeme; it is neither required nor sufficient by itself. Definite determiners which have form paradigms in French and German are not considered lexemes, while English adjectives are lexemes even though their forms do not vary.

We turn now to the two remaining properties (1d–e): belonging to an open-ended category and participating as the output and potentially the input of derivational morphology.

2.2 Morphological input

In French, simple cardinals clearly serve as input for several morphological derivations as summarised in Table 2 below (see Saulnier 2008, Fradin & Saulnier 2009, Saulnier 2010 for a detailed discussion).[7]

As bases for the ordinals, simple cardinals are part of a morphological category in terms of Van Marle (1985) namely the derivational domain of ordinals, but to satisfy (1d), simple cardinals have to belong to a unique morphosyntactic category.

---

[6]In the case of type E, there is also hesitation for the ⊕ form between nœv and nœf as they can both provide an onset for the following trigger unlike in type B.

[7]While belonging to the same series of nouns designating groups of approximate cardinality, *millier* ('thousand'), *million* ('million'), *milliard* ('billion') are derived from *mille* with different suffixes (*-ier*, *-ion*, *-iard*).

Table 2: Some derivations on French cardinals (adapted from Fradin & Saulnier 2009: 201)

Suffix	Derivation	Category
-ième	*deux (2)* ⟶ *deuxième* ('second' ordinal)	Adj
-ième	*cinq (5)* ⟶ *cinquième* ('fifth' part)	Adj/N
-ain	*quatre (4)* ⟶ *quatrain* ('quatrain')	N
-aine	*douze (12)* ⟶ *douzaine* ('dozen')	N
-aire	*trente (30)* ⟶ *trentenaire* ('thirty-year-old')	Adj/N

2.3 Morphosyntactic category

Following Saulnier (2010), we consider simple cardinals to be a sub-category of indefinite determiners, CARD.

Saulnier (2010: 31–40) applies the discriminating contexts defined in Leeman (2004)'s work on French indefinite determiners. She shows that cardinals have the following distribution across the six diagnostic contexts.

(5)   *en* dislocation: + ⟶ il a *deux* solutions = il en a *deux*

'he has 2 solutions = he has 2'

only alone before N: – ⟶ mes *deux* livres (*mes *plusieurs* livres)

'my 2 books (*my several books)'

following the indefinite: – ⟶ *un *deux* livres (un *certain* livre)

'*a 2 books (a certain book)'

following the definite: + ⟶ les *deux* livres (*les *certains* livres)

'the 2 books (*the certain books)'

followed by the definite: – ⟶ *deux* les livres (*tous* les livres)

'*2 the books (all the books)'

followed by *de* NP: + ⟶ *deux* de mes collègues

'2 of my colleagues'

With these criteria in mind for the category CARD, it becomes clear that there are simple cardinals that were not listed in (2) because they do not participate in the formation of complex cardinals.

Zéro '0', for example, is not a construction unit for complex cardinals but it behaves like a CARD in all the contexts in (5). Saulnier (2010: 38) considered *zéro* to depart from the cardinals distribution because she could not find examples for the contexts in (6), expecting *zéro* to be singular.[8]

---

[8]In the same contexts, Saulnier does not examine *un* and the surprising plural number that arises when it follows a definite or a possessive. For example, in *pour ses/son **un** mois* 'for his one month anniversary', the masculine singular form of the possessive *son* is far less common than the plural *ses*; the possessive can take its plural form *ses* despite the presence of the cardinal *un* '1'.

(6)   Examples from the web (26/12/2016)

> *en* dislocation: + ⟶ Il a des tas d'contacts, des tonnes de numéros pour remplir son phone mais des vrais potes il en a *zéro*.[9]
>
> only alone before N: – ⟶ Et il ne nous restera alors que **nos**$_{pl}$ *zéro* euros d'augmentation pour pouvoir demander un crédit.[10]
>
> following the indefinite: – ⟶ *un *zéro* livre/livres.[11]
>
> following the definite: + ⟶ Je vote pour **les**$_{pl}$ *zéro* heures payées trente-cinq.[12]
>
> followed by *de* NP: + ⟶ Mais même les potes des autres viennent ici et *zéro* **de mes potes** sont venus me voir.[13]

> And contra Saulnier (2010), *zéro* also appears in *zéro*+N subject NPs:

> *zéro*+N subject: + ⟶ Pendant ce temps, ***zéro* personnes**$_{pl}$ sont mortes de surdoses de marijuana.[14]

In derivational morphology, *zéro* also gives a corresponding ordinal *zéroième* following the pattern of other simple cardinals.

2.4 Morphological output

Apart from fixed value cardinals, French uses variable cardinals such as *n* 'n' (pronounced [ɛn]) or *x* 'x' (pronounced [iks]). Like *zéro*, these variable cardinals do not participate in complex cardinal formation but they appear in the contexts in (5) and allow a subset of the derivations for fixed value cardinals (e.g. *énième* 'nth' pronounced [ɛnjɛm] and *xième* 'xth' pronounced [iksjɛm]).

(7)   a.   Une solution consiste à rechercher les *N* meilleures solutions pour chaque ville épelée.[15]

> b.   Donc l'installateur fait des bidouilles avec les *X* paramètres qui en [soi] ne sont pas très clairs ou pas forcément adaptés aux diverses situations des clients...[16]

---

[9]'He's got many contacts, tons of numbers to fill his phone, but real mates, he's got zero.'
https://genius.com/Enz-narcisse-and-cassandre-lyrics

[10]'Then we will only have our 0 euros of raise to ask for a credit.'
http://psasochaux.reference-syndicale.fr/files/2015/04/Tract-avril-15.pdf

[11]'*a zero book/books'

[12]'I'm voting for the 0 hours being paid as 35.'
https://fr.toluna.com/opinions/762230/Etes-vous-pour-ou-contre-les-35-heures

[13]'But while even the other guys' pals come here, 0 of mine have come to see me.'
https://twitter.com/MisHyding/status/762360289329307649

[14]'All this while, 0 persons have died of marijuana overdose.'
https://anarchocommunismelibertaire.wordpress.com/

[15]'A solution would be to search for the N best possibilities for every city name.'
http://www.afcp-parole.org/spip.php?article152

[16]'So the installer switches around the X parameters which are a bit obscure or not necessarily adapted to the various customer situations.'
https://www.bricozone.fr/t/reglage-chaudiere-viessman.11296/page-7

    c. Aujourd'hui, je constate que pour la *énième* fois, une voiture est garée devant mon entrée de garage, m'empêchant de sortir.[17]

These cardinals are obtained by converting letter names, usually French or Greek, to cardinals, making them the output of a morphological process and therefore fitting part of criterion (1e).

2.5 Open-ended set

In the general domain or in mathematical contexts this practice is limited to the conversion of a few letter names, but in computer programming names for integer-valued variables are created all the time and behave as simple cardinals , making CARD an open-ended category.[18] Even the derived ordinals appear in computer program descriptions.

(8)   a. Lance le son à partir de la *nbième* [ɛnbejɛm] seconde.[19]

       b. appFunc(NUM): Renvoie l'adresse de la *NUMième* fonction de la page courante[20]

    The preceding discussion shows that French simple cardinals are part of an open-ended set with the productive coinage of integer variables. As we have seen above, ordinal derivation takes simple cardinals as input and letter name conversion gives simple cardinals as output. These three observations indicate that French simple cardinals fit the property (1e).

2.6 Interim conclusion: the lexical status of simple cardinals

In this section, we have shown that simple cardinals in French have all the properties deemed characteristic of lexemes by Fradin (2003). Like typical lexemes, elements of CARD are created by borrowing and arbitrary coining while grammemes emerge through diachronic phenomena. Considering simple cardinals to be lexemes might seem at odds with the fact that we have taken them to be a sub-category of determiners, usually not regarded as a lexeme-based category. In the following section, we argue that CARD, in general, are a part of the syntactic category of determiners but constitute a morphological category of their own.

---

[17]'Today, for the nth time, I see a car parked in front of my garage door, blocking my way.'
https://goo.gl/lOrTuo
[18]Note that the French complex cardinals are not an open-ended set but rather a large set containing one trillion elements, as French speakers can count from 0 to 999,999,999,999.
[19]'Run the soundbite from the nbth second.'
http://www.forum-dessine.fr/index.php?id=06038
[20]'Returns the address of the NUMth function in the current page.'
https://goo.gl/LHh46c

3 French cardinals: Category?

In this section, we examine the status of French cardinals, simple and complex. We start with an overview of 'The Composition of Complex Cardinals' (Ionin & Matushansky 2006), as an example of a completely syntactic view of cardinal derivation. Then we argue that the phonological idiosyncrasies of complex cardinals are best modelled with a morpholexical system.
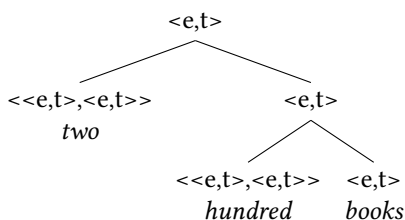
3.1 Complex cardinals in syntax

Ionin & Matushansky (2006: 316) argue that 'complex cardinals are composed entirely in syntax and interpreted by the regular rules of semantic composition'.

3.1.1 Semantics

Their analysis describes the semantics of complex cardinals and their syntax in several languages, focusing particularly on Russian. To allow for the semantic combination of Cards in CardP, they propose that simplex cardinals have the type <<e,t>, <e,t>> so that a series of simplex cardinal followed by a noun predicate of type <e,t> will be able to combine step by step with a parent simplex cardinal as in (9) and result in a type <e,t>.

(9)

```
                        <e,t>
                  _____/    _____
                 /                   \
         <<e,t>,<e,t>>              <e,t>
             two               ____/    \____
                              /              \
                      <<e,t>,<e,t>>         <e,t>
                         hundred            books
```

The actual semantic combination is not described in detail but the authors seem to rely on the packing strategy of Hurford (2007) where complex cardinals are analyzed based on the simple set of syntagmatic rules associated with calculations in (10). Figure 1 gives the corresponding structure for 5,002,600.

(10) · NUMBER $\longrightarrow$ $\begin{Bmatrix} \text{DIGIT} \\ \text{PHRASE (NUMBER)} \end{Bmatrix}$

        value(NUMBER) = value(PHRASE) + value(NUMBER)

    · PHRASE $\longrightarrow$ (NUMBER) M
      value(PHRASE) = value(NUMBER) × value(M)

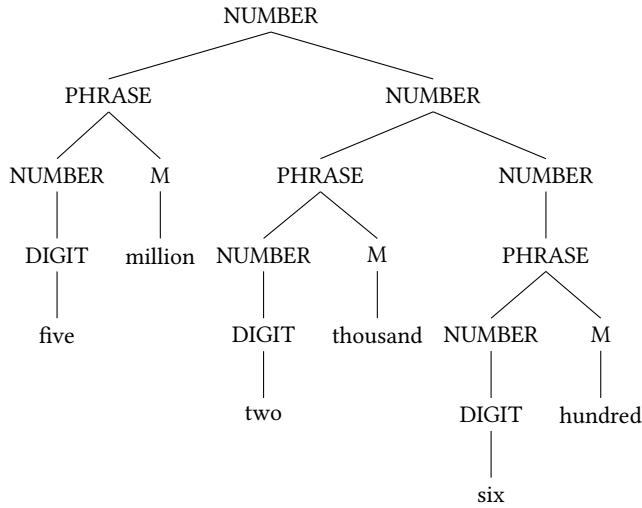Hurford describes the packing strategy as a constraint on the syntagmatic grammar in (10):

```
                              NUMBER
                 ┌──────────────┴──────────────┐
              PHRASE                          NUMBER
            ┌────┴────┐              ┌──────────┴──────────┐
        NUMBER       M           PHRASE                  NUMBER
          │          │         ┌────┴────┐                 │
        DIGIT      million  NUMBER       M               PHRASE
          │                   │          │             ┌────┴────┐
         five               DIGIT     thousand      NUMBER       M
                              │                        │         │
                             two                     DIGIT     hundred
                                                       │
                                                      six
```

Figure 1: Syntagmatic analysis of 5,002,600 from Hurford (2007)

- The sister constituent of a NUMBER must have the highest possible value.[21]

The semantic analysis proposed by Ionin & Matushansky (2006) does not warrant a syntactic view of complex cardinals. From an external perspective, it manages to treat complex cardinals and simple cardinals in the same manner, giving them the same semantic type and the same combinatorial constraints on the counted noun (atomicity and countability).

3.1.2 Syntax

Concerning syntax, Ionin & Matushansky (2006) describe two phenomena relevant to French cardinals: case assignment and number morphology. In Russian, cardinal-containing NPs do not realize the direct cases (nominative & accusative) the same way as other NPs. For example, the NPs in (11) could all be used as subjects or direct objects. In (11a), *šag* 'step' has the nominative/accusative plural form expected for a direct argument but in (11b) it has the genitive singular form (paucal in the terms of Ionin & Matushansky) and, in (11c), the genitive plural form.

(11)　a.　šag-i
　　　　　step-NOM.PL
　　　　　'steps'

　　　b.　četyre šag-á
　　　　　four　　step-GEN.SG
　　　　　'four steps'

---

[21]This constraint is intended to have the same effect as converting time in seconds into complex units such as days/hours/minutes/seconds, maximising the number of days first, then hours, minutes and finally seconds.

c. šest' šag-ov
six step-GEN.PL
'six steps'

The case and number appearing on the head noun depend on the last simple cardinal in CardP. Cardinal 1 does not interfere with direct cases, cardinals 2–4 assign genitive singular and the other cardinals assign genitive plural.

This phenomenon also happens inside CardP in multiplicative contexts such as (12). *Tysjača* '1,000' appears in the nominative singular alone, but in the genitive singular with 4 and in the genitive plural with 5.

(12) a. tysjač-a šag-ov
thousand-NOM.SG step-GEN.PL
'one thousand steps'

b. četyre tysjač-i šag-ov
four thousand-GEN.SG step-GEN.PL
'four thousand steps'

c. pjat' tysjač šag-ov
five thousand.GEN.PL step-GEN.PL
'five thousand steps'

The form variations above do not interfere with the external case and number. The case and number realized internally on the head noun and the multiplied cardinals in the CardP do not affect the case and number of the NP in its relation to the rest of the sentence.

French does not have an inflectional case system similar to Russian but cardinals still display similar properties. In syntax, the CARD category identified for morphology in section 2.3 opposes the cardinals ending with elements *million* and *milliard*, infelicitous in (13a), with all other cardinals infelicitous in (13b).[22]

(13) a. Paul a deux/cent/*un million euros à la banque.
'Paul has X euros in his account.'

b. Paul a *deux/*cent/un million **d**'euros à la banque.
'Paul has X **of** euros in his account.'

The data in (13) could be interpreted as a difference in category, *un million* being considered as a noun rather than a CARD. But while the use of *un million* changes the shape of the NP, it does not affect its external relations to the sentence, just as in Russian. It appears that *millions* and *milliard* assign genitive plural to the head noun resulting in

---

[22]This could be contrived as *million* and *milliard* being classifiers but their behavior in complex numerals shows that they are indeed cardinal construction elements.

(i) un milliard trois cents millions **d**'euros '1,300,000,000 euros'

(ii) un_M million une_F pages_F '1,000,001 pages'

a *de* NP without changing the overall distribution of the cardinal-containing NP. Both structures participate in the contexts (5) used by Saulnier (2010) repeated below.

(14) *en* dislocation: + ⟶ il en a *deux*/*un million*
      'he has 2/1,000,000'
      only alone before N: - ⟶ mes *deux* livres/mes *un million* de livres
      'my 2/1,000,000 books'
      following the definite: + ⟶ les *deux* livres/les *un million* de livres
      'the 2/1,000,000 books'
      followed by *de* NP: + ⟶ *deux*/*un million* de mes collègues
      '2 /1,000,000 of my colleagues'

Including *million*, *milliard* and their combinations in the CARD category with different controlling features captures the external similarity while retaining the appropriate contrast between the different NP structures CARD N vs CARD *de* N in the examples above.

French also displays number morphology inside complex cardinals , like Russian. The marks are visible in liaison contexts before triggers as shown in (15).

(15) a. *cent* ans
          sãt ã
          'one hundred years'

      b. deux *cents* ans
          dø   *sãz* ã
          'two hundred years'

      c. *vingt* ans
          vẽt ã
          'twenty years'

      d. quatre -*vingts* ans
          katʁə -*vẽz* ã
          'eighty years'

The ⊕ forms of simple cardinals *cent* and *vingt* end in t but their final consonant is replaced by z in multiplicative contexts.[23] This change does not seem to be mandated by plural marking as *cent* and *vingt* are already plural controllers.[24]

All in all, Ionin & Matushansky (2006) and Hurford (2007) provide an interesting framework in which to analyze French cardinals as a unique syntactic category. The differentiated control properties and the idiosyncrasic number morphology they propose

---

[23]In liaison contexts, the t-final ⊕ forms alternate with the ⊖ forms depending on collocations. Frequent ones such as *vingt ans* '20 years' and *cent ans* '100 years' are generally pronounced with ⊕ forms (vẽt⊕ã, sãt⊕ã), but rarer collocations like *vingt écureuils* '20 squirrels' and *cent écureuils* '100 squirrels' are often found with the ⊘ forms (vẽ⊘ekyʁœj, sã⊘ekyʁœj). But in any case, the emergence of a z-final ⊕ form outside multiplicative contexts is considered faulty: *vẽz⊕ekyʁœj, *sãz⊕ekyʁœj.

[24]Hurford (2003: Section 3) describes a case in Finnish were number marking on cardinals makes a difference. Plural cardinals count groups of N while singular cardinals count N individuals.

allows for a uniform syntactic analysis where all complex cardinals are constructed in the same way. However, the phonological aspects of French cardinals do not go along with the perfectly predictable semantics and syntax of the complex cardinals on which Ionin & Matushansky (2006) build their syntactic view of the process.

3.2 Complex cardinals and phonology

From a phonological standpoint, idiosyncrasies are everywhere in the construction of French complex cardinals. In the following we review the various combinatorial exceptions in the formation of complex cardinals and argue that it would be difficult to account for these with a purely syntactic analysis.

As we have seen in section 2.1, French simple cardinals are subject to form variation according to liaison contexts. In the derivation of complex cardinals, however, simple cardinals use the same forms but in quite different distributions. For example, *vingt* '20' and *cent* '100' belong to the same type B in Table 1, p. 22: both combine with simple cardinals 2–9, but *vingt* uses the ⊕ form vɛ̃t[25] even though these cardinals are not liaison triggers, while *cent* uses the ⊖ form sɑ̃ in the same context, as shown in Table 3.

Table 3: *vingt* and *cent* combinations with simple cardinals from 2 to 9

	2	3	4	5	6	7	8	9
20⊕	vɛ̃t-dø	vɛ̃t-tʁwa	vɛ̃t-katʁ	vɛ̃t-sɛ̃k	vɛ̃t-sis	vɛ̃t-sɛt	vɛ̃t-ɥit	vɛ̃t-nœf
100⊖	sɑ̃-dø	sɑ̃-tʁwa	sɑ̃-katʁ	sɑ̃-sɛ̃k	sɑ̃-sis	sɑ̃-sɛt	sɑ̃-ɥit	sɑ̃-nœf

Combinations involving *cinq* '5' and *huit* '8' in the construction of multiples of 100 and 1000 are not parallel even though they belong to the same type D of simple cardinals in Table 1, with two alternating realisations for the ⊖ form: sɛ̃/sɛ̃k , ɥi/ɥit. With *cinq* both of the ⊖ forms can be used in the combinations but with *huit* only the short ⊖ form ɥi is felicitous:

(16) a. 500 sɛ̃-sɑ̃/sɛ̃k-sɑ̃ , 5000 sɛ̃-mil/sɛ̃k-mil

   b. 800 ɥi-sɑ̃/*ɥit-sɑ̃ , 5000 ɥi-mil/*ɥit-mil

Moreover, the same simple cardinal *dix* '10' combines with 7–9 and with 1000, none of which are liaison triggers, but it uses the ⊕ form in the first case and the ⊖ in the second:

(17) a. 17 diz-sɛt , 18 diz-ɥit , 19 diz-nœf

   b. 10000 di-mil

Finally, instances of *quatre-vingt* have to be pronounced with an r at the end of *quatre*, even for speakers who usually drop it in word-final complex codas.

---

[25]Note that this holds true independently of the fact that the ∅ form of 20 is subject to diatopic variation between vɛ̃ and vɛ̃t .

(18) a. un arbre frappé par la foudre
          ɛ̃n aʁbʁə fʁape paʁ la fudʁə = ɛ̃n aʁb fʁape paʁ la fud

      b. vingt-quatre francs
          vɛ̃tkatʁə fʁɑ̃ = vɛ̃tkat fʁɑ̃

      c. quatre-vingts francs
          katʁəvɛ̃ fʁɑ̃ ≠ *katvɛ̃ fʁɑ̃[26]

We conclude that even though both the semantic and syntactic dimensions of complex cardinal formation are simple and regular, the combinatory principles at work at the phonological level are far from simple and must be specific to cardinal formation, leading us away from syntax and towards a lexical account of the derivation of complex cardinals.

3.3 Complex cardinals in CARD

As complex cardinals have the same distribution in the Saulnier-Leeman contexts in (5) and serve as input for the ordinal derivation, we analyze numerical cardinals as compounds created by means of a phrase structure grammar similar to those proposed by Hurford (1975, 1994, 2003, 2007). The analysis will be presented in two parts. We first introduce a model limited to the structure of 2-digit cardinals where most of the phonological and syntagmatic idiosyncrasies occur and then generalize it to the rest of the cardinals.

3.3.1 2-digit cardinals

Cardinal components are categorized according to their combinatorial properties (Table 4). To demonstrate the mechanics of the analysis, we use arbitrary categories rather than motivated features to differentiate elements. The category names reflect their purpose in the system. Unit categories start with u for digits (u, u1, u4, u7) and uv (uv, uv1) for units under 20, while categories for multiples of ten begin with d (d, d1, d2, d6).[27]

  The rules in Table 5 generate all 2-digit cardinals (category Digit2). Rule 1 states that simple cardinals are de facto Digit2. Rule 2 generates *dix-sept*, *dix-huit*, *dix-neuf*. Rules 3 and 5 assemble *et un* and *et onze*. Rule 4 produces DixP for number between *vingt* '20' and *cinquante-neuf* '59'.[28] Rule 6 makes the *soixante* compounds from *soixante* '60' to *soixante-dix-neuf* '79' and rules 7 to 9 create the compounds based on *quatre-vingt* for number between *quatre-vingts* '80' and *quatre-vingt-dix-neuf* '99'.[29] Finally, rule 10 elevates all intermediary compounds to Digit2.

---

[26]katvɛ̃ is correct, however, for the decimal number '4.20'.

[27]To account for the Swiss and Belgian cardinal systems, the category d would have to include *septante* '70', *octante/huitante* '80' and *nonante* '90'.

[28]In rule 4, the ⊕ form is selected for the first term: d2.⊕=vɛ̃+t

[29]In rule 7, the ⊖ form is selected for the first term, Dix8X.⊖=katʁəvɛ̃. In rule 9, The liaison consonant for d2 changes to z, ⊕ becomes vɛ̃+z.

Table 4: Categories of cardinal components for 2-digit cardinals

Cat	Components	Example	
u	deux (2), trois (3), cinq (5), six (6)	vingt-**deux**	20+**2**=22
u1	un (1)	vingt-et-**un**	20&**1**=21
u4	quatre (4)	**quatre**-vingts	**4**x20=80
u7	sept (7), huit (8), neuf (9)	dix-**sept**	10+**7**=17
uv	douze (12), treize (13), quatorze (14) quinze (15), seize (16)	soixante-**douze**	60+**12**=72
uv1	onze (11)	soixante-et-**onze**	60&**11**=71
d	trente (30), quarante (40), cinquante (50)	**trente**-deux	**30**+2=32
d1	dix (10)	*****dix**-deux≠douze	**10**+2≠12
		soixante-**dix**	60+**10**=70
d2	vingt (20)	quatre-**vingts**	4x**20**=80
d6	soixante (60)	**soixante**-treize	**60**+13=73
et	et (&)	trente-**et**-un	30**&**1=31

Table 5: Syntagmatic rules for 2-digit cardinals

	Rule	Comment
1	Digit2 ⟶ u/u1/u4/u7/uv/uv1/d/d1/d2/d6	simplex cardinals
2	Dix1P ⟶ d1.⊕ u7	diz (10) for 10+7..9
3	Et1 ⟶ et u1	eœ̃ (&1) for 20/30/40/50&1
4	DixP ⟶ d/d2.⊕ u/u4/u7/Et1	20/30/40/50+2..9/&1
5	Et11 ⟶ et u1/uv1	eœ̃/eɔ̃z (&1/&11) for 60&1/11
6	DixP ⟶ d6 u/u4/u7/d1/Et11/uv/Dix1P	60+2..10/12..16/(10+7..9)/&(1/11)
7	Dix8X ⟶ u4 d2.z	4x20 for 80
8	DixP ⟶ Dix8X	complex 80
9	DixP ⟶ Dix8X.⊖ u/u1/u4/u7/d1/uv/uv1/Dix1P	80+1..16/(10+7..9)
10	Digit2 ⟶ DixP	complex cardinals

The syntagmatic rules in Table 5 integrate constraints stipulating the combining forms:

(19)   a.   Rules 2 and 4 use the linking form ⊕ of the first component;

          b.   Rule 7 changes the liaison consonant of the second component from t to z;

          c.   Rule 9 uses the ⊖ form of the first component.

Figure 2 illustrates the application of rules 1–4, and more particularly the way diz-sɛt and vɛ̃t-sɛt are obtained with d1.⊕ diz and d2.⊕ vɛ̃t .
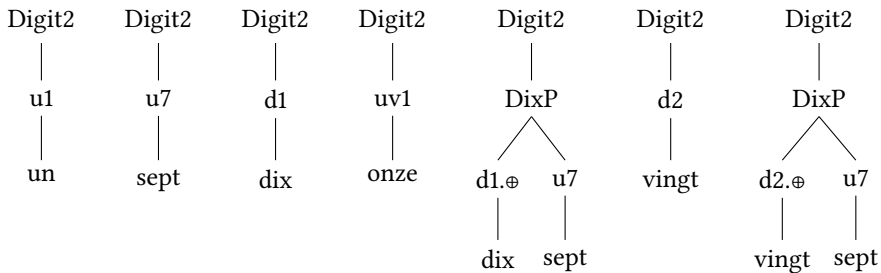
Figure 2: Phrase structures for 1, 7, 10, 11, 17, 20, 27

Figure 3 shows how *et onze* '& 11' and intermediary compounds such as *dix-sept* '17' are combined with *soixante* '60'.
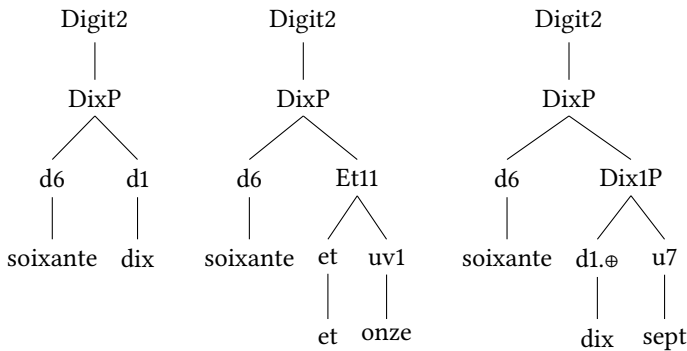
Figure 3: Phrase structures for 70, 71, 77

Finally, Figure 4 displays the combinations involving the *quatre-vingt* intermediary compound. When Dix8X is formed, the linking consonant of *vingt* is changed from t to z, but when the Dix8X is itself combined with another element by means of rule 9, its ⊖ form is selected rendering the previous change invisible. Thus we obtain the ⊕ form katʁə-vɛz for *quatre-vingts* '80' and the forms katʁə-vɛ̃-sɛt and katʁə-vɛ̃-diz-sɛt for *quatre-vingt-sept* '87' and *quatre-vingt-dix-sept* '97'.
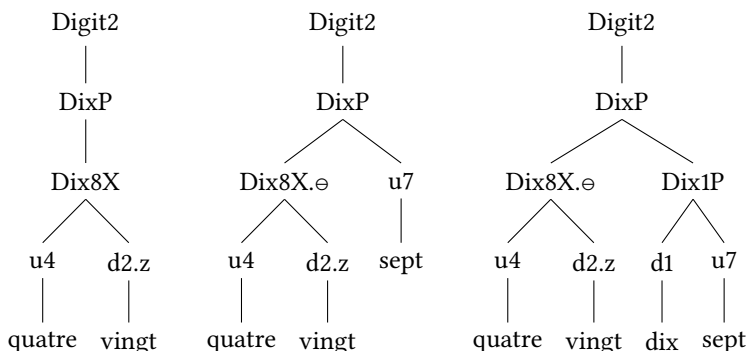
Figure 4: Phrase structures for 80, 87, 97

Even though we provide rules for all Digit2 cardinals in Table 5, most of these compounds are probably lexicalized. The rules are like redundancy generalizations *à la* Lieber (1982) or Koenig (1999), stating observable regularities in existing lexemes.

3.3.2 Numerical cardinals

With most of the idiosyncrasies residing below 100, the fragment in Table 6[30] for the composition of the higher combinations is simpler. It breaks the compounding into four levels corresponding to the counting units *cent* '100', *mille* '1,000', *million* 'million', and *milliard* 'billion'. Each level is composed of two rules, one to multiply the unit level and one to add the units from the level below.

Table 6: Syntagmatic rules for 3-digit+ cardinals

	Rule	Comment
11	CentX ⟶ u/u4/u7.⊖ Cent.z	hundreds
12	CentP ⟶ CentX/Cent.⊖ (Digit2)	adding the Digit2
13	MilleX ⟶ CentP/Digit2P.⊖ Mille	thousands
14	MilleP ⟶ MilleX/Mille (CentP/Digit2)	adding the hundreds
15	MionX ⟶ CentP/Digit2.⊖ Mion.z	millions
16	MionP ⟶ MionX.⊖ (MilleP)	adding the thousands
17	MiardX ⟶ CentP/Digit2.⊖ Miard.z	billions
18	MiardP ⟶ MiardX.⊖ (MionP)	adding the millions

For example, rule 11 assembles the multiples of *cent* '100' and rule 12 adds the units

---

[30]We found no critical data for or against adding a linking z to the ⊕ form of multiplied *million* and *milliard*, rules 15 and 17.

from the level Digit2.[31] In rules 12, 16 and 18, the selection of the ⊖ form[32] happens only in the presence of the optional second term.

Figure 5 shows how the two sets of rules combine in the analysis of numerical cardinals in general.
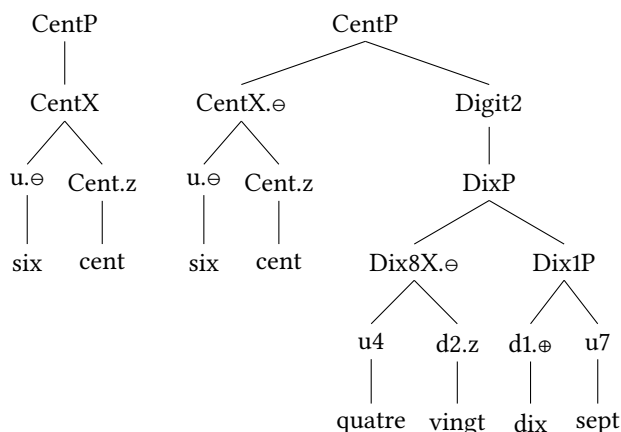
Figure 5: Phrase structure for 600 and 697

The analysis presented here relies on 26 combination elements, the 23 in (2) plus *et*, *million* and *milliard*. All numerical cardinals, including the simple ones, are derived from these elements. So, on the one hand, cardinal elements belong to special categories in the lexicon while, on the other hand, all numerical cardinals, including the simple ones, are CARDs derived from cardinal elements.

4 French cardinals: Paradigm?

In this section, we propose an analysis for a uniform paradigm of simple and complex cardinals. The analysis combines the observations about gender, liaison and compounding to (i) give a set of rules that fills the cells of the paradigm with the appropriate forms and (ii) associate each numerical cardinal with its proper syntactic frame.

As lexemes belonging to the CARD category, French cardinals, simple and complex, undergo inflection with a paradigm based on two features:

- LIAISON: ⊖, ∅, ⊕

- GENDER: M, F

---

[31]These two rules could be modified to generate the 11 to 19 multiples of *cent* (e.g. *dix-huit cents* '1,800'). The rest would also have to be adapted to avoid the generation of aberrations such as **un million dix-huit cents mille* '2,800,000'.

[32]To be more precise, rules 16 and 18 select the M.⊖ form (i.e œ̃ for *un*).

This results in the six-cell paradigm exemplified in Table 7 with simple cardinals.

Table 7: Uniform paradigm of cardinals

'1'	M	F	'2'	M	F	'4'	M	F	'6'	M	F
⊖	œ̃	yn	⊖	dø	dø	⊖	katʁ	katʁ	⊖	si	si
⊘	œ̃	yn	⊘	dø	dø	⊘	katʁ	katʁ	⊘	sis	sis
⊕	œ̃n	yn	⊕	døz	døz	⊕	katʁ	katʁ	⊕	siz	siz

The paradigm of complex cardinals follows the pattern of the rightmost element in the compound. For example, in Table 8, TRENTE-ET-UN, QUATRE-VINGT-UN and CENT-UN share the pattern of UN, and TRENTE-SIX, QUATRE-VINGT-SIX and CENT-SIX inflect like SIX.

The only exception are *vingt* '20' and *cent* '100', which change their linking consonant from t to z in rules 7 and 11 (p. 32 & p. 34).

Not only do the forms of complex cardinals depend on the element on the right edge, but their controlling properties are also derived from the right edge element. This distin-

Table 8: Inflection on the Right Edge

'31'	M	F	'81'	M	F	'101'	M	F
⊖	tʁɑ̃teœ̃	tʁɑ̃teyn	⊖	katʁəvɛ̃œ̃	katʁəvɛ̃yn	⊖	sɑ̃œ̃	sɑ̃yn
⊘	tʁɑ̃teœ̃	tʁɑ̃teyn	⊘	katʁəvɛ̃œ̃	katʁəvɛ̃yn	⊘	sɑ̃œ̃	sɑ̃yn
⊕	tʁɑ̃teœ̃n	tʁɑ̃teyn	⊕	katʁəvɛ̃œ̃n	katʁəvɛ̃yn	⊕	sɑ̃œ̃n	sɑ̃yn
'36'	M	F	'86'	M	F	'106'	M	F
⊖	tʁɑ̃tsi	tʁɑ̃tsi	⊖	katʁəvɛ̃si	katʁəvɛ̃si	⊖	sɑ̃si	sɑ̃si
⊘	tʁɑ̃tsis	tʁɑ̃tsis	⊘	katʁəvɛ̃sis	katʁəvɛ̃sis	⊘	sɑ̃sis	sɑ̃sis
⊕	tʁɑ̃tsiz	tʁɑ̃tsiz	⊕	katʁəvɛ̃siz	katʁəvɛ̃siz	⊕	sɑ̃siz	sɑ̃siz

Table 9: Number morphology on the Right Edge

20	M	F	80	M	F
⊖	vɛ̃	vɛ̃	⊖	katʁəvɛ̃	katʁəvɛ̃
⊘	vɛ̃	vɛ̃	⊘	katʁəvɛ̃	katʁəvɛ̃
⊕	vɛ̃t	vɛ̃t	⊕	katʁəvɛ̃z	katʁəvɛ̃z
100	M	F	200	M	F
⊖	sɑ̃	sɑ̃	⊖	døsɑ̃	døsɑ̃
⊘	sɑ̃	sɑ̃	⊘	døsɑ̃	døsɑ̃
⊕	sɑ̃t	sɑ̃t	⊕	døsɑ̃z	døsɑ̃z

guishes cardinals ending in *million/milliard* from the others as seen below in (20) and in (13) (p. 28).

(20)   a.   un milliard trois cents millions cinq cent *mille* chinois/*de chinois
            '1,300,500,000 Chinese'
       b.   un milliard trois cent *millions* *chinois/de chinois
            '1,300,000,000 Chinese'

Cardinals ending with *million* 'million' or *milliard* 'billion' impose a *de*-NP structure. We use a DE feature to encode this difference: DE = + for (20b), DE = − for (20a).

Both the DE feature and the inflectional paradigm of compound cardinals can be constructed using the Right Edge mechanism introduced by Tseng (2003) and Bonami et al. (2004) to model French phrasal affixes (*à* 'at', *de* 'of') and liaison. The proposed mechanism ensures that the properties of the rightmost element are propagated to the top of the construction by copying the relevant features of the last component to its parent node at every level of compounding represented by the arrows in Figure 6. Rules combining two elements get a specific form from the left paradigm and prefix it to the paradigm on the right.
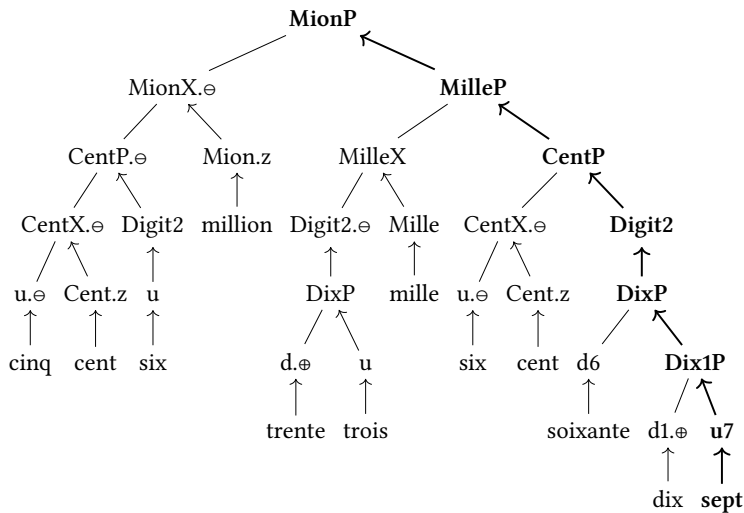
Figure 6: Phrase structure for 506,033,677

For example, on the right side, in (20a), the Dix1P prefixes the M.⊕ form of DIX diz to all forms of SEPT and carries the controlling property DE = − from SEPT. In (20b), the combination selects the M.⊖ form of SIX si and combines it with the modified paradigm of CENT where the linking consonant of the ⊕ forms t has been changed to z .

(21) a.

	'10'			'7'			'17'	
	M	F		M	F		M	F
⊖	di	di	⊖	*sɛt*	*sɛt*	⊖	dizsɛt	dizsɛt
⊘	dis	dis	⊘	*sɛt*	*sɛt*	⊘	dizsɛt	dizsɛt
⊕	**diz**	diz	⊕	*sɛt*	*sɛt*	⊕	dizsɛt	dizsɛt

DE = −   *DE = −*   DE = −

b.

	'6'			'100'			'600'	
	M	F		M	F		M	F
⊖	**si**	si	⊖	*sã*	*sã*	⊖	sisã	sisã
⊘	sis	sis	⊘	*sã*	*sã*	⊘	sisã	sisã
⊕	siz	siz	⊕	*sãz*	*sãz*	⊕	sisãz	sisãz

DE = −   *DE = −*   DE = −

The percolations proceed level by level, and yield a structure at the top with a full paradigm and the appropriate value of the DE feature.[33]

The model outlined here relies on the propagation of ready-made elementary paradigms via a phrase structure grammar rather than rules of exponence or referral based on the inflectional features of the different cardinals as is common with Word and Paradigm syntagmatic frameworks[34] such as A-Morphous Morphology (Anderson 1992), Paradigm Function Morphology (Stump 2001) or the Information-Based Model of Bonami & Crysmann (2013). It is more in line with paradigm-oriented models like Network Morphology (Corbett & Fraser 1993).

5 Conclusion

In this chapter, we set out to discuss the place of cardinals in French morphology with a focus on their status as lexemes, their categories and their inflectional paradigms. Taking into account the number of phonological idiosyncrasies in the formation of French cardinals, we argued that they should be considered as lexemes. Following Saulnier (2008, 2010), Fradin & Saulnier (2009), we examined both their morphotactic properties and their syntactic distribution and concluded that they belonged to a morphosyntactic category CARD inside the determiners. We showed that there are two types of cardinals regarding the way they associate with nouns, the direct type like CINQUANTE-DEUX (*cinquante-deux années*) and the indirect type like UN-MILLIARD-TROIS-CENTS-MILLIONS (*un milliard trois cents millions d'années*). This distinction making no difference on the

---

[33]

'506,033,677'

	M	F
⊖	sɛ̃ksãsimiljɔ̃trãtʁwamilsisãswasãtdizsɛt	sɛ̃ksãsimiljɔ̃trãtʁwamilsisãswasãtdizsɛt
⊘	sɛ̃ksãsimiljɔ̃trãtʁwamilsisãswasãtdizsɛt	sɛ̃ksãsimiljɔ̃trãtʁwamilsisãswasãtdizsɛt
⊕	sɛ̃ksãsimiljɔ̃trãtʁwamilsisãswasãtdizsɛt	sɛ̃ksãsimiljɔ̃trãtʁwamilsisãswasãtdizsɛt

DE = −

[34]See Boyé & Schalchli (2016) for a typology of views on inflectional paradigms in different theories.

outside of the NP, we analyzed them as compounds based on 26 simple elements[35] using a phrase structure grammar, even though the cardinals below 100 are probably lexicalized. Our compounding mechanism propagates the inflectional and syntactic properties of the rightmost component to the entire compound to create its paradigm and percolate its type (DE = ±).

The type of compounds we advocate for is different from the usual two-component ones. It expands the ternary compounds described in the biomedical domain by Namer (2005) to higher levels of composition. The extended compounding mechanism allows to generate all numerical cardinals as CARD without having to cast them into the different subcategories that would be needed to break the compounding process into binary operations. It does not presuppose that complex cardinals are lexicalised but only that they can be created online by morphology, as [+morphological, -lexical] compounds in the sense of Gaeta & Ricca (2009).

The model outlined here should be integrated with the formal analysis of Bonami et al. (2004) of liaison in HPSG (Pollard & Sag 1994). It would be interesting to examine data from the cardinals in other languages to parallel the work of Stump (2010) on the ordinals[36] and from the composition of the decimals and its interference with the integers.[37]

5.1 Remaining questions

Cardinal coordinations do not respect lexical integrity. Examples like (21a) are common, and even stranger coordinations appear with ordinals where the first ordinal is realised as a gender-agreeing cardinal as in (21b).

(21)   a.   Quelques *soixante-dix* ou *quatre-vingt* mille personnages sont passés à la trappe, 35 000 sont à l'ombre.[38]

       b.   Ses débuts, il les fit, dans sa ville natale, au début du siècle dernier dans sa vingt-et-*une* ou vingt-deux*ième* année.[39]

Saulnier (2008) observes that *quelques* follows the syntactic distribution of CARDs and derives the *quelquième* ordinal found in *trente et quelquième* 'thirty-somethingth'.[40]

---

[35]Nothing would prevent French from using more elements. In fact, it has been proposed since the 15th century to expand the counting system by including *billion*, *trillion*, *quadrillion*, etc. (see Saulnier 2010: 147–151 for an overview of the proposals).

[36]French ordinals are derived from their cardinal counterparts by *-ième* suffixation as proposed by Stump (2010: p. 228) with the notable exceptions of *millionième* and *milliardième* which drop the *un* from *un million* and *un milliard*.

[37]Many ill-formed cardinals are in fact well-formed decimals. For example, *cinq vingt* is automatically understood as '5.20'. Furthermore, *un million un* '1,000,001', when not followed by a counted noun, is usually perceived as '1,100,000' with *million* interpreted as a measure unit.

[38]'Some seventy or eighty thousand persons have disappeared, 35,000 are in jail.'
http://plumenclume.org/blog/173-erdogan-consolide-son-emprise-par-israel-adam-shamir

[39]'His debut, he made at the beginning of last century, when he was in his twenty-first or twenty-second year.'
http://www.www.dutempsdescerisesauxfeuillesmortes.net/fiches_bio/darbon/darbon.htm

[40]Fradin & Saulnier (2009) also mention *combien/combientième* 'how many', *quel/quellième* 'which' as potential cardinal/ordinal pairs (*quantième/tantième* look more like fractions than ordinals).

The arguments developed in this chapter for a morphological analysis of the composition of cardinals rely on the idiosyncrasies of complex cardinals below 100. To capture the phenomenon in (21), it would be possible to propose a morphological analysis of lower complex cardinals as compounds and lexemes, while still allowing syntactic composition for higher complex cardinals.

Acknowledgements

References

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Bonami, Olivier, Gilles Boyé & Jesse Tseng. 2004. An integrated approach to French liaison. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Wintner (eds.), *Proceedings of Formal Grammar 2004*, 29–45. Nancy: CSLI Publications.

Bonami, Olivier & Berthold Crysmann. 2013. Morphotactics in an information-based model of realisational morphology. In Stefan Müller (ed.), *Proceedings of HPSG 2013*, 27–47. Stanford: CSLI Publications.

Boyé, Gilles & Gauvain Schalchli. 2016. The status of paradigms. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge handbook of morphology*, 206–234. Cambridge: Cambridge University Press.

Corbett, Greville G. & Norman M. Fraser. 1993. Network morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29(01). 113–142.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Fradin, Bernard & Sophie Saulnier. 2009. Les cardinaux et la morphologie constructionnelle du français. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 199–230. Saint-Denis: Presses Universitaires de Vincennes.

Gaeta, Livio & Davide Ricca. 2009. Composita solvantur: Compounds as lexical units or morphological objects. *Rivista di Linguistica* 21(1). 35–70.

Hurford, James R. 1975. *The linguistic theory of numerals*. Vol. 16. Cambridge: Cambridge University Press.

Hurford, James R. 1994. Typology of numeral-noun interaction. In *Proceedings of the Edinburgh linguistics department conference*, vol. 94, 33–40.

Hurford, James R. 2003. The interaction between numerals and nouns. In Frans Plank (ed.), *Noun phrase structure in the languages of Europe*, 561–620. The Hague: Mouton de Gruyter.

Hurford, James R. 2007. A performed practice explains a linguistic universal: Counting gives the packing strategy. *Lingua* 117(5). 773–783.

Ionin, Tania & Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23(4). 315–360.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

Leeman, Danielle. 2004. *Les déterminants du nom en français: syntaxe et sémantique*. Paris: Presses Universitaires de France.

Lieber, Rochelle. 1982. *On the organization of the lexicon*. Bloomington: Indiana University Linguistics Club.

Miller, Philip. 1992. *Clitics and constituents in phrase structure grammar*. New York: Garland.

Namer, Fiammetta. 2005. Guessing the meaning of neoclassical compounds within LG: The case of pathology nouns. In Pierrette Bouillon & Kyoko Kanzaki (eds.), *Generative approaches to the lexicon*, 175–184.

Plénat, Marc. 2008. La liaison "obligatoire" avec et sans enchaînement. In Jacques Durand, Benoît Habert & Bernard Laks (eds.), *Actes du premier Congrès Mondial de Linguistique Française*, 1657–1667. Paris: Institut de Linguistique Française.

Plénat, Marc & Camille Plénat. 2011. La liaison de l'adjectif sur le nom en français : morphologie, syntaxe, phonologie. *Linguistica* 51(1). 299.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Saulnier, Sophie. 2008. *Les nombres: lexique et grammaire*. Université Paris Ouest Nanterre La Défense dissertation.

Saulnier, Sophie. 2010. *Les nombres: lexique et grammaire* (Rivages Linguistiques). Rennes: Presses Universitaires de Rennes.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2010. The derivation of compound ordinal numerals: Implications for morphological theory. *Word Structure* 3(2). 205–233.

Tseng, Jesse. 2003. Edge features and French liaison. In Jong-Bok Kim & Stephen Wechsler (eds.), *Proceedings of the 9th international conference on HPSG*, 313–333. Stanford: CLSI Publications.

Van Marle, Jaap. 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.

Chapter 3

Word formation and word history: The case of CAPITALIST and CAPITALISM

Franz Rainer

WU Vienna

The treatment of the history of modern vocabulary in historical and etymological dictionaries is generally disappointing, especially with respect to the processes by which the words came into being. The *TLFi*[1] only provides the following information concerning the history of French *capitalisme* and *capitaliste*: "**Capitalisme** [...] Dér. de *capital*²*; suff. *-isme**", "**Capitaliste** [...] Dér. de *capital**; suff. *-iste**". Such a treatment, which is inadequate even from a synchronic point of view (in the sense 'a supporter of capitalism', *capitaliste* is derived from *capitalisme* by affix substitution), does not do justice to the manifold relationships that have developed between these two words and their common base *capital* in the course of the 300 years since the creation of Dutch *Capitalist* in 1621. The present paper retraces in detail the many steps of the unfolding of these two words in French. It is shown that each of their many senses constitutes a separate lexeme and must be provided with an etymology of its own. Particular attention is dedicated to the identification of the exact mechanism (borrowing, semantic extension, word formation) that was at work at each step.

1 In the beginning was the lexeme

Right from the beginning of the study of the internal structure of complex words, scholars have been divided between those who tried to put complex words together from smaller pieces in a bottom-up fashion (the Pāṇinian tradition) and those who tried to account for the internal structure by mapping words onto other words (the Greco-Roman tradition, based on analogy). This fundamental divide is still with us, in the form of an opposition between what we now call "morpheme-based" and "word-based" (or "lexeme-based") approaches to morphology (see Aronoff 2007). In the French linguistic landscape, the morpheme-based approach held some sway before the turn of the millennium due to having been embraced by Danielle Corbin (see Corbin 1987), who played an important role in the renewal of the study of word formation in France. But more recently

---

[1] *Trésor de la langue française informatisé*, available at http://atilf.atilf.fr/.

most French morphologists seem to be quite unanimous in preferring the lexeme-based approach, not least due to the forceful argumentation in its favour in Fradin (2003).

In my contribution, I would like to pour more water on the lexeme-based mill by looking in some detail into the history of the two words CAPITALIST and CAPITALISM, in which semantic change, calques and word formation – suffixation, conversion, but also suffix substitution, a notorious conundrum for morpheme-based approaches – have interacted in a complex manner. It will become apparent that these changes find a natural explanation within a lexeme-based framework, while they seem to be difficult to accommodate without contortions in a morpheme-based one. However, the chapter is meant to be of interest not only to morphologists or lexicologists, who constitute the main intended readership. Both words treated are key concepts of present-day intellectual vocabulary and as such have attracted considerable attention from scholars from other disciplines, mostly historians such as Fernand Braudel, Lucien Febvre, Henri Hauser or Edmond Silberner in France, or Richard Passow, Marie-Elisabeth Hilger and Annette Höfer in Germany. For such readers, the linguistic arguments of this contribution may sometimes seem to be a little far-fetched, while they would probably here and there like to receive more abundant encyclopedic information. This latter type of information, however, must be kept to a minimum here, providing just what is necessary for underpinning the linguistic argumentation. Even so, non-linguists will hopefully appreciate the new facets of the history of these two words, which I was able to add to the existing documentation due to the abundance of new material that we can now dip into thanks to Google Books and Gallica.[2]

In order to avoid misunderstandings, one formal proviso is in order before we start our investigation. It is established practice in linguistics to write lexemes in small caps. In this tradition, the English lexeme CAPITALIST would represent the set of English word forms { *capitalist*, *capitalists* }. I will not follow this usage here, but use small caps instead whenever referring to a word independently from its exact formal realization in individual European languages. Throughout this text, CAPITALIST therefore represents the set {English *capitalist*, German *Kapitalist*, French *capitaliste*, etc.}, and similarly for other words in small caps.

2 CAPITAL, CAPITALIST and CAPITALISM in synchrony and diachrony

For present-day speakers of European languages, CAPITALISM refers to a specific kind of economic system and is undoubtedly felt to be based somehow on CAPITAL, though many speakers will be hard-pressed to specify the exact semantic relationship between base and derivative or will construe it in different ways. This indeterminacy is mainly

---

[2]On the history of CAPITALISTE, see Rainer (1998). A short, updated entry on the history of French *capitaliste*, written together with Jean-Paul Chauveau, can be found on *TLF-Étym*, an etymological online dictionary that can be consulted at http://www.atilf.fr/tlf-etym/. The corresponding entry on French *capitalisme* can be consulted on the same site.

due to the fact that the word CAPITAL itself has various senses, not all of them equally familiar to non-economists, and that it is not obvious which sense is the relevant one for the construal of the meaning of CAPITALISM. The *Free Dictionary*,[3] for example, manages to define CAPITALISM without recourse to CAPITAL: "An economic system in which the means of production and distribution are privately or corporately owned and development occurs through the accumulation and reinvestment of profits gained in a free market." CAPITALIST, on the contrary, will most often be spontaneously analyzed as based on CAPITALISM, referring to a supporter of the particular kind of economic system denoted by this word. 'A supporter of capitalism', in fact, is the first sense in the online dictionary quoted above, which adds two more senses that seem to be less prominent today: 2. 'An investor of capital in business, especially one having a major financial interest in an important enterprise'; 3. 'A person of great wealth'. The foregoing remarks seem to be valid for European languages in general. In other respects, however, individual languages differ, for example, with respect to whether they tolerate the adjectival usage of CAPITALIST, possible in French and English, but not in German. The connotations of the members of this word family will also differ, depending on the stance that a speaker or speech community takes with respect to the economic system called CAPITALISM.

The etymological treatment of CAPITALISM and CAPITALIST in historical dictionaries seems to have been inspired by and large by such intuitions about the synchronic relationship between CAPITAL, CAPITALIST and CAPITALISM. The *TLFi*, for example, writes:

> **Capitalisme** *subst. masc.* [...] Dér. de *capital²**; suff. *-isme**.
>
> **Capitaliste** *adj.* et *subst.* [...] Dér. de *capital**; suff. *-iste**. L'hyp. d'un empr. au néerl. *kapitalist* (BL.-W.⁵) ne semble pas justifiée. Le corresp. all. *Kapitalist* « possesseur d'un capital » est attesté dep. 1694 (WEIGAND).[4]

As we will see, this kind of analysis in no way does justice to the complex interrelationships that have developed over time among the three words of this word family, nor to the inter-European relationships that link corresponding members in different European languages. I will now describe these relationships by following the evolutions of the individual words step by step from the 17th century up to the present time.

3 The evolution of the noun CAPITALIST from the 17th to the 19th century

3.1 CAPITAL

This is not the place to take up the complex history of CAPITAL at full length. Suffice it to say that by the time that the first derivative, CAPITALIST, appeared, CAPITAL gener-

---

[3] http://www.thefreedictionary.com/capitalism.

[4] [**Capitalisme** *masc. noun* [...] Derived from *capital²**; suffix *-isme**. / **Capitaliste** *adj.* and *noun* [...] Derived from *capital**; suffix *-iste**. The hypothesis that it be a loan from Dutch *kapitalist* (BL.-W.⁵) does not seem to be justified. The corresponding German word *Kapitalist* 'owner of capital' has been attested since 1694 (WEIGAND).]

ally referred to the property, not necessarily only money, that a rich person owned. In double-entry bookkeeping, the term referred to the net worth owned by the merchant after taking away the liabilities from the assets. Towards the end of the 18th century economists extended the meaning of the term to include the means of production (buildings, machines, tools) used in agriculture or industry, what is now called PHYSICAL CAPITAL. This more technical sense still has not really penetrated into common language, but it did play a role in the history of CAPITALIST and CAPITALISM, as we will see. More recent extensions of the concept, by contrast, such as HUMAN CAPITAL or SOCIAL CAPITAL, had no influence.

3.2 CAPITALIST: the Dutch origins

As we saw in Section 2, the *TLFi* rejected the hypothesis of a Dutch origin of the French noun *capitaliste*, which had first been put forward by Barbier (1944–1952: nr. XXV). This decision was ill-advised, since the noun *Capitalist* was indeed coined in the Netherlands (then: "United Provinces") back in 1621 by tax authorities in order to designate a wealthy citizen who possessed 2,000 guilders or more:

> Special registers distinguished the taxpayers into two categories: those owning more than 2,000 guilders were called 'capitalists' (from 1621), and those owning 1,000 to 2,000 guilders were the so-called 'half capitalists' (from 1625). People owning less than 1,000 guilders were fully exempt from extraordinary property taxes. A proposal from 1641 to introduce another level, from 20,000 or 30,000 upwards, was not accepted. The word 'capitalist', here used in its earliest meaning, clearly designated someone owning property. ('t Hart 1993: 122–123)

Dutch *Capitalist* was derived from *Capitaal* 'capital' and followed the pattern of formations in -IST that designated persons engaged in some activity, not the supporter pattern, both of which were already well established at that time (see Wolf 1972). In order to understand the choice of suffix, we probably have to assume that the coiner conceived of a *Capitalist* as a money-lender or investor, not as a passive possessor of a huge sum of money or property. Dutch *Capitalist* was a complex concept, designating at the same time a wealthy person, mostly engaged in money-lending or investment activities, as well as a category of the tax authorities. Since both these facets were linked by mutual inference, we should view them as part of one and the same concept, not as two independent concepts, very much like *book* can designate at the same time the object on the table and its content. It is also highly probable that the precise original definition of *Capitalist* on the part of the tax authorities ('a person worth 2,000 guilders or more') was relaxed in common parlance to refer simply to very rich individuals in general.

The 17th century is called the "Golden Age" in Dutch historiography, because the United Provinces at that time were at the forefront of trade, military, science and art. This background, especially their eminent position in international finance, explains how a

Dutch neologism could spread abroad and start an astounding international career. Already by the end of the 17th century, we find loan translations in German and French. German *Capitalist* (today written *Kapitalist*) appears as early as 1671 in a document on the financial system of the United Provinces, where, due to its novelty, it is glossed as 'money-lender' (Rainer 1998: 10). The German word, as far as I can see, had no influence on French, which will be the focus of the rest of this paper.

3.3 French *capitaliste*: its semantic evolution until the Physiocrats

There can be no doubt about the Dutch origin of the French noun *capitaliste*. The oldest example, in fact, comes from the *Mercure Hollandois* of 1678, p. 13 and clearly refers to the very special fiscal meaning which the term had at that time in the United Provinces: "Pour cet effet [i.e. to put up an army of 100,000 men in a fortnight] ils posoient qu'il y avoit dans la Province de Hollande 65 500 Capitalistes, qui étoient taxés sur les Cahiers de l'Etat à 2.4.6.10.20. & 80 000 livres."[5] The few examples that we find in French until the middle of the 18th century (quoted under II.A in the corresponding *TLF-Étym* entry) refer to that same Dutch reality. In the second half of the 18th century, however, the noun *capitaliste* firmly established itself in French with a reference independent from the Dutch context. Here is a quote from the *Dictionnaire domestique portatif* (Paris: Vincent 1765), vol. 3, p. 505: "RENTIERS; ce terme est synonyme à *capitaliste*, c'est-à-dire, à celui qui fait valoir son argent, en le disposant suivant le cours de la place, & qui vit de ses rentes."[6]

The diffusion of the term among a wider public was furthered by its adoption by the Physiocrats, an economic school that began holding much sway at that time, in France and abroad. The following example from Turgot's *Réflexions sur la formation et la distribution des richesses* illustrates the meaning that will be the dominant one throughout the 19th century:

§ XCIII

*Le **capitaliste** prêteur d'argent appartient, quant à sa personne, à la classe disponible.*

Nous avons vu que tout homme riche est nécessairement possesseur ou d'un capital en richesses mobilieres, ou d'un fonds équivalent à un capital. Tout fonds de terre équivaut à un capital ; ainsi tout propriétaire est **capitaliste**, mais tout **capitaliste** n'est pas propriétaire d'un bien fonds ; et le propriétaire d'un capital mobilier a le choix, ou de l'employer à acquérir des fonds, ou de les faire valoir dans des entreprises de la classe cultivatrice ou de la classe industrieuse. Le **capitaliste**, devenu entrepreneur de culture ou d'industrie, n'est pas plus disponible, ni lui ni ses

---

[5][To that effect they assumed that there were in the province of Holland 65,500 capitalists, whose tax charge according to the state's tax lists was 2, 4, 6, 10, 20 or 80 thousand pounds.]

[6][RENTIERS ; this term means the same as *capitalist*, that is, one who invests his money according to the evolution of rates on the market and lives off his private income.]

profits, que le simple ouvrier de ces deux classes ; tous deux sont affectés à la continuation de leurs entreprises. (Turgot, *Réflexions sur la formation et la distribution des richesses*, s.l. 1788, p. 125)[7]

As one can see, the term is now completely detached from its original fiscal context and simply refers to wealthy individuals who try to increase their capital by either lending money at interest or investing it in productive enterprises (directly, or on the stock market). The meaning, therefore, roughly corresponded to both senses 2 and 3 of the *Free Dictionary* quoted in Section 2. It was not really a French innovation: already the Dutch capitalists typically engaged in precisely these two activities. What is new is that the word could now be used without reference to the particular Dutch context and that the fiscal perspective to which the Dutch term was originally tied had sunk into oblivion. By the same token, the original concept was simplified, being stripped of its fiscal facet.

3.4 Capitalist spilling over to the Anglo-Saxon world

Nowadays we strongly associate capitalism with the Anglo-Saxon world, but the truth is that Great Britain and the United States were the last among the big, developed nations to take up the word CAPITALIST. In English, *capitalist* does not make its appearance before 1787, when the following example is attested in Madison's writings (*The Writings of James Madison*, ed. G. Hunt. New York/London: Putnam's Sons 1903, vol. 4, p. 123):[8] "In other Countries this dependence results in some from the relations between Landlords and Tenants in others both from that source and from the relations between wealthy **capitalists** and indigent labourers." Four years later, the word is used in England by Edmund Burke:

> On the policy of that transfer I shall trouble you with a few thoughts. In every prosperous community something more is produced than goes to the immediate support of the producer. This surplus forms the income of the landed **capitalist**. It will be spent by a proprietor who does not labour. (Edmund Burke, *The Political Magazine* 21, 1791, p. 75)

Up to that moment, capitalists were generally referred to as *monied men* in English, an expression that rapidly succumbed to the prestige of the newcomer, but not before giving rise, for a short period of time, to the blend *monied capitalists*. There can be no doubt that French was the donor language for the English calque.

---

[7][§ XCIII / *The money-lending capitalist is part of the available class* / We have seen that any monied man necessarily owns either capital constituted of transferable riches or a property equivalent to capital. Landed properties are always equivalent to capital; therefore all landowners are capitalists, but not all capitalists own property; and the owner of transferable capital can choose to use it to buy property or to invest it in enterprises of the agricultural or industrial class. The capitalist who has become entrepreneur in agriculture or industry is no more available, neither he himself nor his profits, than the simple worker of these two classes; both are engaged in the continuation of their enterprises.]

[8]The first attestation given in the *OED* is from 1792.

3.5 The capitalist as entrepreneur

From the 17th century to the 19th century, the dominant meaning of CAPITALIST in all European languages was that of a wealthy person who made his capital "work" by lending it at interest, buying bonds or shares, or investing it in productive activities. In this last case, a capitalist could easily become an entrepreneur himself, directly engaged in the management of the firm he owned or of which he was an associate. By shifting the attention from the 'monied man' sense to this latter facet of the complex concept 'capitalist', the word eventually also became established in the new sense of 'entrepreneur', defined in the *Free Dictionary* as 'a person who organizes, operates, and assumes the risk for a business venture'. As already observed by Passow (1927: 109–111), this shift in meaning occurred first in English:

> When the manufacturing **capitalist** of Europe shall advert to the many important advantages, which have been intimated, in the course of this report, he cannot but perceive very powerful inducements to a transfer of himself and his capital to the United States. (*The American Museum*, Philadelphia: Carey 1792, Part I, from January to June, Appendix II, p. 19)

> All the laws connected with our manufacturing system, appear to be founded on one erroneous principle, that the **capitalists** or masters are the only part to be protected against combination and injustice, though the artizans or workmen have an equal right to be protected in their property or skill […]. (*The Parliamentary Debates from the Year 1803 to the Present Time*. Vol. 23. London: Longman 1812. July 21, 1812 – column 1165)

> The small farmer has disappeared, and the smaller manufacturers are superseded by large **capitalists**, who alone can afford to purchase expensive machinery. (*Remarks on the Practicability of Mr. Robert Owen's Plan to Improve the Condition of the Lower Classes.* London: Leigh 1819, p. 6)

The new sense may have arisen in English at that time due to the lack of specific word for 'entrepreneur' (*entrepreneur* in the relevant sense dates from the mid-19th century). What is more surprising is that this English usage should be taken over by French, where the word *entrepreneur*, which English was to borrow a few decades later, was already well established. One precocious example which, at least at first sight, seems relevant in our context is the following from Charles Caseaux' *Considérations sur les effets de l'impôt dans les différens modes de taxation*:[9]

> […] on doit toujours distinguer avec le même soin deux espèces de **capitalistes** ou propriétaires ; j'appelle les uns *capitalistes de la terre*, et les autres *capitalistes de l'industrie* : —les capitalistes de la terre ou territoriaux, sont non-seulement les propriétaires du grand capital de la terre mais ceux de toutes les espèces de capitaux

[9]Note that Caseaux lived in London at that time.

> nécessaires pour tirer du grand capital, tout le produit dont il est susceptible : — les capitalistes industriels, ou de l'industrie, sont les différens propriétaires non-seulement du capital en argent qui met journellement le travailleur en action dans l'industrie comme il le met sur la terre, mais de tous ces autres capitaux appelés bâtimens, ustensiles, machines, *crédit* même etc. (Charles Caseaux, *Considérations sur les effets de l'impôt dans les différens modes de taxation*, London: Spilsbury 1794, p. 98)[10]

This use of *capitaliste* by Caseaux straightforwardly ties in with his Physiocratic background: the capitalist, for him, is not simply a money-lender but the person who provides capital in the broad sense of the word, that is, including both fixed (land, buildings, machinery, tools) and circulating (intermediate goods, operating expenses) capital. Jean Baptiste Say, in the fourth edition of his *Traité d'économie politique*, is well aware of the potential dangers of the polysemy of the term CAPITALIST and therefore carefully demarcates the concept 'capitalist' from that of 'entrepreneur':

> CAPITALISTE ; est celui qui possède un *capital* et qui le fait valoir par lui-même, ou bien le prête, moyennant un *intérêt*, à l'*entrepreneur d'industrie* qui le fait valoir, et dès lors en *consomme* le *service* et en retire les *profits*. [...] Un entrepreneur d'*industrie agricole* est *cultivateur* lorsque la terre lui appartient ; *fermier* lorsqu'il la loue. Un entrepreneur d'*industrie manufacturière* est un *manufacturier*. Un entrepreneur d'*industrie commerciale* est un *négociant*. Ils ne sont **capitalistes** que lorsque le *capital*, ou une portion du capital dont ils se servent, leur appartient ; ils sont alors à la fois *capitalistes* et *entrepreneurs*. (Jean Baptiste Say, *Traité d'économie politique*, 4th edition, Paris: Deterville 1819, vol. 2, pp. 456, 469)[11]

Despite Say's efforts at clarifying the meaning of CAPITALIST, some of his French compatriots yielded to the new English semantics, using *capitaliste* in lieu of *entrepreneur* or *patron* 'master' and opposing it with *ouvrier* or *travailleur* 'worker'. The English usage may have crept into the French language through translations such as the following:

> Nouveau système d'association entre les petits **capitaliste**s et les ouvriers, proposé par l'auteur (Babbage, Charles *Traité sur l'économie des machines et des manufactures.* Traduit de l'anglais par Éd. Biot. Paris : Bachelier 1833, p. xiv)

---

[10][one always has to distinguish carefully two types of capitalists or owners; I call the first one *landed capitalists*, and the other *manufacturing capitalists*: —the landed capitalists are not only the owners of the important capital of the land but also of all kinds of capital necessary for deriving from the land all the produce it can yield: —the manufacturing capitalists are owners not only of the money that makes workers become active in the factory as it does on the land, but of all the other capitals called buildings, tools, machines, even *loans*, etc.]

[11][CAPITALIST: one who possesses capital and puts it to use himself or lends it to an entrepreneur on interest who then consumes its service and reaps the profit made. [...] An entrepreneur in *agriculture* is called a *farmer* if he owns the land, a *tenant* if he rents it. An entrepreneur in *industry* is called a *manufacturer*. An entrepreneur in trade is a *merchant*. They are only *capitalists* if they own the *capital*, or part of the capital they use; in that case, they are at the same time *capitalists* and *entrepreneurs*.]

> Les marchandises étant le produit du capital et du travail, sont la propriété commune du **capitaliste** et du travailleur (ici ouvrier). (*Contes de Miss Harriet Martineau sur l'économie politique. Traduits de l'anglais par B. Maurice.* La Haye : Vervloet 1834, p. 179)

From the mid-1830s onwards, this new usage also became quite frequent in texts written by French authors and was to establish itself alongside the more restrictive traditional use (respectively senses II.A and II.B in the *TLFi*):

> Maintenant cherchons la loi qui détermine le taux des profits. Cette loi devra avoir un rapport intime avec celles des salaires, car le **capitaliste** et le travailleur se partagent le même produit. (*Journal général de l'Instruction publique*, nouvelle série, vol. 7, nr. 95 (1838), p. 1005 [Pellegrino Rossi])

> Il s'agissait de la grande question de la lutte établie entre le **capitaliste** et le salarié, entre l'entrepreneur et l'ouvrier, de la question du paupérisme enfin. (*Mélanges Religieux*, vol. 1, nr. 21, 11 juin 1841, p. 331)

> Malheureusement la question du salaire se compliqua de celle de la jouissance de la case et du terrain en dépendant, et, ainsi enchevêtrées, elles donnèrent lieu aux plus grandes difficultés entre le **capitaliste** et le travailleur. (Milliroux, Félix *Demerary, transition de l'esclavage à la liberté.* Paris: Fournier 1843, p. 31)

It is easy to see that the rise of the 'entrepreneur' sense of CAPITALIST goes hand in hand with the progress made by the Industrial Revolution, where France followed England with a certain time lag. It was the Industrial Revolution that provided capitalists with new opportunities to put their wealth to use by engaging in industrial activities, instead of lending money on interest or speculating on sovereign debt or the shares of trading companies. This new opposition between capitalists and workers will be of crucial importance for the further fate of our word family from the mid-19th century onwards.

From a linguistic point of view, this second semantic change of CAPITALIST is another example of a shift of emphasis that took place within a complex concept, mirroring changes that had previously occurred in the extra-linguistic world. Examples such as these make it clear that what we call "semantic" change in historical linguistics cannot be described on the basis of a minimalist semantics as conceived by the structuralists and other semanticists, but needs to take into account concepts in all their encyclopedic richness. It should also be mentioned here that the rise of the 'entrepreneur' sense led to a decrease in transparency of CAPITALIST, since the new technical sense of CAPITAL on which it was based, introduced by the Physiocrats and focusing on land, buildings, machinery, raw materials and intermediate goods more than solely money, had not become familiar to the speech community at large. The relationship between base and derivative, which had been quite transparent in the 'monied man' sense, thereby became somewhat obscured for ordinary speakers.

4 CAPITALISM

Throughout the 17th century and most of the 18th century, the noun CAPITALIST was an "only child", pertaining to a word family with only two members, CAPITAL and CAPITAL-IST. At the beginning of the 19th century, however, this nuclear family started expanding in several directions. With CAPITALISM, a little brother was born, and CAPITALIST itself brought into the world an adjectival progeny, as we will see in Section 5. At the same time, complex incestuous relations developed between CAPITALISM and CAPITALIST, both in their nominal and adjectival uses. In this section, we will follow the development of CAPITALISM from its obscure beginnings to its establishment as one of the key notions of modern economic and political discourse in the mid-19th century.

4.1 CAPITALISM 'condition of being rich' (1753): a ghost word?

Dauzat (1972) claimed that French *capitalisme* was used as early as 1753 in the *Encyclopédie* with the meaning 'état de celui qui est riche'.[12] He was followed on this point by the *TLFi*, while Braudel's search for the text alluded to by Dauzat yielded no result: "Le texte invoqué reste introuvable."[13] (1979, vol. 2, p. 205). I could not find it either in the electronic version of the *Encyclopédie* that we have at our disposal nowadays.[14] It is difficult to imagine that Dauzat should have invented his early first attestation, but something must have gone wrong. In fact, neither can the French word be found with Google Books in the entire second half of the 18th century.

However, this latter source provides one isolated early attestation of German *Kapitalismus*, a clearly jocular occasionalism from Itzehoe's *Komische Romane* (Göttingen: Dieterich 1787, vol. 4, p. 304), in a text full of somewhat contrived neologisms. It seems to express very much the same sense as the one indicated by Dauzat for French: "Der Redakteur dieser Papiere, der, wie aus allen seinen Schreibereyen hervorgeht, sich voll tiefer Ehrerbietung gegen jegliches Menschengesicht fühlt, das nur halbwege mit dem Stempel der Vornehmigkeit und des **Kapitalismus** gemarket ist, sieht sich hier in großer Verlegenheit."[15] Since there are no other examples for German either until around 1840, it is best to leave this potential proto-use of CAPITALISM as a riddle for future research and turn to its first appearance in the 19th century.

4.2 CAPITALISME 'high finance' (*ca.* 1810)

At the time of the French Revolution, the noun *capitaliste* had acquired distinctly negative overtones, referring to individuals who had enriched themselves in the political and economic turmoil of those years, to the detriment of the general good (see Höfer 1986). We should keep this background in mind in order to understand the following passage,

---

[12] [condition of being rich]

[13] [The text alluded to is nowhere to be found]

[14] Cf. http://portail.atilf.fr/encyclopedie/.

[15] [The editor of these papers who, as can be seen in all his writings, feels deference for any human face that somehow expresses high rank and capitalism, faces great embarrassment here.]

written at the moment when Napoleon had reached the climax of his power (around 1810) and drawn from a letter addressed to a statesman by an "agent observateur" whose name is not disclosed:

> Mais qui [sic] dire de cette puissance nouvelle du **Capitalisme**, qui née du commerce qu'elle ruine, a succédé avec toute son immoralité, à la puissance si morale de la fructification du sol qu'elle opprime en détournant ses capitaux ? de cette puissance qui sacrifie l'avenir au présent, et le présent à l'individualité, cette lèpre contemporaine. Cette puissance égoïste, cosmopolite, qui s'empare de tout, ne produit rien et n'est infiniment liée qu'à elle-même ; souveraine des souverains qui ne peuvent sans elle ni faire la guerre ni demeurer en paix ; et qui s'enrichit également de leur prospérité et de leur ruine, des biens du peuple qu'elle partage, de leurs maux qu'elle accroît ? (Alphonse de Beauchamp, *Mémoires tirés des papiers d'un homme d'État.* Paris: Michaud 1836, vol. 11, p. 46)[16]

The 'high finance' sense of *capitalisme*, however, does not seem to have had a wide circulation. We meet it again in 1822 in Georges Laurent Aubert du Petit-Thouars' *Toujours la guerre au cadastre français*, where it is used as antonym of *propriété*, designating a society dominated by rentiers rather than the landowning class:

> Deux individus, l'un capitaliste et l'autre propriétaire, ont chacun vingt-cinq mille livres de rente ; [...]. Ainsi la propriété, seul et véritable soutien des monarchies, perd tous les jours en France de son ascendant au profit du *capitalisme* qui de sa nature tend toujours au républicanisme : chaque jour nous le prouve. (Georges Laurent Aubert du Petit-Thouars, *Toujours la guerre au cadastre français*, Paris: Trouvé 1822, p. 42)[17]

Significantly, the word is written in italics in order to highlight its novelty. Our third example appears three years after the publication of Beauchamp's work, in which the anonymous observer's invective quoted above had been made public, in Pons Louis François de Villeneuve's *De l'agonie de la France*:

> Avec le malaise ou l'instabilité de la fortune privée, concorde le malaise encore plus pénétrant de la fortune sociale : et un mal nouveau, le **capitalisme**, insinuant et dangereux serpent, étouffe en ses plis et replis l'une et l'autre. [...] Autre et plus féconde proie est pour le **capitalisme** la fortune publique. Il en pompe les budgets

---

[16] [What should one say about this new power of capitalism, which arose from the commerce that it ruins and with all its immorality succeeded the highly moral power of agriculture that it oppresses by diverting its capital? About this power which sacrifices the future to the present, and the present to individualism, the leprosy of our days. This egoistical, cosmopolitan power that grabs everything, does not produce anything and is only infinitely tied to itself; sovereign of sovereigns, who cannot without it make war nor remain in peace; and that enriches itself both by their prosperity and their ruin, at the expense of the goods of the people that it divides up, of their troubles that it increases?]

[17] [Two individuals, a capitalist and a landowner, both have an income of 25,000 pounds; [...] In that way ownership, the only true support of monarchies, loses influence day by day to the benefit of *capitalism*, which by its very nature tends towards republicanism: each day proves this to be the case.]

> par la rente ; il fait comme à son gré la paix ou la guerre. (Pons Louis François de Villeneuve, *De l'agonie de la France*, Paris: Perisse 1839, pp. 139-140)[18]

These three examples of *capitalisme* are still transparently tied to the old sense of the word *capitaliste*, referring to a very wealthy individual lending his money at interest or placing it in bonds or shares. What is less immediately obvious is the patterns of word formation by means of which this word came into being. Was it derived from *capitaliste* by affix substitution? Was it an independent derivation from *capital*? Nouns in *-isme*, at any rate, were already in use at that time for designating economic systems, witness *colbertisme* (1775, *TLF-Étym*) and *mercantilisme* (1809, *TLF-Étym*).[19] Thus, from a chronological perspective, these words could have served as models for *capitalisme*. The corresponding nouns in *-iste*, *colbertiste* and *mercantiliste*, designated the supporters of the respective doctrine. Since *capitaliste* did not refer to a supporter, but to a profession or occupation, *capitalisme*, for semantic reasons, could not be derived by affix substitution according to a proportional analogy of the kind *colbertiste* : *colbertisme* = *capitaliste* : *x*. The more plausible solution, therefore, is to consider *capitalisme* to have been an independent derivation on the basis of *capital*, following the general pattern noun + *-isme* '(economic) system somehow related to N'.

4.3 Capitalism as the antonym of socialism

As we saw in Section 3.5, capitalist acquired the sense 'entrepreneur' after having crossed the Channel (and the Atlantic), a sense that migrated back to France from the 1830s onwards, where it has cohabitated with the original sense ever since. *Capitaliste*, in that way, became the antonym of *ouvrier*, *travailleur* (both 'worker') and *prolétaire* 'proletarian', just like *capital* 'capital' had become the antonym of *travail* 'work'. This lexical opposition simply reflected an extra-linguistic phenomenon, namely the well-known social divide created by the Industrial Revolution. In the 1840s, French *capitalisme* was also attracted by this lexical field and thereby was converted into the standard designation of the new economic system characterized by the exploitation of workers in factories owned and often run by a small group of capitalists/entrepreneurs. Here are some of the first examples of this new sense, which are probably attributable to Louis Blanc:[20]

> Une lutte récemment engagée entre Lamartine et L. Blanc a donné naissance à un nouveau mot ; le capitalisme. Ce n'est pas au capital, s'écrie ce dernier, que nous avons déclaré la guerre, mais au capitalisme ; c'est-à-dire, sans doute, aux

---

[18][The difficulties and instability of private fortunes matches the even greater difficulties of public fortune: and a new evil, capitalism, this insinuating and dangerous snake, suffocates in its folds the one and the other. (...) Another, even more fertile prey for capitalism is the public fortune. It sucks the budgets by means of government bonds; it makes war and peace as it pleases.]

[19]Similar formations from outside the economic sphere were already older; see *marianisme* (1665, *TLF-Étym*), *spinozisme* (1685, *TLF-Étym*), etc.

[20]See alreadySilberner & Febvre (1940).

capitalistes. (*Mémoires de l'Académie royale des sciences, belles-lettres et arts de Lyon* 1, 1845, p. 282, n. 1)[21]

L'orateur compare la féodalité ancienne avec le capitalisme actuel. La féodalité protégeait du moins l'exploitation de la terre, et par conséquent le travail de l'ouvrier, tandis que le capitalisme exploite l'ouvrier lui-même. (*L'Ami de la religion* 138, 1848, p. 621)[22]

In this new 'economic system' sense, *capitalisme* became the antonym of an alternative system where the workers themselves would own the capital that forms the basis of their activity. Avril, V. *Histoire philosophique du crédit* (Paris: Guillaumin 1849, vol. 1, p. 153) already explicitly opposed CAPITALISM and SOCIALISM: "la différence radicale qui sépare le capitalisme du socialisme".[23] *Socialisme* (1831, *TLFi*) had already been in use for more than a decade when *capitalisme* in this new sense appeared, and *communisme* (1840, *TLFi*) for a few years. Both may well have served as its immediate models.

The case of *capitalisme* in the sense discussed here aptly illustrates the complex factors that come into play in the creation and diffusion of a neologism. The *TLFi*'s statement that it is composed of a base *capital* and a suffix *-isme* is acceptable as a synchronic, though not particularly revealing, description of the word's internal makeup, but hardly qualifies as an etymology doing justice to the circumstances of the word's creation. At the outset, we have to admit that the lack of documentation does not yet allow us to gain full certainty about how it came into being, the most plausible scenario being the following: Assuming that the 'high finance' sense was known to the coiner, which seems likely, we should consider the process as one of semantic change, a conceptual adaptation of the 'high finance' sense to the new situation of capitalists acting themselves as entrepreneurs, and not just as financiers. From that perspective, the new lexical opposition with *socialisme* and *communisme* could be viewed either as a consequence of this conceptual change, or as its trigger. In fact, the relevant meaning of these two terms, namely an 'economic system where the means of production pertains to the workers or to society as a whole', called for a designation for the opposite concept of an economic system where the means of production was concentrated in the hands of a small group of wealthy individuals. Since this means of production was referred to technically as *capital* and the entrepreneurs had come to be called *capitalistes*, *capitalisme* was a natural choice. This reconstruction of the word's origin also neatly explains why the word was used with negative connotations right from the beginning: it was launched by the opponents of capitalism, while capitalists themselves and circles close to them used to call the then prevailing economic system *libéralisme* (*économie de marché* 'market economy' is of much more recent vintage). The transition from the 'high finance' sense to the 'economic system' sense was therefore essentially a process of conceptual rearrangement within an existing lexeme. Nevertheless, word formation also came into play, namely by

---

[21][A quarrel that recently opposed Lamartine to L. Blanc has given rise to a new word, capitalism. It is not to capital, claims the latter, that we have declared war, but to capitalism; that is, no doubt, to capitalists.]

[22][The speaker compares feudalism with present-day capitalism. Feodalism at least protected the exploitation of the land, and hence the activity of the worker, while capitalism exploits the worker himself.]

[23][the radical difference that opposes capitalism and socialism]

licensing the pattern noun + -*isme* with the overall meaning 'system somehow related to N' (note that both *socialisme* and *communisme* have adjectival bases; therefore, strict proportional analogy with these two words would not suffice).

4.4 The further fate of CAPITALISM

The French neologism *capitalisme* in its 'economic system' sense had an immediate and resounding international success in the wake of the 1848 revolution. I will not describe here the diffusion of the term in different European languages,[24] but concentrate instead on its further development in French.

   By a simple metonymic process, designations of systems and similar abstract entities are routinely taken to refer to the persons who represent or support the system. Such was also the case with *capitalisme*. The first example of Section 4.3 could already be interpreted in that sense. Here is a later and clearer example of this collective sense (Burg, Joseph *De la vie sociale...* Rixheim: Sutter 1885, p. 739): "Le capitalisme, dur et arrogant, coudoie le paupérisme, exaspéré et découragé."[25]

   A more interesting conceptual change occurred at the beginning of the 20th century. At that time, academic circles began using the term not only to refer to the contemporary economic system, what we now call INDUSTRIAL CAPITALISM, but also to economic systems of past times that, in their opinion, presented sufficient similarities with the contemporary system to be called CAPITALISM. Proto-capitalism was located in the Renaissance, in the Middle Ages, or even in Antiquity. This conceptual change, which was the result of conscious conceptual manipulation for scientific purposes, resulted in a more abstract concept of capitalism, freed from some of the more contingent aspects of 19th century industrial capitalism, as well as its negative overtones. In France, the historian Henri Hauser was the first to deal with the origins of capitalism in *Les Origines du capitalisme moderne en France* (Paris: Larose) in 1902. However, the international success of this scientific sense was certainly due to the publication, some months before, of Werner Sombart's monumental *Der modern Kapitalismus* (Leipzig: Duncker & Humblot 1902). If Hauser had been inspired by Sombart, the new sense would have to be classified as a calque.

5 CAPITALIST going adjectival

CAPITALIST, as we saw in Section 3, started out as a noun, and it remained exclusively nominal until the end of the 18th century. It is at that time when French *capitaliste* developed adjectival uses that are still parts of the language. Three different adjectival senses must be distinguished: 1. 'owning (a huge amount of) capital', 2. 'of capitalists', and 3. 'of capitalism'.

---

[24]For German, see Hilger (1982).
[25][Capitalism, hard and arrogant, rubs shoulders with pauperism, exasperated and discouraged.]

5.1 *Capitaliste* adj. ʿowning (a huge amount of) capital'

As early as 1790, Charles-Nicolas Ducloz-Dufresnoy, in his *Observations sur l'état des finances*, quotes a "publiciste" called Cerruti who wrote:

> On ne peut appauvrir la Capitale sans appauvrir les Provinces dont elle assemble, grossit, répartit et multiplie les richesses territoriales et industrielles.
>
>   Voilà la véritable idée d'une Capitale.
>
>   Voilà la véritable idée des Capitalistes.
>
>   Le **peuple Capitaliste** est composé de tous ceux qui par leur économie ou par leur activité, ont formé des trésors disponibles prêts à circuler, prêts à se reposer, prêts à se transformer en papier, prêts à se réaliser en terres. (Charles-Nicolas Ducloz-Dufresnoy, *Observations sur l'état des finances*, Paris: Clousier 1790, pp. 14-15)[26]

In the first half of the 19th century this possessive use of *capitaliste* established itself in wider circles, as the following examples show:

> l'aristocratie territoriale adoucit vis-à-vis des campagnes l'aristocratie **capitaliste** (Laborde, Alexandre de *Des aristocraties représentatives*. Paris: Le Normant 1814, p. 96)[27]

> comme s'il ne suffisait pas [...] d'un imprimeur **capitaliste** ou laborieux pour multiplier ces produits (*Revue encyclopédique*, t. 49, janvier-mars 1831, p. 452)[28]

> [la législation des Émigrés] a rendu le peuple propriétaire et la noblesse **capitaliste** (Lahaye de Cormenin, Louis-Marie de *Droit administratif*. Paris: Thoral 1840, t. 1, p. *xxxvii*)[29]

> La bourgeoisie moderne [...] forme une espèce d'aristocratie **capitaliste** et foncière, [...]. (Proudhon, Pierre-Joseph *Organisation du crédit et de la circulation*. Paris: Garnier 1848, p. 21)[30]

> Ce n'est pas la bourgeoisie qui est boursière, c'est la société tutta quanta qui veut être **capitaliste** en exploitant les éventualités des échanges. (Bianchini, Lodovico *La science du bien-être social*. Bruxelles: Librairie universelle 1857, p. 351)[31]

---

[26] [One cannot make the capital poorer without making poorer the provinces whose agricultural and industrial wealth it assembles, increases, distributes and multiplies. / This is the true idea of a capital. / This is the true idea of capitalists. / The capitalist people is composed of all those who through their savings and activity have formed treasures ready to circulate, ready to lie idle, ready to be transformed into paper, ready to be realized as landed property.]

[27] [the landed aristocracy makes the capitalist aristocracy more acceptable for the countryside]

[28] [as if it were not enough [...] to have a well-capitalized or hard-working type-setter in order to multiply these products]

[29] [[the legislation on emigrants] has turned the people into owners and the aristocracy into capitalists]

[30] [The modern bourgeoisie [...] forms a kind of capitalist and landed aristocracy]

[31] [It is not the bourgeoisie who is crazy about the stock market, it is the entire society that wants to be capitalist by taking advantage of the opportunities of trading.]

From a linguistic point of view, the meaning 'owning (a huge amount of) capital' constitutes a case of noun-adjective conversion, the base being constituted by the noun *capitaliste* with the meaning 'person owning (a huge amount of) capital'. This conversion pattern does not seem to have had any direct model among words in -*iste*, none of which had a possessive meaning, by the way, if we exclude obsolete *actioniste* 'shareholder', which was also of Dutch origin. As argued in Section 3.2, *capitaliste* should be classified as a marginal member of the agentive niche represented by words such as *aubergiste* 'innkeeper', *copiste* 'copyist', *ébéniste* 'cabinetmaker', *latiniste* 'Latin scholar or student', *psalmist* 'psalmist'. Such nouns, however, do not seem to have developed adjectival uses (of the relevant kind), according to the information provided by the *TLFi*.[32] The model must therefore be sought outside derivative patterns in -*iste*.

5.2 *Capitaliste* adj. 'of capitalists'

The second adjectival sense – which, incidentally, the *TLFi* fails to mention – corresponds to a relational use referring to the corresponding noun *capitaliste*. Again, we find one early outlier in 1791, this time in a translation of Adam Smith's *Inquiry into the Nature and Causes of the Wealth of Nations* :

> Lorsque ces compagnies [...] commercent avec des capitaux réunis, et que chacun des membres a sa part dans le bénéfice commun ou dans la perte commune, en proportion des fonds qu'il y a mis ; on les appelle compagnies CAPITALISTES. (Adam Smith, *Recherches sur la nature et les causes de la richesse des nations*, translated by J. A. Roucher, Paris: Buisson 1791, vol. 4, p. 90)

This passage translates the following one from Smith's original (I quote here from the 9th edition, where, as we can see, *joint stock company* corresponds to the translator's *compagnie capitaliste*).

> When they trade upon a joint stock, each member sharing in the common profit or loss in a proportion to his share in this stock, they are called joint stock companies. (Adam Smith, *Inquiry into the Nature and Causes of the Wealth of Nations*, 9th edition, London: Strahan 1799, vol. 3, p. 110)

Compagnie capitaliste must therefore be considered to be a neologism created by the translator. The only other example provided by Google Books until the mid-19th century is the following, which is obviously inspired by the example just quoted:

> La confection ou entretien d'un canal navigable qui ne peuvent guère être exécutés que par des **compagnies capitalistes**, sont des entreprises qui portent avec elles le privilège qui garantit aux entrepreneurs le bénéfice qu'ils doivent en retirer. (Roux,

---

[32] Appositions such as *rabbin cabaliste* 'cabalist rabbi', *moine copiste* 'monk copyist', *ouvrier ébéniste* 'cabinet worker', etc. are classified as adjectival in the *TLFi*, but this is highly questionable. Some of the nouns quoted are indeed used as adjectives, but in a relational sense (e.g. *la tradition ébéniste* 'the tradition of cabinet-making', etc.).

Vital *De l'influence du gouvernement sur la prospérité du commerce*. Paris: Fayolle 1800, p. 257)[33]

Overall, however, Rouchet's neologism did not catch on. The more common way throughout the 19th century of denominating a company composed of various capitalists in French was *compagnie de capitalistes* 'company of capitalists' or *société de capitalistes* 'society of capitalists', both amply attested since the time of the French Revolution.

On a larger scale, the relational sense 'of capitalists' only appears from the second half of the 19th century onwards. These examples, it seems, were independent from the use of *capitaliste* by Roucher in 1791 in the term *compagnie capitaliste*. It is not always easy to distinguish the relational sense 'of capitalists' from the sense 'of capitalism', since *capitalisme* can also be understood metonymically as the totality of capitalists. In the following list, I have chosen examples where reference to capitalists seems more plausible than to capitalism as an economic system.

[...] à fin de se délivrer de l'exploitation **capitaliste** et usuraire, comme ils se sont délivrés de la tyrannie monarchique et jesuitique (Eugène Sue, *Mystères du peuple*, 1851, vol. 2, p. 90, quoted in: *Archiv des Criminalrechts*. Neue Folge. Jahrgang 1851, p. 57)[34]

Comme nous le disions hier, la conjuration **capitaliste**, l'alliance offensive et défensive du privilége contre le prolétariat est formée ; il y a entente cordiale entre tous ces hommes que nous supposons ennemis : [...]. (Proudhon, P.-J. *Mélanges. Articles de journaux 1848-1852*. Premier volume. Paris: Lacroix, Verboeckhoven & Cie 1868, p. 229)[35]

la tyrannie **capitaliste** et mercantile (Colins, Jean Guillaume *L'économie politique source des révolutions et des utopies prétendues socialistes. Paris:* Librairie générale 1856, p. 56)[36]

Ce sera donc bien une association ouvrière. — Ce sera une association **capitaliste** où [...] le travail sera subordonné au capital. (*Journal des économistes, t. 15, juillet à septembre* 1869, p. 172)[37]

la classe **capitaliste** et la classe ouvrière [...] dans le milieu **capitaliste** (Marx, Karl emphLe capital. Tr. de J. Roy revisée par l'auteur. Paris: Lachatre 1872, pp. 248, 285)[38]

---

[33][The building and maintenance of a shipping canal, which can hardly be undertaken but by a capitalist company, are enterprises that come with a privilege that guarantees the entrepreneurs the profit they can make on it.]

[34][in order to free themselves from capitalist and usurious exploitation, as they had freed themselves from monarchic and jesuitic tyranny]

[35][As we said yesterday, the capitalist conspiracy, the offensive and defensive alliance of the privilege against the proletariat already exists; there is an entente cordiale between all these men that we deemed ennemies]

[36][the capitalist and mercantile tyranny]

[37][This will therefore indeed be an association of workers. — This will therefore indeed be a capitalist association where work will be subordinated to capital]

[38][the capitalist class and the working class (...) in capitalist circles]

> Marx est donc bien loin d'appeler subjectivement le profit **capitaliste** un vol (*Revue internationale du socialisme rationnel*, t. 8, 1883, p. 147)[39]

> le député Rasseneur parlerait de "l'oppression **capitaliste** et de la revanche prolétarienne" (Bonnetain, Paul emphL'Opium. Paris: Charpentier 1886, p. 581)[40]

> l'avidité **capitaliste** contraint les mécaniciens des chemins de fer à effectuer des journées de travail de dix-huit et vingt heures (*La Revue socialiste*, t. 10, 1889, p. 685)[41]

> la moyenne de la vie ouvrière est inférieure à la moyenne de la vie **capitaliste** (*La Réforme sociale*, t. 25, 1893, p. 467)[42]

> incapables [...] d'opposer aux **exigences capitalistes** une résistance efficace (*La Société nouvelle*, t. 2, 1894, p. 448)[43]

These examples should suffice to prove the existence of the relational sense 'of capitalists' from the mid-19th century onwards. This relational use followed a pattern of conversion turning personal nouns into relational adjectives that was already quite well established by the middle of the 19th century, even with nouns in -*iste* (see Rainer to appear). Outside nouns in -*iste*, we find the relational use of *ouvrier* in collocations such as *association ouvrière* 'workers' association' and *classe ouvrière* 'working class; lit. workers' class' as early as 1802 in the *TLFi*. The same relational sense is also attested in the *TLFi* for *prolétaire* (in the example from Bonnetain above, though, the synonymous suffixal derivative *proletarien* is used). Since the noun *capitaliste* by the mid-19th century had become the antonym of *ouvrier* and *prolétaire*, it could well be that its relational use was induced by the relational use of these two antonyms. There is no need to choose between these two hypotheses: the influence of *ouvrier* and *prolétaire* may well have worked in tandem with the pattern converting nouns in -*iste* into relational adjectives.

5.3 *Capitaliste* adj. 'of capitalism'

The relational sense 'of capitalism' was also established in the French language in the middle of the 19th century. As we saw in Section 4, *capitalisme* in the relevant sense was itself a neologism at that time. Here are some early examples in which the sense 'of capitalists' definitively seems less adequate than the sense 'of capitalism'.

> Le système **capitaliste** a été établi en France sous des conditions bien moins propices (Sagra, Ramon de la *Révolution économique*. Paris: Capelle 1849, p. 81)[44]

> le plus grand écrivain de vos théories **capitalistes** (Avril, V. *Histoire philosophique du crédit.* Paris: Guillaumin 1849, p. 69)[45]

---

[39] [Marx is therefore far from subjectively calling capitalist profit theft]
[40] [MP Rasseneur was said to speak about "capitalist oppression and proletarian revenge"]
[41] [capitalist greed obliges the train drivers to work for 18 or 20 hours]
[42] [the lifetime of a worker on average is shorter than a capitalist's lifetime]
[43] [unable to counter the demands of capitalists with an efficient opposition]
[44] [The capitalist system has been established in France under much less favourable conditions]
[45] [the greatest writer on your capitalist theories]

la négation du **régime capitaliste**, agioteur et gouvernemental, qu'a laissé après
elle la première révolution (Proudhon, Pierre-Joseph *Idée générale de la révolution
au XIXe siecle*. Paris: Garnier 1851, p. 107)[46]

Le résultat sera donc un accroissement de population dans le **pays capitaliste** B.
(De Laveleye, Emile *Etudes historiques et critiques sur le principe et les conséquences
de la liberté du commerce international*. Paris: Guillaumin 1857, p. 88)[47]

From a present-day perspective, this usage seems straightforward, since most nouns
in *-isme* referring to ideologies and similar notions are flanked by a relational adjective in
-iste: *marxisme/marxiste*, *racisme/raciste*, etc. Morphologically, the relationship between
such pairs is one of affix substitution. What is crucial in our context is whether this rela-
tion of affix substitution was already operative in the middle of the 19th century. The *TLFi*
does not provide reliable evidence bearing on this question, since in most entries a date
of first attestation is only given for the nominal use of *-iste*. However, relevant examples
are not difficult to come by. In many cases, one may waver between the interpretations
'of Xists' and 'of Xism': "mouvement anarchiste" (d'Ivernois, Francis *Les cinq promesses*.
Londres: Cox 1802, p. 149), for example, could be glossed equally naturally as 'movement
of anarchists' and 'movement inspired by anarchism', "journal légitimiste" (*Procès de M.
Gisquet contre* Le Messager. Paris: Pagnerre 1839, p. 1) as 'newspaper of/for legitimists'
and 'newspaper inspired by/defending legitimism'. In "une thèse matérialiste" (Gibon, H.
Fragments philosophiques. Paris: Hachette 1836, p. 69), however, 'a dissertation inspired
by materialism' would seem to be the only reasonable gloss.

We can therefore safely assume that the 'of capitalism' sense could be derived, by the
middle of the 19th century, from *capitalisme* by means of affix substitution. For the sake
of completeness, however, let us still check an alternative possibility which some might
wish to entertain. As we have seen, CAPITALIST already spilled over to the Anglo-Saxon
world at the end of the 18th century and since then it has been a much-used term in the
English language. Could it not be, therefore, that the relational sense in question was
simply due to a calque from English? In order to answer this question, let us observe
the dates of first attestation[48] of the English collocations corresponding to those quoted
above for French: *capitalist country* (1861), *capitalist system* (1862), *capitalist regime* (1863),
capitalist theories (after 1900). As we can see, the English collocations follow the French
ones by a lapse of time of some 10 years. It may therefore safely be assumed that English
imitated French, not vice versa.

---

[46][the negation of the capitalist, speculative and governmental regime left over from the first revolution]

[47][The result will therefore be an increase in population in the capitalist country B.]

[48]Using the first book allowing a full view of the text, front matter included, in Google Books.

6 A 20th-century codicil: CAPITALIST 'supporter of capitalism'

As we saw in Section 3, n the middle of the 19th century, the 'entrepreneur' sense had been added to the 'monied man' sense. In the 20th century, a third sense was added to these two, namely that of 'supporter of capitalism', which has largely superseded the other two. In the second half of the 19th century, CAPITALISM had evolved from a name characterizing an economic system to that of an ideology. Especially after the international success of Marxism, CAPITALISM became the antonym of COMMUNISM, which could also denote both an economic system and an ideology. Due to this status of CAPITALISM as an antonym of COMMUNISM, CAPITALIST followed COMMUNIST in designating a person that embraced the ideology expressed by the corresponding word in -ISM. The following example illustrates this last transformation of CAPITALIST with French *capitaliste*: "Outre la question de l'attitude du Chrétien, un point irrite particulièrement André Gide ; c'est le reproche qui lui est fait d'être à la fois **capitaliste** et communiste et il s'ingénie à retourner l'accusation contre les chrétiens."[49] (Fillon, Amélie *François Mauriac*. Paris: Société Française d'Éditions Littéraires et Techniques 1936, p. 330). What the author wanted to say here is that Gide was accused of having embraced the ideologies of capitalism and communism at the same time, not that he was a financier, investor, or entrepreneur. In this latest sense one can even be a capitalist without possessing any money or property.

From a linguistic point of view, this last transformation of CAPITALIST is to be regarded as a case of affix substitution on the basis of CAPITALISM, as the gloss 'supporter of capitalism' suggests. What is less easy to tell is whether this affix substitution first took place in French or in some other European language, notably English or German. The question is almost impossible to answer since at that time these three languages were already in perfect harmony concerning CAPITALIST and CAPITALISM as well as the -ISM/-IST pattern. In French, for example, this kind of affix substitution could base itself on a sizeable number of potential models: an *anarchiste* was a supporter of *anarchisme*, a *communiste* a supporter of *communisme*, etc. It is worth mentioning that, from a historical perspective, the derivative in -*iste* tended to occur earlier than that in -*isme*, but at some point in time the names of the supporters came to be reinterpreted as dependent on the names of the doctrines.

7 Conclusion

After having accompanied CAPITALIST and CAPITALISM in their unfolding since the 17th century, it is time to draw some general conclusions about the relationship between word history and word formation and to highlight the role of the lexeme in this affair.

---

[49] [Apart from the question of the attitude of the Christian, one point in particular irritates André Gide: the reproach that is addressed to him of embracing at the same time the ideology of capitalism and communism, and he is at pains to return the charge against the Christians.]

As we have seen, semantic change, borrowing and word formation have all substantially contributed to the evolution of these two key words of our politico-economic vocabulary. And in each of these three modes of lexical enrichment the lexeme has been seen to play a key role. What is traditionally called *semantic change* in reality should better be called *conceptual change*, as Andreas Blank convincingly argues in his 1997 book. The semantic changes observed in the history of CAPITALIST and CAPITALISM affected holistic concepts tied to lexemes, in close interaction with changes in extra-linguistic reality, not affixes or roots. Borrowing also repeatedly played a role: in the migration of CAPITALIST from the United Provinces to France, from France to the Anglo-Saxon world and back again, to mention just those involving French. Now, calquing is a process that is also located at the level of the lexeme. It can be conceived of as an analogical process where model and copy are located in different languages (though in the same speaker's mind). If seen in this light, calquing is close to word formation, which is also best conceived of as an analogical, pattern-based process. This is particularly obvious in the case of affix substitution, which played a prominent role in derivatives with -IST and -ISM.

We have also seen that a full understanding of the evolution of our two words requires taking into consideration the structure of the lexicon at the relevant points in time. A lacuna in the lexicon may induce semantic change, as Passow already surmised in relation to the rise of the 'entrepreneur' sense of English *capitalist*. The absence of a specific word for 'entrepreneur' around 1800 may have prompted the English speakers to adapt the meaning of *capitalist*, originally referring to a rich money lender or investor, in order to fill this empty slot. Another case in point may have been the introduction of the 'economic system' sense of French *capitalisme* in the 1840s, which filled the need for an antonym of *socialisme* and *communisme*. Similarly, the specific configuration of a semantic field may induce change, as we have seen in the case of the opposition 'entrepreneur' vs. 'worker', which may have helped to establish the relational use of French *capitaliste* in the 'of capitalists' sense, providing a ready counterpart for the already established relational use of *ouvrier* and *prolétaire*. The same search for formal/semantic parallelism was probably also operative in the rise of the 'supporter' sense of French *capitaliste* in the 20th century. These latter processes can be accounted for straightforwardly as proportional analogies.

At many points in our discussion we have seen that the French historical dictionaries that we have at our disposal, notably the *TLFi*, only provide a shaky basis for detailed investigations into the history of word-formation patterns in post-Renaissance French. In some sense, the *TLFi* is a marvel of a dictionary, second probably only to the *OED*. Nevertheless, it is obvious in many entries that the lexicographers where overwhelmed by the wealth of raw data at their disposal and hampered by the lack of a sound theory of word formation (or an inconsistent application of the theory, if they had one). The relationship between words in *-isme* and the corresponding relational adjectives in *-iste*, for example, is not given a separate etymological treatment but identified with that of nouns in *-iste*, which are themselves handled in different ways in different entries:

Anarchiste: "Dér. du rad. de *anarchie**; suff. *-iste**"

Animiste: "Dér. du rad. du lat. *anima (âme**)*; suff. *-iste**"

Colbertiste: "du rad. de *colbertisme,* suff. *-iste**"

Cubiste: "Dér. de *cube**; suff. *-iste**"

Fétichiste: "Dér. de *fétiche** formé sur le modèle de *fétichisme**; suff. *-iste**"

Piétisme: "Dér. de *piétiste**; suff. *-isme**"

Quiétiste: "Dér. de *quietisme** par substitution du suff. *-iste** à *-isme*"

In a proper etymological treatment, each step in the history of a word, which roughly corresponds to a word's subentries in a well-ordered dictionary, must be provided with a separate etymological explanation, and each explanation should explicitly name the change according to a catalogue of standard mechanisms of lexical change. In the case of semantic change and borrowing, a list of universal mechanisms such as *calque*, metaphor, and metonymy will generally be sufficient, though some of these mechanisms also show language-specific patterns that should then be named explicitly.[50] For word formation, by contrast, it is vital to make sure that the pattern alluded to in a certain etymological explanation was productive at the moment in question.

The rather glaring shortcomings of the *TLFi* in that respect are now being emended by the *TLF-Étym* project, to which I am happy to contribute from time to time. Word histories in the *TLF-Étym* style are a necessary prerequisite for a history of word formation in modern French,[51] which constitutes a great desideratum. At the same time, detailed studies on the history of single word-formation patterns would yield important contributions to historical lexicography. The two fields are so intimately intertwined, that they of necessity must evolve in tandem.

References

Aronoff, Mark. 2007. In the beginning was the word. *Language* 83(4). 803–830.

Barbier, Paul. 1944–1952. Miscellanea lexicographica. *Proceedings of the Leeds Philosophical and Literary Society: Literary and Historical Section.* 41–60, 93–110, 217–252, 327–360, 409–444, 482–505.

Blank, Andreas. 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen.* Tübingen: Niemeyer.

Braudel, Fernand. 1979. *Civilisation matérielle, économie et capitalisme, XVe-XVIIIe siècle.* Vol. 2: *Les jeux de l'échange.* Paris: Armand Colin.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique.* Tübingen: Max Niemeyer Verlag.

Dauzat, Albert. 1972. *Nouveau dictionnaire étymologique.* Paris: Larousse.

Febvre, Lucien & Henri Hauser. 1939. Capitalisme et capitaliste. *Annales d'histoire sociale* 1. 401–406.

---

[50]For example, in French or Spanish the name of the central product can be used to designate the respective economic sector or activity, while this is not an established metonymic pattern in German or English, witness *travailler dans la tomate / trabajar en el tomate* vs. **in der Tomate arbeiten / *to work in the tomato*.

[51]On French *-isme*, see Roché (2007). For Spanish, Muñoz Armijo (2012).

Fradin, Bernard. 2003. *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Hilger, Marie-Elisabeth. 1982. Kapitalismus. In Otto Brunner, Werner Conze & Reinhart Kosellek (eds.), *Geschichtliche Grundbegriffe. Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*, vol. 3, 442–451. Stuttgart: Klett-Cotta.

Höfer, Anette. 1986. Financier, banquier, capitaliste. In Rolf Reichardt & Eberhard Schmitt (eds.), *Handbuch politisch-sozialer Grundbegriffe in Frankreich 1680-1820*, vol. 5. Munich: Oldenburg.

Muñoz Armijo, Laura. 2012. *La historia de los sufijos* -ismo e -ista. San Millán de la Cogolla: Cilengua.

Passow, Richard. 1927. *Kapitalismus. Eine begrifflich-terminologische Studie*. 2nd edition. Jena: Fischer.

Rainer, Franz. 1998. Avatars sémantiques et migrations européennes de *capitaliste* jusqu'au milieu du XIXᵉ siècle. *Zeitschrift für französische Sprache und Literatur* 108. 8–17.

Rainer, Franz. to appear. Étude diachronique sur l'emploi adjectival des noms en *-iste*. *Zeitschrift für französische Sprache und Literatur* 126.

Roché, Michel. 2007. Logique lexicale et morphologie : La dérivation en *-isme*. In Fabio Montermini, Gilles Boyé & Nabil Hathout (eds.), *Selected proceedings of the 5th Décembrettes: Morphology in Toulouse*, 45–58. Somerville: Cascadilla Press.

Silberner, Edmund & Lucien Febvre. 1940. Mots et choses: le mot capitalisme. In *Annales d'histoire sociale*, vol. 2, 133–134.

't Hart, Marjolein. 1993. *The making of a bourgeois state. War, politics and finance during the Dutch revolt*. Manchester: Manchester University Press.

The Oxford English Dictionary. 1991. Clarendon Press.

Thiele, Johannes. 1974. Zur Entwicklung des lexikalischen Feldes *socialisme/socialiste*. In Rudolf Große (ed.), *Beiträge zur Soziolinguistik*, 171–185. Halle: Niemeyer.

Wolf, Hans Jürgen. 1972. It. *romanista, quattrocentista*: Anwendungsbereiche eines gelehrten Suffixes im Italienischen, Spanischen und Französischen. *Romanische Forschungen* 84. 314–367.

Part II

Lexeme Formation Rules

Chapter 4

Nom et/ou adjectif ? Quelle catégorie d'output pour les suffixés en *-iste* ?

Delphine Tribout
Université de Lille

Dany Amiot
Université de Lille

Cet article aborde la question de la catégorie des construits morphologiques, en particulier le cas des suffixés en *-iste*. Ceux-ci ont la particularité, pour la plupart, d'être ambigus du point de vue de la catégorie dans la mesure où ils peuvent être noms et/ou adjectifs. Nous montrons qu'il existe deux types de suffixés en *-iste* : les uns sont fondamentalement des noms, les autres sont fondamentalement des adjectifs, qui peuvent néanmoins être employés comme noms sous certaines conditions. Pour ce dernier cas nous proposons une analyse en termes de coercion.

1 Introduction

Cet article se focalise sur les catégories construites par la suffixation en *-iste*. Celle-ci soulève en effet des questions intéressantes car les dérivés qu'elle sert à former semblent appartenir à deux catégories différentes, celles du nom et de l'adjectif.

Les dérivés en *-iste* ont déjà fait l'objet de plusieurs études, notamment par (Dubois 1962, Corbin 1988, Roché 2011). Notre étude se distingue des précédentes dans la mesure où nous nous focalisons ici sur les catégories d'output de la suffixation en *-iste*. En cela nous adoptons un point de vue différent de celui de (Roché 2011) qui met l'accent sur la sémantique de la suffixation, indépendamment des catégories impliquées. Nous nous intéressons de notre côté aux rapports catégoriels des dérivés en *-iste* qui peuvent souvent être adjectifs et noms. Puisque ces dérivés sont le produit d'une construction morphologique, on peut se demander si une catégorie est première, construite par la morphologie, et à partir de laquelle serait obtenue l'autre catégorie. Si c'est le cas, se posent alors deux questions : l'identification de la catégorie première et le mode de formation de l'autre

catégorie. On peut au contraire envisager une construction des deux catégories en parallèle, ou encore s'interroger sur une indétermination catégorielle des construits. C'est à ces questions que nous nous proposons de répondre.

Nous ne mettrons pas en regard, dans cet article, les dérivés en *-iste* avec les dérivés en *-isme* pour différentes raisons. D'une part, la question des rapports entre les suffixations en *-iste* et en *-isme* a déjà été traitée, notamment par (Corbin 1988) et plus récemment et avec beaucoup de détails par (Roché 2011). D'autre part, pour la question qui nous intéresse, c'est-à-dire celle des rapports entre catégories adjectivale et nominale des dérivés en *-iste*, analyser les suffixés en *-iste* comme dérivés ou construits parallèlement aux suffixés en *-isme* ne résout pas le problème. Enfin, il existe un certain nombre de dérivés en *-iste* qui ne présentent aucun correspondant en *-isme*, par exemple CHIMISTE, FLEURISTE, GARAGISTE, PIANISTE, ce qui nous semble justifier l'étude des suffixés en *-iste* indépendamment de leur relation avec la suffixation en *-isme*.

Dans un premier temps nous présentons notre méthodologie de constitution du corpus et d'identification des catégories (§ 2). Puis nous présentons notre analyse des suffixés en *-iste* (§ 3 et 4) et montrons qu'il existe deux cas de figure distincts, tant du point de vue du sens que du point de vue des catégories. Nous montrons que dans le deuxième cas la catégorie adjectivale est première et la catégorie nominale seconde (§ 5). Pour ce dernier cas, après avoir envisagé deux analyses possibles, l'ellipse et la conversion, nous proposons notre propre analyse, en termes de coercion (§ 6).

2 Méthodologie

2.1 Constitution du corpus

Notre étude des noms et adjectifs suffixés en *-iste* se fonde sur les données de *Lexique 3* (http://www.lexique.org/). Ce lexique comprend 135 000 formes fléchies correspondant à 55 000 lemmes. À chaque forme sont associées différentes informations telles que la catégorie, le genre et le nombre pour les noms et adjectifs, le temps, le mode, la personne et le nombre pour les verbes, la transcription phonétique, etc. En plus des informations morphosyntaxiques, *Lexique 3* fournit la fréquence des formes fléchies et des lemmes dans deux corpus, l'un étant un sous-ensemble de textes littéraires récents tirés de Frantext, et l'autre étant un corpus de sous-titres de films.

Pour mener notre étude des noms et adjectifs en *-iste*, nous avons dans un premier temps extrait de *Lexique 3* tous les lemmes se terminant formellement par *-iste* et catégorisés comme noms ou adjectifs, avec leurs fréquences dans les deux corpus. Ces deux fréquences ont été additionnées pour chaque lemme de façon à ne conserver qu'une seule information de fréquence. Dans un second temps, les noms et adjectifs extraits ont été mis en regard de manière automatique afin d'identifier les noms en *-iste* sans correspondant adjectival, les adjectifs en *-iste* sans correspondant nominal, et les cas de paires nom-adjectif. Enfin, nous avons validé manuellement les données afin d'écarter les lexèmes se terminant par *-iste* mais qui ne sont pas construits (par exemple LISTE, PISTE, TRISTE), ainsi que les lexèmes qui sont bien formés au moyen du suffixe mais dont la suffixation

en *-iste* ne correspond pas à la dernière opération morphologique effectuée (par exemple CHIRURGIEN-DENTISTE, EX-GAUCHISTE, PHOTOJOURNALISTE, NÉO-COMMUNISTE). Au terme de la validation manuelle notre corpus d'étude contient, selon l'étiquetage de *Lexique 3* : 277 noms en *-iste* sans adjectif correspondant, 64 adjectifs en *-iste* sans correspondant nominal, et 153 paires nom-adjectif.

Lors de l'examen des données issues de *Lexique 3* l'étiquetage catégoriel des formes en *-iste* nous a paru parfois discutable. En effet, parmi les noms sans correspondant adjectival dans la ressource nous avons trouvé plusieurs lexèmes pour lesquels un adjectif nous semble parfaitement possible et est de surcroît attesté, dans le *TLFi* ou ailleurs. C'est le cas par exemple de ABSTENTIONNISTE, CARRIÉRISTE, CHAUVINISTE, POUJADISTE ou UTOPISTE. À l'inverse, les 64 formes en *-iste* étiquetées comme uniquement adjectivales dans la ressource nous ont semblé pouvoir également être employées comme des noms. Par exemple des lexèmes tels que DUALISTE, FÉDÉRALISTE, RÉFORMISTE ou STRUCTURALISTE peuvent avoir un emploi nominal comme le montrent les exemples (1)-(4) tirés de *Frantext*.

(1) il sollicita les ministres en leur confessant sa vieille amitié pour le **fédéraliste** expiré [...] (Balzac, 1843)

(2) et les **réformistes** ne furent pas les moins acharnés à défendre les formules anciennes [...] (Sorel, 1912)

(3) accréditant ainsi pour longtemps, chez les **structuralistes**, la thèse de l'univocité [...] (Hagège, 1985)

(4) Les résultats que l'on peut obtenir sont très différents de ceux auxquels visaient les **dualistes** anciens [...] (David, 1965)

Nous avons donc eu besoin d'établir des critères afin de déterminer la catégorie des formes en *-iste*.

2.2 Critères catégoriels

Si la distinction entre nom prototypique et adjectif prototypique est clairement établie, il existe néanmoins une zone de flou entre ces deux classes, où les oppositions sont moins tranchées et où la distinction entre catégorie et emploi est plus difficile à établir. Nous présenterons d'abord, très rapidement, les critères des noms et adjectifs prototypiques, puis nous listerons les contextes qui peuvent être ambigus entre les deux catégories.

La grammaire traditionnelle convoque généralement trois critères pour distinguer les catégories nominale et adjectivale : des critères morphosyntaxiques, sémantiques et syntaxiques (distribution et fonctions). Ces différents critères sont résumés dans le tableau 1.

Relativement opératoires pour distinguer les cas prototypiques, ces critères ont souvent été critiqués (cf. par exemple Wierzbicka (1998), Croft (2001, 2002), Dixon & Aikhenvald (2002), Haspelmath (2007) pour ne citer que quelques travaux récents) car ils laissent dans l'ombre de nombreux cas d'usage courant qui enfreignent l'un ou l'autre de ces critères, en particulier les constructions prédicatives (5), que la prédication soit première (5a) ou seconde (5b), et l'épithète détachée (6).

Tableau 1 : Critères catégoriels

Critère	Nom prototypique (ex. *table*)	Adjectif prototypique (ex. *grand*)
morphosyntaxique	genre inhérent, variation en nombre	variation en genre et en nombre
sémantique	dénote des entités	dénote des propriétés
distribution	est précédé d'un déterminant, peut être expansé par un adjectif, un syntagme prépositionnel ou une relative	peut être modifié par un adverbe de degré, peut être suivi d'une expansion sous la forme d'un syntagme prépositionnel ou d'une complétive
fonctions typiques	sujet, COD, COI	attribut, épithète

(5)   a.   Pierre est {intelligent/avocat}.

      b.   J'ai un ami {intelligent/avocat}.

(6)   Pierre, {intelligent comme toujours/avocat de renom}, a signalé que…

Dans ces constructions, en effet, un nom comme AVOCAT peut s'employer sans déterminant et manifeste ainsi le même comportement qu'un adjectif comme INTELLIGENT. Tous les noms ne peuvent cependant pas entrer dans ce type de constructions : un nom comme TABLE ne présente pas la même capacité que AVOCAT, comme le montrent les exemples (7).

(7)   a.   * Ceci est table.

      b.   * J'ai un meuble table.

      c.   * Ce meuble, table nouvellement achetée, est vraiment superbe.

Les noms de profession et de fonction sociale, comme AVOCAT, forment de ce fait une classe spécifique. Ce sont sans doute des noms non prototypiques, mais ils répondent néanmoins à tous les autres critères caractérisant les noms, notamment les critères syntaxiques, distributionnel et fonctionnel. D'autre part, ce comportement caractéristique des noms de profession ou de fonction sociale ne les assimile pas non plus pleinement à des adjectifs : ils ne peuvent notamment pas être coordonnés avec un adjectif qualificatif comme le montre l'exemple (8).

(8)   ?? Pierre est grand et avocat.

Nous avons ainsi considéré comme des noms toutes les formes en *-iste* qui remplissent les critères des noms prototypiques (tableau 1), mais aussi celles qui peuvent être employées sans déterminant dans les contextes (5) et (6) mais ne peuvent pas être coordonnées avec un adjectif qualificatif comme dans le contexte (8).

L'application de ces critères nous a permis d'identifier deux types de formes en *-iste* : celles qui ne sont employées que comme des noms et celles qui sont doublement catégorisées, nom et adjectif.[1] Les Sections 3 et 4 décrivent ces deux cas de figure.

3 Les formes en *-iste* nominales

Les formes en *-iste* ayant un emploi uniquement nominal forment un ensemble relativement homogène du point de vue morphologique. En effet, ces noms en *-iste* dérivent quasiment tous de noms (9). Deux exemples (9b) sont construits formellement sur des adjectifs mais dérivent en réalité d'unités polylexicales de catégorie nominale : le criminaliste étudie le droit criminel (sous-type de droit), et l'interniste étudie la médecine interne (sous-type de médecine).

(9) a. AUBERGISTE (<AUBERGE), CAVISTE (<CAVE), DENTISTE (<DENT), MACHINISTE (<MACHINE), NOUVELLISTE (<NOUVELLE), VIOLONISTE (<VIOLON)

   b. CRIMINALISTE (<DROIT CRIMINEL), INTERNISTE (<MÉDECINE INTERNE)

Dans quelques cas la base est ambiguë entre nom ou verbe (10) mais du point de vue du sens une analyse à partir du nom est toujours possible lorsque le nom est associé à une activité.

(10) ARCHIVISTE (<ARCHIVE/ARCHIVER), CARICATURISTE (<CARICATURE/CARICATURER), CONTORSIONNISTE (<CONTORSION/CONTORSIONNER), COPISTE (<COPIE/COPIER), ILLUSIONNISTE (<ILLUSION/ILLUSIONNER), POLÉMISTE (<POLÉMIQUE/POLÉMIQUER), VOCALISTE (<VOCALISE/VOCALISER)[2]

Enfin, nous avons trouvé un nom formé sur un sigle, CIBISTE (<CB = CITIZEN-BAND), et un autre dérivé d'un verbe ou du nom en *-isme* correspondant : EXORCISTE (<EXORCISER/ EXORCISME).

Du point de vue du sens ces noms sont fondamentalement des noms de métier ou de fonction sociale. Ils correspondent à l'une des deux catégories identifiées par (Wolf 1972), l'autre étant celle des noms de partisans. De façon générale ces noms de métiers en *-iste* n'acceptent pas l'emploi adjectival :

(11) ?? ils sont nombreux à vouloir choisir le **métier garagiste**

En (11) *garagiste* ne semble pas fonctionner comme un adjectif en fonction d'épithète dont le rôle serait de qualifier le nom recteur, mais plutôt comme un nom. Il existe en

---

[1]Bien que menée dans un cadre radicalement différent, cette distinction en deux sous-ensembles rejoint les deux cas de figure identifiés par (Dubois 1962) et (Dubois & Dubois-Charlier 1999).

[2]Dans certains cas la finale du lexème base est tronqué devant le suffixe *-iste*, *a fortiori* si elle comprend déjà un [i]. Ainsi pour POLÉMISTE le segment final *ique* (si la base est nominale) ou *iquer* (si la base est verbale) est tronqué. Cette troncation n'est pas liée à l'ambiguïté catégorielle de la base : elle se retrouve également dans FATALISTE, dérivé de FATALITÉ (ou FATALISME). Elle n'est pas davantage spécifique au suffixe *-iste* et s'observe assez fréquemment en français et avec différents suffixes. À ce sujet voir (Corbin & Plénat 1992).

effet une relation d'hypéronymie/hyponymie entre *métier* (l'hypéronyme) et *garagiste* (l'hyponyme).[3] Il est cependant possible d'en trouver des exemples, comme en (12) :

(12)   Je ne suis pas d'un **tempérament archiviste** (*Le Monde*, 9 février 2008)[4]

Selon Rainer (2016), la possibilité d'employer un nom de métier (nom d'agent dans ses termes) en *-iste* comme adjectif qualificatif dépend de la facilité avec laquelle on peut associer au référent du nom une qualité stéréotypique. Dans le cas de ARCHIVISTE on peut assez facilement attribuer au référent la qualité d'être conservateur et ordonné. Cependant, tous les noms de métier en *-iste* n'offrent pas aussi aisément prise aux stéréotypes. De ce fait, nous ne suivrons pas Rainer (2016) qui considère qu'il existe, en français actuel, un patron bien établi de formation d'adjectifs en *-iste* par conversion morphologique N > A. Une telle analyse ne nous convainc pas dans la mesure où l'emploi d'un nom de métier ou de fonction sociale en position adjectivale ne concerne qu'un petit nombre de noms, et ne semble pas être un processus productif et régulier.

4 Les formes en *-iste* doublement catégorisées

Les suffixés en *-iste* présentant les deux catégories, nominale et adjectivale, forment en revanche une classe moins homogène du point de vue morphologique. En effet, comme l'a remarqué Roché (2011), ils peuvent dériver de noms communs (13a), de noms propres (13b), d'adjectifs (13c) ou de verbes (13d). Ils peuvent également avoir pour base autre chose qu'un lexème, comme des sigles (13e) ou des syntagmes (13f).

(13)   a. ANARCHISTE (<ANARCHIE), CENTRISTE (<CENTRE), CAPITALISTE (<CAPITAL), CYCLISTE (<CYCLE), GAUCHISTE (<GAUCHE$_N$), IDÉALISTE (<IDÉAL$_N$), HUMORISTE (<HUMOUR), NOMBRILISTE (<NOMBRIL), PROGRESSISTE (<PROGRÈS), SEXISTE (<SEXE), TERRORISTE (<TERREUR)

   b. BOUDDHISTE (<BOUDDHA), CALVINISTE (<CALVIN), FRANQUISTE (<FRANCO), GAULLISTE (<DE GAULLE), MARXISTE (<MARX), ORLÉANISTE (<ORLÉANS), SIONISTE (<SION), TROTSKISTE (<TROTSKY)

   c. COMMUNISTE (<COMMUN), LOYALISTE (<LOYAL), MODERNISTE (<MODERNE), POSITIVISTE (<POSITIF), SIMPLISTE (<SIMPLE)

   d. ARRIVISTE (<ARRIVER), CONFORMISTE (<SE CONFORMER), DIRIGISTE (<DIRIGER)

   e. CÉGÉTISTE (<CGT), VÉTÉTISTE (<VTT)

   f. FIL-DE-FÉRISTE (<*fil de fer*), JUSQU'AU-BOUTISTE (<*jusqu'au bout*)

Pour un certain nombre de lexèmes, comme ceux présentés ci-dessous, la base est ambiguë entre nom et adjectif (14) ou entre verbe et nom (15). Selon Roché (2011), dans les cas sous (14) la base formelle (le radical dans les termes de l'auteur) est l'adjectif tandis que le lexème en *-iste* dériverait sémantiquement du nom.

---

[3]Sur les ambiguïtés nom *vs* adjectif en position épithète, cf. par ex. Noailly (1999).
[4]Exemple emprunté à (Rainer 2016).

(14) ABSENTÉISTE (<ABSENCE/ABSENT), EXISTENTIALISTE (<EXISTENCE/EXISTENTIEL), FÉ-
MINISTE (<FEMME/FÉMININ), INDIVIDUALISTE (<INDIVIDU/INDIVIDUEL), ROYALISTE
(<ROI/ROYAL)

(15) ALARMISTE (<ALARMER/ALARME), RÉCIDIVISTE (<RÉCIDIVER/RÉCIDIVE), SÉPARATISTE
(<SÉPARER/SÉPARATION), TRANSFORMISTE (<TRANSFORMER/TRANSFORMATION)

D'un point de vue sémantique les suffixés en *-iste* doublement catégorisés sont en
revanche plus homogènes : en tant qu'adjectifs ils renvoient à des propriétés compor-
tementales, idéologiques, morales ou philosophiques. En tant que noms ils désignent
soit des partisans ou pratiquants d'une idéologie, une philosophie, une discipline ou une
activité (16), soit des habitués d'un certain comportement (17).

(16) AUTONOMISTE, CÉGÉTISTE, CYCLISTE, GRÉVISTE, JANSÉNISTE, MONARCHISTE

(17) ABSENTÉISTE, ALTRUISTE, FATALISTE, MATÉRIALISTE, JE-M'EN-FOUTISTE

Roché (2011) a également mentionné la possibilité pour les dérivés en *-iste* de désigner
des gentilés, comme NORDISTE, et un cas inclassable, celui de UNIJAMBISTE, auquel on
peut ajouter SIMPLISTE.

Enfin, d'un point de vue syntaxique, ces formes en *-iste* doublement catégorisées
semblent se comporter à la fois comme de vrais noms et de vrais adjectifs. Ce sont de
vrais adjectifs par les fonctions qu'elles sont capables d'assumer (cf. critères présentés
en 2.2), mais aussi par la capacité qu'elles ont à prendre les marques de degré, comme
en (18).

(18) a. Le jeune Du Camp devient très **socialiste**. (Flaubert, 1850)

  b. Qui est cet électeur frondeur dans ce territoire fortement **socialiste** ? (Web)

  c. Fournière ne connaissait pas d'âme plus **socialiste** et de cerveau plus fécond
que Leroux. (Web)

En tant que noms, ces formes se comportent également comme de vrais noms : elles
peuvent prendre tout type de déterminant : défini (19a-b), indéfini (19c-d), démonstra-
tif (19e) ou numéral (19f), et peuvent être employées au singulier comme au pluriel.
D'autre part, elles ne semblent manifester aucune « déficience catégorielle » selon les cri-
tères de (Lauwers 2014c) et sont pleinement comptables comme le montre la possibilité
d'une détermination par *plusieurs* (19d) ou par un numéral (19f).

(19) a. Aux élections, il voterait pour le **socialiste**. (Aragon 1936)

  b. du côté de la Bastille où les **socialistes** organisaient un grand rassemblement
(Osmont 2012)

  c. Un **socialiste** se leva, mais un second extravagant l'arrêta de la main. (Malraux,
1937)

  d. plusieurs **socialistes** de Londres étaient venus nous voir pour dissuader Georges
de se marier à l'église (Torrès 1939-1945)

e. il faudrait qu'il devînt le point de ralliement de ces **socialistes** (Barrès 1918)

f. chez les Alsaciens : on avait repéré deux **socialistes** (Sartre 1949)

En outre, ces formes nominales en *-iste* peuvent, comme n'importe quel nom, être modifiées par un adjectif, un syntagme prépositionnel ou une relative (20) et assumer toutes les fonctions nominales (21).

(20) a. le milord Link est détesté de ses collègues pour être partisan de ce **terroriste** anglais (Stendhal, 1835)

b. espérons que ce **réaliste** de profession n'est pas trop romanesque (Sand, 1866)

c. les **fétichistes** qui vénéraient certaines parties de son corps (Duvignaux, 1957)

(21) a. S   les **communistes** suscitent l'admiration (Jablonka, 2012)

b. COD   néanmoins il aimait bien les **communistes** (Osmont, 2012)

c. CdN   au fond de toutes les théories des **communistes** (Proudhon, 1840)

d. CdA   Celui-là […] roide comme un **communiste** (Balzac, 1846)

Les formes en *-iste* doublement catégorisées semblent donc être autant adjectifs que noms. Par conséquent la question du rapport entre les catégories nominale et adjectivale se pose de manière cruciale. La section suivante est consacrée à l'analyse de cette question.

5 Orientation catégorielle

Les formes en *-iste* étant morphologiquement dérivées, plusieurs analyses du rapport catégoriel entre adjectif et nom sont possibles : soit l'une des deux catégories est construite par la suffixation en *-iste* et l'autre est dérivée, et il s'agit alors de déterminer quelle catégorie est première ; soit les deux catégories sont formées en parallèle par la règle de suffixation. Roché (2011 : 92) considère quant à lui que les dérivés en *-iste* sont sous-spécifiés pour la catégorie et que leur emploi nominal ou adjectival est déterminé par le contexte. Nous ne souscrivons pas à cette analyse par indétermination catégorielle et pensons au contraire que les dérivés en *-iste* sont non seulement catégorisés, mais sont en premier lieu des adjectifs et que leur emploi nominal est second. Pour arriver à ce résultat nous explorons deux critères : la fréquence des emplois adjectivaux et nominaux (§ 5.1) et l'émergence de ces deux emplois en diachronie (§ 5.2). Nous analysons ensuite les caractéristiques sémantiques des emplois en tant que noms et en tant qu'adjectifs pour montrer l'antériorité de la catégorie adjectivale (§ 5.3).

5.1 Fréquences

Afin de déterminer l'orientation de la relation entre deux formes homonymes et de catégories différentes, Marchand (1964) propose de se fonder sur la fréquence d'emploi des deux formes. Selon lui, la forme la plus fréquente est première et la moins fréquente est

dérivée. Nous avons donc examiné les fréquences adjectivales et nominales des formes en *-iste* doublement catégorisées dans *Lexique 3*. Ce critère n'a été appliqué qu'aux 153 paires nom-adjectif issues du lexique. Les formes que nous considérons comme doublement catégorisées selon les critères présentés en 2.2 mais qui sont enregistrées dans *Lexique 3* uniquement comme adjectifs n'ont pas pu être prises en compte, leur fréquence en emploi nominal étant évidemment absente de la ressource.

Nous avons tout d'abord regardé la fréquence moyenne en tant que nom et en tant qu'adjectif pour l'ensemble des 153 formes en *-iste* doublement catégorisées : elle est de 1.51 pour les formes adjectivales et de 1.73 pour les formes nominales. La différence est minime, d'autant plus qu'une forme joue un rôle perturbateur, ARTISTE, qui a une fréquence en tant que nom de 86.66, alors qu'elle n'est que de 13.2 en tant qu'adjectif.[5] Dans la majorité des cas en effet (cf. le tableau 2, qui regroupe les fréquences des neuf premières formes en *-iste* de notre corpus), une même forme possède une fréquence plus ou moins identique en tant que nom ou en tant qu'adjectif (ABSENTÉISTE, ANABAPTISTE), sachant que dans certains cas c'est l'emploi nominal qui est un peu plus fréquent (ACTIVISTE), alors que dans d'autres c'est l'emploi adjectival (ALTRUISTE).

Tableau 2 : Fréquence des emplois A et N pour une même forme en *-iste*

Lexème	fréq_A	fréq_N
abolitionniste	0.1	0.24
absentéiste	0.01	0.01
activiste	0.75	1.42
adventiste	0.08	0.27
affairiste	0.15	0.28
alarmiste	0.51	0.28
altruiste	0.86	0.12
anabaptiste	0.42	0.34
anarchiste	4.53	6.48

Du point de vue des fréquences, rien ne nous permet donc d'affirmer qu'une catégorie serait plus fondamentale que l'autre.

5.2 Émergence des catégories en diachronie récente

Nous avons ensuite mené une petite étude en diachronie récente afin de déterminer si les formes doublement catégorisées avaient un emploi préférentiel de nom ou d'adjectif dans leurs premières attestations. L'hypothèse que nous avons faite est que si une forme en *-iste* possède fondamentalement une catégorie conférée par son mode de formation

---

[5]L'évolution diachronique de ARTISTE en fait un lexème tout à fait à part dans la série des termes doublement catégorisés.

morphologique, l'autre catégorie devrait être attestée plus tardivement, et son acquisition devrait se faire progressivement. Pour le vérifier, nous avons sélectionné dans le corpus doublement catégorisé huit formes attestées après 1800, soit FÉTICHISTE (1824), GAUCHISTE (1839), COMMUNISTE (1840), ABSENTÉISTE (1853), PACIFISTE (1902), ROUSSEAUISTE (1912), CENTRISTE (1922) et FRANQUISTE (1936), pour lesquelles nous avons récupéré leurs cent premiers contextes d'apparition dans *Frantext*.

L'analyse des contextes des huit formes étudiées nous a permis de constater que pour chaque forme en -*iste*, les deux catégories sont attestées quasiment simultanément, comme le montrent les exemples (22)-(24). Précisons que ces exemples sont pris dans les toutes premières attestations de ces formes relevées dans *Frantext*.

(22)   a.   la république **communiste** de Platon suppose […] (Proudhon, 1840)

        b.   au fond de toutes les théories des **communistes** (Proudhon, 1840)

(23)   a.   les penseurs théologistes, et même **fétichistes**, l'appliquèrent mieux (Comte 1852)

        b.   la naïve situation des vrais **fétichistes**. (Comte, 1852)

(24)   a.   soutenait un point de vue ultra-**gauchiste** (Queneau, 1937)

        b.   sa bande de petits **gauchistes** (Beauvoir, 1951)

En outre, ces formes en -*iste* se comportent pleinement à la fois comme des noms et comme des adjectifs dès les premières attestations. L'analyse des emplois adjectivaux et nominaux en diachronie ne nous permet donc pas davantage que les fréquences de déterminer si une catégorie est antérieure à l'autre.

5.3 Contraintes sémantiques

Pour finir, nous avons étudié les caractéristiques sémantiques des emplois adjectivaux et nominaux et celles-ci nous conduisent à considérer que la catégorie adjectivale est première et la catégorie nominale seconde. En effet, nous avons observé que l'emploi nominal est beaucoup plus contraint sémantiquement que l'emploi adjectival. Un adjectif comme FANTAISISTE, par exemple, peut s'appliquer à différents types de noms : des noms d'humains (*une personne fantaisiste*) ou d'objets abstraits (*une idée fantaisiste*), et même, bien que plus rarement, des noms d'objets concrets (*un meuble fantaisiste*). Cependant il ne peut être employé comme nom que pour référer à un humain. On pourra dire en effet *un fantaisiste* pour désigner un homme fantaisiste, mais on ne dira jamais, nous semble-t-il, *un fantaisiste* pour parler d'un comportement, ni *une fantaisiste* pour désigner une idée ou une théorie fantaisiste. Ce comportement n'est pas spécifique à FANTAISISTE, il s'observe au contraire de manière systématique pour tous les adjectifs en -*iste* : on peut dire *un {personnage/projet/bâtiment} futuriste*, mais *un futuriste* ne peut désigner qu'un homme ; *les {personnes/thèses} progressistes* sont tous deux possibles mais *les progressistes* désigne uniquement un groupe d'humains… Cette contrainte, très forte, justifie à nos yeux l'antériorité de la catégorie adjectivale et l'orientation adjectif > nom.

Se pose alors la question du passage d'adjectif à nom. On peut en effet se demander quel type de procédé permet ce changement de catégorie. La section suivante passe en revue les différentes analyses possibles du phénomène avant de présenter celle que nous proposons.

6 Formation des noms désadjectivaux

6.1 Ellipse

Une première possibilité serait de considérer que les noms en *-iste* issus d'adjectifs sont formés par ellipse, sur le modèle de l'analyse traditionnelle. C'est également le traitement proposé plus récemment par Borer & Roy (2010), Alexiadou & Iordăchioaia (2013) ou McNally & de Swart (2015) dans le cadre d'analyses plus larges concernant les noms désadjectivaux, qu'ils soient ou non suffixés. Selon cette analyse, *un humaniste* serait obtenu à partir de *un homme humaniste* par ellipse du nom *homme*. Une telle analyse pose toutefois plusieurs problèmes.

Le premier problème est celui du nom ellipsé. Dans les cas clairement identifiés comme de l'ellipse, le nom ellipsé varie selon le contexte. Or, dans le cas des noms en *-iste* désadjectivaux, seul un petit nombre de noms pourraient être ellipsés tels que *homme, femme* ou *personne*.

Se pose ensuite la question du genre du nom ellipsé. Lors de l'ellipse d'un nom dans un syntagme nominal, le genre du nom ellipsé est conservé et est visible sur le déterminant et l'adjectif, comme le montrent les exemples en (25).

(25)   a.   Il y a plusieurs robes dans la vitrine. J'aime beaucoup la bleue.
         b.   À l'animalerie, Paul a choisi une souris grise, et Marie une blanche.

Pour les noms en *-iste*, l'interprétation qui semble la plus naturelle est 'personne qui…'. Or, on ne pourrait expliquer le genre masculin de *un humaniste* si le nom ellipsé était *personne*.

Enfin, l'interprétation des noms en *-iste* pose également problème : ces noms dénotent systématiquement des humains et ne semblent pas pouvoir désigner un autre type d'objet concret. Or, si les noms en *-iste* étaient obtenus par ellipse, ceux-ci devraient pouvoir dénoter n'importe quel type d'entité, comme dans les exemples en (25) où *la bleue* désigne un artefact, tandis que *une blanche* dénote un animé.

Il semble donc que l'analyse par ellipse d'un nom ne permette pas d'expliquer la formation de ces noms en *-iste* issus d'adjectifs.

6.2 Conversion

Une autre possibilité est d'analyser ces noms comme des converts. En effet, la conversion adjectif > nom existe en français (Corbin 1987, Kerleroux 1996) comme dans le cas des exemples en (26).

(26) CALME$_A$ >CALME$_N$, BLEU$_A$ >BLEU$_N$

Corbin (1988) analyse d'ailleurs les noms en *-iste* comme des convertis à partir d'adjectifs. Cette analyse se justifie dans la mesure où les noms en *-iste* montrent toutes les propriétés des noms, comme cela a été présenté en Section 4. Toutefois, ils manifestent aussi des propriétés adjectivales, notamment la possibilité d'être modifiés par un adverbe de degré, comme le montrent les exemples en (27) trouvés sur le Web.

(27)   a. Il reste que Letizia d'Espagne, critiquée par **les très royalistes** et admirée par les moins conventionnels, incarne la princesse moderne par excellence

   b. **les très idéalistes** ne se retrouvent pas facilement ensemble et au contraire se trouvent souvent en plein contentieux

   c. Seuls les esprits étriqués ont jamais pensé que le réel se limitait à ce que nous en percevions ! clament **les plus idéalistes**.

   d. En tête de liste, l'enseignement. **Les plus alarmistes** pourraient imaginer des professeurs purement et bonnement remplacés par des ordinateurs

Or, un nom ne peut normalement pas être modifié par un adverbe, sauf s'il est coercé par une construction prédicative (Lauwers 2014b) comme *femme* dans *Marie fait très femme maintenant*, qui sera discuté dans la section suivante (exemple (29)). Cette faculté à être modifiés par un adverbe montre que les noms de partisans en *-iste* ne sont pas des noms ordinaires. De ce fait, une analyse par conversion ne nous paraît pas satisfaisante car elle ne saurait expliquer cette faculté. En effet, un convert présente toutes les propriétés de la catégorie à laquelle il appartient, comme l'a souligné Kerleroux (1996), mais ne présente normalement pas les propriétés syntaxiques de sa base. C'est pourquoi nous présentons dans la section suivante une analyse alternative.

6.3 Coercion

Pour rendre compte des propriétés à la fois adjectivales et nominales des noms de partisans en *-iste* nous proposons une analyse par coercion. Pour cela nous présentons d'abord les différents types de coercion (§ 6.3.1) avant de montrer en quoi l'« override coercion » permet de rendre compte des particularités du passage d'adjectif à nom qui résistaient aux analyses par ellipse ou par conversion (§ 6.3.2)[6]. Précisons que cette analyse par coercion est similaire à celle proposée par Lauwers (2008, 2014a) pour les noms de propriété désadjectivaux.

6.3.1 Différents types de coercion

Depuis les années 1990, une abondante littérature a été consacrée à la coercion. On peut se reporter par exemple à (Pustejovsky 1991, Jackendoff 1997, Michaelis 2003, Francis & Michaelis 2003, Lauwers & Willems 2011). Comme l'ont établi (Lauwers & Willems 2011 : 1219) « at the basis of coercion, there is a mismatch (cf. Francis & Michaelis 2003) between

---

[6]Nous choisissons de traduire le terme *override* par *forçage*, c'est en effet le terme qui nous a semblé le mieux correspondre à la définition donnée ; cf. *infra*.

the semantic properties of a selector (be it a construction, a word class, a temporal or aspectual marker) and the inherent semantic properties of a selected element, the latter being not expected in that particular context. ».

Audring & Booij (2016) distinguent trois types de coercion : la coercion par sélection, la coercion par enrichissement et la coercion par forçage. Les deux premiers types sont fondamentalement des adaptations contextuelles de traits sémantiques ; la coercion par forçage quant à elle, qui est le type de coercion le plus fort et celui qui possède la portée la plus large, est fondée sur l'« override principle » de (Michaelis 2003 : 9) : « *Override principle. If lexical and structural meanings conflict, the semantic specifications of the lexical element conform to those of the grammatical structure with which that lexical item is combined.* ». Dans la coercion par forçage en effet, c'est le contexte qui prend le pas sur les propriétés (sémantiques, catégorielles ou syntaxiques) de l'item coercé et lui impose son interprétation.

En français, F. Kerleroux, dès le début des années 1990 (cf. notamment Kerleroux 1991, 1996), a proposé une analyse relativement similaire par le biais de la notion de « distorsion catégorielle ». En s'appuyant sur l'opposition opérée par Milner (1989) entre terme et position, elle a en effet rendu compte de cas comme celui de l'exemple (28) où l'adjectif *élégant* est utilisé en position nominale.

(28)   Il est d'un élégant !

Pour elle, c'est l'inadéquation entre la catégorie du terme lui-même (un adjectif) et la position dans laquelle il est employé (dans un syntagme nominal après un déterminant) qui rend compte du comportement et de l'interprétation particulière de *élégant* dans ce contexte. Une telle analyse correspond aussi plus ou moins à celle que propose Lauwers (2014a) pour certains noms abstraits désadjectivaux.

6.3.2 La coercion par forçage (*override coercion*)

L'analyse en termes de coercion est fréquente en Grammaire de Construction pour rendre compte de cas comme celui sous (29) :

(29)   Marie fait très femme

Dans cet exemple, un nom (*femme*) est employé en contexte typiquement adjectival, c'est-à-dire un contexte prédicatif, avec modification par l'adverbe d'intensité *très* (cf. § 2.2). *femme* ne devient pas réellement un adjectif, mais son interprétation, dans un contexte comme celui-ci, va être semblable à celle d'un adjectif : ce qui importe ici, ce sont les propriétés qui lui sont prototypiquement associées.

Une telle analyse peut être facilement transposées aux adjectifs en *-iste* employés comme noms. Nous faisons donc l'hypothèse que ces adjectifs sont coercés en étant intégrés à un syntagme nominal (SN), c'est-à-dire un contexte fait pour être saturé par un nom :

(30)   a.   cas prototypique : [$_{SNcompt}$ Dét N] ⟷ 'SN comptable'
         b.   coercion par forçage : [$_{SNcompt}$ Dét A] ⟷ 'SN comptable'

La représentation, très simplifiée, emprunte aux Grammaires de Construction (notamment (Booij 2010)), pour lesquelles une construction, par exemple un SN, est une association forme/sens : à gauche de la double flèche, entre crochets droits, figure la forme, alors qu'à la droite, encadré par des guillemets simples, figure le sens. En (30b), le fait de placer un adjectif dans une place normalement dévolue à un nom (cf. le cas prototypique illustré par (30a)) confère donc à l'ensemble une interprétation nominale, identique à celle qu'elle aurait si le terme était un nom. Nous avons choisi de préciser que le SN est un SN comptable pour justifier de la sémantique des N en *-iste* −ils dénotent des individus−, et pour justifier de la possibilité qu'ils ont d'être précédés de tous types de déterminants (cf. les ex. (19)).[7]

Avant de mentionner les avantages d'une telle analyse, nous voudrions revenir sur un point, qui concerne leur éventuelle lexicalisation. Le procédé de coercion tel que nous venons de le décrire explique l'apparition des formes nominales issues des adjectifs correspondants. Certaines formes nominales peuvent cependant être utilisées à une fréquence importante et être consacrées par l'usage.[8] Il en est ainsi par exemple de *communiste*, dont la fréquence d'emploi en tant que nom (37.28 si on se base sur *Lexique 3*) est sensiblement identique à celle qu'il a en tant qu'adjectif (36.17 dans *Lexique 3*). Certaines formes nominales sont même devenues nettement plus fréquentes que les formes adjectivales correspondantes, c'est le cas de *terroriste* (N : 19.12 *vs* A : 8.87). De telles formes peuvent alors être lexicalisées en tant que noms, et figurer à ce titre dans des dictionnaires. *Optimiste* (A : 8.4 *vs* N : 1.84), *réaliste* (A : 12.86 *vs* N : 1.02), *intimiste* (A : 0.24 *vs* N : 0.07) ou *fantaisiste* (A : 2.56 *vs* N : 0.97) restent en revanche assez fondamentalement associées à la catégorie de l'adjectif, et la coercion joue sans doute encore pleinement son rôle lorsqu'ils sont employés en tant que noms.

Cette analyse par coercion possède, selon nous, au moins deux grands avantages :

(i) d'une part elle permet d'expliquer la facilité avec laquelle il est possible d'obtenir ces formes désadjectivales : la coercion étant un phénomène syntaxique, cela rend compte du caractère systématique de leur formation ;

(ii) d'autre part elle permet aussi, et surtout, d'expliquer pourquoi les noms désadjectivaux en *-iste* peuvent encore avoir des propriétés adjectivales, notamment être modifiés par un adverbe de degré (par ex. *les très idéalistes* en (27)) : en tant que formes coercées, les désadjectivaux en *-iste* ne sont pas pleinement des noms mais des adjectifs en emploi nominal.

Cette analyse des noms issus d'adjectifs en *-iste* s'intègre à une analyse plus large de l'alternance adjectif/nom, un phénomène présent dans l'ensemble du lexique et que nous avons décrit dans Amiot & Tribout (à paraître) : n'importe quel adjectif, qu'il soit simple (JEUNE, GRAND), morphologiquement construit (AMBITIEUX, PARLEMENTAIRE) ou issu d'un participe (BLESSÉ, PERDANT) peut être employé comme nom pour désigner un

---

[7]Pour rendre compte de la formation et des caractéristiques des noms de propriété désadjectivaux, Lauwers (2014a) avait quant à lui fait l'hypothèse que les adjectifs étaient intégrés à des SN massifs.

[8]Sur le rôle et la fonction de la fréquence, voir par exemple Bybee (2006), Bybee & Thompson (1997), Ellis (2002), Gries (2013).

humain à condition que la propriété dénotée par l'adjectif soit susceptible de caractériser l'humain. L'ambition, par exemple, peut caractériser une personne (ex. *un homme ambitieux*) c'est pourquoi l'adjectif AMBITIEUX peut être utilisé comme nom pour référer à un être humain (*un ambitieux*). À l'inverse, un adjectif comme ARGILEUX semble difficilement pouvoir caractériser un être humain et ne peut donc être employé comme nom d'humain (*??un argileux*).

Par rapport à ce cas général, la spécificité de la suffixation par *-iste* réside dans ses affinités particulières avec l'humain : en témoigne le sémantisme des dérivés nominaux, qui dénotent des noms de métier et de fonction sociale (par ex. DENTISTE, GARAGISTE) ; en témoigne aussi le sémantisme des dérivés adjectivaux, qui dénotent généralement des propriétés relatives à des comportements (ABSENTÉISTE, ALARMISTE, INDIVIDUALISTE), des croyances (BOUDDHISTE, CALVINISTE, JANSÉNISTE), des idéologies (MARXISTE, CAPITALISTE, PROGRESSISTE) etc., c'est-à-dire des propriétés qui sont toutes aptes à caractériser, directement ou indirectement, l'humain. C'est la raison pour laquelle tous les adjectifs en *-iste* sont propres à l'emploi nominal à référence humaine, contrairement à d'autres types de suffixations, comme *−eux* ou *−aire*, dont les dérivés ne possèdent pas tous cette capacité (par ex. ARGILEUX, BUDGÉTAIRE).

7 Conclusion

Dans cet article nous nous sommes intéressées à la suffixation en *-iste* puisque ce procédé de formation de lexèmes soulève des questions peu étudiées jusqu'à présent et qui concernent la relation entre lexèmes construits de forme identique mais de catégories différentes, ici adjectivales et nominales.

Nous avons montré qu'il existe deux types de suffixés en *-iste* :

(i) les formes nominales auxquelles ne correspondent pas d'adjectif. Ces formes sont sémantiquement très homogènes, elles dénotent des métiers ou fonctions sociales. Même si l'emploi adjectival n'est pas totalement exclu pour certains de ces noms, celui-ci reste assez souvent ambigu (apposition ou adjectif épithète ?).

(ii) les formes adjectivales auxquelles correspondent des noms. Celles-ci présentent elles aussi une grande homogénéité d'un point de vue sémantique : en tant qu'adjectifs elles renvoient à des propriétés comportementales, idéologiques, morales ou philosophiques ; en tant que noms elles dénotent les partisans ou les pratiquants d'une idéologie, une philosophie, une discipline, une activité, ainsi que les habitués d'un certain comportement. Pour ces noms, nous n'avons pas repris l'analyse de Roché (2011), à savoir la sous-spécification entre catégories adjectivale et nominale. Nous considérons quant à nous que ces noms, ayant nécessairement une référence humaine, sont issus de la forme adjectivale correspondante. En outre, nous avons montré que ces noms, qui présentent toujours des propriétés adjectivales, sont obtenus par coercion.

Ce traitement par coercion des noms issus d'adjectifs en *-iste* s'intègre à une analyse plus large d'un phénomène observé dans tout le lexique, à savoir que tout adjectif est employable comme nom pour désigner un humain si la propriété qu'il dénote peut caractériser l'humain (Amiot & Tribout, à paraître). Par ailleurs, les noms abstraits issus d'adjectifs homonymes tels que *le beau, l'utile, l'humanitaire...* ont été traités par Lauwers (2008, 2014a) comme des adjectifs coercés dans des emplois nominaux. Notre analyse des noms d'humains s'articule donc parfaitement avec celle de Lauwers et vient ainsi compléter la description des noms homonymes d'adjectifs en français. Enfin, il existe également des noms d'objets obtenus à partir d'adjectifs homonymes tels que COMMODE, COLLANT ou BLEU. Ils diffèrent toutefois des noms d'humains sur deux points : i) ils ne peuvent pas être modifiés par un adverbe ; ii) l'emploi nominal pour désigner des artefacts n'est pas aussi systématique que pour désigner des humains. Ces noms d'artefacts restent à étudier afin de déterminer comment leur description s'articule avec celle que nous avons proposée pour les noms d'humains, ainsi qu'avec celle proposée par Lauwers pour les noms abstraits.

Références

Alexiadou, Artemis & Gianina Iordăchioaia. 2013. Two syntactic strategies to derive (abstract) deadjectival nominalizations. Communication au Workshop on *Adjectives and their nominalizations*, Stuttgart.

Amiot, Dany & Delphine Tribout. à paraître. De-adjectival human nouns in French. In Geert Booij (éd.), *Advances in construction morphology*. Cham, Switzerland : Springer.

Audring, Jenny & Geert Booij. 2016. Cooperation and coercion. *Linguistics* 54. 617–637.

Booij, Geert. 2010. *Construction morphology*. Oxford : Oxford University Press.

Borer, Hagit & Isabelle Roy. 2010. The name of the adjective. In Patricia Cabredo-Hofherr & Ora Matushansky (éds.), *Adjectives : Formal analyses in syntax and semantics*, 85–114. Amsterdam : John Benjamins.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Bybee, Joan & Sandra A. Thompson. 1997. Three frequency effects in syntax. In *23rd annual meeting of the berkley linguistics society*, 378–388.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Max Niemeyer Verlag.

Corbin, Danielle. 1988. Une hypothèse à propos des suffixes *-isme*, *-ique*, *-iste* du français : la troncation réciproque. In Ronald Landheer (éd.), *Aspects de linguistique française. Hommage à Q.I.M. Mok*, 63–76. Amsterdam : Rodopi.

Corbin, Danielle & Marc Plénat. 1992. Note sur l'haplologie des mots construits. *Langue Française* 96. 101–112.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective.* Oxford : Oxford University Press.

Croft, William. 2002. *Typology and universals.* Cambridge : Cambridge University Press.

Dixon, Robert & Alexandra Aikhenvald. 2002. Word: A typological framework. In Robert Dixon & Alexandra Aikhenvald (éds.), *Word: A cross-linguistic typology*, 1–41. Cambridge : Cambridge University Press.

Dubois, Jean. 1962. *Étude sur la dérivation suffixale en français moderne et contemporain.* Paris : Larousse.

Dubois, Jean & Françoise Dubois-Charlier. 1999. *La dérivation suffixale en français.* Paris : Nathan.

Ellis, Nick C. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24(2). 143–188.

Francis, Elaine & Laura Michaelis (éds.). 2003. *Mismatch: Form-function incongruity and the architecture of grammar*. Stanford : CSLI Publications.

Gries, Stefan. 2013. Data in construction grammar. In Thomas Hoffmann & Graeme Trousdale (éds.), *The Oxford handbook of construction grammar*, 93–108. Oxford : Oxford University Press.

Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11(1). 119–132.

Jackendoff, Ray S. 1997. *The architecture of the language faculty*. Cambridge : The MIT Press.

Kerleroux, Françoise. 1991. « il est d'un élégant ! et d'un calme ! » un phénomène de distorsion entre l'appartenance catégorielle des termes et l'étiquette des positions syntaxiques. *Bulletin de linguistique appliquée et générale* 17. 79–116.

Kerleroux, Françoise. 1996. *La coupure invisible : études de syntaxe et de morphologie*. Lille : Presses Universitaires du Septentrion.

Lauwers, Peter. 2008. The nominalization of adjectives in French: From morphological conversion to categorial mismatch. *Folia Linguistica* 42(1-2). 135–176.

Lauwers, Peter. 2014a. Between adjective and noun: Category/function mismatch, constructional overrides and coercion. In Francesca Masini & Raffaele Simone (éds.), *Word classes*, 203–226. Amsterdam : Benjamins.

Lauwers, Peter. 2014b. Copular constructions and adjectival uses of bare nouns in French: A case of syntactic recategorization? *Word* 60. 89–122.

Lauwers, Peter. 2014c. Les pluriels «lexicaux». Typologie quantifiée des déficits de dénombrabilité. *Langue Française* 183. 117–132.

Lauwers, Peter & Dominique Willems. 2011. Coercion: Definition and challenges, current approaches, and new trends. *Linguistics* 49(6). 1219–1235.

Marchand, Hans. 1964. A set of criteria for the establishing of derivational relationship between words unmarked by derivational morphemes. *Indogermanische Forschungen* 69. 10–19.

McNally, Louise & Henriëtte de Swart. 2015. Reference to and via properties: The view from Dutch. *Linguistics and Philosophy* 38. 315–362.

Michaelis, Laura. 2003. Headless constructions and coercion by construction. In Elaine Francis & Laura Michaelis (éds.), *Mismatch : Form-function incongruity and the architecture of grammar*, 259–310. Stanford : CSLI Publications.

Milner, Jean-Claude. 1989. *Introduction à une science du langage*. Paris : Le Seuil.

Noailly, Michèle. 1999. *L'adjectif en français.* Paris : Ophrys.

Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics* 17(4). 409–441.

Rainer, Franz. 2016. *Étude diachronique sur l'emploi adjectival des noms en* -iste. Conférence au Laboratoire de Linguistique Formelle, Paris, septembre 2016.

Roché, Michel. 2011. Quel traitement unifié pour les dérivations en *-isme* et en *-iste*? In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.), *Des unités morphologiques au lexique*, 69–143. Paris : Hermès Science-Lavoisier.

Wierzbicka, Anne. 1998. Anchoring linguistic typology in universal semantic primes. *Linguistic Typology* 2(2). 191–195.

Wolf, Hans Jürgen. 1972. It. *romanista, quattrocentista* : Anwendungsbereiche eines gelehrten Suffixes im Italienischen, Spanischen und Französischen. *Romanische Forschungen* 84. 314–367.

Chapter 5

Les adverbes en *-ment* du français : Lexèmes ou formes d'adjectifs ?

Georgette Dal

Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France

Cet article cherche à déterminer le statut des adverbes en *-ment* du français : s'agit-il de lexèmes résultant de l'application d'une règle de construction de lexèmes, ou de mots-formes relevant du paradigme de l'adjectif ? Contrairement à d'autres langues comme l'anglais, ou, pour ce qui est des langues romanes, l'espagnol ou l'italien, la question a été peu débattue en français dans des travaux récents, à l'exception de Dal (2007). Or, un examen attentif des propriétés de ces adverbes et, dans le même temps, de la règle dont ils sont le produit fait clairement opter pour une analyse flexionnelle. La conclusion est par conséquent que les adverbes en *−ment* constituent des variantes contextuelles d'adjectifs, dont ils sont des mots-formes.

1 Introduction

La séquence *−ment* présente dans des adverbes du français pouvant être mis en relation formelle et sémantique avec un adjectif comme *joyeusement* / *joyeux, prestement* / *preste* ou *timidement* / *timide* est en général tenue pour dérivationnelle, au point qu'elle figure comme telle en bonne place dans les ouvrages universitaires à visée pédagogique (par exemple, Huot 2006, Niklas-Salminen 2015, Gardes-Tamine et al. 2015), sans parler des manuels ou ressources en ligne à destination de jeunes publics où, bien souvent, la formation d'adverbes en *−ment* constitue l'exemple archétypal de dérivation.

Le statut dérivationnel de la règle dont *−ment* est l'exposant –par conséquent, le caractère lexématique des adverbes qu'elle permet de former–, n'est pas davantage remis en cause dans les travaux de recherche, y compris chez les morphologues (voir par exemple Corbin 1982, 1987, van Willigen 1983, Bonami & Boyé 2005, Roché 2010, Boyé & Plénat 2015, Detges 2015, Rainer 2016), même dans un cadre comme celui de la morphologie naturelle dans lequel l'opposition flexion / dérivation n'est pas discrète mais scalaire (pour des points récents sur ce courant, cf. Dressler 2005, Luschützky 2015). Or, si l'on considère attentivement les caractéristiques des adverbes en *−ment* du français, il apparaît

que le caractère dérivationnel de la règle dont cette séquence est l'exposant n'a aucun caractère d'évidence. C'est ce que cherche à (re)mettre en lumière cette recherche, dans le prolongement de Dal (2007).

Le présent chapitre débutera par un état de l'art sur le traitement de quelques homologues des adverbes en *–ment* du français dans plusieurs langues romanes et en anglais. Cet état de l'art sera l'occasion de poser quelques jalons pour la suite. Dans un deuxième temps, j'examinerai si les adverbes en *–ment* du français répondent aux attendus des produits d'une règle de construction de lexèmes. À l'issue de cet examen, il apparaîtra que la réponse est négative sur tous les plans et qu'à l'instar de leurs homologues dans d'autres langues romanes et en anglais, ces adverbes peuvent être tenus pour des variantes contextuelles d'adjectifs instanciant une case du paradigme des adjectifs auxquels ils sont morpho-sémantiquement appariables.

2 État de l'art

Si peu, pour ne pas dire pas, de travaux récents, excepté Dal (2007), s'interrogent sur la nature de la règle associant à un adjectif donné un adverbe en *–ment* en français (son statut dérivationnel est en général asserté sans discussion), celle des règles produisant des adverbes à partir d'adjectifs a fait l'objet de davantage de questionnement dans plusieurs langues du monde. On se concentrera ici sur la suffixation en -MENTE[1] dans plusieurs langues romanes en dehors du français et en *–ly* en anglais, et l'on verra que la question est loin d'être résolue, même dans les travaux les plus récents[2].

2.1 Les adverbes en -MENTE dans les langues romanes (hors français)

La question du statut de la séquence -MENTE des adverbes des langues romanes, en dehors du français, a été abordée dans de nombreux travaux. Quatre hypothèses ont été formulées : l'hypothèse compositionnelle (§ 2.1.1), l'hypothèse dérivationnelle (§ 2.1.2), l'hypothèse de l'affixe syntagmatique (§ 2.1.3) et l'hypothèse flexionnelle (§ 2.1.4).

2.1.1 L'hypothèse compositionnelle

Une hypothèse récurrente est que les adverbes en -MENTE seraient des composés, partant, que -MENTE serait un nom conformément à son étymon latin *mens, mentis* (« esprit »). L'hypothèse a été développée pour l'espagnol (cf., parmi d'autres Bello 1847, Hockett 1958, Seco 1972, Zagona 1990, Kovacci 1999). On la trouve aussi formulée en filigrane pour le catalan et le portugais dans Chircu (2007).

Outre l'argument étymologique, l'argument majeur sur lequel se fondent les partisans de cette hypothèse est la possibilité que présente -MENTE dans certaines langues romanes

---

[1]La notation en capitales -MENTE neutralise ici les réalisations sous les formes *-mente* ou *–ment* selon les langues concernées.

[2]On trouvera dans Ricca (2015) une synthèse très documentée de la question pour d'autres langues du monde.

d'être élidé et mis en facteur commun en cas de coordination d'adverbes, -MENTE étant porté par le premier ou le dernier adverbe de la série selon les langues. La possibilité est attestée au moins en espagnol, catalan et portugais, comme l'indiquent les exemples (1-3) empruntés à la Toile :

(1) (esp.) Inspira **lenta y profundamente** [lentement et profondément].

(2) (cat.) **Ràpidament i silenciosa** [rapidement et silencieusement], l'Elena se'ls va acostar.

(3) (port.) Quantas pessoas foram, **severa e cruelmente** [sévèrement et cruellement], torturadas por se oporem ao regime ?

Certains linguistes, comme Saporta (1990), ont tiré argument de cette possibilité pour voir dans les adverbes en -MENTE des composés endocentriques dont la tête serait le nom -MENTE.

La double accentuation des adverbes en -MENTE, une première fois sur l'adjectif repérable dans leur structure, une seconde sur la séquence -MENTE, est un autre des arguments parfois avancés en faveur de la composition (Saporta 1990, Detges 2015). C'est particulièrement vrai de l'espagnol (cf. 4) où, normalement, un lexème issu d'un processus de dérivation ne comporte qu'un seul accent, tandis que les composés permettent une double accentuation :

(4) (esp.) literàlmènte ; ràpidamènte ; cuidadósamènte

Le dernier argument parfois invoqué, à vrai dire davantage contre l'hypothèse dérivationnelle qu'en faveur de l'hypothèse compositionnelle, est celui de la forme féminine de l'adjectif à laquelle s'adjoindrait -MENTE. Si ce dernier était un suffixe dérivationnel, il ne pourrait pas s'appliquer postérieurement à une règle flexionnelle (on reviendra ultérieurement sur ce point) : or, si la séquence -MENTE n'est pas un suffixe dérivationnel, les adverbes en -MENTE ne peuvent être que des composés, et -MENTE un nom, comme son étymon.

2.1.2 L'hypothèse dérivationnelle

L'hypothèse dérivationnelle, que formulent entre autres Karlsson (1981), Bosque (1989), Varela Ortega (1990) ou Rainer (1996, 2016) à propos de l'adjonction de -MENTE à un adjectif pour former un adverbe, est en général une réponse aux faiblesses de l'hypothèse compositionnelle. Les arguments, dont on trouve une synthèse récente dans Torner 2016, sont en substance les suivants :

(i) la séquence -MENTE présente dans les adverbes des langues romanes n'a plus la valeur pleine du nom latin *mens, mentis* « esprit », et les adverbes qui en sont pourvus peuvent avoir des types sémantiques variés : au moins pour l'italien et l'espagnol, adverbes de manière (*lentamente*), ou de point de vue (*economicamente*), adverbes orientés sujet (*francamente*), etc. ;

(ii) si les lexèmes en -ᴍᴇɴᴛᴇ étaient des composés endocentriques à tête nominale, ils devraient être des noms et non pas des adverbes (cf. aussi Fábregas 2007) ;

(iii) l'adverbe en -ᴍᴇɴᴛᴇ hérite la structure argumentale de l'adjectif que l'on repère dans sa structure (Bosque 1989, Cifuentes Honrubia 2002), comme l'indiquent les exemples espagnols sous (5) empruntés à Fábregas (2007), ce que ne permet pas de prédire l'hypothèse compositionnelle :

> (5)   a. paralelo **a esto** / paralelamente **a esto**
>   b. independiente **de ello** / independientemente **de ello**
>   c. proporcional **al resultado** / proporcionalmente **al resultado**

(iv) le type sémantique de l'adjectif détermine le type sémantique de l'adverbe en -ᴍᴇɴᴛᴇ : les adjectifs relationnels produisent des adverbes de domaine ou de point de vue ; les adjectifs exprimant la manière d'agir d'un agent produisent des adverbes orientés agent ; etc. De la même manière, les restrictions de sélection lexicale de l'adverbe sont corrélées à celles de l'adjectif.

L'hypothèse dérivationnelle n'entre pas en conflit avec la catégorisation adverbiale des séquences en -ᴍᴇɴᴛᴇ à base adjectivale (cf. l'argument (ii) ci-dessus), et est davantage en conformité avec l'héritage, du lexème-base par le lexème-dérivé, de propriétés syntaxiques (cf. iii) et sémantiques (cf. iv). Si elle ne résout pas la variété des types sémantiques d'adverbes en −*ment* (cf. i), du moins n'est-elle pas incompatible avec elle.

2.1.3 L'hypothèse de l'affixe syntagmatique

Reprenant une notion mise au jour par Zwicky (1987), Nevis (1985) et Miller (1992) et principalement appliquée aux clitiques, Torner (2005, 2016) voit dans le statut d'affixe syntagmatique une alternative aux hypothèses compositionnelle et dérivationnelle.

L'hypothèse de l'affixe syntagmatique se fonde sur le caractère hybride de la séquence -ᴍᴇɴᴛᴇ des adverbes de l'espagnol. L'argument majeur réside dans l'application de cette séquence à (ce qui se donne à voir comme) la forme féminine de l'adjectif, autrement dit à une forme flexionnelle construite en syntaxe. Or, selon l'universel 28 de Greenberg (1963)[3], que réinvestit à sa manière l'hypothèse de la morphologie scindée (*split morphology*) développée par Anderson (1977, 1982, 1992) et Perlmutter (1988), la flexion est réputée s'appliquer après la dérivation.

Même si, à la suite de Rainer (1996 : 87), Torner (2005 : 131) convient que ce choix d'une forme féminine est davantage vestigial, étant donné l'étymon de -ᴍᴇɴᴛᴇ, que requis par la syntaxe, il s'agit pour lui d'un argument décisif, qui explique en outre la possibilité, soulignée plus haut, d'une élision de la séquence en cas de coordination d'adverbes. Dans l'hypothèse de l'affixe syntagmatique, il n'y a en fait pas d'élision, mais plutôt un attachement de -ᴍᴇɴᴛᴇ à un syntagme adjectival (Torner 2005 : 132), autrement dit à

---

[3]"If both derivation and inflection follow the root, or they both precede the root, the derivation is always between the root and the inflection" (Greenberg 1963 : 93).

une séquence syntaxique (d'où la notion d'affixe syntagmatique), formée par conséquent postérieurement à l'application d'une marque flexionnelle à l'adjectif.

2.1.4 L'hypothèse flexionnelle

L'hypothèse flexionnelle semble avoir été moins explorée que les hypothèses compositionnelle et dérivationnelle pour expliquer le statut de -MENTE dans les langues romanes.

Pour l'espagnol, on la trouve néanmoins formulée dans Hjelmslev (1928), et, à sa suite, dans Alarcos Llorach (1951 : 85), pour qui la « forma adverbial del adjetivo en *-mente* debe considerarse como un 'casus adverbialis', pues su morfema es exigido por el 'verbo' regente ». Pottier (1966) considère pareillement qu'en espagnol, les adverbes en *-mente* ne sont rien d'autre que la forme que revêt l'adjectif sous rection verbale et, donc, que *-mente* y est une marque casuelle.

En ce qui concerne l'italien, on peut citer Scalise (1990) et Ricca (1998, 2004), même si, au terme de leur examen, ni l'un ni l'autre ne retiennent l'hypothèse flexionnelle.

Selon Scalise (1990), le principal écueil auquel elle se heurte en italien réside dans la productivité limitée de la suffixation en *-mente*, où *productif* est à entendre comme « apte à s'appliquer dès que sont réunies les conditions catégorielles favorables à l'application »[4]. En effet, là où la flexion passe pour être entièrement productive – par exemple, en français, tout adjectif peut être fléchi en nombre –, la dérivation le serait moins. Ce contraste figure en bonne place parmi les très nombreux travaux s'interrogeant sur les critères cherchant à opposer flexion et dérivation (cf., entre autres, Dressler 1989, Scalise 1988, Haspelmath 1996, Blevins 2001, Kilani-Schoch & Dressler 2005, Stump 2005[5], ten Hacken 2014, Štekauer 2015). Or, s'agissant de *-mente* en italien, Scalise (1990) recense plusieurs catégories d'adjectifs qui seraient rétifs à son adjonction. Si, comme lui, l'on exclut le cas des possessifs, démonstratifs, indéfinis, numéraux au motif que leur statut adjectival est discutable, il s'agit, pour l'essentiel (le marquage par un astérisque est le fait de Scalise 1990) :

(a) des adjectifs exprimant des propriétés physiques (*calvo* 'chauve' / **calvamente*), dont les adjectifs chromatiques (*giallo* 'jaune' / **giallamente*),

(b) de l'acception littérale des adjectifs possédant une acception littérale et une acception métaphorique (*aridamente* ne serait possible qu'avec l'interprétation figurée de *arido* 'dépourvu de sentiment'),

(c) de l'acception spatiale des adjectifs présentant une lecture spatiale et une lecture temporelle : ainsi, à l'adjectif *lungo* 'long (dans le temps ou dans l'espace)' correspond bien un adverbe, *lungamente*, mais ce dernier exprime une propriété exclusivement temporelle,

---

[4]On distingue ici cet emploi de la notion de productivité de celui qu'en fait Schultink (1961) (en substance : possibilité, pour les locuteurs d'une langue, de former, de façon non intentionnelle, un nombre en principe infini de nouveaux mots morphologiquement complexes à l'aide d'un procédé donné). Pour un point récent sur la notion de productivité, cf. Gaeta & Ricca (2015) et Dal & Namer (2016).

[5]Stump (2005 : 54) préfère utiliser le terme de *completeness* à celui de *productivity*.

(d) d'un certain nombre d'adjectifs construits : évaluatifs (*leggerino* 'assez léger' / **leggerinamente*), adjectifs de relation en −*acco* (*polacco* 'polonais' / **polaccamente*), en −*ale* (*postale* 'postal' / **postalmente*), en −*ano* (*isolano* 'insulaire' / **isolanamente*), etc., adjectifs en −*bile* à base verbale sous leur forme positive (*utilizzabile* 'utilisable' / **utilizzabilamente*). Pour S. Scalise, 45 des 65 suffixes formant des adjectifs en italien bloqueraient ainsi l'application postérieure de la suffixation en -*mente*, sans qu'il ne s'agisse toutefois d'une impossibilité structurelle catégorique, comme en attestent *naturalmente*, *temporaneamente*, *barbaricamente*, *amabilmente*, etc., que cite Scalise (1990)[6].

Les adjectifs résultant d'un processus de composition seraient pareillement impropres à donner lieu à un adverbe en -*mente* en italien : **dolceamaramente*, **storicocriticamente*, etc.

Se fondant sur ce qu'il considère comme une applicabilité limitée de la suffixation en -*mente*[7], Scalise (1990) rejette par conséquent l'hypothèse flexionnelle et lui préfère l'hypothèse dérivationnelle.

Pour ce qui est de D. Ricca, son rejet de l'hypothèse flexionnelle pour expliquer la suffixation en -*mente* en italien est moins irrémédiable. En effet, à l'issue de l'examen des différentes caractéristiques de cette suffixation, Ricca (1998) conclut qu'elle constitue un bon exemple de cas intermédiaire entre flexion et dérivation, et ce, autant d'un point de vue synchronique que d'un point de vue diachronique. Dans Ricca (2004), il nuance cette position et considère qu'au sein du système morphologique de l'italien, du fait des restrictions de natures morphologique et sémantique auxquelles elle est sujette et malgré sa productivité très élevée (cf. aussi Gaeta 2008), la suffixation en -*mente* relève de la dérivation, même s'il ne s'agit pas là d'une dérivation prototypique (Ricca 2004 : 473).

2.2 Les adverbes en −*ly* de l'anglais

En anglais, la question du statut de la séquence −*ly* figurant dans des adverbes comme *beautifully* ou *rapidly* sous (6) a été abordée de façon récurrente :

(6) a. She sings beautifully.
   b. The birds moved rapidly.

Les discussions portent sur le statut dérivationnel ou flexionnel de la règle à laquelle est associée la séquence −*ly*, à l'exclusion de toute autre hypothèse. Contrairement à ce qu'on a vu pour -MENTE, l'hypothèse compositionnelle n'est en effet pas explorée, malgré l'étymon nominal de −*ly*, *lic*, signifiant « forme, apparence, corps » en vieil anglais (cf. notamment Jespersen 1954, Ricca 2015).

---

[6]On relève également sur la Toile des occurrences de ces séquences marquées comme impossibles par Scalise (1990). Par exemple *polaccamente* (litt. « polonaisement ») : « (...) e il segretario particolare di Giovanni Paolo, un prete polacco dal nome **polaccamente** impossibile ».

[7]Les mêmes impossibilités ont été peu ou prou signalées pour l'espagnol : cf. Egea (1993), Garcia Page (1991), Kovacci (1999), Fábregas (2007).

2.2.1 L'hypothèse flexionnelle

Pour les tenants de la piste flexionnelle, que défendent entre autres Hockett (1958 : 110), Lyons (1968), Sugioka & Lehr (1983), Miller (1991 : 95), Haspelmath (1996 : 49–50), Baker (2003 : 230–235), ou, plus récemment, Giegerich (2012) et Pittner (2015), les arguments sont en substance les suivants[8] :

— la productivité réputée très élevée de la suffixation en −*ly*, que s'accordent à reconnaître tous les travaux qui lui sont consacrés indépendamment du statut qui lui est dévolu. En anglais, tout adjectif est en effet susceptible de donner lieu à un adverbe en −*ly* (cf. notamment Bybee 1985 : 84, Scalise & Guevara 2005 : 159) , sauf quelques cas régulièrement cités : adjectifs auxquels correspond un adverbe irrégulier (ex. : *good* / *well* / **goodly*) et, surtout, adjectifs terminés par −*ly*[9], encore que l'exemple emblématique *sillily* (sur *silly* « bête, stupide »), régulièrement cité comme impossible, soit attesté dans l'*Oxford English Dictionary* ainsi qu'entre autres, dans le *Wester's Online Dictionary*, à côté de, notamment, *burlily, chillily*, *cleanlily, comelily, deadlily, friendlily* ou *ghastlily*. Cette contrainte morphophonologique que Štekauer (2005 : 216) impute au phénomène de dissimilation prohibant la consécution de deux séquences /li/ de part et d'autre d'une frontière constructionnelle, pour autant qu'elle soit avérée, est donc pour le moins faible ;

— le fait que le recours à un adverbe en −*ly* plutôt qu'au lexème adjectival auquel il est apparenté est motivé par la syntaxe, l'adverbe apparaissant dans des contextes non nominaux, autrement dit dans des contextes impropres à accueillir un modifieur dont le statut adjectival est évident. Par exemple, si les verbes *sing* et *move* plus haut en (6) requièrent un adverbe en −*ly*, les noms *song* et *movement* en (7) ne peuvent cooccurrer qu'avec un authentique adjectif :

> (7) a. She sings beautifully. / Her song is beautiful. A beautiful song.
>   b. The birds moved rapidly. / Their movements are rapid. Rapid movements.

— le traitement différencié que nécessiteraient les adverbes en −*ly* relativement aux adverbes dépourvus d'un affixe en anglais, si l'on traitait les premiers comme dérivationnels. Selon Giegerich (2012), un tel traitement aurait pour effet de masquer des généralisations pouvant être faites sur la classe des adverbes non affixés, dans lesquels il voit des adjectifs (on reviendra sur ce point dans le § 3.3.1.) ;

— le caractère mutuellement exclusif de la suffixation en −*ly* et de celles en −*er* et −*est* marquant la comparaison et le haut degré. Cette observation fait conclure

---

[8]On peut encore citer Emonds (1976), Radford (1988), Plag (2003) ou Bassac (2004), qui sont des manuels ayant contribué, en tant que tels, à disséminer la thèse flexionnelle.

[9]Cf. Bybee (1985 : 84–85), Anderson (1992 : 195). Sur l'évitement des adverbes se terminant par la séquence −*lily*, cf. Bauer (1983, 1992, 2001). Pour un examen détaillé des adjectifs se terminant par la séquence /ly/, cf. Bauer et al. (2013 : chap. 15).

à Hockett (1958) que les formes adverbiales en −*ly* relèvent du même paradigme que les formes adjectivales en −*er* et en −*est*, donc que, comme −*er* et −*est*, −*ly* est flexionnel (cf. aussi Giegerich 2012). S'agissant du point d'achoppement que peut constituer la catégorisation comme adverbes des mots en −*ly* relativement aux adjectifs auxquels ils sont liés − la flexion est en effet réputée conserver intègre la catégorie lexicale du lexème sur lequel elle opère −, deux explications sont en concurrence parmi les partisans de la thèse flexionnelle :

— Haspelmath (1996) fait l'hypothèse de l'existence de procédés flexionnels pouvant agir sur la catégorie lexicale des inputs (il parle de *transpositional inflection*), tels les participes présents en allemand ou les adjectifs en position épithète en turc, et étend précisément cette hypothèse aux adverbes en −*ly* de l'anglais (cf. aussi Bybee 1985) ;

— Sugioka & Lehr (1983) mettent en question la pertinence même de la catégorie de l'adverbe (cf. aussi entre autres Fábregas 2014), et considèrent que ce qu'on a coutume de nommer « adverbe en −*ly* » n'est rien d'autre que la forme du paradigme de l'adjectif qui est sélectionnée dans un contexte non nominal.

On reviendra plus longuement sur cette question de la catégorisation comme adverbe dans le § 3.3.1, lorsqu'il s'agira de déterminer le statut dérivationnel ou flexionnel de la règle dont −*ment* est l'exposant en français.

2.2.2 L'hypothèse dérivationnelle

Pour les tenants de la piste dérivationnelle dont font partie Zwicky (1995) − en réponse à Sugioka & Lehr (1983) −, et Payne et al. (2010), que reprend Ricca (2015), les arguments sont les suivants :

— il est faux qu'adjectifs et adverbes en −*ly* correspondants apparaissent dans des environnements syntaxiques mutuellement exclusifs : noms pour les adjectifs, autres environnements pour les adverbes. C'est là l'argument central de Payne et al. (2010), qui considèrent que, à condition d'être postposés aux noms qu'ils modifient, les adverbes en −*ly* sont aptes à figurer dans la fonction de modifieurs de noms, comme, par exemple, *globally* et *internationally* dans les exemples sous (8), repris de cette étude :

(8) a. [The unique role **globally** of the Australian Health Promoting Schools Association], as a non-government organization specifically established to promote the concept of the health promoting school, is described.

      b. The NHS and [other health organisations **internationally**] clearly need methodologies to support benefit analysis of merging healthcare organisations.

— les adverbes en −*ly* n'héritent pas tous de la structure argumentale des adjectifs auxquels ils sont apparentés, comme l'indique le contraste *proud of his daugther* / **proudly of his daughter* ;

— certains types sémantiques d'adjectifs ne donnent pas lieu à la production d'un adverbe en −*ly*. On retrouve, mentionnées ici, les catégories d'adjectifs citées par Scalise (1990) pour l'italien : adjectifs chromatiques ; adjectifs exprimant une propriété sensorielle, sauf à ce que l'adverbe ait une valeur métaphorique (ex. *warmly*) ; etc.

2.3 Discussion

L'état de l'art qui précède a mis en évidence au moins un point : le statut des règles morphologiques produisant des adverbes à partir d'adjectifs dans plusieurs langues romanes et germaniques a donné lieu à des discussions nourries, parfois virulentes, et la question n'est toujours pas résolue. À cet égard, on ne peut qu'être surpris qu'en français, peu de travaux se soient penchés sur le statut de la règle à laquelle ressortit la séquence adverbiale *-ment.*

Il est par ailleurs remarquable que, dans les travaux dont il a été question dans cet état de l'art, l'hypothèse dérivationnelle n'ait jamais été abordée positivement : soit elle constitue une réponse aux faiblesses de l'hypothèse compositionnelle (cf. § 2.1.2), soit elle tempère les généralisations de l'hypothèse flexionnelle (cf. § 2.1.4 et § 2.2), mais elle met rarement, pour ne pas dire jamais, en avant d'arguments irréfutables montrant que les adverbes en -MENTE ou en −*ly* résultent de l'application d'une règle de construction de lexèmes, partant, que ces adverbes sont des lexèmes à part entière.

3 Quel statut pour les adverbes en −*ment* du français ?

Étant admis que l'hypothèse compositionnelle est exclue en français – l'argument de l'élision ou de la mise en facteur commun de −*ment* entre plusieurs adverbes, jugé décisif par les partisans de cette hypothèse en espagnol, ne tient pas pour le français moderne[10] –, l'alternative est la même que pour −*ly* en anglais : flexion ou dérivation ?, à moins que les adverbes en −*ment* du français ne relèvent de l'une de ces « zones grises » (Bybee 1985), indécidables entre flexion et dérivation.

Pour tenter d'apporter des éléments de réponse à cette question, je me propose de reprendre dans ce qui suit les attendus d'une règle de construction de lexèmes. Auparavant, je discuterai de la forme du radical de l'adjectif à laquelle s'attache −*ment* afin d'évacuer cette question de la discussion.

---

[10]Meyer-Lübke (1894 : 638) signale cette possibilité en ancien français au travers de l'exemple « Ainzi fu la guere maintenue Si cruel e si longuement », également cité dans Karlsson (1981 : 60).

3.1 Forme du radical de l'adjectif

La séquence –*ment* du français est réputée s'appliquer à la forme féminine de l'adjectif auquel est apparenté l'adverbe (cf. entre autres Guimier 1996, Molinier & Levrier 2000 : 28–29), autrement dit à une forme fléchie. Comme on l'a vu précédemment, ce même constat effectué pour, entre autres, l'espagnol et l'italien a été porté au crédit de l'hypothèse flexionnelle et de celle de l'affixe syntagmatique, dans la mesure où une règle dérivationnelle est supposée ne pas pouvoir s'appliquer postérieurement à une opération de flexion.

Or, la notion aronovienne de morphome (Aronoff 1994), selon laquelle certaines unités morphologiques n'expriment aucune propriété morphosyntaxique ou sémantique – ce sont de pures formes, ou, selon les termes de Bonami & Boyé (2005 : 82), de « purs objets morphologiques » –, offre une explication élégante et neutre vis-à-vis de l'attribution d'un quelconque statut à la règle à laquelle est associée la séquence –*ment.*

Recourant à la notion de morphome, Bonami & Boyé (2005) font l'hypothèse que les adjectifs du français possèdent un espace thématique constitué de deux thèmes, identiques ou distincts, qui n'expriment aucune propriété morphosyntaxique, et qui servent à construire les cinq formes de leur paradigme : les quatre formes traditionnelles faisant intervenir les catégories de genre et de nombre, plus une forme de liaison du masculin singulier en position prénominale. Le tableau 1, emprunté à Bonami & Boyé (2005), donne les thèmes de quelques adjectifs du français :

Tableau 1 : Espace thématique de quelques adjectifs en français

Lexème	Thème 1	Thème 2
LIVIDE	/livid/	/livid/
SEC	/sɛk/	/sɛʃ/
VIF	/vif/	/viv/
VIEUX	/vjø/	/vjɛj/
NOUVEAU	/nuvo/	/nuvɛl/

En flexion, le thème 1 est utilisé pour le masculin, hors liaison (*arbre sec* ; *regard vif* ; *vieux fauteuil* ; *nouveau manteau*) ; le thème 2 l'est pour le féminin (*branche sèche* ; *riposte vive* ; *vieille ferme* ; *nouvelle tenue*). Pour ce qui est de la forme de liaison au masculin singulier en position prénominale, selon les adjectifs, sont mobilisés le thème 1 (*sec entretien* : [sɛkɑ̃trətjɛ]) ou le thème 2 (*vieil avion* : [vjɛjavjɔ̃]).

Pour rendre compte de la forme du radical des adverbes en –*ment*, la solution, amorcée dans Dal (2007) et largement développée dans Boyé & Plénat (2015), consiste à ajouter un troisième thème à l'espace thématique de l'adjectif en français. Selon les cas, ce troisième thème peut être (i) homophone du thème 2, (ii) homophone du thème 1, (iii) différent des thèmes 1 et 2. Majoritairement, les adverbes en –*ment* mettent en jeu un thème homophone du thème 2 (par exemple, /sɛʃmɑ̃/ relativement à /sɛʃ/), autrement dit le thème qui sert aussi très majoritairement à former le féminin des adjectifs. Cette observation

explique l'assertion récurrente selon laquelle, dans les adverbes, la séquence −*ment* s'appliquerait à une forme fléchie au féminin singulier ainsi, d'ailleurs, que les hésitations des scripteurs lorsque les formes de masculin et féminin sont homophones sans être homographes : à titre d'exemple, /ʒolimã/, que les scripteurs orthographient *joliment* (9 millions d'occurrences sur la Toile au moyen du moteur de recherche Google fin juin 2017) ou *joliement* (300 000 occurrences). D'autres choix de radicaux sont toutefois possibles, comme le montrent Boyé & Plénat (2015) :

— lorsque le thème 1 de l'adjectif se termine par /ã/, c'est à un homophone de ce thème que s'applique le plus souvent −*ment*, modulo la dénasalisation de la voyelle finale (cf. Pagliano 2003) : cf. *méchamment* préféré à *méchantement* (qui compte quand même une petite centaine d'occurrences sur la Toile fin juin 2017), *violemment* préféré, en français moderne, à *violentement*, sauf quand cette finale est précédée d'une nasale ou d'une labiale, auquel cas le thème sélectionné est préférentiellement homophone du thème 2 (cf. *charmantement* préféré à *charmamment*[11]) ;

— dans d'autres cas, le thème 3 est différent des thèmes 1 et 2, et se caractérise par l'émergence d'un /e/ concaténé au thème 2 servant à former le féminin (*obtusément*)[12], ou est simplement imprédictible (*brièvement*) du point de vue synchronique.

Le tableau 2, adapté de Boyé & Plénat (2015), récapitule ces résultats.

Tableau 2 : Proposition d'ajout d'un thème 3 dans l'espace thématique de quelques adjectifs en français.

Lexème	Thème 1	Thème 2	Thème 3
JOLI	/ʒoli/	/ʒoli/	/ʒoli/
SEC	/sɛk/	/sɛʃ/	/sɛʃ/
CHARMANT	/ʃaʁmã/	/ʃaʁmãt/	/ʃaʁmãt/
MÉCHANT	/meʃã/	/meʃãt/	/meʃãm/
OBTUS	/ɔpty/	/ɔptyz/	/ɔptyse/
BREF	/bʁɛf/	/bʁɛv/	/bʁijɛv/

---

[11]Si l'on exclut les emplois en mention et les pages redondantes, *charmantement* (parfois sous la forme *charmentement*) compte environ 160 occurrences sur la Toile au 1er octobre 2016 (ex. : « La pluie continuait de tomber. J'étais **charmantement** abritée »), contre une dizaine pour *charmamment / charmament* (ex. : « Évidemment un peu vieux jeu, **charmamment** démodé »).

[12]L'émergence de ce /e/ n'est pas aléatoire : il apparaît, de façon récurrente, après une consonne nasale (*cochonnément, communément, conformément, opportunément, uniformément...*) ou après une fricative, le plus souvent sifflante, sonore (*concisément, confusément, précisément...*) ou sourde (*densément, expressément...*), plus rarement liquide (*aveuglément*) ou vibrante (*obscurément*). S'agissant du premier cas, l'émergence de ce /e/ pourrait avoir pour objectif de satisfaire la contrainte dissimilative déjà citée. L'option prise ici, comme dans Boyé & Plénat (2015), est que /e/ fait partie du radical. Je renvoie à ce travail pour une argumentation.

La solution de l'ajout d'un troisième thème au paradigme de l'adjectif pour former des adverbes en −*ment*, résumée ici à partir de Boyé & Plénat (2015), est orthogonale à la question du statut de la règle à laquelle est associé l'exposant −*ment*, dans la mesure où tant les règles flexionnelles que les règles dérivationnelles peuvent sélectionner tel ou tel thème de l'espace thématique , de façon exclusive ou privilégiée (cf. Bonami et al. 2009). Elle permet par conséquent d'évacuer de la discussion la forme du radical à laquelle s'adjoint la forme −*ment* et évite de tirer argument de cette forme identifiée, à tort, comme étant un féminin : plus exactement, si le radical affecte le plus souvent la forme d'un féminin, c'est parce que la formation d'adverbes en −*ment*, quel qu'en soit le statut, et la formation du féminin de l'adjectif en français opèrent toutes deux de façon privilégiée sur le thème 2, ou sur un homophone de ce thème.

On note du reste que, sans toutefois mobiliser explicitement la notion de morphème, ten Hacken (2014 : 19) considère pareillement que, pour concilier les données du français et l'universel 28 de Greenberg, une solution est de considérer que, dans *lentement*, *lente* est une variante du radical de l'adjectif. Pour sa part, Ricca (2015 : 1392) recourt à la notion de morphome pour expliquer la voyelle /a/ qui clôt le radical de certains adverbes en -MENTE en italien, portugais et espagnol.

3.2 Attendus d'une Règle de Construction de Lexèmes

Une Règle de Construction de Lexèmes (désormais, RCL) peut être schématiquement définie comme un ensemble de régularités observables entre deux séries de lexèmes dont les uns, les outputs, ont un degré de complexité supérieur aux autres, les inputs.

Selon Fradin (2003), le schéma de représentation d'une RCL relevant du procédé de dérivation est le suivant (Tableau 3) :

Tableau 3 : Schéma de représentation d'une RCL relevant du procédé de dérivation selon Fradin (2003).

Inputs		Outputs
Phonologie 1		Phonologie 2
Syntactique 1	⟺	Syntactique 2
Sémantique 1		Sémantique 2

Ce schéma revient à dire qu'une RCL opère sur trois plans : le plan phonologique, le plan syntaxique et le plan sémantique.

De façon générale, des contraintes de différents types peuvent opérer sur les inputs et sur les outputs. Si l'on exclut les contraintes phonologiques qui opèrent davantage au niveau de tel ou tel lexème (ou ensemble de lexèmes) particulier qu'au niveau de la règle en tant que telle, pour l'essentiel, il s'agit :

— de contraintes sémantiques : chaque procédé constructionnel s'applique à un type sémantique de bases (par exemple, bases exprimant des propriétés, référant à des événements, des parties naturelles, etc.), ou demande des bases qu'il sélectionne

qu'elles-mêmes relèvent (ou ne relèvent pas) d'un certain type sémantique. Pareillement, le sens des outputs est une fonction du sens des inputs, cette fonction se caractérisant par une constante – celle, précisément, qui enregistre la contribution sémantique de la RCL – et par une variable, représentée par le sens de l'input ;

— de contraintes syntaxiques – une RCL s'applique sur un certain type catégoriel de bases et forme un certain type catégoriel de dérivés –, qui peuvent être vues comme une conséquence des contraintes sémantiques (cf., notamment, Dal 2004).

D'autres contraintes peuvent jouer (contraintes historiques, pragmatiques, notamment), nous les laissons de côté ici.

S'agissant de la règle qui forme les adverbes en –*ment* à partir d'adjectifs en français, une fois la question de la forme le plus souvent féminine du radical résolue grâce au recours à la notion de morphome et l'ajout d'un troisième thème dans l'espace thématique de l'adjectif, il s'agit désormais de déterminer si les contraintes en entrée et en sortie dont elle s'assortit satisfont ce que demande une RCL.

3.3 Examen

3.3.1 Contraintes syntaxiques

3.3.1.1 Contraintes syntaxiques d'entrée

La règle dont –*ment* est l'exposant prend très majoritairement en entrée des d'adjectifs (notons cette propriété P_1).

Pour donner un ordre d'idée, le corpus réuni par Pagliano (2003) compte 2746 adverbes dont 2725 formés à partir d'adjectifs ou de participes, soit plus de 99%.

Le 1% restant est constitué d'adverbes figurant :

(i) dans des séquences formulaires du type *X-ment vôtre* ou *X-ment parlant*, comme dans les exemples sous (9) relevés sur la Toile[13] :

    (9)   a. Internet'ment vôtre ; rock'n'roll'ment vôtre ; jazz'ment vôtre ; meuh..ment vôtre

               b. Le script est crade **HTML ment** parlant.

               c. Il n'est pas bizarre, **marketing-ment** parlant, de faire ça.

(ii) dans des créations ludiques, comme *ordinateurement* ou *mousquetairement* sous (10), également empruntées à la Toile :

    (10)   a. Protection contre les maladies **ordinateurement** transmissibles.

              b. Blafard de teint, ses cheveux aplatis, sa barbe pointue et sa moustache « **mousquetairement** » retroussée rutilent comme l'or.

---

[13]Sur la morphologie des séquences en *X-ment parlant* et *X-ment vôtre*, cf. Boyé & Plénat (2015), ainsi que, pour ces dernières, Mora (2007).

(iii) dans des adverbes désanthroponymiques, comme les exemples littéraires *baudelairement* ou *lamartinement* sous (11) (cf. Amiot & Flaux 2005) :

> (11)   a.   Je suis dans un jour où je vois tout idéalement et douloureusement, et enfin, s'il m'est possible de m'exprimer ainsi, **lamartinement** (Sainte-Beuve, Portr. Littér.)
>
> b.   Une manière de fatalité (…) qu'à présent il nomme moins **baudelairement** le train-train de l'existence (Verlaine, *Œuvres posthumes*)

Une hypothèse est que ces séquences soient formées par analogie (cf. Dal 2003) avec des séquences mettant en jeu un adverbe à support adjectival[14] :

— il est assez probable que les formules de politesse ludiques telles celles sous (9a) fassent écho à des formules à support indéniablement adjectival comme *cordialement vôtre* ou *amicalement vôtre* ;

— dans les séquences en *Xment parlant*, les adverbes sont principalement formés à partir d'adjectifs relationnels (cf., relevés sur la Toile, *grammaticalement parlant*, *philosophiquement parlant*, *financièrement parlant*, *culturellement parlant*). S'il est plus difficile de trouver un chef de file pour les séquences sous (9b) que pour celles sous (9a), on peut néanmoins considérer que ces séquences à support adjectival leur aient servi de modèle ;

— dans une séquence comme celle sous (10a), on ne peut pas ne pas remarquer le jeu échoïque avec la séquence quasi-figée *maladies sexuellement transmissibles* (ce même jeu échoïque avec une séquence quasi-figée s'observe pareillement dans, par exemple, « Paysage *ordinateurement* modifié »[15], également relevé sur la Toile) ;

— enfin, tant avec l'exemple sous (10b) qu'avec les désanthroponymiques sous (11), la suffixiformité adjectivale de la finale du nom support (*mousquetaire*, *Baudelaire*, *Lamartine*) est un facteur favorisant l'émergence de l'adverbe (Amiot & Flaux, font une remarque analogue pour les désanthroponymiques) [16]. Lorsque l'anthroponyme n'a pas de finale suffixiforme, la tendance est de transiter par un adjectif relationnel (c'est le cas dans cet exemple relevé sur la Toile : « (…) en mettant **molièresquement** tous les rieurs de son côté »).

---

[14]Dans le cadre de la grammaire de construction, une autre explication, non incompatible avec celle qui est proposée ici, serait que *−ment* sous (9)/(11) force une lecture adjectivale de l'item auquel il est concaténé (cf. Audring & Booij 2016).

[15]L'analogue est bien sûr ici *génétiquement modifié*.

[16]Dans certaines langues, la séquence finale de séquences paraphrasables par « à la manière de X », où X est un nom, est traitée comme un marqueur du cas essif, donc comme flexionnelle (par exemple, en hongrois *−kent* dans *turistakent* « à la manière d'un touriste » ; cf. Ricca 2015 : 1399).

En somme, ce 1% résulterait d'une pression lexicale, et serait formé par analogie avec des adverbes (ou des séquences comportant un adverbe) à support authentiquement adjectival.

Relativement au statut de la règle à laquelle −*ment* est associé, la contrainte d'entrée P_1 − -*ment* s'applique massivement à des adjectifs − n'est pas décisive : si les adverbes en −*ment* sont produits par une règle dérivationnelle, cette dernière prendrait des adjectifs en entrée ; s'ils le sont par une règle de réalisation de lexème (par la flexion, donc), on s'attend à ce qu'ils soient des mots-formes d'une catégorie unique, qui serait en l'occurrence celle des adjectifs.

3.3.1.2 Contraintes syntaxiques de sortie

Admettons donc que les supports des mots en −*ment* soient des adjectifs. Il n'en reste pas moins que ces mots sont catégorisés comme adverbes. Appelons cette propriété P_2. Or, l'une des propriétés régulièrement invoquées pour différencier la flexion de la dérivation est que seules les règles dérivationnelles peuvent former des lexèmes relevant d'une catégorie différente de celle des lexèmes qu'elles prennent en entrée : on tiendrait là l'argument décisif en faveur du caractère dérivationnel de la règle ayant −*ment* pour exposant.

Toutefois, on a vu plus haut que, pour Haspelmath (1996) qui suit en cela la proposition amorcée dans Bybee (1985), la flexion peut avoir un effet sur la catégorie des sorties et que, selon lui, en anglais, la suffixation en −*ly* serait précisément l'une de ces règles flexionnelles transpositionnelles (Scalise 1988 envisage également le cas de règles flexionnelles dont les outputs ne relèveraient pas de la catégorie des inputs). La formation d'adverbes en −*ment* du français pourrait être passible de la même explication.

Par ailleurs, même si l'on récuse cette possibilité, on a déjà souligné plus haut la difficulté à cerner de façon satisfaisante la catégorie de l'adverbe, qui se caractérise, pour le moins, par une très grande hétérogénéité (Ricca 2015), au point que certains linguistes remettent en question son existence même, parfois de façon péremptoire. C'est le cas d'Aronoff (1994 : 10), qui affirme : « I assume without argument that adverbs are adjectives ».

Reprenons les principaux arguments avancés, ou pouvant l'être, en faveur de la remise en cause, totale ou partielle, de la catégorie de l'adverbe.

Pour Giegerich (2012), les arguments sont d'abord morphologiques. Pour lui, en anglais, ce qu'il est convenu d'appeler « adverbes » ne présente aucune propriété morphologique qui distinguerait cette catégorie de celle des adjectifs : il en conclut que les adverbes sont des formes d'adjectifs. Cette « single-category claim », qui vaut tant pour les adverbes en −*ly* que pour les adverbes dépourvus de marque affixale (il fait de ces derniers des adjectifs non fléchis), expliquerait le fait que, contrairement aux catégories du nom, de l'adjectif et du verbe, la catégorie de l'adverbe ne puisse pas servir d'input à une quelconque règle dérivationnelle, compte tenu de l'ordre d'application dérivation, puis flexion[17]. Parallèlement, l'hypothèse d'une catégorie unique réunissant adjectifs et

---

[17] Les contre-exemples apparents qu'il reprend à Payne et al. (2010 : 63) tels *soonish, soonness, seldomness, unseldom* mettent en jeu des affixations qui, précisément, s'appliquent typiquement à des adjectifs.

adverbes expliquerait que si, pour le français, l'on excepte les cas à la marge comme *baudelairement* vus plus haut, aucun adverbe ne dérive de nom ou de verbe, là où, pour les catégories lexicales majeures authentiques que sont les noms, les adjectifs et les verbes, toutes les combinaisons sont deux à deux possibles.

De surcroît, alors que les noms, adjectifs et verbes peuvent servir d'inputs à plus d'une règle dérivationnelle, dans l'hypothèse de l'attribution d'un statut dérivationnel à la suffixation en *−ly*, l'adverbe serait atypique en ceci qu'outre la conversion d'adjectif à adverbe (on reviendra plus loin sur ce point), il ne mettrait en jeu que cette seule suffixation.

La situation est stricto sensu transposable au français : il apparaît que la catégorie de l'adverbe ne sert pas d'input au système constructionnel du français et qu'en sortie, une seule marque, *−ment*, appliquée à la seule catégorie de l'adjectif, serait possible, en plus de la conversion.

Comme pour l'anglais, en faisant de l'adverbe un cas d'espèce de l'adjectif et de *−ment* une marque flexionnelle, la position atypique des adverbes dans le système dérivationnel du français trouve une explication : l'adverbe ne peut pas servir d'input à une règle dérivationnelle, parce que c'est un mot-forme et non pas un lexème, et il ne constitue la sortie que de la catégorie adjectivale, parce qu'il occupe une case du paradigme de cette catégorie.

Pour Giegerich (2012), du point de vue de la flexion, l'adverbe en anglais ne présente pas davantage de propriétés qui le distingueraient de l'adjectif. La variation morphologique en degré est possible pour l'adverbe, mais elle n'affecte que les adverbes dépourvus de *−ly*, et les marques flexionnelles utilisées sont précisément celles que connaît également l'adjectif (*big* : *bigger*, *biggest* ; *soon* : *sooner*, *soonest*). Comme on l'a déjà vu, pour sa part, le fait que les adverbes en *−ly* n'acceptent pas de marquage en degré au moyen de marques flexionnelles s'explique dans l'hypothèse flexionnelle défendue par Giegerich, puisque, en tant que mots-formes, ils occupent une case du paradigme de l'adjectif : les exposants *−er*, *−est* et *−ly* permettant d'instancier des mots-formes du même paradigme, ils sont mutuellement exclusifs.

Pour ce qui est du français, la situation est comparable, au moins en partie, dans la mesure où l'adverbe y est réputé invariable. Hummel (2013, 2014) remet en effet en cause l'invariabilité des « short adverbs », en même temps que celle de l'appartenance de ces derniers à la catégorie de l'adverbe. Pour lui comme pour Abeillé & Godard (2004), *gras* dans *manger gras* ou *direct* dans *Pierre et Marie vont direct au café* ne sont pas des adverbes, mais des « adjectifs non marqués » ou « adjectifs en fonction adverbiale ». Son argumentation tout à la fois convoque des arguments diachroniques et exploite des données de corpus actuelles, dans une perspective variationniste. En effet, dans les langues qui connaissent la flexion de l'adjectif comme le français, une tendance observée dans la langue contemporaine dans des emplois non standard renoue avec celle qui a eu cours jusqu'au XVIIᵉ siècle d'accorder les adverbes courts. Cet accord s'observe avec le sujet ou avec l'objet interne, comme on le voit sous (12a), relevé sur la Toile, et (12b), emprunté à Hummel & Gazdik (2014) :

(12) a. Ils jouent forts, et souvent faux, ponctuent les chansons d'exclamations en espagnol, sont d'une bonne humeur resplendissante et communicative.

 b. Je suis sur le point d'arrêter nette ma conso de cannabis.

L'hypothèse de M. Hummel est qu'il s'agit là d'une stratégie destinée à maintenir la cohésion thématique au sein de la prédication avec l'un des arguments, interne ou externe, du verbe. On observe toutefois que cet accord est favorisé par une homophonie de l'adverbe court et de la forme de masculin de l'adjectif. Ainsi, si l'on relève sur la Toile des exemples comme ceux sous (13) :

(13) a. Ce que la nouvelle recherche suggère, c'est que les bénéfices de la course à pieds pourraient s'arrêter nets plus tard dans la vie.

 b. En juin 2011, un généalogiste amateur originaire de l'Aude et résidant depuis quelques années dans l'Hérault a vu ses recherches piétiner pour s'arrêter nettes.

des requêtes telles « joue(nt) forte(s) », « joue(nt) fausse(s) » ramènent beaucoup moins de résultats utiles[18].

Quoi qu'il en soit, l'adverbe court ne se distingue en français par aucune marque flexionnelle qui lui serait exclusive : soit, dans une perspective normée de la langue, il est invariable ; soit, dans une perspective plus en prise avec l'usage, il recourt aux marques flexionnelles de genre et nombre de l'adjectif.

Du point de vue de la syntaxe, lorsque le degré est exprimé syntaxiquement, de nouveau, adjectifs et adverbes partagent les mêmes marqueurs. Ce qui vaut de l'anglais – les deux peuvent remplacer X dans, par exemple, le comparatif « more X than », et admettent les mêmes modifieurs adverbiaux : par exemple, *very expensive* / *very quickly* ; *too big* / *too slowly* – vaut aussi du français. Dans les exemples attestés ci-dessous, les marqueurs *très, plutôt*, *un peu, extrêmement* portent aussi bien sur des adjectifs (14) que sur des adverbes, avec ou sans *−ment* (15) :

(14) a. Il faut généralement agir de façon **très** stupide pour se retrouver exilé ici.

 b. Même s'il était **plutôt** maigre, **plutôt** petit et ma foi **un peu** ridicule, je pouvais imaginer que (...)

 c. Pourquoi mes muscles sont **extrêmement** douloureux après l'exercice ?

(15) a. Nous nous levâmes **très** tôt, nous fûmes **très** rapidement habillées.

 b. L'ensemble contrastait **plutôt** désagréablement avec le reste de la demeure.

 c. On s'est engagé **un peu** vite, sans évaluation suffisante des impacts sur la santé.

 d. J'ai été affecté **extrêmement** douloureusement par tout cela.

---

[18]À titre d'exemple, en juillet 2017, « jouent fausses » ramène une trentaine de résultats utiles contre environ 450 pour « arrêtent nettes ».

En conclusion, il apparaît que, pas plus que P_1, P_2 n'est irréfutablement décisive quant au statut dérivationnel de la règle à laquelle ressortit l'exposant *-ment* : certes, les séquences en *−ment* sont des adverbes, mais on vient de voir que la pertinence même de la catégorie de l'adverbe comme catégorie distincte de celle de l'adjectif peut être mise en cause sous de nombreux aspects, et que, si l'on considère qu'en récuser l'existence est excessif, l'hypothèse transpositionnelle, qui pose que la flexion peut produire des séquences ne relevant pas de la catégorie de ce sur quoi elle s'applique, affaiblit l'hypothèse P_2.

Examinons dans ce qui suit si les contraintes sémantiques sont davantage décisives.

3.3.2 Contraintes sémantiques

3.3.2.1 Contraintes sémantiques d'entrée

La règle qui forme des adverbes en *−ment* en français peut s'appliquer à des types sémantiques d'adverbes variés :

— adjectifs qualificatifs exprimant une propriété : *étrange / étrangement* ; *sale / salement* ;

— adjectifs dits relationnels, mettant en relation le référent du nom sur lequel ils sont construits et le référent de leur nom recteur, comme en témoignent sous (16) les adverbes relevés sur la Toile pouvant être mis en relation avec un adjectif en *−al, −aire*, *−el*, *−esque*, *−ien*, *−ique* et *−if*[19] :

(16)   a. La France n'est-elle pas déjà **présidentiellement** rayonnante ?

        b. Il n'y a pas de frontières, du moins pas de frontières définies **géographiquement**.

        c. Si j'avais su que commander à La Redoute impliquait de se faire spammer à ce point, **électroniquement** et **postalement**, je dormirais encore sur mon matelas.

        d. Les 10 Chefs qui ont marqué **mondialement** l'Année gastronomique 2014.

        e. En effet, c'est un mandarin qui a vécu **insulairement** (un peu comme le français de Québec par rapport à la France).

        f. (…) en mettant **molièresquement** tous les rieurs de son côté.

        g. (…) Ou si, **rabelaisiennement** nourri d'un savoir immense, (…)

        h. Un nouveau fléau guetterait les jeunes : les maladies transmises **auditivement**.

S'agissant des adjectifs qualificatifs, il a toutefois été souligné, notamment pour l'italien (cf. § 2.1.4) et pour l'anglais (cf. § 2.2.2), que certains types sémantiques d'adjectifs sont rétifs à l'adjonction d'un exposant adverbialisateur. L'observation a été faite en

---

[19]Sur la productivité des adverbes en *−ment*, cf. Molinier (1992).

particulier pour les adjectifs chromatiques et, plus généralement, pour les adjectifs exprimant des propriétés physiques ou sensorielles.

En premier lieu, pour se limiter ici aux seuls chromatiques, on remarquera qu'il ne s'agit pas là d'une impossibilité structurelle, comme le montrent les exemples relevés sur la Toile sous (17)[20], dans lesquels, contrairement à des adjectifs lexicalisés comme *vertement*, *blanchement* ou *noirement* qu'atteste le *Trésor de la Langue française*, les séquences en *−ment* présentent bien la valeur chromatique de leur adjectif support :

(17) a. Bientôt la machine [la guillotine] aura sans doute déclenché son couperet : la vie d'une vieillarde et de deux gamins se répandra **rougement**.

     b. Les puces de Cugnat avaient dû aller chercher ailleurs un abri et le charbonnier ne montrerait plus jamais le bout **violettement** épaté de son nez.

     c. Tout jeune, il avait trouvé sa voie : vagabonder sur les fortifications dont les talus, **jaunement** verdis de gazon brûlé par le soleil, viennent mourir près du viaduc.

En second lieu, plutôt que de considérer, comme Scalise (1990) ou Ricca (2015) pour l'italien, que l'obstacle vient d'une incompatibilité entre le sens de l'adjectif et les contraintes sémantiques que fait peser sur ses inputs la règle dont *−ment* est l'exposant, je réitère l'hypothèse faite dans Dal (2007) que la rareté d'adverbes en *−ment* à valeur chromatique et, plus généralement, en lien avec un adjectif exprimant une propriété physique ou sensorielle, tient au fait que, si l'on admet que la caractéristique des adverbes en *−ment* est d'émerger dans des contextes non nominaux, dans la mesure où ce à quoi renvoient une phrase, un verbe, un adjectif ou un adverbe n'a pas d'extension spatiale, on peut difficilement lui associer des propriétés physiques ou sensorielles. En somme, je rejoins Fábregas (2007), qui considère que, les adjectifs de couleur ou de forme étant fortement associés à des entités physiques (Quine 1960), il est attendu que les adverbes en *−ment* correspondants, voués de ce fait eux aussi à exprimer des propriétés chromatiques ou physiques, trouvent peu de contextes non nominaux dans lesquels émerger. La contrainte ne tient donc pas à la morphologie en tant que telle, mais est purement sémantique. Elle ne diffère guère de l'impossibilité d'utiliser un adjectif chromatique avec un nom ne référant pas à une entité physique, en préservant la valeur chromatique initiale de l'adjectif : le fait qu'une délibération puisse difficilement être dite violette ou un exploit marron ne signifie pas pour autant que *violet* ou *marron* ne sont pas des adjectifs.

La règle à laquelle ressortit *−ment* ne semble donc pas faire peser de contraintes sémantiques sur les lexèmes qu'elle prend en entrée, les impossibilités, toutes relatives, pointées pour certains types sémantiques d'adjectifs pouvant s'expliquer sans en faire supporter la responsabilité à la morphologie.

3.3.2.2 Contraintes sémantiques de sortie

Du point de vue des sorties, il ne semble pas davantage que l'on puisse définir de fonction sémantique qui soit commune à l'ensemble des adverbes en *−ment*. En effet, comme

---

[20]Sur les adverbes à valeur chromatique, cf. Mora Millan (2005).

le remarquent Plag (2003 : 196) pour l'anglais et Fábregas (2007 : 6) pour l'espagnol, la règle à laquelle la séquence *−ment* est associée n'encode pas de signification lexicale particulière, et l'adverbe garde intègre le sens de l'adjectif auquel il correspond. Plus précisément, aux adjectifs exprimant des qualités correspondent des adverbes classiquement rangés parmi les adverbes de manière (18) ; aux adjectifs à sens relationnel correspondent des adverbes de point de vue ou de domaine (cf. (19) qui reprend (16b)) :

(18)   Il déploie joyeusement sur la toile ses émotions et ses visions avec une belle énergie.

(19)   Il n'y a pas de frontières, du moins pas de frontières définies géographiquement.

Le cas des adverbes dits de phrase peut sembler démentir cette constante.

Molinier (1990) définit les adverbes de phrase, desquels il propose une typologie[21], comme croisant les deux propriétés suivantes : (i) pouvoir figurer en tête de phrase négative ; (ii) ne pas pouvoir être extraits dans *c'est … que.* Ainsi, dans (20), *sincèrement* et *étrangement* sont des adverbes de phrase :

(20)   a. Sincèrement, je ne pensais pas qu'un groupe pareil s'intéresserait un jour à moi.
        b. Étrangement, le chasseur ne semblait pas du tout gêné par l'odeur.

Certains adverbes de phrase peuvent être homomorphes d'un adverbe de manière. C'est le cas des adverbes de (20), comme le montrent les exemples relevés sur la Toile sous (21) :

(21)   a. Si tu t'estimes sincèrement dans ton bon droit, (…)
        b. À l'accueil de l'hôtel, la réceptionniste le regarde étrangement.

D'autres, tel *certainement,* ne semblent pouvoir être utilisés que comme adverbes de phrase, même si, pour Molinier (1990), ils ont pu connaitre un emploi comme adverbes de manière jusqu'au XIXe siècle.

La difficulté que posent ces adverbes relativement à l'assertion selon laquelle l'adverbe garde intègre le sens de l'adjectif auquel il correspond et, en particulier, qu'à un adjectif qualificatif fait écho un adverbe de manière est qu'elle ne prédit pas l'existence des adverbes de phrase, ni la possibilité d'adverbes présentant un double emploi comme ceux sous (20) et (21). Une façon de résoudre cette difficulté est de considérer que, de quelque type qu'elle soit, l'opération d'ajout de la séquence *−ment* à un adjectif est transparente sémantiquement, mais qu'une autre opération, indépendante de la première, permet d'employer les adverbes en *−ment* comme des adverbes de phrase. Pour Lamiroy & Charolles (2004), cette seconde opération relève du phénomène de pragmaticalisation,

---

[21]Il opère une première dichotomie entre adverbes conjonctifs, qui requièrent un contexte gauche (*subséquemment, semblablement...*) et adverbes disjonctifs, qui n'imposent pas cette condition. Ces derniers sont à leur tour répartis entre disjonctifs de style (*honnêtement, franchement*), disjonctifs d'attitude – eux-mêmes classés en disjonctifs d'habitude, évaluatifs et modaux –, disjonctifs d'attitude orientés sujet.

qu'ils définissent comme le passage de la composante grammaticale à la composante pragmatique ou discursive du langage.

Quoi qu'il en soit, si l'hypothèse flexionnelle n'offre pas de meilleure explication à ce phénomène, l'hypothèse dérivationnelle y achoppe tout autant.

En bref, l'adjonction de la séquence −*ment* à un adjectif ne s'assortit pas d'une fonction sémantique repérable, qui serait dévolue à une RCL.

3.4 La formation d'adverbes en −*ment* en français : une règle flexionnelle

3.4.1 Les adverbes en −*ment* : des formes d'adjectifs dans des contextes non nominaux

Au terme de l'examen qui précède, il apparaît que la règle morphologique permettant de former des adverbes en −*ment* en français ne possède, de façon irréfutable, aucune des propriétés attendues d'une règle de construction de lexèmes, aussi bien du point de vue syntaxique que du point de vue sémantique : l'existence même de la catégorie de l'adverbe peut être mise en question, et, sans aller jusqu'à nier la pertinence de cette catégorie, pour le moins, on pourrait être ici face à un cas de transposition flexionnelle ; du point de vue du système, tous les types d'adjectifs semblent pouvoir se voir associer un adverbe en −*ment* ; sémantiquement, l'adjonction de −*ment* préserve le sens de l'adjectif, les adverbes de phrase en −*ment* pouvant être considérés comme constituant des emplois spécifiques d'adverbes de manière.

A contrario, une fois levées les objections auxquelles elle semble achopper, la formation d'adverbes en −*ment* passe avec succès l'ensemble des critères permettant de distinguer la flexion de la dérivation qu'on peut trouver dans, entre autres, Bauer (1997), Dressler (2005), Stump (2005) ou Štekauer (2005) : parmi ces critères, on retiendra ici le fait qu'à tout adjectif peut correspondre un adverbe en −*ment* sans que l'application de cette séquence ne s'assortisse d'une opération sémantique constante repérable.

La conclusion qui s'impose est par conséquent que la formation d'adverbes en −*ment* relève de la flexion, et, partant, que ces adverbes sont la forme que peuvent revêtir les adjectifs dans des contextes non nominaux. Autrement dit, il s'agit là d'un cas d'espèce de flexion contextuelle, pour reprendre la terminologie de Booij (cf. entre autres 1994, 1996 et 2000). Dans un cadre théorique différent, ce résultat rejoint ceux, anciens, de Kuryłowicz (1936 : 83), qui voit en −*ment* un « morphème syntaxique », donc une marque flexionnelle, et de Moignet (1963), dans la perspective de la psychomécanique.

À l'appui de ce résultat, on peut convoquer les exemples sous (22), relevés sur la Toile et/ou partiellement repris de Dal (2007), que l'adverbe soit interne au domaine verbal ou qu'il fonctionne comme modificateur d'un adjectif ou d'un adverbe. Ainsi, le choix de *soigneux* vs *soigneusement* en (a/a') est lié à la catégorie du lexème sur lequel portent ces formes, selon qu'il s'agit d'un nom (a) ou d'un verbe (a'). La remarque vaut en (b/b') avec *réponse rapide* vs *répondre rapidement*, en (c/c') avec *applaudissements bruyants* vs *applaudir bruyamment* et en (d/d') avec *marcheur lent* vs *marcher lentement*. En (e/e'), c'est le contexte adjectival qui déclenche l'émergence de l'adverbe *rapidement* en (e'),

tandis qu'en (f/f'), le déclencheur est l'adverbe *vite* (plus probablement adjectif si on suit Giegerich 2012[22]) . Dans ces divers exemples, les adverbes en *–ment* satisfont la définition, communément admise, qu'Anderson (1992 : 83) donne de la flexion selon laquelle « Inflection thus seems to be just the morphology that is accessible to and/or manipulated by rules of syntax » :

(22)   a.   « Sac à dos » à roulettes d'une grande capacité est semi-rigide afin de permettre un **rangement soigneux** et une protection optimum.

   a'.   L'album photo 26x30 est l'outil parfait pour **ranger soigneusement** vos précieux clichés.

   b.   Vous recevez une **réponse anonyme** et gratuite à vos questions.

   b'.   Plus de 16 000 collégiens et lycéens de 12 à 18 ans **ont répondu anonymement** à un questionnaire détaillé.

   c.   Alors tout le bois résonne des **applaudissements bruyants** des spectateurs et des cris ardents des supporters.

   c'.   Il savait qu'ils ne pouvaient plus remonter, lui répondit Harry, en criant lui aussi pour couvrir le vacarme, mais sans cesser d'**applaudir bruyamment**.

   d.   Je suis un **marcheur lent** qui ne cherche pas la performance mais le plaisir de la marche dans un cadre sublime.

   d'.   Commencez à **marcher lentement**, puis accélérez le pas et marchez rapidement pour les 5 prochaines minutes.

   e.   Il m'avait laissée tomber pour une fille qui se prenait pour un gars et qui était d'une **laideur abominable**.

   e'.   Autant le dire tout de suite, c'est **abominablement laid**.

   f.   Ce qui m'ennuie plutôt c'est la **vitesse atroce** et la stabilité … emm… très « délicate » … mais je réserve mon jugement pour plus tard …

   f'.   Je suis désolée d'avoir mis si longtemps à donner de mes nouvelles mais le temps passe **atrocement vite** non ??

On relève bien sur la Toile quelques exemples marginaux similaires à ceux dont se servent Payne et al. (2010) pour récuser le fait que les adjectifs et les adverbes apparaissent en distribution complémentaire, donc l'hypothèse flexionnelle en anglais (cf. supra, § 2.2.2). Ainsi en (23), l'adverbe émerge dans un contexte nominal et il semble commutable avec un adjectif :

(23)   Dans une pure tradition franco-britannique et dans la signature de cet hommage résolu à l'absurde du comique anglais, nous nous attaquons sans commune mesure à un pan entier de la culture d'une *île* **insulairement** sans frontière terrestre ni avec la Hollande…

---

[22] *Vite* a d'ailleurs été longtemps catégorisé comme adjectif en français, cette catégorisation étant confirmée par le nom de propriété *vitesse*, ainsi que la citation suivante de Vialar, que mentionne le Trésor de la Langue Française *(1971–1984)* : « En tête, c'est Pandore : un chien vite et solide, et qui prend bien les erres sur la feuille ».

Toutefois, si tant est que, dans (23), *insulairement* fonctionne bien comme modifieur post-nominal du nom *île*[23], il n'en demeure pas moins que, dans la grande majorité des cas, adjectifs et ce qu'il est convenu d'appeler adverbes en −*ment* figurent en distribution complémentaire, comme le note pareillement Giegerich (2012 : 356) : les quelques exemples de ce type ne suffisent pas à invalider l'hypothèse flexionnelle.

3.4.2 Quelques autres propriétés

Au moins trois propriétés remarquables des adverbes en −*ment* du français trouvent en outre une explication sous l'éclairage de l'hypothèse flexionnelle :

— la position en clôture de mot de la séquence −*ment.* La remarque a été faite pour l'italien par Ricca (1998), et, pour l'anglais, notamment par Geuder (2000) ainsi qu'indirectement, par l'ensemble des travaux qui listent −*ly* parmi les affixes de niveau 2, selon la généralisation de Siegel (1979)[24]. Or, les règles de réalisation de lexèmes sont réputées s'appliquer postérieurement aux règles de construction de lexèmes – cf. de nouveau l'universel 28 de Greenberg et son incarnation dans l'hypothèse de la morphologie scindée –, du moins quand il s'agit de flexion contextuelle.

Si les adverbes en −*ment* du français constituent la réalisation d'adjectifs dans un contexte non nominal, on comprend que −*ment* se situe en clôture de mot et, puisqu'il s'agit de flexion contextuelle, qu'une forme en −*ment* ne puisse pas servir d'input à une RCL. Faire de −*ment* l'exposant d'une RCL revient en revanche à entériner cette propriété sans l'expliquer ;

— le fait que, pour ce qui est des adverbes de manière, ils ne diffèrent des adjectifs correspondants ni par leur fonction sémantique – les uns et les autres expriment des propriétés, d'individus et événements pour l'adjectif *vs* événements seulement pour l'adverbe[25] –, ni par leur fonction pragmatique, pour reprendre les distinctions opérées par Croft (2003 : 185). En effet, les adjectifs et les adverbes de manière en −*ment* correspondants assument la même fonction pragmatique de modification, même s'il est probable qu'il faille faire une distinction selon que cette modification s'exerce sur un référent de type objet ou de type événement (Croft, c.p.) ;

— le fait que la classe des adverbes en −*ment* soit une classe ouverte, comme l'est, du reste, celle des « short adverbs », au contraire des autres sous-catégories

---

[23]On peut aussi considérer qu'il fonctionne comme adverbe de point de vue glosable par « du point de vue insulaire » et portant sur le syntagme prépositionnel qui suit.

[24]Selon l'*Affix Ordering Generalization* de Siegel (1979), les affixes se répartissent en affixes de niveau 1 et affixes de niveau 2 : selon ce principe, très discuté (par ex. Fabb 1988), un lexème résultant d'une affixation de niveau 2 ne peut pas servir de base à une affixation de niveau 1.

[25]Sans entrer dans le détail, s'agissant des adverbes orientés agents (par ex. *soigneusement*), l'hypothèse a été faite qu'ils possèdent aussi un argument de type individu. La remarque vaut pour les adverbes résultatifs (par ex. *confortablement*), dont l'argument individu serait constitué de l'objet implicite, résultant de l'événement. Pour une argumentation, cf. Geuder (2000) repris en partie dans Bonami et al. (2004).

d'adverbes, réputées fermées. À l'échelle des langues du monde, on oppose en effet les catégories des noms, verbes et adjectifs, qui constituent des classes ouvertes, à toutes les autres (adpositions, conjonctions, articles, etc.), qui constituent des classes fermées, cette partition ouvert / fermé allant de pair avec l'opposition lexème / grammème (catégorie lexicale majeure / catégorie lexicale mineure ; *content word* / *function word* ; etc. Pour une remise en cause partielle, cf. Croft 2000) . Or, dans les langues connaissant la catégorie de l'adverbe, toutes les sous-classes de la catégorie de l'adverbe sont fermées, sauf précisément celle des adverbes de manière (cf. pour l'anglais Haspelmath 2001 : 16544 ; pour le français, Fradin 2003 : 18) . L'hypothèse qui consiste à faire des adverbes de manière et de domaine, avec ou sans −*ment*, des formes d'adjectifs a ceci d'intéressant qu'elle vide la catégorie de l'adverbe de sa seule sous-classe présumément ouverte, et que, dès lors, la catégorie de l'adverbe, si on la maintient, s'homogénéise et devient clairement une catégorie lexicale mineure. On tient, en même temps, une explication plausible au fait que le nombre des adverbes de manière puisse s'accroître : ils tiennent cette possibilité du fait que ces adverbes (avec ou sans marque affixale) instancient une case du paradigme des adjectifs, donc du paradigme d'une catégorie elle-même ouverte.

3.4.3 Conséquence pour l'organisation de la catégorie de l'adjectif

On a vu plus haut que la notion de morphome résolvait la question de la forme le plus souvent apparemment féminine du radical sur lequel −*ment* s'applique, à condition d'ajouter un troisième radical à l'espace thématique de l'adjectif, le plus souvent homophone du thème 2, auquel s'applique l'exposant −*ment*.

Dans l'hypothèse flexionnelle défendue ici, la conséquence est que l'adjectif connaît deux modes de variation : l'un premier en contexte nominal, l'autre second en contexte non nominal, et que le paradigme de l'adjectif en français passe de cinq à six cases :

— en contexte nominal, l'adjectif varie en français selon les catégories traditionnelles du genre et du nombre, avec, en outre, une forme spécifique dédiée à la forme de liaison au masculin singulier (FLMS) selon l'hypothèse Bonami & Boyé (2005) rappelée plus haut (§ 3.1) ;

— en contexte non nominal, si l'on intègre au dispositif les hypothèses d'Abeillé & Godard (2004) et de Hummel (2013, 2014) qui font des adverbes courts des formes d'adjectifs (cf. supra, § 3.3.1), deux formes seraient en compétition dans une même case du paradigme : une forme longue avec −*ment*, une forme courte, sans −*ment*. Sur ce dernier point, en flexion, il existe en effet des cas avérés *d'overabondance* (cf. Thornton 2012), autrement dit de compétition entre plusieurs formes pour une même case de paradigme. Pour le français, c'est par exemple le cas du verbe *asseoir* que citent Apothéloz & Boyé (2004) et qui possède les trois formes [asɛj], [asje], [aswa] pour la même structure de traits {ind, prés, 3pl}. La différence, ici, serait que cette compétition ne serait pas occasionnelle, mais systématique pour la catégorie

de l'adjectif dans son ensemble. Il resterait à explorer plus en avant la compétition en contexte non nominal, ce qui déborde le propos du présent article[26].

Le tableau 4 propose une représentation du paradigme qui intègre la proposition qui précède. Dans la langue standard, l'adverbe court est homomorphe de l'adjectif fléchi au masculin, singulier, hors liaison :

Tableau 4 : Paradigme de l'adjectif en français

	Contexte nominal	
	Singulier	Pluriel
Masculin	A {masc., sg, -liaison } A {masc., sg, +liaison prénominale }	A {masc., plur.}
Féminin	A {fem., sg }	A {fem., pl }
	Contexte non nominal	
	Adverbe en *−ment*	
	Adverbe court	

4 Conclusion

En première intention, dans une théorie qui prend le lexème pour unité de base, la réponse à la question de déterminer le statut des séquences adverbiales en *−ment* est a priori aisée à établir : si ce sont des lexèmes, ce sont des produits d'une règle de construction de lexèmes formant, en tant que telle, des lexèmes différents de ceux qu'elle prend en entrée ; si ce sont des mots-formes, ils résultent d'une règle flexionnelle, servant par conséquent à réaliser des mots-formes des lexèmes sur lesquelles elle opèrent.

S'agissant des adverbes en *−ment* du français, il est apparu que ce qui est cité comme le cas de dérivation par excellence chez de nombreux linguistes et dans de nombreux manuels à vocation pédagogique mérite largement discussion. À la lumière des travaux menés pour d'autres langues, un faisceau d'arguments donne à penser que leurs propriétés sont davantage celles de mots-formes que de lexèmes, et que « adverbe en *−ment* » est une étiquette commode pour nommer la forme que peut revêtir un adjectif dans un contexte non nominal : l'adjonction de *−ment* du français, loin de constituer une zone grise entre flexion et dérivation, serait ainsi pleinement une règle flexionnelle.

Il resterait toutefois quelques points à étayer, énoncés ici sous forme de questions, pour que l'hypothèse flexionnelle emporte définitivement l'adhésion :

---

[26]Une piste à explorer, que me souffle Dany Amiot, serait une distribution complémentaire tendancielle entre les formes courtes, préférentiellement affectées aux adjectifs exprimant une propriété perceptible par les sens (*parler haut / fort / bas ; jouer gros / petit,* etc.) et les formes longues, qui ont peu d'affinité avec ce type sémantique d'adjectifs.

— selon quel(s) critère(s) le choix entre la forme courte et la forme en *−ment* de l'adverbe s'effectue-t-il ?

— existe-t-il d'autres cas avérés, en flexion, de mots-formes s'émancipant du lexème au paradigme duquel ils relèvent ?

Remerciements

J'adresse tous mes remerciements à Dany Amiot, Olivier Bonami, Stéphanie Lignon et Fiammetta Namer, pour leur relecture attentive d'une version précédente du présent chapitre et pour leurs diverses suggestions d'amélioration, dont j'ai tenté de tirer profit au mieux.

Références

Abeillé, Anne & Danièle Godard. 2004. Les adjectifs invariables comme compléments légers en français. In Jacques François (éd.), *L'adjectif en français et à travers les langues*, 209–224. Caen : Presses Universitaires de Caen.

Alarcos Llorach, Emilio. 1951. *Gramática structural*. Madrid : Gredos.

Amiot, Dany & Nelly Flaux. 2005. Les adverbes en *-ment* dérivés des noms propres de personne. In Jan Goes (éd.), *L'adverbe : un pervers polymorphe*, 69–87. Arras : Artois Presses Université.

Anderson, Stephen R. 1977. On the formal description of inflection. In *Proceedings of the 13th meeting of the Chicago linguistics society*, 15–44.

Anderson, Stephen R. 1982. Where's Morphology ? *Linguistic Inquiry* 13(4). 571–612.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge : Cambridge University Press.

Apothéloz, Denis & Gilles Boyé. 2004. Remarques sur la compositionnalité en morphologie. *Verbum* XXVI(4). 375–385.

Aronoff, Mark. 1994. *Morphology by itself : Stems and inflectional classes*. Cambridge : MIT Press.

Audring, Jenny & Geert Booij. 2016. Cooperation and coercion. *Linguistics* 54. 617–637.

Baker, Mark C. 2003. *Lexical categories : Verbs, nouns and adjectives*. Cambridge : Cambridge University Press.

Bassac, Christian. 2004. *Principes de morphologie anglaise*. Pessac : Presses Universitaires de Bordeaux.

Bauer, Laurie. 1983. *English word formation*. Cambridge : Cambridge University Press.

Bauer, Laurie. 1992. Scalar productivity and *lily* adverbs. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1991*, 185–191. Dordrecht : Kluwer Academic Publishers.

Bauer, Laurie. 1997. Derivational paradigms. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1996*, 243–256. Dordrecht : Kluwer Academic Publishers.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge : Cambridge University Press.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford : Oxford University Press.

Bello, Andrés. 1847. *Gramática de la lengua castellana destinada al uso de los americanos. edición crítica de ramón trujillo (1981)*. Tenerife : Instituto de Lingüística Andrés Bello.

Blevins, James P. 2001. Paradigmatic derivation. *Transactions of the Philological Society* 99(2). 211–222.

Bonami, Olivier & Gilles Boyé. 2005. Construire le paradigme d'un adjectif. *Recherches Linguistiques de Vincennes* 34. 77–98.

Bonami, Olivier, Gilles Boyé & Françoise Kerleroux. 2009. L'allomorphie radicale et la relation dérivation-construction. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 103–125. Saint-Denis : Presses Universitaires de Vincennes.

Bonami, Olivier, Danièle Godard & Brigitte Kampers-Manhe. 2004. Adverb classification. In Francis Corblin & Henriëtte de Swart (éds.), *Handbook of French semantics*, 143–184. Stanford : CSLI Publications.

Booij, Geert. 1994. Against split morphology. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1993*, 27–49. Dordrecht : Kluwer Academic Publishers.

Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1995*, 1–16. Dordrecht : Kluwer Academic Publishers.

Booij, Geert. 2000. Inflection and derivation. In Geert Booij, Christian Lehmann & Joachim Mugdan (éds.), *Morphology. An international handbook on inflection and word-formation*, t. 1, 360–369. Berlin : Walter De Gruyter.

Bosque, Ignacio. 1989. *Las categorías gramaticales*. Madrid : Editorial Síntesis.

Boyé, Gilles & Marc Plénat. 2015. L'allomorphie radicale dans les lexèmes adjectivaux en français. Le cas des adverbes en *-ment*. In Eulália Bonet, Maria-Rosa Lloret & Joan Mascaró (éds.), *Understanding allomorphy. Perspectives from optimality theory*, 70–106. London : Equinox Publishing.

Bybee, Joan. 1985. *Morphology. A study of the relation between meaning and form*. Amsterdam : John Benjamins.

Chircu, Adrian. 2007. *L'adverbe dans les langues romanes. Études étymologique, lexicale et morphologique (français, roumain, italien, espagnol, portugais, catalan, provençal)*. Cluj-Napoca : Casa Cărții de Știință.

Cifuentes Honrubia, José Luis. 2002. Sobre la gramaticalización preposicional de los adverbios en *-mente. Archivo de Filología Aragonesa* LIX-LX. 325–337.

Corbin, Danielle. 1982. Le monde étrange des dictionnaires (2) : sur le statut lexicographique des adverbes en *-ment. Lexique* 1. 25–64.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Max Niemeyer Verlag.

Croft, William. 2000. Lexical and grammatical meaning. In Geert Booij, Christian Lehmann & Joachim Mugdan (éds.), *Morphology. An international handbook on inflection and word-formation*, t. 1, 257–263. Berlin : Walter De Gruyter.

Croft, William. 2003. *Typology and universals.* 2nd edition. Cambridge : Cambridge University Press.

Dal, Georgette. 2003. Analogie et lexique construit : quelles preuves ? *Cahiers de Grammaire* 28. 9–30.

Dal, Georgette. 2004. À propos de *Nouvelles approches en morphologie* de Bernard Fradin ; à quelle dimension des catégories les règles de construction de lexèmes sont-elles sensibles ? *Lexique* 16. 231–263.

Dal, Georgette. 2007. Les adverbes de manière en -*ment* du français : dérivation ou flexion ? In Nabil Hathout & Fabio Montermini (éds.), *Morphologie à Toulouse. actes du colloque international de morphologie 4e Décembrettes*, 121–147. Münich : Lincom Europa.

Dal, Georgette & Fiammetta Namer. 2016. Productivity. In Andrew Hippisley & Gregory T. Stump (éds.), *The Cambridge handbook of morphology*, 70–90. Cambridge : Cambridge University Press.

Detges, Ulrich. 2015. The romance adverbs in -*mente* : A case study in grammaticalization. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, 1824–1841. Berlin : de Gruyter Mouton.

Dressler, Wolfgang U. 1989. Prototypical differences between inflection and derivation. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 42. 3–10.

Dressler, Wolfgang U. 2005. Word formation in Natural Morphology. In Pavol Štekauer & Rochelle Lieber (éds.), *Handbook of word formation*, 267–284. Dordrecht : Springer.

Egea, Esteban Rafael. 1993. Restricciones lexicológicas en el uso de los adverbio en -*mente*. In Soledad Varela (éd.), *La formación de palabras*, 282–299. Madrid : Taurus.

Emonds, Joseph E. 1976. *A transformational approach to English syntax : Root, structure-preserving, and local transformations.* New York : Academic Press, Inc.

Fabb, Nigel. 1988. English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* 6. 527–539.

Fábregas, Antonio. 2007. Adverbios en -*mente* y la estructura del adjetivo en español. *Estudios de Lingüística* 21. 103–124.

Fábregas, Antonio. 2014. Adjectival and adverbial derivation. In Rochelle Lieber & Pavol Štekauer (éds.), *The Oxford handbook of derivational morphology*, 276–295. Oxford : Oxford University Press.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie.* Paris : Presses Universitaires de France.

Gaeta, Livio. 2008. Productivity in inflectional morphology. *Morphology* 17. 181–205.

Gaeta, Livio & Davide Ricca. 2015. Productivity. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, 842–858. Berlin : de Gruyter Mouton.

Garcia Page, Mario. 1991. El adverbio en -*mente*. motivación contextual en formaciones léxicas 'anómalas'. *Anuario de Estudios Filológicos* 14. 149–181.

Gardes-Tamine, Joëlle, Antoine Gautier & Florence Mercier-Lecas. 2015. *Cours de grammaire française.* Paris : Armand Colin.

Geuder, Wilhelm. 2000. *Oriented adverbs. Issues in the lexical semantics of event adverbs.* Universität Tübingen thèse de doct.

Giegerich, Heinz. 2012. The morphology of *-ly* and the categorial status of 'adverbs' in English. *English Language and Linguistics* 16. 341–359.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, 73–113. Cambridge : MIT Press.

Guimier, Claude. 1996. *Les adverbes du français.* Paris : Ophrys.

Haspelmath, Martin. 1996. Word-class-changing inflection and morphological theory. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1995*, 43–66. Dordrecht : Kluwer Academic Publishers.

Haspelmath, Martin. 2001. Word classes/parts of speech. In Paul B. Baltes & Neil Smelser (éds.), *International encyclopedia of the social and behavioral sciences*, 16538–16545. Amsterdam : Pergamon.

Hjelmslev, Louis. 1928. *Principes de grammaire générale.* Copenhagen : Bianco Lunos Bogtrykkeri.

Hockett, Charles F. 1958. *A course in modern linguistics.* New York : Macmillan.

Hummel, Martin. 2013. Attribution in Romance : Reconstructing the oral and written tradition. *Folia Linguistica Historica* 34. 1–42.

Hummel, Martin. 2014. The adjective-adverb interface in Romance and English. In Petra Sleeman, Freek Van de Velde & Harry Perridon (éds.), *Adjectives in Germanic and Romance*, 35–71. Amsterdam : Benjamins.

Hummel, Martin & Anna Gazdik. 2014. Le dictionnaire historique de l'adjectif-adverbe : de *aimer haut* à *baiser utile*. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberger, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4ième congrès mondial de linguistique française*, 587–603. Institut de Linguistique Française.

Huot, Hélène. 2006. *Morphologie. Forme et sens des mots du français.* 2e édition. Paris : Armand Colin.

Jespersen, Otto. 1954. *A modern English grammar on historical principles.* London : Allan & Unwin.

Karlsson, Keith E. 1981. *Syntax and affixation. the evolution of* MENTE *in latin and romance.* Tübingen : Max Niemeyer Verlag.

Kilani-Schoch, Marianne & Wolfgang U. Dressler. 2005. *Morphologie naturelle et flexion du verbe français.* Tübingen : Gunter Narr Verlag.

Kovacci, Ofelia. 1999. El adverbio. In Ignacio Bosque & Violeta Demonte (éds.), *Gramática descriptiva de la lengua española*, 705–786. Madrid : Espasa Calpe.

Kuryłowicz, Jerzy. 1936. Dérivation lexicale et dérivation syntaxique. Contribution à la théorie des parties du discours. *Bulletin de la Société de Linguistique de Paris* 37. 79–92.

Lamiroy, Béatrice & Michel Charolles. 2004. Des adverbes aux connecteurs : *simplement, seulement, malheureusement, heureusement. Travaux de Linguistique.* 57–79.

Luschützky, Hans Christian. 2015. Word formation in Natural Morphology. In Peter O. Müller, Ingeborg Ohnheiser & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, 123–144. Berlin : de Gruyter Mouton.

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge : Cambridge University Press.

Meyer-Lübke, Wilhelm. 1894. *Grammatik der romanischen Sprachen. Zweiter Band : Formenlehre*. Leipzig : Reisland.

Miller, George A. 1991. *The science of words*. New York : Scientific American Library.

Miller, Philip. 1992. *Clitics and constituents in phrase structure grammar*. New York : Garland.

Moignet, Gérard. 1963. L'incidence de l'adverbe et l'adverbialisation des adjectifs. *Travaux de Linguistique et de Littérature* 1. 175–194.

Molinier, Christian. 1990. Une classification des adverbes en *-ment*. *Langue française* 88. 28–40.

Molinier, Christian. 1992. Sur la productivité adverbiale des adjectifs. *Langue française* 96. 65–73.

Molinier, Christian & Françoise Levrier. 2000. *Grammaire des adverbes*. Genève : Droz.

Mora Millan, Maria Luisa. 2005. Des adverbes de couleur. In Injoo Choi-Jonin, Myriam Bras, Anne Dagnac & Magali Rouquier (éds.), *Questions de classification en linguistique : méthodes et descriptions*, 227–240. Berlin : Peter Lang.

Mora, Luisa. 2007. L'autocitation à travers la structure *adv-ment parlant*. *Verbum Analecta Neolatina* IX. 279–290.

Nevis, Joel A. 1985. *Finnish particle clitics and general clitic theory*. Ohio State University thèse de doct.

Niklas-Salminen, Aïno. 2015. *La lexicologie*. Paris : Armand Colin.

Pagliano, Claudine. 2003. Frontière morphologique, voyelle nasale et chemin dérivationnel : les adverbes en *-ment* en français. In Lew Zybatow (éd.), *Europa der Sprachen : Sprachkompetenz − Mehrsprachigkeit − translation. Akten des 35. linguistischen Kolloquiums in Innsbruck 2000*. Sonderdruck : Peter Lang.

Payne, John, Rodney Huddleston & Geoffrey K. Pullum. 2010. The distribution and category status of adjectives and adverbs. *Word Structure* 3. 31–81.

Perlmutter, David M. 1988. The split morphology hypothesis : Evidence from Yiddish. In Michael Hammond & M. Noonan (éds.), *Theoretical morphology : Approaches in modern linguistics*, 79–100. San Diego : Academic Press.

Pittner, Karin. 2015. Between inflection and derivation : Adverbial suffixes in English and German. In Karin Pittner, Daniela Elsner & Fabian Barteld (éds.), *Adverbs : Functional and diachronic aspects*, 133–156. Amsterdam : John Benjamins.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge : Cambridge University Press.

Pottier, Bernard. 1966. *Introduction à l'étude de la morphosyntaxe espagnole*. Paris : Ediciones Hispanoamericanas.

Quine, W. V. O. 1960. *Word and object*. Cambridge : The MIT Press.

Radford, Andrew. 1988. *Transformational grammar*. 2nd edition. Cambridge : Cambridge University Press.

Rainer, Franz. 1996. Inflection inside derivation : Evidence from Spanish and Portuguese. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1995*, 83–91. Dordrecht : Kluwer Academic Publishers.

Rainer, Franz. 2016. Derivational morphology. In Adam Ledgeway & Martin Maiden (éds.), *The Oxford guide to the Romance languages*, 513–522. Oxford : Oxford University Press.

Ricca, Davide. 1998. La morfologia avverbiale tra flessione e derivazione. In Paolo Ramat, Giuliano Bernini, Pierluigi Cuzzolin & Piera Molinelli (éds.), *Ars linguistica. Studi offerti da colleghi ed allievi a Paolo Ramat in occasione del suo 60° compleanno*, 447–466. Rome : Bulzoni.

Ricca, Davide. 2004. Derivazione avverbiale. In Maria Grossmann & Franz Rainer (éds.), *La formazione delle parole in italiano*, 472–479. Tübingen : Max Niemeyer Verlag.

Ricca, Davide. 2015. Adverbial categories. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, 1390–1406. Berlin : de Gruyter Mouton.

Roché, Michel. 2010. Base, thème, radical. *Recherches Linguistiques de Vincennes* 39. 95–134.

Saporta, Soledad. 1990. The status of Spanish forms in *-mente*. *Hispanic Linguistics* 4. 181–183.

Scalise, Sergio. 1988. Inflection and derivation. *Linguistics* 26. 561–581.

Scalise, Sergio. 1990. Constraints on the Italian suffix *-mente*. In Wolfgang U. Dressler, Hans Christian Luschützky, Oskar E. Pfeiffer & John R. Rennison (éds.), *Contemporary morphology*, 87–98. Berlin : Mouton de Gruyter.

Scalise, Sergio & Emiliano Guevara. 2005. The lexicalist approach to word-formation and the notion of lexicon. In Pavol Štekauer & Rochelle Lieber (éds.), *Handbook of word-formation*, 147–188. Dordrecht : Springer.

Schultink, Henk. 1961. Produktiviteit als morfologisch Fenomeen. *Forum te Letteren* 2. 110–125.

Seco, Manuel. 1972. *Gramática esencial del español*. Madrid : Aguilar.

Siegel, Dorothy. 1979. *Topics in English morphology*. New York : Garland.

Štekauer, Pavol. 2005. *Meaning predictability in word formation*. Amsterdam : John Benjamins Publishing company.

Štekauer, Pavol. 2015. The delimitation of derivation and inflection. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, 218–234. Berlin : de Gruyter Mouton.

Stump, Gregory T. 2005. Word-formation and inflectional morphology. In Pavol Štekauer & Rochelle Lieber (éds.), *Handbook of word formation*, 49–72. Dordrecht : Springer.

Sugioka, Yoko & Rachel Lehr. 1983. Adverbial *-ly* as an inflectional affix. *Papers from the parasession on the interplay of phonology, morphology and syntax. Chicago Linguistic Society* 19(2). 293–300.

ten Hacken, Pius. 2014. Delineating derivation and inflection. In Rochelle Lieber & Pavol Štekauer (éds.), *The Oxford handbook of derivational morphology*, 10–25. Oxford : Oxford University Press.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5(2). 183–207.

Torner, Sergi. 2005. On the morphological nature of Spanish adverbs ending in *-mente*. *Probus* 17(1). 115–144.

Torner, Sergi. 2016. Adverbio. In Guttiérrez-Rexach (éd.), *Enciclopedia de lingüistica hispánica*, t. 1, 380–392. London : Routledge.

Trésor de la Langue Française. 1971–1984. Paris : Gallimard.

van Willigen, Marieke. 1983. Remarques sur la dérivation des adverbes en -ment en français moderne. *Cahier de Lexicologie* 42. 63–71.

Varela Ortega, Soledad. 1990. *Fundamentos de morfología*. Madrid : Síntesis.

Zagona, Karen. 1990. *Mente* adverbs, compound interpretation and the projection principle. *Probus* 2. 1–30.

Zwicky, Arnold M. 1987. Suppressing the Zs. *Journal of Linguistics* 23. 133–148.

Zwicky, Arnold M. 1995. Why English adverbial *-ly* is not inflectional. In *Chicago linguistic society*, t. 31, 523–535. Chicago : Chicago Linguistic Society.

Chapter 6

Des lexèmes à forme unique : comment le créole réanalyse les dérivations du français

Florence Villoing

Modèles, Dynamiques, Corpus (MoDyCo) CNRS : UMR7114, Université Paris Nanterre

Maxime Deglas

Structures Formelles du Langage (SFL), CNRS :UMR7023, Université Paris VIII - Vincennes Saint-Denis

Le présent article présente les conditions d'apparition de deux schémas morphologiques en créole guadeloupéen, la suffixation verbale dénominale en –é (N-$é_v$) (ex : *biké* 'se réfugier' ← *bik* 'refuge' ; *miganné* 'mélanger ← *migan* 'purée') et la parasynthèse verbale dénominale (*dé*-N-$é_v$) (ex : *déchèpiyé* 'mettre en charpie' ← *chépi* 'charpie', *dépyété* 'retirer les pattes (crabes)' ← *pyèt* 'pattes'). Nous montrons que ces shémas ont émergé via la réanalyse de paires morphologiques Verbe / Nom, massivement héritées du français, langue lexificatrice, issues soit de conversions (*bròs* 'brosse' / *brosé* 'brosser') soit de préfixations (*bwa* 'bois' / *débwazé* 'déboiser'). L'article défend l'hypothèse que c'est notamment la spécificité des lexèmes créoles de n'apparaître que sous une forme unique qui a conduit à ces réanalyses : les verbes créoles ne variant pas flexionnellement, la finale flexionnelle française /e/ héritée est réanalysée comme suffixe dérivationnel, suivant ainsi un processus de déflexionnalisation propre au changement linguistique.

1 Introduction

La réflexion menée ces cinquante dernières années sur l'identité lexicale et la notion de lexème, notamment par les morphologues, a permis d'éclairer l'analyse de dérivés français impliquant des verbes. Ainsi, les verbes dénominaux, traditionnellement traités comme suffixés au moyen de la marque de l'infinitif (*boiser, plumer, neiger*) ou comme parasynthétiques par adjonction simultanée d'un préfixe et d'un suffixe d'infinitif (*embarquer, désosser, décourager*) ont pu être analysés comme des convers (*boiser, plumer, neiger*)

ou des préfixés (*embarquer, désosser, décourager*) sur base nominale à partir du moment où une réflexion théorique sur l'identité du lexème a été menée (cf. § 2). Mais une telle analyse de ces dérivés français est remise en cause une fois qu'ils intègrent les langues créoles à base française, et on voit s'opérer comme un retournement de situation par rapport aux analyses traditionnelles. En effet, bien que ces créoles aient hérité d'une bonne partie des dérivés verbaux dénominaux convers et préfixés du français, l'analyse morphologique que l'on peut en faire en créole est radicalement différente : là où les paires nom/verbe relèvent de conversions en français, elles sont formées au moyen d'une opération de suffixation en créole ; et là où les paires s'interprètent comme des préfixations en français, on doit y voir des parasynthèses en créole. Cette réanalyse des paires nom/verbe construites et héritées du français a fait système en créole conduisant à la création de nouveaux schémas morphologiques qui sont devenus parfaitement disponibles.

Le présent article présente les conditions d'apparition de ces deux schémas morphologiques en créole, la suffixation verbale dénominale en −*é* (désormais N-*é*$_\text{v}$[1]) et la parasynthèse verbale dénominale (désormais *dé*-N-*é*$_\text{v}$[2]), en défendant l'hypothèse que c'est notamment la spécificité des lexèmes créoles de n'apparaître que sous une forme unique qui a conduit à ces réanalyses (§ 3).

L'analyse que nous présentons est pertinente pour plusieurs créoles à base française (au moins le martiniquais, le haïtien et le saint-lucien), mais s'appuie uniquement sur des données du créole guadeloupéen. Les ressources disponibles pour la constitution d'une base de données de grande ampleur du lexique guadeloupéen font largement défaut, tant du point de vue lexicographique que numérique (cf. Villoing & Deglas 2016a, § 2.). Devant l'absence de ressource fiable et directement exploitable, nous avons basé notre étude sur un corpus original établi par Maxime Deglas, locuteur natif, à partir de plusieurs ressources :

(i) les dictionnaires existants du Guadeloupéen (Ludwig et al. 2012, Poullet & Telchid 1984, Tourneux & Barbotin 1990) dont les entrées ont été filtrées grâce à des enquêtes de terrain vérifiant leur attestation auprès de locuteurs natifs ;

(ii) des enquêtes de terrain réalisées auprès d'une quarantaine de locuteurs natifs issus de toutes les îles de la Guadeloupe ;

(iii) d'un corpus issu d'une activité de veille terminologique réalisée au sein d'ouvrages littéraires en langue créole, d'émissions de télévision et de chansons traditionnelles (cf. Villoing & Deglas 2016, § 2. pour plus de détails).

Le corpus ainsi constitué est composé de 7680 unités lexicales du créole guadeloupéen, soit une envergure équivalente à celle des dictionnaires existants. Il comprend 1805 verbes et 4643 noms qui ont permis l'étude spécifique des relations morphologiques

---

[1]La représentation N-*é*$_\text{v}$ de la structure des verbes dénominaux suffixés en −*é* s'interprète comme suit : N représente la base nominale, −*é* le suffixe, et v la classe syntaxique (V pour verbe) du dérivé.

[2]La représentation *dé*-N-*é*$_\text{v}$ de la structure des verbes dénominaux affixés en *dé*−...−*é* s'interprète comme suit : N représente la base nominale, *dé*−...−*é* l'affixe parasynthétique dont la forme phonologique comprend un préfixe *dé*− associé à un suffixe −*é*, et v la classe syntaxique (verbe) du dérivé.

Nom/Verbe dans le cadre de la suffixation verbale dénominale en −*é* et de la parasynthèse verbale dénominale. Le corpus est enregistré sous format électronique dans une base de données interrogeable selon plusieurs critères, phonologiques, sémantiques, syntaxiques, qui permettent une étude fine.

Nous menons l'étude de ce corpus en suivant une approche théorique relevant de la morphologie lexématique (cf. par ex. Matthews 1991, Aronoff 1994, Anderson 1992, Fradin 2003, Booij 2010), envisageant que les unités de base de la morphologie sont les lexèmes (et non les morphèmes). Nous nous inscrivons dans une perspective qui reconnaît aux langues créoles une morphologie dynamique (tout au moins pour ce qui concerne la morphologie lexicale), nous inscrivant en faux relativement aux détracteurs du contraire (Valdman 1978, Seuren & Wekker 1986, McWhorter 1998, par exemple). La démonstration commencera par une présentation des débats autour des analyses des paires Nom/Verbe convers et préfixés du français (§ 2.) pour ensuite développer notre hypothèse de leur réanalyse en créole qui a conduit à la création de nouveaux schémas morphologiques, la suffixation N-*é*$_V$ et la parasynthèse *dé*-N-*é*$_V$ (§ 3).

2 Analyse des paires N/V en français

Les créoles à base française ont hérité une partie du lexique du français, qui est encore aujourd'hui largement représenté dans la langue créole (par exemple, pour le Guadeloupéen, 90% de mots d'origine française, issus principalement du français populaire du 17$^{\text{ème}}$ siècle, mais également d'emprunts contemporains, selon Hazaël-Massieux 2002). Ce lexique hérité, clairement reconnaissable malgré quelques divergences phonologiques avec l'origine française, comprend des paires de lexèmes morphologiquement construits en français tels que (1) et (2).

(1) a. bò / débòdé ('bord' / 'déborder')
   b. bwa / débwazé ('bois' / 'déboiser')
   c. figi / défigiré ('figure' / 'défigurer')
   d. fòwm / défòwmé ('forme' / 'déformer')
   e. kras / dékrasé ('crasse' / 'décrasser')
   f. rasin / dérasiné ('racine' / 'déraciner')

(2) a. adisyon / adisyonné ('addition' / 'additionner')
   b. bav / bavé ('bave' / 'baver')
   c. bròs / brosé ('brosse' / 'brosser')
   d. divòs / divòsé ('divorce' / 'divorcer')
   e. fèt / fété ('fête' / 'fêter')
   f. savon / savonné ('savon' / 'savonner')

Ces paires Nom/Verbe héritées sont prises dans une relation morphologique en français que l'on ne peut plus leur reconnaître en créole. Les paragraphes qui suivent donnent

un rapide aperçu des analyses morphologiques auxquelles elles répondent en français, pour présenter, ensuite, l'analyse morphologique que nous en proposons en créole guadeloupéen.

2.1 Les paires du type *bois* / *déboiser*

La formation en français des verbes de (1) a été l'objet de grandes discussions. Une tradition qui remonte au 19ème siècle les a analysés comme des construits morphologiques par parasynthèse, c'est-à-dire comme relevant d'une construction morphologique où une base est simultanément préfixée et suffixée. Cette analyse remonte au moins à Arsène Darmesteter.

> « Cette sorte de composition[3] est très riche : les verbes qu'elle forme, et que l'on désigne sous le nom de parasynthétiques, offrent ce remarquable caractère d'être le résultat d'une composition et d'une dérivation agissant ensemble sur un même radical, de telle sorte que l'une ou l'autre ne peut être supprimée sans amener la perte du mot. C'est ainsi que de barque l'on fait em-barqu-er, dé-barqu-er, deux compositions absolument uns et dans lesquelles on ne retrouve ni des composés débarque, embarque, ni un dérivé barquer, mais le radical barque. » Darmesteter (1894 : 24)

L'analyse est largement reprise au 20ème siècle par Nyrop (1936 : 215), et a rencontré encore beaucoup de succès à partir des années 70 dans d'autres théories, comme la Grammaire Générative Transformationnelle (Dubois 1962, Guilbert 1975, Zribi-Hertz 1972, Scalise 1994) ou encore dans le cadre lexicaliste (Booij 1977). Elle s'est également étendue aux grammaires traditionnelles (Grevisse & Goose 1988 : 253) et scolaires en France (cf. par exemple, Chevalier et al. 1964 : 54, Béchade 1992 : 119), voire aux manuels de morphologie du français (Gardes-Tamine 1988 : 65, Apothéloz 2002 : 91, Huot 2006 : 121-122) . Malgré sa popularité, l'analyse parasynthétique est remise en cause pour ces verbes par Dell (1970 : 201–202) puis plus largement par Corbin (1987 : 121–139), et à leur suite Fradin (2003 : 288-307). La critique s'appuie unanimement sur l'erreur d'analyse récurrente qui est faite de la forme du verbe prise métalinguistiquement : l'affixe d'infinitif (qui apparaît de façon conventionnelle dans la forme de citation du verbe) est assimilé à un suffixe dérivationnel. Cette erreur provient en partie d'une confusion entre la langue et la métalangue (Corbin 1987 : 124) et en partie de ce que les cadres théoriques ne définissent pas théoriquement l'individu lexical. Une double confusion est ainsi à l'œuvre (Kerleroux 2000) : une première confusion entre la forme de citation métalinguistique du verbe (qui est traditionnellement l'infinitif en français) et sa forme phonologique, et une seconde confusion entre la forme phonologique du verbe avec l'individu lexical. Ainsi,

> « le rapport catégoriel N>V va être vu comme une suffixation, puisque la forme d'infinitif (dans son rôle citationnel) est prise pour le verbe lui-même, et que l'infinitif français présente un suffixe (à la différence de l'anglais). [...] Tout le problème

---

[3]Darmesteter parle de composition pour caractériser la préfixation, témoignant par-là du fait que certains préfixes sont issus de prépositions latines.

est que cela implique de voir dans le suffixe flexionnel d'infinitif un suffixe qui soit également dérivationnel… » (Kerleroux 2000 : 9)

Or il a été clairement démontré que l'affixe d'infinitif ne peut être identifiable à un suffixe dérivationnel, comme le prouve le fait qu'il n'apparaisse jamais en dérivation, où seul le radical sert toujours de base (Corbin 1987 : 129, Lyons 1977 : 19, Fradin 2003 : 93, Fradin et al. 2009 : 9, par exemple). Ainsi, il aura fallu plus d'un siècle pour montrer que le suffixe d'infinitif de la forme citationnelle n'appartient pas au lexème, en tant qu'unité lexicale.

Il résulte de cette remise en cause une nouvelle analyse selon laquelle « les pseudo-parasynthèses verbales ne sont en fait que des préfixations » (Corbin 1987 : 129) : la base est nominale et le dérivé verbal. Ainsi, selon cette perspective, les données de (1) sont-elles analysées, en français, comme des verbes préfixés sur bases nominales dont la structure correspond à (3) :

(3) $[\, d\acute{e}\text{–} \, [\mathrm{N}] \,]_V$

Ces préfixes dénominaux verbalisateurs présentent, selon Corbin, une propriété originale au regard de la majorité d'entre eux, ils entraînent un changement de catégorie de la base, au même titre que la plupart des suffixes. Cette propriété des préfixes n'ayant pas été reconnue par toute une tradition, a également, selon Corbin, largement contribué à l'analyse en terme de parasynthèse.

Les paires morphologiques Nom/Verbe en (2) ci-dessus ont subi une erreur d'analyse du même type.

2.2 Les paires du type *brosse/brosser*

La formation des verbes du français en (2) a également fait l'objet de grandes discussions. L'analyse de ces paires s'est heurtée, dans la littérature sur la morphologie du français, aux mêmes blocages que les verbes dénominaux préfixés : le suffixe d'infinitif de la forme citationnelle du verbe a été interprété par toute une tradition comme un suffixe dérivationnel.

> C'est cette même prétendue suffixation qui apparaît dans la formation de verbes dénominaux non préfixés comme *clouer*, ou dans les déadjectivaux comme *brunir*, *rougir*. (Dell 1970 : 200–202)

Selon l'orientation de l'opération morphologique (de nom à verbe ou de verbe à nom), la disparition (orientation V → N) ou l'apparition du suffixe (orientation N → V) a été vue comme relevant de deux mécanismes différents,

— la « dérivation régressive » (terminologie que l'on retrouve chez Nyrop (1936), dans les grammaires traditionnelles (Grevisse & Goose 1988) et certains manuels de morphologie (Gardes-Tamine 1988)), rend compte d'une apocope du suffixe d'infinitif, permettant de former un nom à partir d'un verbe (par exemple *voler* → *vol*) ;

— un mécanisme de suffixation de l'infinitif permettant à un nom de devenir un verbe (*plante* → *planter*). Cependant, ce rapport entre nom et verbe n'est pas clairement reconnu par les premiers grammairiens comme relevant de la morphologie comme l'atteste le flou dans lequel il est traité par exemple par Nyrop (1936), Meyer-Lübke (1894) et plus tard par les grammaires traditionnelles (cf. par exemple Grevisse & Goose 1988 : 238).

Là encore, le défaut de ces analyses est l'absence de questionnement théorique quant à l'identité du lexème, confondant forme citationnelle et unité lexicale. Les approches plus contemporaines répondent à ces analyses erronées en voyant dans les paires en (2) des construits ressortissant à une opération de conversion de nom à verbe ou de verbe à nom (cf. pour le français, Corbin 1987, 2004, Mel'čuk 1996, Kerleroux 2000, Fradin 2003, Namer 2009, Tribout 2010). L'apparente différence phonologique entre le nom et le verbe n'est liée qu'à la convention que l'on adopte en français de citer les verbes au moyen de leur forme d'infinitif et les noms à partir de leur forme de singulier. Mais les formes phonologiques des lexèmes bases et dérivés (en d'autres termes, leurs radicaux), sont bien en tous points identiques, ce qui autorise à reconnaître entre eux une relation morphologique de conversion.

Ainsi, les paires en (2) sont-elles analysables soit selon la structure (4a), soit selon la structure (4b), sans qu'aucune sorte d'affixe ne soit en jeu :

(4)   a. $[N]_V$
      b. $[V]_N$

3 Analyses des paires N/V en créole

Les données en (1) et (2) formées par préfixation ou conversion verbale dénominale en français et héritées, ne peuvent pourtant pas recevoir la même analyse en créole. Dans les paragraphes qui suivent, nous argumentons en faveur de la double hypothèse qu'en créole,

(i)   la relation morphologique entre les noms et les verbes en −*é* de (2) correspond à une suffixation verbale sur base nominale (N-*é*$_V$) et non à une conversion comme en français ;

(ii) la relation morphologique entre les noms et les verbes en *dé*-N-*é*$_V$ de (1) correspond à une parasynthèse plutôt qu'à une préfixation comme c'est le cas en français.

Ces résultats nous amènent à conclure que ces paires morphologiques Nom/Verbe ont subi une réanalyse du français au créole[4], réanalyse due en grande partie à la spécificité des lexèmes créoles de n'apparaître que sous une unique forme. C'est sur cette spécificité des verbes en créole guadeloupéen que s'ouvre le § 3.1.

---

[4]Nous entendons "réanalyse" au sens général de Langacker (1977 : 58), à savoir un changement dans la structure (morphologique) d'un lexème qui n'implique pas pour autant de modification dans sa forme phonologique de surface. Voir aussi le recours qu'en fait DeGraff (2001 : 67–68).

3.1 Les verbes du créole guadeloupéen

3.1.1 Morphologie

Les verbes du créole guadeloupéen, comme toutes les autres unités lexicales, ne présentent pas de morphologie flexionnelle, ce que la littérature pointe en évoquant soit l'absence de flexion dans les langues créoles, soit une morphologie pauvre, voire inexistante. Les propriétés liées au Temps-Aspect-Mode sont prises en charge par des particules qui précèdent le verbe, comme on l'observe en général dans les créoles à base française (cf. Valdman 1978, Bernabé 1987, Mufwene & Djikhoff 1989, Hazaël-Massieux 2002 : 71 ; voir aussi Germain 1976 : 109–134, pour le guadeloupéen) .

Lorsque les verbes sont hérités du français, une seule forme du verbe est conservée en créole. Il s'agit, *a priori*, soit de la forme de l'infinitif soit de la forme du participe passé, soit d'une de celles du présent indicatif ou impératif (Germain 1976 : 110). Pour les verbes du 1er groupe et 2ème groupe, l'origine de la forme héritée n'est pas décidable puisque les formes du participe passé et de l'infinitif sont homonymes à l'oral avec une finale en

— /e/ pour les verbes du 1er groupe,

— /i/ pour les verbes du 2ème groupe (sachant qu'au 17ème siècle, époque où la majorité du lexique français est hérité, le /r/ final des infinitifs en −*ir* ne se prononçait plus avant d'être réhabilité ultérieurement sous l'influence des grammairiens et des poètes).

La table 1 présente les différentes finales verbales des verbes créoles hérités des verbes français et les formes fléchies supposées originelles.

Chacune des finales n'a pas la même représentativité au sein du lexique guadeloupéen, et on note une très large majorité de verbes à finale en −*é* (toute origine confondue, hérités, construits en créole ou autre, cf. Table 2)[5]. Nous supposons que cette très forte proportion est liée à un héritage massif de verbes français à finale en −*é*, héritage qui aurait eu un impact important dans la morphologie du créole (cf. § 3.1.2. ci-dessous).

3.1.2 Verbes hérités *versus* verbes créoles

La discrimination, au sein du lexique créole, entre verbes hérités et verbes créoles –ou « indigènes », pour reprendre la terminologie de Lefebvre (2003) et Brousseau (2011)– suscite discussion, dans la mesure où rares sont les cas où l'héritage est total. En effet, les verbes, en passant du français au créole, peuvent avoir subi des modifications phonologiques, sémantiques ou syntaxiques. Une position consiste à considérer comme non français tout lexème hérité ayant subi une variation en créole : par exemple, pour Brousseau (2011 : 68), les lexèmes *pitiab* 'pitoyable' et *lonvi* 'longues-vues' en Saint-Lucien, sont considérés comme des bases inexistantes en français à cause de l'écart phonologique entre les deux langues, et *kouvé* 'couvrir' à cause de la différence sémantique avec

---

[5]Dans la table 2, la classe « autres » inclut principalement des verbes à finale consonantique dont une bonne part sont construits par composition d'un verbe et d'un nom (*bat chat* 'battre en retraite', *pèd lakat* 'perdre la tête).

Tableau 1 : Verbes hérités du français.

Groupe en français	Finale créole	Finale en français	Issus de formes d'infinitifs	Issus de participes passés	Issus de formes d'indicatif/ impératif présent	Glose
1	-é	/e/	karésé			'caresser'
			brosé			'brosser'
			gomé			'gommer'
			blagé			'blaguer'
2	-i	/ir/~/i/	asòrti			'assortir'
			chwazi			'choisir'
			fléri			'fleurir'
			nwasi			'noircir'
3	-è	/ɛr/		dékouvè		'découvrir'
				ofè		'offrir'
				wouvè		'ouvrir'
				soufè		'souffrir'
		/ɛ/~/ɛr/	fè	fè	fè	'faire'
			plè		plè	'plaire'
	-an	/ã/~/ãdr/	aprann		aprann	'apprendre'
			défann		défann	'prendre la défense'
			étann		étann	'étendre'
			fann		fann	'fendre'
	-wè	/wa/~/war/	bwè		bwè	'boire'
			kwè		kwè	'croire'
			pèsivwè		pèsivwè	'apercevoir'
			wousouvwè		wousouvwè	'recevoir'
	-i	/ir/~/i/	bouyi			'bouillir'
			fui	fui		'fuir'
			manti			'mentir'
			rèdi	rèdi		'redire'
			sòti			'sortir'
	-enn	/ɛ̃/~/ɛ̃dr/	détenn	détenn	détenn	'déteindre'
			étenn	étenn	étenn	'éteindre'
				krenn	krenn	'craindre'
					soutyenn	'soutenir'
			tenn	tenn	tenn	'teindre'
	-èt	/ɛtr/	admèt			'admettre'
			disparèt			'disparaître'
			pwomèt			'promettre'
			rèkonnèt			'reconnaître'
		/ɛt/		défèt		'défaire'

Tableau 2 : Proportion des verbes guadeloupéens selon leur finale.

Verbes à finale en	nbr	Total % des V totaux
-é	1451	80%
-i	147	8,1%
-è	25	1,4%
-ann	30	1,6%
-wè	10	0,5%
-enn	11	0,6%
-èt	15	0,8%
-an	10	0,5%
-o	8	0,4%
autres	98	5,4%
Total	1805	100%

le verbe *couver*. Nous nous distinguerons de cette position en considérant comme hérité du français tout verbe dont l'origine française est reconnaissable, phonologiquement et sémantiquement, malgré les modifications subies en créole. Ainsi, parmi les exemples de Brousseau, seul *kouvé* 'couvrir' ne serait pas reconnu comme d'origine française à cause du sens trop éloigné du verbe *couver* du français. Notre choix repose sur le fait (i) d'une part qu'il est extrêmement difficile de connaître précisément la phonologie et la sémantique des lexèmes hérités d'un état ancien ou régional du français, et en conséquence, de déterminer, avec certitude, l'écart entre le supposé verbe français et son correspondant hérité en créole ; (ii) d'autre part que quasiment tout lexème hérité du français a subi une modification phonologique voire sémantique, même mineure, et qu'il serait difficile d'établir des critères départageant les lexèmes suffisamment altérés pour être classés créoles et les autres.

Afin de déterminer l'origine française d'un lexème créole, nous nous sommes appuyés sur leur attestation en entrée d'un dictionnaire de français, tout dictionnaire, registre de langue et variétés dialectales confondus (voir aussi Brousseau 2011 : 68 sur l'utilité des dictionnaires du 16ème au 20ème siècle). La recherche est largement facilitée par la Toile qui met à notre disposition plusieurs types de dictionnaires du français, permettant notamment de retrouver des verbes aujourd'hui perdus mais relevant d'un état de langue ancien ou d'un dialecte du français, dont on suppose qu'ils constituent le fond du lexique créole (cf. par exemple Thibault 2012 : 12).

Ces critères nous permettent de distinguer les verbes hérités de deux autres types de verbes :

(i) les verbes morphologiquement construits en créole, tout procédé morphologique et toutes bases confondues (bases non héritées (5), bases héritées (6), bases héritées avec changement phonologique (7) ou sémantique (8)).

(5) a. bik   → biké
        'refuge'   'se réfugier'

     b. fifin   → fifiné
        'bruine'   'bruiner'

     c. migan   → miganné
        'purée' 'mélange'   'mélanger'

     d. plich   → pliché
        'correction'   'donner une correction'

     e. vonvon   → vonvonné
        'bourdon'   'bourdonner'

(6) a. balkon → balkonné
        'balcon'   'être au balcon'

     b. garé   → dégaré
        'garer, stationner'   'sortir de la place de garage, de stationnement'

     c. lang   → langé
        'langue'   'embrasser'

     d. pyé   → dépyété
        'pattes'   'retirer les pattes (crabe)'

     e. tik   → détiké
        'tique'   'retirer les tiques'

(7) a. fouch   → fouchté
        'fourche'   'bêcher'

     b. katyé   → dékatyé
        'morceau'   'couper en quartier'

     c. nwèl → nwélé
        'noël'   'fêter Noël'

     d. pengné → dépengné
        'peigner'   'défaire une coiffure'

     e. vès   → vèsté
        'veste'   'mettre sa veste'

(8) a. kabann → kabanné
        'lit'   'traîner au lit'

     b. kaz   → dékazé
        'maison'   'déplacer une maison à l'aide
                        d'un véhicule pour l'installer ailleurs'

     c. loup   → loupé
        'boursouflure'   'enfler'

 d. parad   → paradé
    'étalage'   'parader'

(ii) les verbes ne répondant à aucun de ces critères, ni hérités du français, ni construits en créole, et dont l'origine peut être connue (cf. un emprunt à l'anglais, à l'espagnol, aux langues africaines ou autre), ou non.

(9) a. bénékaki
       'hésiter'

   b. griji
       's'égratigner'

   c. kòviyé
       'tordre'

   d. lolé
       'remuer'

   e. tòtòy
       'agacer'

Sur la base de cette répartition tripartite des verbes en créole (verbe hérité, verbe construit en créole, verbe autre), nous obtenons les proportions suivantes (cf. Table 3 qui ne représente que les trois finales les plus représentées, les finales verbales en −é, −i, et −ann).

Tableau 3 : Proportion de verbes hérités, construits ou autres selon leur finale.

	Verbes en créole guadeloupéen							
Verbes à finale en	Total		Verbes hérités		Verbes construits en créole		Autres	
	nbr	% des V totaux	nbr	% des V à finale en …	nbr	% des V à finale en …	nbr	% des V à finale en …
-é	1451	**80%**	1230	84 %	153	10,5%	66	4,5%
-i	147	8,1%	122	83 %	17	11,5 %	8	5,5%
-ann	30	1,6%	27	90%	3	10%	0	0%
Total	1805	100%	1468	81% des V	248	14% des V	86	5% des V

Notre corpus comprend ainsi une part majeure de verbes hérités du français : sur les 1805 verbes listés, 1468 sont hérités, soit 81 % des verbes du créole. Parmi ces verbes hérités, la majorité sont des verbes à finale en −é (soit 84 %). Loin derrière se trouvent les verbes hérités à finale en −i qui ne représentent que 8,3% des verbes hérités (122 verbes hérités à finale en −i parmi 1468 verbes hérités). Les verbes présentant d'autres finales

(*−ann, −è, −wè* etc.) sont encore moins nombreux et très peu représentés. Cet ordre de préférence se reflète largement dans les verbes construits en créole : là encore, les verbes à finale en *−é* sont les plus représentés (61,5% correspondant à 153 verbes construits en *−é* relativement à 248 verbes construits), suivis de loin par les verbes à finale en *−i* (moins de 7%). Les autres verbes restent de l'ordre de l'épiphénomène. Ce parallèle entre finale des verbes hérités et verbes construits en créole conduit raisonnablement à faire l'hypothèse que le lexique hérité a fortement pesé sur la formation morphologique des verbes créoles. Ainsi, dans la mesure où la majorité des verbes hérités sont ceux à finale en *−é* et que les verbes créoles dérivés sur base nominale présentent également majoritairement cette finale, nous émettons l'hypothèse que la finale flexionnelle en *−é* des verbes hérités a été réanalysée, dans certaines circonstances, comme un suffixe dérivationnel en créole. Le paragraphe 3.2. présente des hypothèses sur les conditions de cette réanalyse. Nous n'examinerons pas plus avant ici la possible réanalyse des finales de verbes hérités en *−i*, mais remarquons néanmoins qu'en dépit de la très faible proportion de ces verbes dans le lexique créole (8,1%), la part des verbes construits en *−i* est proportionnellement équivalente à celles des verbes construits en *−é* (11,5 % contre 10,5 % pour les verbes en *−é*), ce qui conduirait à rendre crédible l'hypothèse de la création d'un suffixe verbalisateur *−i* en créole guadeloupéen.

3.2 Réanalyses des paires N/V de convers comme suffixations

Selon notre hypothèse, la réanalyse des verbes à finale en *−é* du français en créole n'a été possible que dans le contexte lexical créole où ces verbes français sont hérités avec les noms français en relation de conversion avec eux, soit une conversion de nom à verbe (N→ V) soit une conversion de verbe à nom (V→ N) (cf. (10)). Ainsi, le lexique du créole guadeloupéen comprend des paires de convers Nom/Verbe héritées du français, pour lesquels l'analyse en terme de conversion n'est pas valide en créole.

3.2.1 De la conversion en français à la suffixation en créole

La raison principale qu'une relation de suffixation soit perçue en créole entre ces paires Nom/Verbe tient au fait que le *−é* final du verbe apparaît comme du matériel phonologique supplémentaire par rapport à la forme phonologique du nom base (10). Y voir une conversion de nom à verbe serait alors contraire à la notion de conversion puisque les radicaux ici se différencient phonologiquement.

(10) a. adisyon   / adisyonné
          'addition' 'additionner'

      b. bav   / bavé
          'bave' 'baver'

      c. bròs   / brosé
          'brosse' 'brosser'

      d. divòs   / divòsé
          'divorce' 'divorcer'

    e. fèt   / fété
        'fête' 'fêter'

    f. mank   / manké
        'manque' 'manquer'

    g. savon / savonné
        'savon' 'savonner'

Comme les verbes créoles n'ont qu'une forme, les verbes en (10) ne présentent donc que la forme comprenant un *−é* final. Ce *−é* final, de fait, appartient bien au verbe en tant qu'unité lexicale et n'est pas le marqueur du mode infinitif apparaissant dans la forme citationnelle du verbe français. Ainsi, les paires Nom/Verbe en (10) héritées du français ne peuvent subir la même analyse en français et en créole. Elle se distinguent des paires de Nom/Verbe en (11) qui, au contraire, entretiennent bien une relation morphologique de conversion en créole (de type N→ V ou V→ N). En effet, en créole, comme dans toutes les autres langues, les noms et les verbes en relation de conversion sont phonologiquement en tous points identiques (cf. en (11a) des paires de convers Nom/Verbe à finale en *−é* et en (11b) des paires de convers Nom/Verbe présentant une autre finale vocalique).

(11)   a.   i. balyé$_N$ / balyé$_V$
               'balai' 'balayer'

          ii. chanté$_N$ / chanté$_V$
              'chanson' 'chanter'

          iii. goumé$_N$ / goumé$_V$
              'combat' 'se battre'

          iv. lélé$_N$ / lélé$_V$
              'touillette' 'touiller'

          v. manjé$_N$ / manjé$_V$
              'repas, mets' 'manger'

          vi. tété$_N$ / tété$_V$
              'sein' 'téter'

      b.   i. anvi$_N$ / anvi$_V$
               'envie' 'avoir envie'

          ii. bobi$_N$ / bobi$_V$
              'assoupissement' 'somnoler'

          iii. kaka$_N$ / kaka$_V$
              'excrément' 'déféquer'

          iv. mò$_N$ / mò$_V$
              'mort' 'mourir'

          v. travay$_N$ / travay$_V$
              'travail' 'travailler'

Par ailleurs, on ne peut, en aucun cas, tenir l'hypothèse de la conversion en traitant le −*é* final des verbes en (10) comme une marque spécifiquement verbale :

(a) soit une marque de classe flexionnelle (une voyelle thématique).

(b) soit une marque identifiant la catégorie verbe.

En effet, aucune des deux hypothèses ne tient : l'hypothèse (a) d'une voyelle thématique tombe car le créole n'a pas de système flexionnel pour les verbes, et il n'y aurait aucune pertinence à exploiter une voyelle thématique ; et l'hypothèse (b) tombe aussi parce que les finales vocaliques des verbes sont variées (finale en /i/, /e/, /wɛ/ présentées ci-dessus Table 1), auxquelles on peut ajouter celles en /o/, /j/, /õ/ en (12), et on peut difficilement imaginer que la langue dispose d'autant de marqueurs verbaux, en particulier parce que les noms, aussi, présentent des finales vocaliques en /e/, qu'ils soient ou non hérités (cf. (13a) pour les noms hérités, et (13b) pour les noms créoles) :

(12) a. bo,   mò,   cho,   fè kò
      'embrasser', 'être mort', 'avoir chaud', 'unifier'

    b. tòtòy,   kay
       'agacer', 'aller'

    c. mawon,   fè-fon
       's'échapper', 'compter sur'

(13) a. chaplé,   bondyé, fiyansé, pyé, zyé,   sòsyé
       'chapelet', 'dieu',   'fiancé', 'pied', 'yeux', 'sorcier'

    b. bankoulélé, kyolé,   matété,   wélélé
       'vacarme',   'ribambelle', 'plat à base de riz et crabe', 'brouhaha'

L'hypothèse d'une conversion ne tient donc dans aucun cas. Comme le −*é* qui apparaît sur le verbe correspond à du matériel phonologique supplémentaire par rapport au nom, et que la relation catégorielle et sémantique change, tout porte à croire que le verbe est morphologiquement plus complexe que le nom. Il faut donc faire l'hypothèse d'une formation impliquant une suffixation verbale en −*é* sur bases nominales.

3.2.2 L'impossible règle de formation des noms par suppression du −*é*

Une autre hypothèse aurait également pu être envisagée, celle d'une règle de construction de noms sur base verbale, par suppression du −*é* final du verbe (ou une « rétroformation »). Mais cette hypothèse rencontre plusieurs difficultés :

(a) la première tient à ce que ce mode de formation est jugé traditionnellement rare dans les langues (sur la « subtractive morphology » ou « deletion » et sa rareté, voir ce qu'en disent les manuels, comme Anderson 1992 : 64-66 ; Haspelmath 2002 : 24 ; Fradin 2003 : 47)

(b) la deuxième s'appuie sur des paires Nom/Verbe dont le nom est hérité du français mais pas le verbe qui est nécessairement construit en créole (14) ; or ce verbe laisse apparaître un *−é* final supplémentaire.

(14) a. alyans → alyansé
     'alliance' 'se lier'

   b. bwa → bwaré
     'bras' 'enlacer'

   c. chikann → chikanné
     'contestation' 'contester'

   d. fèr → féré
     'fer à cheveux' 'défriser les cheveux'

   e. janm → janbé
     'jambe' 'enjamber'

   f. tij → tijé
     'bourgeon' 'bourgeonner'

Comme le nom est hérité du français, et le verbe construit en créole, le nom ne peut pas être dérivé du verbe par une règle de suppression du *−é* final du verbe ; c'est bien le verbe qui est formé par suffixation sur la base du nom.

(c) le troisième argument s'appuie sur l'absence de noms déverbaux créoles construits par suppression du *−é* d'un verbe hérité. En effet, notre corpus ne fournit aucun nom dérivé à partir de verbes hérités par simple suppression de la finale en *−é*. La disparition de la finale en *−é* des verbes hérités peut avoir lieu à l'occasion d'une dérivation, mais uniquement lorsque la dérivation se fait par suffixation (voir par exemple, (15) pour la suffixation V→ N en *−è/−ez*, (16) pour la suffixation V→ N en *−aj*, et (17) pour la suffixation V→ N en *−asyon*).

(15) a. fiyansèz ← fiyansé
     'fiancée' 'se fiancer'

   b. kouyonnèz ← kouyonné
     'celle qui couillonne' 'couillonner'

   c. soutirèz ← soutiré
     'celui qui couvre les bêtises de qqun' 'couvrir les bêtises de qqun'

(16) a. bokantaj ← bokanté
     'échange' 'échanger'

   b. diraj ← diré
     'qui dure' 'durer'

   c. konblaj ← konblé
     'comblement' 'combler'

(17) a. pwofitasyon ← pwofité
   'profit'   'profiter'

  b. anmerdasyon ← anmerdé
   'emmerdement' 'emmerder'

  c. poursuivasyon ← poursuiv
   'poursuite par le diable' 'poursuivre'

Une dérivation par conversion (18) n'imposera pas, quant à elle, la disparition de la finale vocalique du verbe.

(18) a. déboulé / déboulé
   'défilé' 'défiler rapidement'

  b. lélé / lélé
   'touillette' 'touiller'

  c. mayé / mayé
   'mariage' 'se marier'

  d. pété / pété
   'pet' 'faire un pet'

La voyelle finale du verbe disparaissant uniquement dans le contexte d'une dérivation dont le suffixe est à initiale vocalique, tout porte à croire qu'une contrainte morphophonologique est en jeu (contrainte d'évitement du hiatus, contrainte de taille…) et invalide l'hypothèse de l'existence d'une règle dérivationnelle de suppression.

3.2.3 Conditions d'apparition

Ces arguments conduisent à envisager que les paires de convers Nom/Verbe du français ont subi une réanalyse de telle sorte qu'en créole, la relation morphologique entre les noms et les verbes en −é de (13) ne relève pas d'une conversion, comme en français, mais d'une suffixation verbale sur base nominale (N→ V). Ces paires ont été héritées en nombre suffisant pour avoir fait système et permis de former productivement, par analogie, d'autres verbes dénominaux suffixés par -*é* sur des bases françaises ou non françaises comme en (19).

(19) a. bòk / boké
   'affront' 'faire un affront'

  b. chiktay / chiktayé
   'émiettage' 'émietter'

  c. fèr   / féré
   'fer à cheveux' 'défriser'

  d. lyann / lyanné
   'liane' 'se servir d'un tuteur pour grimper'

   e. dousin / dousiné
     'câlin'   'câliner'

   f. djòb   / djobé
     'petit boulot' 'faire un petit boulot'

   g. plòk   / ploké
     'cloque' 'avoir des cloques'

Ainsi, la réanalyse de ces paires Nom/Verbe héritées a abouti à la création d'un suffixe verbal −*é* en créole, inexistant dans la langue lexificatrice. Ce schéma morphologique est représenté sous (20) où X est mis pour le lexème base (et non le radical qui peut subir des modifications phonologiques lors de la suffixation comme nous le présentons en § 3.3) :

(20) $X_N \longrightarrow X\acute{e}_V$

La création de ce schéma morphologique n'a rien d'inédit à travers les langues ; il peut s'apparenter à ce que la littérature dédiée aux mécanismes et aux motivations du changement dans la formation des mots appelle « secretion » (Rainer 2015 : 1771). Ce concept repris à Jespersen (1922 : 384) , réfère à un processus par lequel une séquence purement phonologique acquiert le statut de « morphème » (phénomène déjà signalé, selon Rainer 2015, par Bloomfield 1891, ou Lass 1990 qui parle de « exaptation »[6]).

> By secretion I understand the phenomenon that one portion of an indivisible word comes to acquire a grammatical signification which it had not at first, and is then felt as something added to the word itself. (Rainer 2015 : 1771)

Il peut également s'apparenter à un cas de « degrammaticalization » ou de « deinflectionalization » (Rainer 2015 : 1768–69) dans la mesure où la finale flexionnelle du verbe français héritée (/e/) devient un suffixe dérivationnel.

Quoi qu'il en soit, les conditions requises pour aboutir à la naissance du suffixe verbal dénominal −*é* en créole lui sont spécifiques. Nous stipulons qu'elles sont les suivantes :

1) d'une part, la très forte représentativité, dans le lexique créole, de couples morphologiques Nom/Verbe hérités du français où ils entretiennent une relation de conversion ;

2) d'autre part, au sein de ces couples, une très forte majorité de verbes à finale en −*é* ;

---

[6]Ce cas est à distinguer de ce que Haspelmath (1995 : 8–10) appelle « secretion » qui fait référence à une extension d'un affixe par l'incorporation d'une partie non affixale de la racine (schématisé sous (a))

(a)   Affix secretion
    Xyz → xyz-a
    R ⟹ -za
    ⟹ new suffix −za, e.g. klm → klm-za

3) et enfin, la propriété des lexèmes verbaux créoles de n'apparaître que sous une forme unique : ainsi la marque flexionnelle des verbes hérités n'a pu être interprétée comme flexionnelle en créole.

C'est la conjonction de ces trois conditions qui a rendu possible la création de ce suffixe en créole guadeloupéen. Si l'une de ces conditions n'avait pas été remplie, il y a fort à parier qu'aucun nouveau schéma morphologique n'aurait pu voir le jour. Par exemple, tous les verbes créoles hérités du français remplissent la condition 3), mais seules les finales en −*é* des verbes hérités du français ont été réanalysées comme une règle de suffixation de verbes dénominaux. Cela tient aux conditions 1) et 2) réunies : seules les paires héritées du français Nom / Verbe à finale en −*é* ont été héritées en grand nombre, à l'exception d'autres finales verbales. Toutes les autres paires Nom/Verbe apparaissent en nombre infime et la deuxième condition présentée ci-dessus n'est pas remplie. En effet, même si le guadeloupéen compte un certain nombre de verbes hérités présentant une autre finale que −*é* (cf. la table ci-dessus), ces verbes soit ne sont reliés à aucun nom (comme (21) pour les verbes en −*i*), soit ils le sont, mais uniquement dans une relation de conversion (22 pour les verbes en (−*i*)), soit le nom relié est difficile à mettre en relation morphologique avec le verbe à cause d'une variation phonologique entre les deux trop importante (cf. (23) pour les verbes en −*i*).

(21)   a. abouti
       'aboutir'

       b. aji
       'agir'

       c. dégarni
       'dégarnir'

       d. flétri
       'flétrir'

       e. konstwi
       'construire'

(22)   a. amòrti   / amòrti
       'amortir' 'amorti'

       b. anvi   / anvi
       'avoir envie' 'envie'

       c. griji   / griji
       's'égratigner' 'égratignure'

       d. jwi   / jwi
       'jouir' 'sperme'

       e. vèrni   / vèrni
       'vernir' 'verni'

(23)   a. chwa   / chwazi
       'choix' 'choisir'

b. fen / fini
'fin' 'finir'

c. flè / fléri
'fleur' 'fleurir'

d. kous / kouri
'course' 'courir'

e. trèt / trayi
'traître' 'trahir'

Finalement, les verbes hérités qui ne remplissent pas les conditions 1) et 2) ne donnent lieu à aucune création créole. Pour reprendre l'exemple des verbes en −*i*, les seuls de notre corpus qui ne soient pas hérités ne sont pas dérivés par un suffixe verbalisateur −*i* (24) :

(24) a. bigidi
'faiblir'

b. bénékaki
'hésiter'

c. siri
'devenir aigre'

d. tini
'avoir'

Les trois conditions nécessaires à la création du suffixe −*é* ne sont pas propres au guadeloupéen et se sont retrouvées dans d'autres créoles à base française. En effet, plusieurs créoles ont suivi le même processus et la suffixation en −*é* compte parmi les schémas morphologiques disponibles du Haïtien (DeGraff 2001, Lefebvre 1998, 2003) et du Saint-Lucien (Bhatt & Nikiema 2000, Brousseau 2011). Elle n'a néanmoins jamais fait l'objet d'études de détails dans les travaux portant sur ces créoles.

3.2.4 Propriétés du suffixe verbal dénominal −*é* en créole

3.2.4.1 Forme phonologique du suffixe

Nous postulons que la forme phonologique du suffixe verbal dénominal ainsi créé est /e/ (orthographié −*é*). Cet affixe vocalique apparaît dans certains contextes précédé d'une consonne, /t/ par défaut (cf. (25)) et il y a lieu de se demander si cette consonne à la frontière entre le radical et le suffixe n'appartient pas au suffixe. Tout porte à croire néanmoins que la consonne intercalaire est de nature épenthétique, permettant, dans un contexte lexical, d'éviter la succession de deux voyelles à la frontière entre la base et l'affixe.

(25) a. konplo → konploté
'complot' 'comploter '

b. niméwo → nim* *éroté
   'numéro'   'numéroter'

c. soulyé   → soulyété
   'chaussures'   'mettre des chaussures'

Un premier argument en ce sens est le fait que l'évitement du hiatus en créole gua-
deloupéen s'observe régulièrement à la frontière morphologique dans les cas de dériva-
tion : citons, à titre d'exemple, la formation de dérivés suffixés dont le suffixe à initiale
vocalique entraîne la suppression de la finale vocalique du verbe en −*é*. Un deuxième
argument est le développement d'autres stratégies d'évitement du hiatus en contexte
morphologique, comme le recours à des règles de dérivation permettant de contourner
le problème, en l'occurrence la conversion ou la préfixation. On peut ainsi affirmer que
la suffixation en −*é* entraîne des changements phonologiques sur les bases nominales,
dont les épenthèses ne sont qu'un exemple (voir Villoing & Deglas 2016a pour plus de
détails) .

La présence de toute autre consonne entre le radical et le suffixe relève de cas différents
de l'épenthèse consonnantique ou de l'allomorphie suffixale. Ainsi,

(i) une réalisation spécifique des voyelles nasales en contexte de dérivation en guade-
loupéen, comme dans d'autres créoles à base française (cf. Bhatt & Nikiema 2000),
laisse apparaître une consonne nasale à la suite de la voyelle nasale du radical lors
de la suffixation en −*é* (cf. (26)) ;

(ii) la réalisation de consonnes lexicales héritées des lexèmes français qui se révèlent
uniquement dans ce contexte dérivationnel (le suffixe protégeant la consonne),
puisqu'elles ont par ailleurs disparu en finale (cf. (27)) :

(26) a. boukan   → boukanné
         'feu de brindille'   'griller au feu de bois'

      b. dirèksyon → dirèksyonné
         'direction'   'montrer la direction'

      c. gidon   → gidonné
         'guidon'   'mener'

      d. losyon → losyonné
         'lotion'   'se parfumer'

      e. migan → miganné
         'purée'   'mélanger'

(27) a. arbit   → arbitré
         'arbitre'   'arbitrer, trancher'

      b. chalè   → chaléré
         'chaleur'   's'inquiéter'

      c. janm   → janbé
         'jambe'   'enjamber'

    d. tè   → téré
        'terre'   'enterrer'

    e. penti   → pentiré
        'peinture'   'peindre'

3.2.4.2 Propriétés sémantiques de la règle

La relation sémantique entre le nom de base (désormais Nbase) et le verbe dénominal suffixé en −é apparaît, pour une part, typique de ce type de construction morphologique en français et pour une autre part originale.

    Elle est typique dans les cas où le Nbase renvoie aux actants du verbe comme l'instrument en (28) (qui comprend aussi bien les artefacts (28a) que les parties du corps (28b)), à un agent en (29), à une entité déplacée (*locatum verbs, figure verbs*) en (30a), au lieu du procès (*location verbs, grounds-verbs*) en (30a), et à l'objet résultant du procès en (31).

(28)   N : instrument

    a.   i. fak   → faké
              'bêche'   'bêcher'

        ii. kòn   → koné
             'klaxon'   'klaxonner'

        iii. graj → grajé
             'rape'   'raper'

        iv. pikwa → pikwaté
             'pioche' 'piocher'

    b.   i. lang → langé
              'langue'   'embrasser avec la langue'

        ii. bwa → bwaré
             'bras'   'enlacer'

        iii. zig   → zigé
             'position des doigts   'faire une pichenette'
             pour faire une pichenette'

        iv. zyé   → zyété
             'yeux'   'surveiller'

(29)   N : agent

    a. mako   → makoté
       'mouchard'   'moucharder'

    b. makrèl   → makrélé
       'celle qui se mêle de tout'   'surveiller'

    c. mandyan → mandyanné
       'mendiant'   'mendier'

(30) a. N : entité déplacée

    i. bonda → bondaté
      'fesses' 'poser ses fesses'

    ii. janb → janbé
      'jambe' 'enjamber'

    iii. pyé → pyété
      'pied' 'poser le pied'

    iv. soulyé → soulyété
      'chaussures' 'mettre des chaussures'

  a. N : lieu final du procès

    i. balkon → balkonné
      'balcon' 'être au balcon'

    ii. kabann → kabanné
      'lit' 'traîner au lit'

    iii. kan → kanté
      'côté' 'se mettre sur le côté, sur le flanc'

(31) N : objet résultant

  a. flang → flangé
    'entaille' 'entailler'

  b. migan → miganné
    'purée' 'mélanger'

  c. fifin → fifiné
    'bruine' 'bruiner'

  d. tij → tijé
    'bourgeon' 'bourgeonner'

La relation sémantique entre le Nbase et le verbe dérivé suffixé en –*é* est néanmoins atypique dans les exemples (32) où le Nbase dénote une situation dynamique (voir Villoing & Deglas 2016a pour une présentation des tests d'événementialité) :

(32) a. bonbans → bonbansé
    'fête' 'faire la fête'

  b. chikann → chikanné
    'contestation' 'contester'

  c. chiktay → chiktayé
    'émiettage' 'émietter'

  d. dousin → dousiné
    'caresse' 'caresser'

  e. driv → drivé
    'promenade' 'promener'

      f. kalbann → kalbanné
         'culbute'   'culbuter'

En effet, en français, les « noms d'événément » sont prototypiquement déverbaux et les cas de noms d'événement servant de base à la formation d'un verbe dérivé restent minoritaires. Par exemple, Corbin (2004) note, en français, quelques verbes suffixés construits sur des noms simples dénotant des procès (*guerroyer* et *satiriser* construits sur les noms processifs, *guerre* et *satire*). Mais ces exemples sont forcément très peu nombreux,

— d'une part parce que les noms simples dénotant un événement restent rares dans le lexique du français (*le concert, l'orage*; ils représentent 8,1% des noms simples selon Tribout et al. 2014) et sont, en général, issus de noms déverbaux en latin;

— d'autre part, parce que si les bases des verbes en −*iser* peuvent être processives, ce cas se présente rarement, aussi bien en anglais (Plag 1999) qu'en français (Namer 2013);

— et enfin, la suffixation en −*oyer* apparaît peu productive.

Cette rareté vient confirmer l'hypothèse de Croft (1991) selon laquelle les noms dénotent prototypiquement des objets.

La situation semble être différente lorsque les bases nominales processives sont elles-mêmes complexes morphologiquement. En effet, quelques travaux récents sur le français ont mentionné la relative disponibilité de certains noms construits dénotant des événements à servir de base à la formation d'un verbe. Tribout (2010), par exemple, montre qu'un nombre non négligeable de verbes dénominaux convers sont formés sur des noms événementiels déverbaux (33) :

(33)   a. louer → louange → louanger

       b. vider → vidange → vidanger

       c. recevoir → réception → réceptionner

       d. frotter → friction → frictionner

       e. partir → partage → partager

Tribout (2010) l'explique par le fait que le nom base a perdu sa motivation morphologique et que la perception de sa construction sur base verbale n'existe plus (par exemple, (33c), (33d), (33e)). Mais pour d'autres paires, la relation entre le nom abstrait et son verbe base reste tout à fait transparente (par exemple, (33a), (33b)).

C'est un résultat que partagent Lignon & Namer (2014) sur d'autres cas de conversion du français, les noms abstraits suffixés en −*ion* servant de bases à la formation de verbes convers, alors que ces noms sont construits sur des bases verbales facilement reconstructibles (34) :

(34)   a. attirer → attraction → attractionner
       b. intercéder → intercession → intercesser
       c. soumettre → soumission → soumissionner
       d. voir → vision → visionner

Parallèlement, une autre formation permet de construire des verbes sur des bases nominales événementielles, la rétroformation à partir de composés néoclassiques (Namer 2012) (cf. (35)).

(35)   a. photoémission → photoémettre
       b. hydromassage → hydromasser
       c. hydroextraction → hydroextraire

Ainsi, la formation d'un verbe ayant pour base un nom d'événement en français (i) n'est disponible que pour des bases nominales morphologiquement construites et (ii) la règle impliquée est préférentiellement la conversion. Cette configuration spécifique ne se retrouve pas dans les données du créole guadeloupéen étudiées ci-dessus qui font état d'une règle de suffixation sur base nominale événementielle morphologiquement simple. Le créole présente donc une originalité sémantique par rapport au français tout à fait intéressante. Nous l'attribuons à la formation très spécifique de la règle de suffixation en −*é* qui est issue de la réanalyse de paires Nom/Verbe du français relevant de deux règles de conversion : la conversion V→ N et N→ V.

3.3 Réanalyse des paires N/V-préfixé en parasynthétiques

L'absence de flexion verbale en créole guadeloupéen et l'héritage d'une forme unique du verbe français (en l'occurrence, pour les verbes qui nous intéressent, la forme de l'infinitif ou du participe passé en /e/) entraînent d'autres réanalyses morphologiques. Ainsi, les paires héritées en (36), dont le verbe est formé en français par préfixation, ne peuvent s'analyser en créole qu'en terme de parasynthèse.

(36)   a. bò   / débòrdé
          'bord' 'déborder'

       b. frich   / défriché
          'friche' 'défricher'

       c. kras   / dékrasé
          'crasse' 'décrasser'

       d. mayo   / démayoté
          'étoffe d'emmaillotage de bébé'   'démailloter'

       e. rasin   / dérasiné
          'racine' 'déraciner'

Les paragraphes qui suivent argumentent en faveur de cette hypothèse et présentent les propriétés phonologiques et sémantiques associées à ce schéma morphologique qui est propre au créole.

3.3.1 *dé*-N-*é*_V parasynthétiques

Les exemples de paires morphologiques Nom/Verbe en (36) héritées du français ne supportent pas la même analyse morphologique en créole guadeloupéen et conduisent à envisager un nouveau cas de réanalyse morphologique. Là où, en français, l'analyse reconnaît un dérivé verbal au moyen d'une préfixation en *dé-* sur base nominale, le créole, quant à lui, forme un verbe par parasynthèse sur base nominale.

Le raisonnement qui conduit à ce résultat est proche de celui qui a mené à identifier la création du suffixe dénominal verbalisateur *−é* : les verbes créoles ne se réalisant que sous une forme unique, la finale en *−é* appartient bien à la forme lexicale du verbe et ne correspond pas à l'affixe d'infinitif apparaissant dans la forme citationnelle du verbe. Ainsi, entre la base nominale et le verbe dérivé, du matériel phonologique supplémentaire apparaît aux deux extrémités : à gauche de la base, un préfixe *dé−*, et à droite de la base, le suffixe verbalisateur *−é*. Or ces affixes ne relèvent pas de l'application successive de deux règles morphologiques. En effet, ni le verbe en *−é* (37) ni le nom en *dé−* (38) n'existent indépendamment l'un de l'autre.

(37)   a.   * bòdé$_V$ $\longrightarrow$ débòdé$_V$
                            'déborder'

        b.   * friché$_V$ $\longrightarrow$ défriché$_V$
                            'défricher'

        c.   * krasé$_V$ $\longrightarrow$ dékrasé$_V$
                            'décrasser'

        d.   * mayoté$_V$ $\longrightarrow$ démayoté$_V$
                             'démailloter'

        e.   * rasiné$_V$ $\longrightarrow$ dérasiné$_V$
                            'déraciner'

(38)   a.   * débòd$_N$ $\longrightarrow$ débòdé$_V$
                            'déborder'

        b.   * défrich$_N$ $\longrightarrow$ défriché$_V$
                            'défricher'

        c.   * dékras$_N$ $\longrightarrow$ dékrasé$_V$
                            'décrasser'

        d.   * démayo$_N$ $\longrightarrow$ démayoté$_V$
                            'démailloter'

        e.   * dérasin$_N$ $\longrightarrow$ dérasiné$_V$
                            'déraciner'

Ainsi, les exemples en (36) ne peuvent ni être analysés comme des préfixés en *dé−* sur base verbale (le verbe n'existe pas), ni comme des verbes suffixés en *−é* sur base nominale (ces bases n'existant pas non plus). Ces propriétés rappellent les critères traditionnellement avancés pour reconnaître une parasynthèse (cf. Darmesteter 1894 : 24 présentés

ci-dessus au § 2.1, Corbin 1987 : 121-125, Fradin 2003 : 288-306). Comme la seule relation morphologique possible est celle existant entre le Nom base et le Verbe dérivé, et qu'elle se manifeste par une préfixation et suffixation simultanée (préfixation en *dé*– et suffixation en *–é*), alors on est en droit de faire l'hypothèse d'une réanalyse en guadeloupéen des paires Nom/Verbe-préfixé du français en parasynthétiques créoles.

De même que les paires Nom/Verbe à finale en *–é* présentées en section 3.2, les paires Nom/Verbe à initiale en *dé*– héritées l'ont été en grand nombre et le schéma morphologique créé à l'issue de cette réanalyse est devenu productif en créole, comme l'attestent les créations de (39) :

(39) a. chèpi → déchèpiyé
         'charpie' 'mettre en charpie'

      b. chouk → déchouké
         'souche' 'déraciner'

      c. pat → dépaté
         'main de banane' 'retirer les mains du régime de banane'

      d. tik → détiké
         'tique' 'retirer les tiques'

      e. zo → dézosé
         'os' 'désosser'

A l'image des paires héritées réanalysées de (36), les créations créoles de (39) s'analysent comme des formations verbales parasynthétiques dans la mesure où ni le verbe en *–é* (40) ni le nom en *dé*– (41) n'existent indépendamment l'un de l'autre :

(40) a. * chèpiyé → déchèpiyé
                      'mettre en charpie'

      b. * chouké → déchouké
                      'déraciner'

      c. * paté → dépaté
                      'retirer la main de bananes du régime'

      d. * tiké → détiké
                      'retirer les tiques'

      e. * zosé → dézosé
                      'désosser'

(41) a. * déchèpi → déchèpiyé
                       'mettre en charpie'

      b. * déchouk → déchouké
                        'déraciner'

      c. * dépat → dépaté
                     'retirer la main de bananes du régime'

    d. * détik → détiké
                  'retirer les tiques'

    e. * dézo → dézosé
                  'désosser'

Ainsi, les conditions requises pour aboutir à la naissance du schéma morphologique (42) en créole guadeloupéen, que nous avons posées au § 3.2.3 sont ici aussi respectées :

1) la très forte représentativité, dans le lexique créole, de couples morphologiques Nom/Verbe à initiale en *dé–* hérités de préfixations verbales dénominales du français ;

2) une quasi-totalité de verbes à finale en *−é*, au sein de ces couples ;

3) et enfin, la propriété des lexèmes verbaux créoles de ne présenter qu'une forme, la marque flexionnelle des verbes hérités n'ayant pas été interprétée comme telle en créole.

Nous pouvons ainsi poser que le créole guadeloupéen dispose d'un schéma morphologique de parasynthèse du type (42), où X représente le lexème de base, de type nominal, et *dé–…–é* l'affixe parasynthétique (circumfixe) formant des verbes. Ce schéma rend aussi bien compte des paires Nom/Verbe héritées du français de (36) que de celles construites en créole en (39) :

(42)   Dé-X_N-*é*$_V$

3.3.1.1 Forme phonologique de l'affixe

La forme phonologique de l'affixe parasynthétique est /de-X-e/ (que nous orthographions *dé-X-é*), où X représente la base nominale et *dé–* … *−é* l'affixe. Les possibles consonnes qui s'intercalent à droite, entre le radical de base et le suffixe *−é* sont à analyser comme des consonnes épenthétiques dans un contexte lexical gauche vocalique, à l'image de ce que nous avons observé pour la suffixation en *−é* (cf. § 3.2.3), que ce soient pour les paires héritées (cf. (43a)) ou pour les paires créoles pour lesquelles nous n'observons qu'un exemple (43b) :

(43) a. i. bò   → débòdé
               'bord'   'déborder'

          ii. figi   → défigiré
              'visage' 'défigurer'

         iii. ma   → dématé
               'mât de bateau'   'démâter (bateau), renverser, retourner'

         iv. mayo   → démayoté
               'maillot, étoffe pour   'démailloter'
               emmailloter un nouveau-né'

v. zo → dézosé
'os'   'désosser'

b. chèpi → déchèpiyé
'charpie'   'mettre en charpie'

L'allomorphie typique que présente le préfixe *dé–* en français et dont a hérité le préfixe *dé–* créole (*dé–* devant verbe à initiale consonantique et *déz–* devant verbe à initiale vocalique ; cf. (44a) pour les paires héritées du français et (44b) pour les exemples de création créole) ne se retrouve pas dans notre corpus de parasynthétiques *dé-X-é*.

(44) a. i. dézabityé   ← abityé
'déshabituer'   's'habituer'

ii. dézakordé   ← akordé
'désaccorder'   'accorder'

iii. dézanbalé ← anbalé
'déballer'   'emballer'

iv. dézanbwaté   ← anbwaté
'désemboîter'   'emboîter'

v. dézankonbré   ← ankonbré
'défaire ce qui était encombré'   'occuper à l'excès un lieu'

b. i. dézantòtiyé ← antòtiyé
'détortiller'   'entortiller'

ii. dézanbaglé ← anbaglé
'débarasser'   'encombrer une table, un meuble'

iii. dézanrajé   ← anrajé
'ne plus être fâché, enragé'   'avoir la rage'

iv. dézapiyé   ← apiyé
'interrompre l'action de s'appuyer'   'appuyer'

En effet, nous ne relevons aucun verbe parasynthétique construit sur base à initiale vocalique. Les seules données qui auraient pu paraître pertinentes sont les hérités *dézosé* 'désosser' et *dézèrbé* 'désherber', mais ils sont analysables en créole sur les bases nominales *zo* 'os' et *zèb* 'herbe' à initiale consonantique.

3.3.1.2 Propriétés sémantiques : sens privatif

Le sens le plus saillant associé à cette formation parasynthétique est ce que la littérature sur les créoles appelle couramment le « sens privatif » régulièrement reconnu pour les formations identiques dans d'autres créoles (cf. Chaudenson 1996 : 27 ; Filipovich 1987 : 44 ; DeGraff 2001 : 78-80, Lefebvre 2003 : 6-8 ; Brousseau 2011 : 70-71). Cette valeur sémantique peut être considérée comme héritée du français où elle est déjà identifiée

comme propre au préfixe verbalisateur *dé–* sur base nominale (cf. Corbin 1987 : 62–63 et 252, par exemple). Plus précisément, ce sens privatif s'inscrit dans une relation spatiale entre le nom de base et le verbe dérivé, relation que les auteurs francophones représentent au moyen de la terminologie cible/site de Vandeloise 1986 (qui correspondent aux oppositions figure/ground ou trajector/landmark de la sémantique cognitive, cf. Fradin 2003 : 298, Amiot 2008 : 10, Jalenques 2014 : 1783). La base nominale de la préfixation en *dé–* du français peut aussi bien dénoter la cible que le site de la relation.

(i) Lorsque la base dénote le site de la relation, le verbe désigne l'action de « sortir de ce que désigne la base » (Jalenques 2014 : 1782) (ce que Corbin 1987 paraphrase par « enlever de X ») : *déterrer, dérailler, débarquer* etc.

(ii) Lorsque la base dénote la cible de la relation, le verbe désigne l'action « d'enlever ce que désigne la base » (Jalenques 2014 : 1782) (ce que Corbin 1987 paraphrase par 'enlever X') : *désosser, déneiger, dépoussiérer, déminer* etc.

Le créole guadeloupéen, en réanalysant les paires Nom/Verbe-préfixé-en-*dé* héritées du français, construit de façon privilégiée des parasynthétiques *dé*-N-$é_V$ dans lesquels le nom de base (désormais Nbase) dénote la cible de la relation (45) :

(45) a. chouk → déchouké
          'souche' 'déraciner'

      b. jouk → déjouké
          'joug' 'enlever le joug'

      c. pat   → dépaté
          'main de banane' 'retirer les mains du régime de banane'

      d. pyèt → dépyété
          'pattes' 'retirer les pattes (crabe)'

      e. tik → détiké
          'tique' 'retirer les tiques'

Comparativement, les parasynthétiques créoles *dé*-N-$é_V$ dont le N dénote le site de la relation sont très faiblement représentés dans notre corpus qui ne comprend que les exemples (46) :

(46) a. bous → débousé
          'bourse' 'dépenser'

      b. tab → détablé
          'table' 'enlever les couverts d'une table'

Cette tendance est largement confirmée par les triplets N / N-$é_V$/*dé*-N-$é_V$ (hérités ou créoles) dont le schéma de construction n'est pas immédiatement transparent (V

→ déV$_V$ ou N → *dé*-N-*é*$_V$?)[7] mais dont les *dé*-N-*é*$_V$ sont compatibles avec une interprétation privative où le nom (N) serait la cible de la relation (47) : là encore, ils sont bien plus nombreux que ceux dont le nom serait le site de la relation (cf. les exemples uniques de (48)) :

(47) a. bach   / baché   / débaché
          'bache' 'bâcher' 'débâcher'

      b. grès   / gresé   / dégrésé
          'graisse' 'graisser' 'dégraisser, enlever la graisse'

      c. kabòs   / kabosé   / dékabosé
          'bosse' 'déformer' 'débosseler'

      d. nat   / naté   / dénaté
          'natte' 'natter des cheveux' 'enlever les nattes'

      e. sèl / salé / désalé
          'sel' 'saler' 'dessaler'

(48) a. kof   / kofré   / dékofré
          'coffre' 'coffrer' 'décoffrer'

      b. kouch / kouché   / dékouché
          'lit'   'se coucher' 'découcher'

      c. plas   / plasé / déplasé
          'place' 'placer' 'déplacer'

      d. tè   / téré   / détéré
          'terre' 'enterrer' 'déterrer'

      e. kwen / kwensé / dékwensé
          'coin' 'coincer' 'décoincer'

La raison de cette nette préférence tient certainement au fait que les paires héritées du français présentent aussi majoritairement cette relation sémantique entre le nom et le verbe (49) comme l'atteste la très faible représentation (3 paires uniquement), au sein de notre corpus, de paires de parasynthétiques *dé*-N-*é*$_V$ dont le N désigne le site de la relation (50)

(49) a. fèy   → déféyé
          'feuilles'   'ôter les feuilles'

      b. fòwm   → défòwmé
          'forme'   'déformer'

      c. kouraj   → dékourajé
          'courage'   'décourager'

---

[7]En effet, dans le cas des triplets, la difficulté tient à ce que l'on ne parvient pas toujours à identifier si le dérivé s'est construit sur le verbe par préfixation ou sur le N par parasynthèse ; comme l'a noté Corbin (1987 : 63) et Amiot (2008 : 12), il existe des « cas d'ambiguïté catégorielle » dont l'interprétation sémantique est compatible avec les deux constructions (par exemple : *débwasé* 'inverse de boiser' ou 'enlever le bois').

    d. kras   → dékrasé
        'crasse'   'décrasser'

    e. mayo   → démayoté
        'maillot, étoffe pour emmailloter le nouveau-né'   'démailloter'

(50) a. bò   → débòdé
          'bord' 'déborder'

      b. bous   → débousé
          'bourse'   'dépenser'

      c. moul   → démoulé
          'moule du gâteau'   'sortir du moule un gâteau'

3.3.1.3 Propriétés sémantiques : autres sens minoritaires

Parallèlement, d'autres sens émergent en créole mais en très faible proportion, reflétant là encore leur faible représentativité dans les paires et les triplets hérités du français :

  (i) le Nbase représente l'objet résultant du procès

(51) a. chèpi   → déchèpiyé
          'charpie'   'mettre en charpie'

      b. gout   → dégouté
          'goutte'   'couler goutte à goutte'

      c. kal   → dékalé
          'raclée'   'tabasser'

      d. katyé   → dékatyé
          'morceau'   'couper en quartier'

  (ii) le Nbase représente l'objet déplacé lorsque le verbe réfère à une localisation ((52a) pour les paires créoles, (52b) pour les paires héritées du français)

(52) a. kaz   → dékazé
          'maison'   'déplacer une case à l'aide d'un véhicule
                       pour l'installer ailleurs'

      b. ménaj   → déménajé
          'ensemble des meubles, des objets   'déménager'
          nécessaires à la vie domestique'

3.3.2 Dé-V préfixés

Ces formations par parasynthèse doivent être distinguées des préfixations en *dé-* sur base verbale qui (i) soit réfèrent au procès inverse de celui que désigne la base (53), (ii) soit ne déclenchent aucun changement sémantique relativement à la base verbale (54).

(53) a. ankayé   → dézankayé
          'se prendre dans les récifs   'enlever des récifs
          (pour un hameçon)'   coraliens'

      b. baké   → débaké
          'embarquer' 'débarquer'

      c. faché   → défaché
          'être faché' 'ne plus être fâché'

      d. manché   → démanché
          'mettre un manche' 'ôter le manche'

      e. rèspèkté   → dérèspèkté
          'respecter' 'manquer de respect'

(54) a. chalviré → déchalviré
          'chavirer'

      b. chiktayé → déchiktayé
          'émietter, mettre en charpie'

      c. libéré → délibéré
          'libérer (qqun de prison)'

      d. rifizé → dérifizé
          'refuser'

      e. viré → déviré
          'tourner en sens inverse'

Bien qu'elles présentent *a priori* des segments phonologiques initiaux et finaux identiques (le préfixe *dé-* et la finale verbale en −*é*) les préfixations sur base verbale se distinguent des parasynthétiques par le fait de ne dériver d'aucun nom. Concomitante à cette différence de construction, se retrouve la relation sémantique entre la base et le dérivé.

3.3.2.1 Préfixation *dé*-V à sens inversif

Dans la majorité des cas, la préfixation en *dé*-V construit un sens non pas privatif mais inversif, comme le reconnaissent les travaux sur les créoles haïtien et saint-lucien. Le sens inversif est différemment appréhendé par les auteurs ayant travaillé sur le français. Si l'on s'en tient aux travaux les plus récents, par exemple de Jalenques (2014 : 1778) qui suit la description proposée par Gerhard-Krait (2000), les verbes préfixés par *dé*− et construits sur base verbale présentent trois acceptions :

a) inversion du résultat du procès exprimé par la base verbale (en lien à ses compléments éventuels) : *dénouer sa cravate* = agir de telle sorte qu'on annule le résultat de « nouer la cravate » ;

b) l'inverse du procès (non résultatif) exprimé par la base : *décroître* = l'inverse de croître ;

c) la négation du procès (non résultatif) exprimé par la base : *déplaire* = ne pas plaire.

Les paires Verbe / *dé*-V$_v$ héritées du français par le créole sont très largement majoritairement du type a) ou b) (55).

(55) a. débatizé   / batizé
          'débaptiser' 'baptiser'

      b. déchosé   / chosé
          'déchausser' 'chausser'

      c. dégonflé   / gonflé
          'dégonfler' 'gonfler'

      d. démayé   / mayé
          'démarier' 'marier'

      e. dézabiyé   / abiyé
          déshabiller 's'habiller'

      f. dézankastré   / ankastré
          'défaire ce qui était encastré' 'encastrer'

      g. dézantòtiyé / antòtiyé
          'détortiller' 'entortiller'

Les paires créoles sont aussi largement de type a) :

(56) a. dégaré   / garé
          'sortir de la place de garage, de stationnement' 'garer, stationner'

      b. dékouvè   / kouvè
          'découvrir' 'couvrir'

      c. dékòviyé   / kòviyé
          'détordre, remettre en position initiale' 'tordre'

      d. dépayé   / payé
          'annuler un pari' 'parier'

      e. dézanbaglé   / anbaglé
          'débarrasser' 'encombrer une table, un meuble'

On ne recense dans le corpus qu'un exemple de type c) cf. (57)

(57) dérèspèkté   / rèspekté
      'manquer de respect, y compris sexuellement' 'respecter'

Les données nous conduisent donc à envisager que le créole, ayant hérité des paires V/ *dé*-V$_V$ les plus disponibles du français –celles à valeur inversive–, a formé sur ces paires, par analogie, les dérivés créoles. Le sens inversif est donc probablement hérité de la préfixation en *dé*– du français. Néanmoins, cette valeur inversive reste cantonnée aux préfixés sur base verbale et n'est représentée dans aucun exemple de parasynthétiques en *dé*-N-*é*$_V$. Ainsi, les deux schémas morphologiques semblent s'être spécialisés sémantiquement en créole :

— le sens privatif est réservé à la parasynthèse *dé*-N-*é*$_V$ (même si d'autres valeurs sémantiques sont possibles);

— le sens inversif est spécifique à la préfixation *dé*-V.

Cette spécialisation sémantique pourrait permettre de trancher l'analyse des triplets N / V / *dé*-N-*é*$_V$ qui apparaissent en bien plus grand nombre dans notre corpus que les parasynthétiques *dé*-N-*é*$_V$ et les préfixés *dé*-V$_V$, tant pour ceux hérités du français (58) que ceux construits en créole (59).

(58) a. apui   / apiyé   / dézapiyé
          'appui'   's'appuyer' 'ne pas s'appuyer'

      b. bwa   / bwazé   / débwazé
          'bois' 'boiser' 'déboiser'

      c. klou   / klouwé / déklouwé
          'clou' 'clouer' 'enlever les clous'

      d. pengn   / pengné   / dépengné
          'peigne' 'peigner' 'dépeigner'

      e. tach   / taché   / détaché
          'tache' 'tacher' 'détacher'

(59) a. bonda   / bondaté   / débondaté
          'fesses'   's'asseoir'   'se lever'

      b. bwa   / bwaré   / débwaré
          'bras' 'enlacer' 'désenlacer'

      c. grij   / griji   / dégriji
          'fronce' 'faire des fronces' 'retirer les fronces'

      d. lyann   / lyanné / délyanné
          'union'   's'unir' 'se désunir'

      e. janm   / janbé   / déjanbé
          'jambe' 'enjamber' 'procès inverse d'enjamber'

3.3.2.2 Préfixation *dé*-V sans changement sémantique

Les formations par parasynthèse *dé*-N-*é*$_v$ doivent, également, être distinguées des pré-fixations en *dé*- sur base verbale (*dé*-V$_v$) qui, à la différence des précédentes ne s'accompagnent d'aucun changement sémantique (cf. en (60) les paires V/*dé*-V$_v$ héritées du français et en (61) celles construites en créoles) :

(60)   a. partajé   → départajé
           'partager'

       b. plimé   → déplimé
           'plumer'

       c. tranpé   → détranpé
           'tremper'

       d. vidé   → dévidé
           'vider'

       e. pozé → dépozé
           'déposer, remettre à sa place'

(61)   a. bwété → débwété
           'boîter, marcher en boitant'

       b. chiktayé → déchiktayé
           'émietter, mettre en charpie'

       c. rifizé   → dérifizé
           'refuser'

       d. sòti   → désòti
           'sortir'

       e. viré → déviré
           'tourner en sens inverse'

Cette absence de variation sémantique associée à la préfixation n'a rien de particulier au créole puisqu'elle est observée en français (Muller 1990, Gerhard-Krait 2000, Apothéloz 2007, Jalenques 2014) (62) et dans d'autres créoles à base française comme le haïtien (Filipovich 1987, Lefebvre 2003, Valdman 1981) (63) ou le saint-lucien (Brousseau 2011 : 74) .

(62)   a. couper → découper
       b. doubler → dédoubler
       c. marquer → démarquer
       d. passer → dépasser
       e. verser → déverser

(63)   a. chiré → déchiré
           'déchirer'

b. chifonnen → déchifonnen
   'froisser'

c. gengole → dégengole
   'se précipiter'

d. grennen → dégrennen
   'égrener'

Une analyse souvent évoquée, tant pour le français que pour le créole, est l'éventualité d'une valeur intensive du préfixé en *dé–* relativement au verbe de base. Bien que cette valeur soit justifiée ponctuellement, elle ne peut tenir pour l'ensemble des cas (voir critique de Jalenques (2014 : 1779) pour le français et de DeGraff (2001) pour le créole). Quoiqu'il en soit, cette propriété ne touche pas les parasynthétiques *dé*-N-*é*$_V$.

4 Conclusion

Le développement, en créole guadeloupéen, de deux schémas morphologiques de formation de verbes par affixation (la suffixation verbale dénominale en −*é* (N-*é*$_V$) et la parasynthèse verbale dénominale *dé*-N-*é*$_V$), est issu de la réanalyse de paires Nom / Verbe héritées du français. Les conditions nécessaires à ces réanalyses s'ancrent crucialement dans la propriété des lexèmes guadeloupéens de ne se réaliser que sous une forme unique. En effet, la majorité des verbes hérités du français présentent un −*é* final probablement issu des formes fléchies de l'infinitif ou du participe passé du verbe français d'origine. Or, c'est ce −*é* final, qui, dans le contexte des paires Nom/Verbe où il apparaît, est réanalysé comme un suffixe dérivationnel, faisant ainsi émerger deux nouveaux schémas morphologiques en créole, inexistants en français. En somme, l'application de la notion de lexème à l'analyse des données créoles permet de reconnaître la validité de ces schémas morphologiques en guadeloupéen alors qu'elle avait conduit à remettre en cause la pertinence de ces mêmes schémas pour les données correspondantes en français.

Ces deux exemples de réanalyse nous conduisent à réfuter la position qui soutient que la dérivation n'émerge que *via* une grammaticalisation graduelle (cf. par exemple McWhorter 1998). Les données du créole guadeloupéen que nous avons examinées nous incitent plutôt à suivre la proposition de Rainer (2015) selon lequel la grammaticalisation n'est qu'un des mécanismes du changement morphologique parmi d'autres, la réanalyse en étant un autre.

Le mécanisme de la réanalyse, qui n'est pourtant pas propre aux langues créoles, y prend néanmoins une place importante du fait de la part massive qu'y occupe le lexique hérité du français. En témoignent d'autres schémas morphologiques tels que la suffixation en −*asyon* en guadeloupéen (*anmerdasyon* 'tracas' ' ← *anmerdé* 'emmerder' ; *pwofitasyon* 'action d'abuser de la faiblesse de qqun' ← *pwofité* 'profiter de la faiblesse de l'autre'), dont la forme phonologique du suffixe est le résultat de l'amalgame de la finale du radical du verbe de base et du suffixe −*ion* des verbes hérités du français (*admirasyon* 'admiration' / *admiré* 'admirer' ; *ògmantasyon* 'augmentation' / *ògmanté* 'organiser') (Villoing & Deglas 2016b).

Références

Amiot, Dany. 2008. La catégorie de la base dans la préfixation en *dé-*. In Bernard Fradin (éd.), *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, 1–15. Amsterdam : John Benjamins.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge : Cambridge University Press.

Apothéloz, Denis. 2002. *La construction du lexique français*. Paris : Ophrys.

Apothéloz, Denis. 2007. Procès centrifuge et procès centripète : RE-, DÉS- et l'antonymie directionnelle en français. In Nabil Hathout & Fabio Montermini (éds.), *Morphologie à Toulouse. actes du colloque international de morphologie 4emes Décembrettes*, 149–169. Munich : Lincom Europa.

Aronoff, Mark. 1994. *Morphology by itself : Stems and inflectional classes*. Cambridge : MIT Press.

Béchade, Hervé. 1992. *Phonétique et morphologie du français moderne et contemporain*. Paris : Presses Universitaires de France.

Bernabé, Jean. 1987. *Grammaire créole. Fondas kréyol-la. Éléments de base des créoles de la zone américano-caraïbe*. Paris : L'Harmattan.

Bhatt, Parth & Emmanuel Nikiema. 2000. Le statut de la nasalité en créole de Ste-Lucie. *Revue Québécoise de Linguistique* 28. 23–45.

Bloomfield, Maurice. 1891. *On adaptation of suffixes in congeneric classes of substantives*. Baltimore : Johns Hopkins Press.

Booij, Geert. 1977. *Dutch morphology*. Dordrecht : Foris.

Booij, Geert. 2010. *Construction morphology*. Oxford : Oxford University Press.

Brousseau, Anne-Marie. 2011. Mesure de la productivité morphologique des créoles : audelà des méthodes quantitatives. *Canadian Journal of Linguistics* 56. 61–86.

Chaudenson, Robert. 1996. Démystification de la relexification. *Études créoles* 19. 93–109.

Chevalier, Jean-Claude, Claire Blanche-Benveniste, Michel Arrivé & Jean Peytard. 1964. *Grammaire Larousse du français contemporain*. Paris : Larousse.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Max Niemeyer Verlag.

Corbin, Danielle. 2004. Français (Indo-européen : roman). In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopetas (éds.), *Morphology. An international handbook on inflection and word-formation*, 1285–1299. Berlin : Walter de Gruyter.

Croft, William. 1991. *Syntactic categories and grammatical relations : The cognitive organization of information*. Chicago : The University Press of Chicago.

Daladier, Anne. 1999. Auxiliation des noms d'action. *Langages* 135. 87–107.

Darmesteter, Arsène. 1894. *Traité de la formation des mots composés dans la langue française*. 2ᵉ éd. Paris : Champion.

DeGraff, Michel. 2001. Morphology in Creole genesis : Linguistics and ideology. In *Ken Hale : A life in language*, 53–121. Cambridge : MIT Press.

Dell, François. 1970. *Les règles phonologiques tardives et la morphologie dérivationnelle du français*. Cambridge : MIT thèse de doct.

Dubois, Jean. 1962. *Étude sur la dérivation suffixale en français moderne et contemporain.* Paris : Larousse.

Filipovich, Sandra. 1987. *La morphologie de l'haïtien.* Université du Québec à Montréal Mémoire de Master.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie.* Paris : Presses Universitaires de France.

Fradin, Bernard, Françoise Kerleroux & Marc Plénat (éds.). 2009. *Aperçus de morphologie du français.* Saint-Denis : Presses Universitaires de Vincennes.

Gardes-Tamine, Joëlle. 1988. *La grammaire. Tomes 1 & 2.* Paris : Armand Colin.

Gerhard-Krait, Francine. 2000. *La préfixation en* dés- *: formes construites et interprétations.* Université Mac Bloch, Strasbourg II Thèse de doctorat.

Gerhard-Krait, Francine. 2012. Du nom *place* aux verbes *déplacer* et *replacer* : quelques questions de legs et d'appropriations sémantiques. *CORELA* HS-12.

Germain, Robert. 1976. *Grammaire créole.* Paris : L'Harmattan.

Grevisse, Maurice & André Goose. 1988. *Le bon usage. Grammaire française.* 12ème édition refondue par A. Goose. Louvain-la-Neuve : Duculot.

Guilbert, Louis. 1975. *La créativité lexicale.* Paris : Larousse.

Haspelmath, Martin. 1995. The growth of affixes in morphological reanalysis. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1994*, 1–29. Dordrecht : Kluwer Academic Publishers.

Haspelmath, Martin. 2002. *Understanding morphology.* London : Arnold.

Hazaël-Massieux, Marie-Christine. 2002. Les créoles à base française : une introduction. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)* 21. 63–86.

Huot, Hélène. 2006. *Morphologie. Forme et sens des mots du français.* 2e édition. Paris : Armand Colin.

Jalenques, Pierre. 2014. Le problème de l'opacité sémantique dans les verbes préfixés en DÉ- : Pour une approche sémantique constructiviste. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4ᵉ congrès mondial de linguistique française (CMLF 2014)*, t. 8, 1775–1796. Institut de Linguistique Française.

Jespersen, Otto. 1922. *Language : its nature, development and origin.* Londres : Allen & Unwin.

Kerleroux, Françoise. 1996. *La coupure invisible : études de syntaxe et de morphologie.* Lille : Presses Universitaires du Septentrion.

Kerleroux, Françoise. 2000. Identification d'un procédé morphologique : la conversion. *Faits de langues* 14. 79–100.

Langacker, Ronald W. 1977. Syntactic reanalysis. In Charles N. Li (éd.), *Mechanisms of syntactic change*, 57–139. Austin : University of Texas Press.

Lass, Roger. 1990. How to do things with junk : Exaptation in language evolution. *Journal of Linguistics* 26. 79–102.

Lefebvre, Claire. 1998. *Creole genesis and the acquisition of grammar : The case of Haitian Creole.* Cambridge : Cambridge University Press.

Lefebvre, Claire. 2003. The emergence of productive morphology in creole languages : The case of Haitian Creole. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 2002*, 35–80. Dordrecht : Kluwer Academic Publishers.

Lignon, Stéphanie & Fiammetta Namer. 2014. Les noms de procès en -*ion* : quand le verbe appelle le verbe. In Florence Villoing, Sophie David & Sarah Leroy (éds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 205–228. Nanterre : Presses Universitaires de Paris Ouest.

Ludwig, Ralph, Danièle Bernini-Montbrand, Hector Poullet & Sylviane Telchid. 2012. *Dictionnaire créole-français Guadeloupe, avec un abrégé de grammaire créole, un lexique français-créole, les comparaisons courantes, les locutions et plus de 1000 proverbes.* Paris : Orphie.

Lyons, John. 1977. *Semantics*. Cambridge : Cambridge University Press.

Matthews, P. H. 1991. *Morphology*. 2nd edition. Cambridge : Cambridge University Press.

McWhorter, John H. 1998. Identifying the creole prototype : Vindicating a typological class. *Language* 74(4). 788–818.

Mel'čuk, Igor. 1996. *Cours de morphologie générale. Troisième partie : moyens morphologiques. Quatrième partie : Syntactiques morphologiques.* Montréal : Presses de l'Université de Montréal.

Meyer-Lübke, Wilhelm. 1894. *Grammatik der romanischen Sprachen. Zweiter Band : Formenlehre.* Leipzig : Reisland.

Mufwene, Salikoko S. & Martha B. Djikhoff. 1989. On the so-called 'infinitive' in Atlantic Creoles. *Lingua* 77. 297–330.

Muller, Claude. 1990. Contraintes de perception sur la productivité de la préfixation verbale en *dé-* négatif. *Travaux de linguistique et de philologie* XXVIII. 172–191.

Namer, Fiammetta. 2009. *Morphologie, lexique et traitement automatique des langues.* London : Hermès Science Publishing.

Namer, Fiammetta. 2012. Nominalisation et composition en français : d'où viennent les verbes composés ? *Lexique* 20. 169–201.

Namer, Fiammetta. 2013. Adjectival Bases of French -*aliser* and -*ariser* verbs : Syncretism or Under-specification ? In Nabil Hathout, Fabio Montermini & Jesse Tseng (éds.), *Morphology in Toulouse. Selected Proceedings of Décembrettes 7 (Toulouse 2-3 December 2010)*, 185–210. München : Lincom Europa.

Nyrop, Kristoffer R. 1936. *Grammaire historique de la langue française.* 2nd edition. T. 3. Formation des mots. Copenhague : Nordisk Forlag.

Plag, Ingo. 1999. *Morphological productivity : Structural constraints in English derivation.* Berlin : Mouton de Gruyter.

Plag, Ingo. 2003. The morphology of Creole languages. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 2002*, 1–2. Dordrecht : Kluwer Academic Publishers.

Poullet, Hector & Sylviane Telchid. 1984. *Dictionnaire des expressions du créole guadeloupéen.* Fort-de-France : Hatier-Antilles.

Rainer, Franz. 2015. Mechanisms and motives of change in word-formation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An*

international handbook of the languages of Europe, 1761–1781. Berlin : de Gruyter Mouton.

Scalise, Sergio. 1994. *Morfologia*. Bologna : Il Mulino.

Seuren, Pieter & Herman Wekker. 1986. Semantic transparency as a factor in creole genesis. In Pieter Muysken & Norval J. Smith (éds.), *Substrata versus universals in creole genesis*, 57–70. Amsterdam : Benjamins.

Thibault, André (éd.). 2012. *Le français dans les antilles : études linguistiques.* Paris : L'Harmattan.

Tourneux, Henry & Maurice Barbotin. 1990. *Dictionnaire pratique du créole de Guadeloupe.* Paris : Karthala.

Tribout, Delphine. 2010. *Les conversions de nom à verbe et de verbe à nom en français.* Université Paris Diderot - Paris 7 Thèse de doctorat.

Tribout, Delphine, Lucie Barque, Pauline Haas & Richard Huyghe. 2014. De la simplicité en morphologie. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4ᵉ congrès mondial de linguistique française (CMLF 2014)*, t. 8 (Institut de Linguistique Française), 1879–1890.

Valdman, Albert. 1978. *Le créole : structure, statut et origine.* Paris : Klincksieck.

Valdman, Albert. 1981. *Haitian Creole English French dictionary.* Bloomington : Indiana University Creole Institute.

Vandeloise, Claude. 1986. *L'espace en français.* Paris : Le Seuil.

Villoing, Florence & Maxime Deglas. 2016a. La formation de verbes dénominaux en guadeloupéen : la part de l'héritage et de l'innovation. In Franck Neveu, Gabriel Bergounioux, Marie-Hélène Côté, Jean-Marc Fournier, Linda Hriba & Sophie Prévost (éds.), *Actes du 5e Congrès Mondial de Linguistique Française. Tours, 4-8 juillet 2016.*

Villoing, Florence & Maxime Deglas. 2016b. La suffixation en -*asyon* et en -*é* du créole guadeloupéen : deux cas de réanalyse de schémas morphologiques du français. *Études créoles* XXXIV(1 & 2). 32–63.

Zribi-Hertz, Anne. 1972. *Remarques sur quelques préfixes du français.* Université Paris 8 Thèse de doctorat.

Chapter 7

Some remarks on clipping of deverbal nouns in French and Italian

Pavel Štichauer

Charles University, Prague

This chapter deals with the restricted class of clipped deverbal nominals in French (e.g. *introduction → intro*) and especially in Italian (e.g. *giustificazione → giustifica*) and aims to show that subtle semantic restrictions seem to constrain such clipping, although there are some differences between the two languages. First, I introduce the well-known distinction between event (E) and result/referential (R) nouns that has been further elaborated by Melloni (2006, 2007, 2011). I then proceed to discuss a class of formations where clipping seems to be sensitive to a special result/object meaning which is very close to what Pustejovsky (1991: 174; see Melloni 2011: 109, 111, 142) calls *information object*. On the basis of a limited class of examples (both attested and hypothetical, e.g. *quantificazione → quantifica*), I argue that where there is such an information object reading available to the relevant nominal, the clipping rule may apply. I take these phenomena to be relevant for Fradin & Kerleroux's (2009: 84–86) *Maximal Specification Hypothesis*, according to which word-formation rules can apply, especially in the case of polysemous lexemes, to specific semantic features inherent in the overall meaning of the base. I demonstrate that clipping can have access to precisely these semantic features.

1 Introduction

It is widely held that morphological phenomena such as clipping (or truncation and blending) can be well explained within a sociolinguistic or pragmatic framework where specific stylistic, diaphasic and/or diastratic factors are at work. Under this view, the only morphologically relevant issue would be that of phonological conditions and constraints on the bases. Nevertheless, there have recently been some attempts to show that there might also be specific semantic constraints that, in some cases, rule out the possibility of such morphological reduction, regardless of any pragmatically constrained context. Such studies demonstrate that truncation may operate in a highly systematic way that involves access to specific semantic information of a given base.

In this chapter, I intend to show that, within the restricted class of clipped dever-bal nominals in French (e.g. *introduction → intro*) and especially in Italian (e.g. *giustificazione → giustifica*), which will be the focus of the present text, special and subtle semantic restrictions seem to constrain the availability of these formations, though the two languages do not cover exactly the same group of formations.

In what follows, I will assume the traditional, though much debated, distinction between inflection and derivation (see, e.g., Spencer 2013: 38-43). Such a distinction is fundamental in that it posits two different roles of morphology: inflectional morphology is supposed to realize the inflected forms of a given lexeme, while derivational morphology serves to create new lexemes.[1] However, the difficulty of the topic to be tackled in the following pages lies precisely in the fact that *clipping* (or *truncation*) does not always seem to deliver an entirely new lexeme.

I shall argue, following Fradin & Kerleroux's (2009: 84–86) *Maximal Specification Hypothesis*, that word-formation rules apply, especially in the case of polysemous lexemes, to specific semantic features inherent in the overall meaning of the base, and that clipping can have access to precisely these semantic features.

The text is organized as follows. In Section 2, I first lay out the well-known distinction between event (E) and result/referential (R) nouns that has been further elaborated by Melloni (2006, 2007, 2011) and that, at first sight, seems to capture some of the known cases. In Section 3, I briefly comment on the French data taken from Kerleroux (1997), Fradin & Kerleroux (2003), and Fradin (2003). In Section 4, I take up the Italian data, based on Thornton (1990, 2004), Štichauer (2006), and Montermini & Thornton (2014) which are, in some fundamental aspects, different with respect to French. In Section 5, I conclude by putting forward a (falsifiable) hypothesis according to which such deverbal nouns are liable to undergo clipping only when special semantic and pragmatic conditions are met. I point out that, contrary to what is usually assumed (especially for Italian), the shortened forms may not always be completely synonymous with their "full" parental nominals.

2 Event/Referential nouns and clipping

Since Grimshaw (1990), the distinction between *complex event nouns*, *simple event nouns* and *result nouns* has become widely accepted, though there has been much critical discussion about the various criteria that Grimshaw herself proposed to individuate the three groups (see Melloni 2011: 21–34).

It has also been thought that only *complex event nouns* can give rise to various result interpretations where the result reading is normally associated with the outcome of the

---

[1]Inflectional morphology provides the word forms inhabiting the cells in the lexeme's paradigm. […], a derivational process defines a new lexeme, which may well have a completely new set of inflectional properties. Therefore, derivational morphology cannot be defined using the same machinery as inflectional morphology, because a derived lexeme is not paradigmatically related to its base and cannot be considered a word form of anything. Rather, it defines an entirely new set of (possibly inflected) word forms. (Spencer 2013: 2).

corresponding complex event noun. Traditional examples of such event/result (E/R) ambiguity are given in (1), where the English examples (1a., 1b.) are given an equivalent version in Italian (1c., 1d.) and French (1e., 1f.).

(1) a. The construction of that house (by the company) took place forty years ago → E

    b. The construction is breathtaking → R

    c. La costruzione di quella casa (da parte dell'impresa) ebbe luogo quarant'anni fa → E

    d. La costruzione è molto bella → R

    e. La construction de la maison (de la part de la compagnie) a eu lieu il y a quarante ans → E

    f. La construction est très belle → R

Simple event nouns (e.g. *party*), instead, do not have an associated event structure, so that the event/result polysemy is not available. Moreover, simple event nouns are said to pattern with result nouns in that they share the same set of properties (see Melloni 2011: 24–25). In what follows, I will assume the general divide between an event-based reading and result-based reading of the derived nominals, discussing various problems in due course.

When it comes to clipping, the general divide between E/R nominals turns out to be relevant as there are specific constraints on the semantic status of the deverbal noun. In fact, as Kerleroux claims (1997: 155), "nouns denoting complex events may not be apocopated".

However, as we shall see, the situation is more complicated since there are more subtle semantic conditions that allow for clipping. More precisely, the clipping rule seems to eliminate the possibility of event noun interpretation (E) regardless of the fact whether the affected noun is a complex event or simple event nominal. Rather, what is required is a specific result/object – or *referential* (R) denotation of the corresponding deverbal noun, as illustrated in (2).

(2) French

    a. La récupération des naufragés fut longue → E

    b. * La récupe des naufragés fut longue → *E[2]
        'The rescue operation of the shipwrecked was long'

    c. J'ai des récupérations à prendre avant Noël → R

    d. J'ai des récupes à prendre avant Noël → R
        'I have some extra days of holiday to take before Christmas'

    e. Il s'oppose à l'introduction du loup à Paris → E

---

[2] Georgette Dal (p.c.) observes that, on the Internet, we can easily find some examples of the eventive reading as well, such as *"La recup(e) a été longue car j'avais une centaine de courriers à récupérer."*

  f. * Il s'oppose à l'intro du loup à Paris → *E
   'He is against the introduction of the wolf into Paris'

  g. Il a apprécié l'introduction de ton livre → R

  h. Il a apprécié l'intro de ton livre → R
   'He enjoyed the introduction of your book'

As far as Italian is concerned, the situation is more intricate. Following Thornton (2004) and Montermini & Thornton (2014), a distinction must be made between those deverbal nouns in *-a* which are the result of the unproductive process of conversion (e.g., *la conquista, la sosta, la firma* etc.), and the apparently identical deverbal nouns in *-a* such as *bonifica, condanna, conferma* whose (diachronic) origin is to be sought in the truncation of the actional suffix *-zione* (see Montermini & Thornton 2014: 187 ff.).

Although the diachronic account is surely on the right track, synchronically the behaviour of pairs of full vs. clipped formations is far from being identical. As I will argue below, it is worth drawing a distinction between three groups.

The first group comprises the pairs of formations which seem to be totally interchangeable displaying (purportedly) absolute synonymy, such as *modificazione / modifica* (3), where both forms display regular E/R ambiguity:

(3) Italian

  a. La modificazione del testo (da parte dell'autore) è stata molto lunga → E

  b. La modifica del testo (da parte dell'autore) è stata molto lunga → E[3]
   'The modification of the text (by the author) took a long time'

  c. La modificazione del testo sarebbe subito saltata fuori → R

  d. La modifica del testo sarebbe subito saltata fuori → R
   'The modification of the text would have surfaced immediately'

  e. La modificazione (del testo) è sul tavolo → R

  f. La modifica (del testo) è sul tavolo → R
   'The modification (of the text) is on the table'

The second group involves partly synonymous formations in which the difference is claimed to lie exclusively at the stylistic level, such as *giustificazione / giustifica* (4), but which may display deeper semantic differences, as I will show, especially when it comes to the difference between an event vs. referential reading. In fact, as the examples in (4) show, the event reading of the clipped form tends to be rather unacceptable.

(4) Italian

  a. Le ripetute giustificazioni dell'assenza (da parte degli studenti) sono intollerabili → E

---

[3]In French, the clipped form *la modif* would also seem to be possible as some examples from the Internet show, such as "*ceux qui sont grisés apparaissent comme dégrisés après la modif du texte.*" (Georgette Dal, p.c.).

b. * Le ripetute giustifiche dell'assenza (da parte degli studenti) sono intollerabili → *E
  'The frequent justifications for absence (on the part of the students) are intolerable'

c. La giustificazione dell'assenza è falsa → R

d. La giustifica dell'assenza è falsa → R
  'The justification for absence is false'

e. La giustificazione è sul tavolo → R

f. La giustifica è sul tavolo → R
  'The justification is on the table'

Finally, a third group, explicitly not addressed in the literature, would involve impossible, unacceptable formations where the clipping of the suffix is disallowed even when the full noun in *-zione* displays some referential reading. The examples in (5) illustrate.

(5) a. La riunificazione delle due Germanie è stata un processo complesso → E

b. * La riunifica delle due Germanie è stata un processo complesso → *E
  'The reunification of the two Germanies was a complex process'

c. Questo sedimento è la stratificazione di rocce diverse → R

d. * Questo sedimento è la stratifica di rocce diverse → *R[4]
  'This sediment is a (result of the) stratification of various rocks'

In what follows, I shall concentrate precisely on these two groups where we find, on the one hand, some attested pairs of full vs. clipped formations with presumably slightly different semantics, and, on the other hand, unattested, yet possible or impossible clipped forms. To begin with, I posit that what the two clipping rules, in French and in Italian, respectively, seem to have in common is a sort of (partial) elimination of event reading of the deverbal noun in favour of a salient referential interpretation. At the same time, a specific semantic condition on the kind of object (i.e. the type of referential reading) is required for the rule in question. In the next sections, after first considering some French and – in more detail – Italian examples, I will argue that a special typology of result nominals (elaborated by Melloni 2011) is needed in order to account for the phenomena in question. I intend to show that a lexical semantic typology of the base verbs will be able to predict, to a large extent, the possibility of clipping.

3 Clipped deverbal nominals in French

In this section, I briefly review the French data, taken from the literature, focusing on the general condition for the clipping rule, which will turn out to be useful in the discussion of the Italian examples as well.

---

[4]I owe this example to Fabio Montermini.

In French, the clipping rule, as far as deverbal nouns with the suffix *-tion* are concerned, may apply to a number of formations.[5] When clipped, the noun receives a special result/object reading although some aspects of event interpretation are maintained. The clipped nouns thus become similar to *simple event nouns*. The internal arguments of the base verb are, in such a formation, excluded (see Kerleroux 1997: 171):

(6)   a.   La manifestation de la vérité aura pris cinquante ans → E

     b.   * La manif de la vérité aura pris cinquante ans → E
        'The demonstration of the truth will have taken fifty years'

     c.   La manifestation (des étudiants) a duré cinq heures → E

     d.   La manif (des étudiants) a duré cinq heures → E
        'The demonstration (of the students) took five hours'

According to Kerleroux (1997: 155), already cited above, the difference lies precisely in the complex / simple event dichotomy. Complex event nominals, which maintain their internal argument structure, cannot undergo clipping, whilst in the case of simple event nouns, such as *manifestation* in the sense of 'demonstration', clipping is allowed.

In the following example, the possibility of clipping is limited to a more concrete (and not *eventive*) interpretation of 'introduction', that of *information-object*.[6] This notion will be of great importance in the discussion of the Italian data.

(7)   a.   L'introduction du lynx dans le massif du Vercors par les responsables de l'ONF → E

     b.   * L'intro du lynx dans le massif du Vercors par les responsables de l'ONF → *E
        'The introduction of the lynx into Vercors Massif by the authorities of the ONF (National Forest Office)'

     c.   L'introduction (de ton livre) compte quatre pages → R

     d.   L'intro (de ton livre) compte quatre pages → R
        'The introduction (of your book) has four pages'

The important point is that clipping in French does not seem to eliminate eventive readings altogether. In the case of event nouns, the difference between pure transpositions (complex event nominals) and what we might call "names of specific events" is relevant. Indeed, as Fradin states, the condition on clipping seems to be that

---

[5]I deliberately leave aside the general context for truncation which, in French, is not limited to complex words (having as its target only the suffix) but may be applied to a wide range of bases, such as *documentation – doc, information – info, actualité – actu,* etc. As Montermini & Thornton (2014: 183) point out, in cases where the truncated material coincides with the suffix (e.g., *invitation – invite*), the coincidence is to be taken as purely fortuitous.

[6]As Fabio Montermini notes (p.c.), such an information-object feature does not prevent, in principle, an event-based reading, as witnessed by the acceptability of *l'intro de son discours a duré une heure*, where *discours* 'speech', being a simple event noun, enables clipping.

(…) d'une manière générale, ne peuvent être accourcies que des expressions nominales fonctionnant comme des dénominations (*names*) d'entités diverses (individu, objet, comportement…). [Generally, what can be shortened are the expressions functioning as denominations, names of various entities such as individuals, objects, behaviour]. (Fradin 2003: 250)

I now turn to the Italian data in order to see further semantic constraints on what kind of entities these generally need to be for clipping to take place.

4 Clipped deverbal nominals in Italian

According to Thornton (1990, 2004: 519), the Italian shortened forms are to be taken simply as stylistic variants of their corresponding full nominals. Furthermore, as Montermini & Thornton (2014: 193–194) show on the basis of corpus frequency, many shortened forms (especially those in *-ifica*) have by now become far more frequent than their full counterparts.

Štichauer (2006) proposes, as already mentioned above, to distinguish three groups of such clipped nominals that behave differently with respect to the original deverbal nouns with the suffix *-zione*.

The first group comprises the pairs such as *modificazione-modifica* (3) or *verificazione-verifica* (8) in which the clipped form has already assumed the same syntactic distribution; moreover, in this case of *verificazione/verifica*, the clipped form is far more acceptable because of its increasing frequency of use.

(8) a. La verificazione della teoria (da parte degli scienziati) è stata affrettata → E
    b. La verifica della teoria (da parte degli scienziati) è stata affrettata → E
       'The verification of the theory (by the scholars) was hasty'
    c. La verificazione (della teoria) va pubblicata su una rivista importante → R
    d. La verifica (della teoria) va pubblicata su una rivista importante → R
       'The verification (of the theory) is to be published in an important journal'
    e. La verificazione (della teoria) è sul tavolo → R
    f. La verifica (della teoria) è sul tavolo → R
       'The verification is on the table'

In the second group of formations we should take into consideration cases in which, on the contrary, we find a shortened form that has a specialized meaning with respect to the noun in *-zione*, e.g. *permutazione - permuta*. While the former noun is a normal event nominal, the latter refers to a specialized type of property exchange.[7] (9):[8]

---

[7] Montermini & Thornton (2014: 196-198) rectify Štichauer's (2006: 33) incorrect claim about the loss of a transpositional relation between the verb *permutare* and *permuta*. In fact, *permuta* clearly functions as an event noun being thus similar to the relation between, say, the French verb *manifester* with respect to *manifestation* and *manif*. Moreover, Montermini & Thornton (2014: 198) suggest that *permuta* is to be taken as a converted form rather than a clipped formation.

[8] The examples are taken from the corpus *La Repubblica* and slightly modified.

(9)   a.   Questo poemetto (…) si fonda sulla permutazione dei ruoli tra l'uomo e l'ani-
          male
      b.   Questo poemetto (…) si fonda sulla *permuta*[9] *dei ruoli tra l'uomo e l'animale*
          'This short poem is based on the permutation of roles between man and ani-
          mal'
      c.   Che dire poi di coloro che cedono la propria auto in permuta?
      d.   Che dire poi di coloro che cedono la propria auto in *permutazione?
          'What can we say then about those who trade in their cars?'

Finally, the third group of nouns would be the one in which clipping is impossible. Al-
though this question is not directly addressed in the literature, I maintain that it is inter-
esting to uncover the constraints that seem to regulate the possibility or impossibility
of a hypothetical nonce-formation. In fact, if only stylistic constraints were at work,
we should find many more examples in various administrative texts than we actually en-
counter.[10] Moreover, if only such diaphasic differences were responsible for the clipping
rule, many a nonce-formation, e.g. *la continua desertificazione del pianeta – la continua
desertifica del pianeta ('the continuous desertification of the planet'), might become ac-
ceptable under specific stylistic circumstances. However, this does not seem to be the
case.

I will limit my analysis to a narrow sample of nouns in *-ificazione* that seem to be the
most frequent deverbal nominals that might, under specific conditions to be stated below,
undergo clipping of the suffix *-zione*. For the present, I will assume that where clipping
is allowed, a special result/object denotation is required or imposed by the mechanism
in question; at the same time, the complex or simple event reading is, in some cases,
partially eliminated.

I shall consider the following six examples: *riunificazione, mercificazione, reificazione,
quantificazione, giustificazione* and *falsificazione*. I will employ roughly the same "diag-
nostic" contexts also used by Melloni 2011. This step is obviously problematic for the
simple reason that the diagnostic contexts do not always yield an entirely natural ex-
ample, attested or "attestable" in the corpora. I attempt to remedy this shortcoming
by modifying or integrating the examples according to real data present in the corpus
CORIS/CODIS[11], *La Repubblica*,[12] or on the Internet (by a general search on google.it).
When necessary, I also add a clarifying footnote (especially when native speakers' judge-
ments tend to give variable results).

---

[9]In fact, web search on google.it (http://www.ilcovile.it/news/archivio/00000420.html) provides one exam-
ple of the shortened form *permuta* in precisely this context. The sequence *permuta dei ruoli* can be found
in the Italian translation of Jankélévitch's book *Le Paradoxe de la morale.*

[10]For instance, in the corpus of *La Repubblica* (330 million tokens), we find about 150 different types in
-ificazione, and about 90 forms ending in *-ifica*, where after careful post-processing, about a dozen for-
mations remain and virtually no *hapax* qualifying as a real neologism can be found (*la chiarifica* being
probably the only exception).

[11]Accessible at: http://corpora.dslo.unibo.it/TCORIS/. Accessed September-October, 2016.

[12]Accessible at: http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica. Accessed Sep-
tember-October, 2016.

I begin with *riunificazione*. In (10), we see that the only available reading is that of an event, all possible result/referential readings are excluded simply because *riunificare* does not belong to any product-oriented verbs (in the sense of Melloni 2011: 184 ff.):

(10) a. La riunificazione delle due Germanie ha richiesto molto tempo → E
      b. * La riunifica delle due Germanie ha richiesto molto tempo → *E
          'The reunification of the two Germanies took a long time'
      c. * La riunificazione è falsa → *R
      d. * La riunifica è falsa → *R
          (intended) 'The reunification is false'
      e. * La riunificazione è sul tavolo → *R
      f. * La riunifica è sul tavolo → *R
          (intended) 'The reunification is on the table'

In the case of *mercificazione* ('commodification') we find essentially the same situation.

(11) a. Questo processo di (continua) mercificazione del corpo femminile → E
      b. * Questo processo di (continua) mercifica del corpo femminile → E
          'This process of (continuous) commodification of the female body'
      c. * Le presenti mercificazioni del corpo femminile non sono affatto belle → *R
      d. * Le presenti mercifiche del corpo femminile non sono affatto belle → *R
          (intended) 'The present commodifications of the female body are not nice at all'
      e. * La mercificazione è sul tavolo → *R
      f. * La mercifica è sul tavolo → *R
          'The commodification is on the table'

It could be argued, however, that the verb *mercificare* is semantically close to verbs of creation (by modification). The impossibility of having an R-reading might be due to the same reasons for which *edificazione* from *edificare*, as a typical creation verb, does not display any result/object interpretation. Melloni (2011: 189) suggests that a possible R-interpretation is blocked by the existing lexeme *edificio*.

Analogous behaviour is also exhibited by *reificazione* (12) ('reification'), which is acceptable only in the eventive reading.

(12) a. Le osservazioni di L. C. sulla (costante) reificazione dei bambini meritano...
          → E
      b. Le osservazioni di L. C. sulla (costante) *reifica dei bambini meritano... → *E
          'L. C.'s remarks on the (constant) reification of children deserve...'
      c. * La reificazione è interessante → *R
      d. * La reifica è interessante → *R
          (intended) 'The reification is interesting'
      e. * La reificazione è sul tavolo → *R

f.   * La reifica è sul tavolo → *R
    (intended) 'The reification is on the table'

In the nouns in (10-12) we thus find that the only possible interpretation is the one associated with event nominals, the result reading of the *construction*-type nouns being ruled out. Arguably, the absence of such a result/object aspect is the factor that does not allow for further clipping of the formation. Indeed, the result/object reading seems to be a necessary, albeit not sufficient, condition. As we will see in the examples below (13-17), clipping seems to be sensitive to a special result/object meaning which is very close to what Pustejovsky (1991: 164; see Melloni 2011: 109, 111, 142) calls *information object*. It thus appears that where there is such an information object reading available to the relevant nominal, the clipping rule may apply.

I now pass to the discussion of such nouns. I start with *quantificazione*. In example (13b), we can see that the shortened form is less acceptable in the eventive reading.[13] The referential reading – conveying an information-object interpretation – allows for clipping giving rise to a possible nonce-formation °*la quantifica*.[14]

(13)   a.   La quantificazione dei costi deve essere effettuata al più presto → E

    b.   ?* La quantifica dei costi deve essere effettuata al più presto → ?*E
        'The quantification of the costs must be carried out immediately'

    c.   La quantificazione (dei costi) contiene un errore → R

    d.   ° La quantifica (dei costi) contiene un errore → R
        'The quantification (of the costs) contains an error'

    e.   La quantificazione è sul tavolo → R

    f.   ° La quantifica è sul tavolo → R
        'The quantification is on the table'

I argue that the pair *giustificazione / giustifica*, seen above in example (4), repeated here as (14), shows essentially the same behaviour despite Montermini & Thornton's (2014: 192) claim about its total synonymy.

(14)   a.   Le frequenti giustificazioni dell'assenza (da parte degli studenti) sono intoller-abili → E

    b.   * Le frequenti giustifiche dell'assenza (da parte degli studenti) sono intollera-bili → *E
        'The frequent justifications for absence (on the part of the students) are in-tolerable'

---

[13]Some speakers tend to accept the shortened form even in this eventive context (Fabio Montermini finds it totally acceptable without perceiving any difference whatsoever). Thus, it would be necessary to see whether all possible *eventive* contexts, offered below for *giustifica*, would equally yield a more or less acceptable formation. The corpora offer no example. However, an internet search conducted in July 2017 found 7 hits, including an example where the author puts the formation within quotation marks in order to signal its peculiar (neological?) status: *Secondo me è una discreta opportunità di lavoro con contratto biennale, ma ho bisogno di una "quantifica" dei costi che io non so proprio fare.*

[14]I follow here Corbin's (1987) use of the ° sign to mark possible, yet unattested formations. However, as we have seen, the formation *quantifica* is modestly attested (albeit to a very limited extent).

    c.   La giustificazione dell'assenza è falsa → R

    d.   La giustifica dell'assenza è falsa → R
          'The justification for absence is false'

    e.   La giustificazione è sul tavolo → R

    f.   La giustifica è sul tavolo → R
          'The justification is on the table'

The example thus deserves more discussion. Montermini & Thornton (2014: 192) claim that *giustificazione* and *giustifica* are absolutely synonymous (differing only in the register, the latter being typical of a school jargon). To support this apparently indubitable fact, they adduce not only their native speaker judgements but also some corpus evidence, such as the (fixed) sequence *libretto delle giustificazioni / libretto delle giustifiche* which appears in a large number of official school rules and regulations. However, I argue that the synonymy of this pair is limited to just the *referential* reading where, indeed, the two formations are wholly interchangeable. Yet, in the eventive readings, the synonymy is far less obvious.

First, as shown above in examples (14a, 14b), if subjected to different tests of actionality, the form *giustifica* turns out to be ruled out. Drawing (loosely) on Anscombre's (1986) tests of actionality, I point out that the following constructions highlight the problems at hand.

(15)   a. Gli studenti hanno sempre trovato un metodo di giustificazione /*giustifica delle loro assenze
          'The students have always found a method of justification of their absences'

    b. In caso di mancata giustificazione / ?? giustifica dell'assenza da parte degli alunni, verrà attivata un'azione disciplinare [15]
          'Failure on the part of the student to provide justification of the absence may result in disciplinary action'

    c. Ora procediamo alla giustificazione/*giustifica delle assenze
          'Now let's move on to justifying the absences'

    d. Non si può accettare una giustificazione/*giustifica così frettolosa
          'It's not possible to accept such a hasty justification'

What I stress is that the clipped form, displaying a clear information-object meaning (*la giustifica* is primarily a written document), is far less acceptable in all eventive readings enhanced by the constructions of the type seen in (15). I argue that such a semantic condition, though being probably just a slight tendency, can be best seen in the example of *falsificazione*. The underlying verb, *falsificare*, can have two meanings, a material one of *falsificare la moneta, la carta di credito* etc. (to falsify the money, the credit card) and a Popperian sense of *falsificare un'ipotesi* (to falsify a hypothesis). When *falsification* is

---

[15]For some speakers, in fact, *giustifica* is acceptable even in this dynamic reading, while for others it tends to be ruled out.

understood in the "material" sense, clipping seems to be ruled out (16), but when it comes to the other meaning, an information-object reading appears to be more acceptable (17) given that the falsification of a hypothesis may in fact be a written document.

(16) a. La falsificazione delle carte di credito (da parte di alcune persone) è sempre stata facile → E

   b. * La falsifica delle carte di credito (da parte di alcune persone) è sempre stata facile → *E
      'The falsification of the credit cards (by some people) was always easy'

   c. Questa carta di credito è una falsificazione → *R

   d. * Questa carta di credito è una falsifica → *R
      'This credit card is a falsification'

   e. La falsificazione (della carta di credito) è sul tavolo → *R

   f. * La falsifica (della carta di credito) è sul tavolo → *R
      'The falsification (of the credit card) is on the table'

(17) a. La falsificazione di quella ipotesi (da parte dello studioso) non ha richiesto molto tempo → E

   b. * La falsifica di quella ipotesi (da parte dello studioso) non ha richiesto molto tempo → *E
      'The falsification of that hypothesis (by the scholar) didn't take much time'

   c. La falsificazione (di quella ipotesi) è geniale → R

   d. ° La falsifica (di quella ipotesi) è geniale → R
      'The falsification (of that hypothesis) is brilliant'

   e. La falsificazione è sul tavolo → R

   f. ° La falsifica è sul tavolo → R
      'The falsification is on the table'

What the two contexts have in common is a possibility of having a result-object interpretation. But while in (16c–16f) the referential reading is more "material", in (17c–17f), the information-object reading of *falsificazione* strongly favours the acceptability of the clipped variant *falsifica* (see also Montermini & Thornton 2014: 196, note 16 on *falsifica*).[16] I take this case, along with the others discussed above, as an example of Frazdin's hypothesis that hypothesis according to which

> [...] un procédé dérivationnel donné opère de manière discriminante sur l'une ou l'autre de ces significations. [a given derivational process operates differentially on one or the other of these meanings.] (Fradin & Kerleroux 2009: 86)

---

[16]The form *falsifica* is in fact attested on the Internet only a couple of times, in a context that seems to be due to strong analogy with *verifica*: "…sostituire alle procedure rigorose di verifica e **falsifica di** proposizioni scientifico-sperimentali un metodo simile a quello storico-comprensivo…"; "…i dati sperimentali sono il fondamento della verifica/**falsifica di** ogni ipotesi scientifica…"; "…isolare e selezionare quei fatti, e quei modi di viverli, che consentono la verifica (o **falsifica) di** date ipotesi…"; "…Gli epistemologi hanno così iniziato a riflettere e a cercare situazioni di verifica o di **falsifica di** queste ipotesi…"

5 Concluding remarks

On the basis of the data so far analyzed – which represent only a very limited sample – I now conclude by summarizing my main proposal.

I maintain that the clipping rule is sensitive to the information-object meaning of the construction in -*zione*. Such an information-object meaning can be predicted from the general semantics of the base verb.

What Melloni (2011: 108) considers to be the core meaning of what she calls the R nominals may be captured in the following four or five classes based on the semantics of the underlying verb: the *product*, *means*, *path and measure*, *entity in state* verbs and the *sense extensions*. She shows that inside of the *product-oriented* nominals a further division is to be made between *creation/result-object* verbs (such as *costruire*), *creation-by-representation verbs* (such as *tradurre*) and *creation-by-modification verbs* (such as *correggere*). The *representation* (and also *modification*) class of *creation verbs* can, as Melloni puts it

> […] undergo a metonymic displacement and convey the concrete interpretation of its container object, (a piece of paper, for instance) […]. (Melloni 2011: 201)

Furthermore, still inside this class of *creation verbs*, there is another non-prototypical group of *speech act verbs* (see Melloni 2011: 213–214) which convey a proposition that can be, once again, understood as *information object* à la Pustejovsky (1991), as, for example, *confessione, communicazione* etc. In such a perspective, we could also reconsider the already lexicalized nouns as, for instance, *condanna, confisca, deroga, proroga, ratifica, nomina* etc. (see Thornton 2004: 519). But this is, of course, a matter of future research. For the present, I only wished to show that a general *information-object* meaning can indeed be a relevant factor in a (marginal) process of clipping of the Italian nouns in -*ificazione*.

Acknowledgments

References

Anscombre, Jean-Claude. 1986. L'article zéro en français: un imparfait du substantif? *Langue française* (72). 4–39.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Max Niemeyer Verlag.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Fradin, Bernard & Françoise Kerleroux. 2009. L'identité lexémique. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 83–102. Saint-Denis: Presses Universitaires de Vincennes.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge: The MIT Press.

Kerleroux, Françoise. 1997. L'apocope et les déverbaux. *Cahiers de grammaire* (22). 155–186.

Melloni, Chiara. 2006. Logical polysemy in word formation: E and R suffixes. *Lingue e linguaggio* V(2). 281–308.

Melloni, Chiara. 2007. *Polysemy in word formation: The case of deverbal nominals*. Università di Verona dissertation.

Melloni, Chiara. 2011. *Event and result nominals*. Bern: Peter Lang.

Montermini, Fabio & Anna M. Thornton. 2014. Su alcune differenze tra deverbali formati per conversione e deverbali formati per riduzione in italiano e francese. In *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 179–203. Paris: Presses Universitaires de Paris Ouest.

Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics* 17(4). 409–441.

Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press.

Štichauer, Pavel. 2006. Codifica versus codificazione, condanna versus condannazione: riflessioni su alcuni sostantivi italiani nati per troncamento di *-zione*. *Linguistica Pragensia* 16(1). 28–36.

Thornton, Anna M. 1990. Vocali tematiche, suffissi zero e "cani senza coda" nella morfologia dell'italiano contemporaneo. In Monica Berretta, Piera Molinelli & Ada Valentini (eds.), *Parallela 4, morfologia/morphologie*, 43–52. Tübingen: Gunter Narr Verlag.

Thornton, Anna M. 2004. Femminili in *-a* nati per troncamento di suffisso (Troncamento di *-zione*: il tipo *revoca*). In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 518–520. Tübingen: Niemeyer.

Part III

Troubles with lexemes

Chapter 8

Lexeme and flexeme in a formal theory of grammar

Olivier Bonami

Laboratoire de linguistique formelle, Université Paris Diderot

Berthold Crysmann

Laboratoire de linguistique formelle, CNRS

This paper deals with the role played by the notion of a lexeme in a constraint-based lexicalist theory of grammar such as Head-driven Phrase Structure Grammar. Adopting a Word and Paradigm view of inflection, we show how the distinction between lexemes, individuated by their lexical semantics, and flexemes, individuated by their inflectional paradigm, can fruitfully be integrated in such a framework. This allows us to present an integrated analysis of stem spaces, inflection classes, heteroclisis and overabundance.

It is often observed by morphologists that contemporary work in theoretical morphology has little impact on formal theories of grammar, which on average are content with a view of morphology quite close to that of offered by the post-Bloomfieldian morphemic toolkit. A notable exception to this situation is the pervasive use in Head-driven Phrase Structure Grammar (henceforth HPSG) of the distinction between WORDS and LEXEMES familiar from Word and Paradigm approaches to morphology (see among many others Robins 1959, Hockett 1967, Matthews 1972, Zwicky 1985, Anderson 1992, Aronoff 1994, Stump 2001, Blevins 2016). In this paper we reevaluate the role of the lexeme in HPSG in the light of 20 years of research, and in particular of recent attempts to integrate a truly realisational theory of inflection within the HPSG framework (Crysmann & Bonami 2016). We conclude that current theorizing conflates two distinct notions of an abstract lexical object: lexemes, which are characterised in terms of their syntax and semantics, and FLEXEMES (Fradin & Kerleroux 2003), which are characterised in terms of their inflectional paradigm. We propose distinct formal representations for lexemes and flexemes, and explore the benefits of the distinction for a formally explicit theory of morphology and the morphology-syntax interface.

The structure of the paper is as follows. In Section 1, we present the standard view of the lexeme in contemporary HPSG, and show that lexemes are given a dual representation, as a distinct type of signs and as the value of the feature LID. In Section 2, we present

Information-based Morphology (IbM),[1] an HPSG-compatible realisational approach to inflection, and show that lexemes-as-signs have no role to play in an HPSG using IbM as its inflectional component. In Section 3 we discuss Fradin and Kerleroux's distinction between lexemes and flexemes, and argue that this should be encoded by distinguishing a feature LID and the values it can take from *pid* objects: while the former reside in syntactic/semantic representations, the latter are found in inflection proper. Finally in section 4 we discuss the consequences of the distinction between LID and *pid* for the modelling of heteroclisis and overabundance.

1 The lexeme in standard HPSG

1.1 Lexemes as a distinct type of lexical signs

Most current work in Head-driven Phrase Structure Grammar (henceforth HPSG; Pollard & Sag 1994, Ginzburg & Sag 2000, Sag et al. 2003) and its variant Sign-Based Construction Grammar (henceforth SBCG; Boas & Sag 2012) embraces the notion of a lexeme, familiar from Word-and-Paradigm approaches to morphology. Under this view, a lexeme is an abstract lexical object encapsulating what is common to the collection of words belonging to the same inflectional paradigm. Although the details are complex and disputed, it is uncontroversial enough to assume that a lexeme may be comprised of some amount of phonological information (in the form of a stem, a collection of stem alternants, a consonantal pattern, etc.), morphological information (e.g. inflection class information), syntactic information (at the very least part of speech and valence information), and semantic information corresponding to a notion of 'lexical meaning' (plus linking of semantic roles to syntactic dependents). Inflection is then concerned with the relation between (abstract) lexemes and (concrete) words,[2] while 'word formation', more adequately called LEXEME FORMATION (Aronoff 1994), is concerned with morphological relations between lexemes.

Since the late 1990s a growing consensus has emerged within HPSG that lexemes should be treated as signs on a par with words.[3] That is, the hierarchy of linguistic objects includes the subhierarchy in Figure 1. Syntactic rules may form phrases by combining signs of type *syn-sign*, while rules of morphology manipulate only signs of type *lex-sign*.

This is intended to implement the notion of strong lexicalism. First, words constitute the interface of morphology and syntax, since they belong to both types. Second, morphology and syntax are discrete components of grammar inasmuch as some aspects of

---

[1] The framework is presented and elaborated in Bonami & Crysmann (2013, 2016), Crysmann (2017), Crysmann & Bonami (2016). The name is intended as a reference to Pollard & Sag's (1987) *Information-based Syntax and Semantics*. In IbM, the notion of information in the sense of feature logic plays a central role in determining morphological wellformedness, defined in terms of exhaustive expression of morphosyntactic properties. Furthermore, IbM implements Paninian competition on the basis of subsumption, a measure of informativity in feature logic.

[2] Alternatively, within an *abstractive* conceptualisation of morphology (Blevins 2006), where words are seen as primitives rather than derived objects, inflection is concerned with the relation between words in a paradigm, and the abstract notion of a lexeme captures what is common between these words.

[3] See Bonami & Crysmann (2016) for a thorough overview of work on morphology in HPSG.
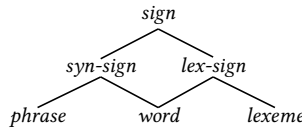
Figure 1: A standard HPSG subhierarchy of signs

the feature geometry of signs will be specific to phrases or lexemes; likewise, this architecture allows for the possibility that the kind of combinatory rules relating phrases to their component parts be very different from the kind of combinatory rules relating words to their component parts.

Although this is by no means an obligation, as we will see below, standard practice in HPSG and SBCG in the past two decades has been to assume an Item and Process view of morphology (Orgun 1996, Riehemann 1998, Koenig 1999, Müller 2002, Sag et al. 2003, Sag 2012), where the word-lexeme opposition captures the difference between inflection and lexeme formation. Rules of inflection map a lexeme to a word, rules of derivation map a lexeme to a lexeme, rules of composition map two lexemes to a lexeme. The three toy rules in Figures 2, 3 and 4 illustrate the basic architecture.
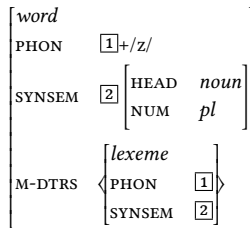
Figure 2: Simplified rule of regular English plural formation.
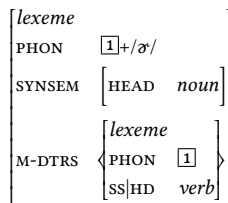
Figure 3: Simplified rule of English Agent noun formation.

Formally, morphological rules are modeled on a par with phrase-structure rules, except for the fact that, in inflectional and derivational rules, the relation between the phonology of the mother (the output lexical sign) and the phonology of the daughter (the input lexical sign) is specified syncategorematically: affixes are not signs, but bits

$$
\begin{bmatrix}
lexeme \\
\text{PHON} & \boxed{1}+\boxed{2} \\
\text{SYNSEM} & \begin{bmatrix} \text{HEAD} & noun \end{bmatrix} \\
\text{M-DTRS} & \left\langle \begin{bmatrix} lexeme \\ \text{PHON} & \boxed{1} \\ \text{SS}|\text{HD} & noun \end{bmatrix}, \begin{bmatrix} lexeme \\ \text{PHON} & \boxed{2} \\ \text{SS}|\text{HD} & noun \end{bmatrix} \right\rangle
\end{bmatrix}
$$
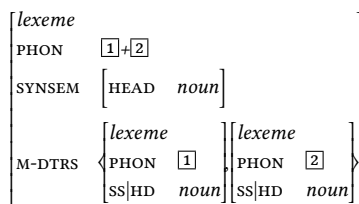
Figure 4: Simplified rule of English noun-noun compound formation

of phonology added by rule.[4] The main difference between inflection and lexeme forma-
tion rules lies in the fact that inflection does not modify the SYNSEM value, but merely
expresses some of its aspects. The main specificity of composition is that the input (the
daughter signs) consists of two lexemes rather than one. Figures 5 and 6 illustrate typical
morphological analyses within such a framework.

$$
\begin{bmatrix}
word \\
\text{PH} & /l\text{ʌ}v\text{ɚ}z/ \\
\text{SS}|\text{HD} & \boxed{1} \begin{bmatrix} \text{NUM} & pl \end{bmatrix} \\
\text{M-DTRS} & \left\langle \begin{bmatrix} lexeme \\ \text{PH} & /l\text{ʌ}v\text{ɚ}/ \\ \text{SS}|\text{HD} & \boxed{1}\ noun \\ \text{M-DTRS} & \left\langle \begin{bmatrix} lexeme \\ \text{PH} & /l\text{ʌ}v/ \\ \text{SS}|\text{HD} & verb \end{bmatrix} \right\rangle \end{bmatrix} \right\rangle
\end{bmatrix}
$$
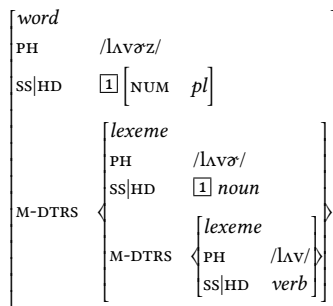
Figure 5: Analysis of the noun *lovers* under an Item-and-Process view of mor-
phology

1.2 Lexeme identifiers

It is sometimes necessary for a lexical entry or syntactic construction to be able to select a
particular lexical item in its environment. One clear case of this is that of flexible idioms.
Consider the idiom *pull strings* 'try something'. As the examples in (1) make clear, while
the idiomatic meaning is present only when the object of *pull* is headed by the lexeme
strings, the noun may occur in either singular or plural form, and combine with a variety
of determiners and modifiers (Bargmann forthcoming).

(1)   a.   There I learned whom [*sic*] my secret advocate was, the man who had pulled
           strings to get me the teaching job in the midst of a terrible economy, and who

---

[4]For a dissenting view see Emerson & Copestake (2015).
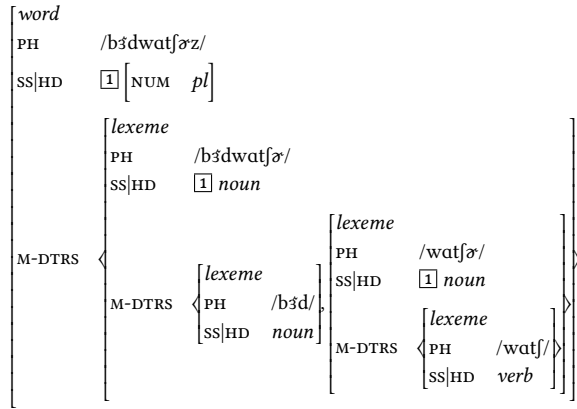
Figure 6: Analysis of the noun *birdwatchers* under an Item-and-Process view of morphology

had pulled more strings to allow me to keep it, and who had then pulled even more strings to have my commission assigned to the Abwehr.[5]

b. You'll never know the trouble I had, and the strings I had to pull to get you back from Berlin.[6]

c. We have to remember that Jacob was at their wedding. Just how many strings did he pull?[7]

d. So I didn't pull any string. Didn't need to.[8]

e. When I got the job, I thought to myself, "Someone upstairs finally pulled a string for me".[9]

f. No string was pulled, it was based on merit.[10]

This type of situation motivated the introduction of the feature LID (or LEXEME IDENTI-FIER) as a head feature projecting to phrasal level information as to which lexeme heads a phrase (Sag 2007, 2012).[11] Simplifying matters considerably, one can see the constructions above as licensed by the two idiomatic lexical entries in Figure 8, which contrast with the two ordinary entries in Figure 7: a special lexical entry of *pull* with idiomatic meaning selects specifically for an object headed by a form of *strings* with idiomatic meaning. The postulation of a specific LID value for idiomatic *string* allows idiomatic *pull*

---

[5]K. Ryan, *The Somnambulist*, New York: iUniverse, 2006.

[6]K. McDermott, *The time of the corncrake*, Victoria: Trafford, 2004.

[7]http://www.losttv-forum.com/forum/showthread.php?t=65542. Accessed on November 26, 2016.

[8]http://www.justusboys.com/forum/archive/index.php/t-437037.html. Accessed on November 26, 2016.

[9]http://ultraphrenia.com/2016/10/02/a-cigarette-break-behind-heavens-gate. Accessed on November 13, 2016.

[10]http://obafemayor02.blogspot.fr/2013_03_24_archive.html. Accessed on November 26, 2016.

[11]Note that a very similar role is played by the feature LISTEME in Soehn (2006) and Richter & Sailer (2010).

to select for a specific combination of an inflectional paradigm with an idiomatic meaning, while abstracting away from inflectional and syntactic variability in the makeup of the object of *pull*.

$$
\begin{bmatrix}
\textit{lexeme} \\
\text{PHON} & /p\upsilon l/ \\
\text{HEAD} & \begin{bmatrix} \text{LID} & \textit{pull-lid} \end{bmatrix} \\
\text{VAL} & \langle \text{NP}_i, \text{NP}_j \rangle \\
\text{CONT} & \begin{bmatrix} \textit{pull-rel} \\ \text{ACT} & i \\ \text{UND} & j \end{bmatrix}
\end{bmatrix}
\quad
\begin{bmatrix}
\textit{lexeme} \\
\text{PHON} & /st\text{ɹ}\text{ɪ}\text{ŋ}/ \\
\text{HEAD} & \begin{bmatrix} \text{LID} & \textit{string-lid} \end{bmatrix} \\
\text{CONT} & \textit{string-rel}
\end{bmatrix}
$$

Figure 7: Ordinary lexical entries for *pull* and *strings*

$$
\begin{bmatrix}
\textit{lexeme} \\
\text{PHON} & /p\upsilon l/ \\
\text{HEAD} & \begin{bmatrix} \text{LID} & \textit{idiomatic-pull-lid} \end{bmatrix} \\
\text{VAL} & \langle \text{NP}_i, \text{NP}_j[\text{LID} \; \textit{idiomatic-string-lid}] \rangle \\
\text{CONT} & \begin{bmatrix} \textit{use-rel} \\ \text{ACT} & i \\ \text{UND} & j \end{bmatrix}
\end{bmatrix}
\quad
\begin{bmatrix}
\textit{lexeme} \\
\text{PHON} & /st\text{ɹ}\text{ɪ}\text{ŋ}/ \\
\text{HEAD} & \begin{bmatrix} \text{LID} & \textit{idiomatic-string-lid} \end{bmatrix} \\
\text{CONT} & \textit{influence-rel}
\end{bmatrix}
$$

Figure 8: Idiomatic lexical entries for *pull* and *strings*

The feature LID provides a useful mechanism for spreading lexical information in syntactic structures that has been used since in the analysis of complex predicates (Müller 2010) and periphrastic inflection (Bonami & Webelhuth 2012, Bonami & Samvelian 2015, Bonami 2015, Bonami et al. 2016). It also provides a direct encoding of lexemic identity. Since LID is a head feature, and inflected words share the HEAD value of the lexeme they are derived from, all inflected forms of a lexeme will have the same LID. Under the natural assumption that all lexemes have a distinct LID value, whether two words instantiate the same lexeme can thus be deduced by inspection of their LID values, without examining their derivation history.

2 The lexeme in a Word and Paradigm version of HPSG

2.1 Going Word and Paradigm

While an Item and Process view of morphology has been dominant in the HPSG literature, over the last 20 years a number of authors have become more vocal in advocating the incorporation into HPSG of a Word and Paradigm view of inflection (see among others Erjavec 1994, Miller & Sag 1997, Ackerman & Webelhuth 1998, Crysmann 2002,

Bonami & Boyé 2006, Bonami & Webelhuth 2012, Bonami 2015, Bonami & Samvelian 2015, Crysmann & Bonami 2016). Under such a view, rules of inflection do not incrementally specify how a basic sign is augmented with morphosyntactic information and phonological exponents; rather, a full morphosyntactic specification of the word is given as input to a system of rules of exponence indicating how such a specification is partially realised by exponents in various positions with respect to the basic stem. The arguments in favour of such a move are the usual ones (Matthews 1974, Zwicky 1985, Anderson 1992, Stump 2001, Brown & Evans 2012): systems of exponence depart too strongly from a one-to-one correspondence between form and content for the Item and Process view to make sense in the general case. We will not rehearse these arguments here, and simply make the sociological observation that Word and Paradigm approaches have over the last two decades become the *de facto* standard for theoretical and typological reasoning on inflection systems.

Recent attempts at implementing Word and Paradigm inflection in HPSG come in two flavors. One the one hand, Bonami & Webelhuth (2012), Bonami (2015), Bonami & Samvelian (2015) explicitly interface Paradigm Function Morphology (Stump 2001, 2016) with HPSG through a set of relational constraints. On the other hand, Crysmann & Bonami (2016) design a realisational framework for inflection native to the HPSG architecture, *Information-based Morphology* (IbM), making heavy use of the underspecification techniques provided by a typed feature structure formalism.

Figure 9 illustrates the main features of IbM by way of the analysis of a rather simple inflected word, the French verb *buvions* 'we drank'. IbM specifies the inflectional system of a language as a set of constraints relating a word's SYNSEM value to its PHONOLOGY. In the present example, a word realising the past imperfective of the verb BOIRE in the context of a 1PL subject is constrained to have the string /byvjɔ̃/ as its phonological realisation. The specification of these constraints makes use of three intermediate, strictly morphological, representations. The feature MS (standing for 'morphosyntactic properties') encodes those syntactic and semantic properties of the word that are relevant to inflection, in a format suitable for the expression of constraints on exponence. The feature MPH (standing for 'morphs') indicates the set of morphs making up the word, indexed for their position within the word (PC, standing for 'position class'). Finally, the feature RR (standing for 'realisation rules') indicates which generalisations on the relationship between morphosyntactic properties and morphs license the particular association between form and content instantiated in that word. Importantly, realisation rules relate a *set* of morphosyntactic properties (listed under MUD, standing for 'morphology under discussion') to a *set* of morphs (listed under MPH). Thus, while in this simple example, there is a one-to-one mapping between properties and morphs, IbM realisation rules can just as easily accommodate cumulative exponence (m properties : 1 morph), extended exponence (1 property : n morphs), overlapping exponence (m properties : n morphs), and zero exponence (m properties : 0 morphs).

The relationship between the various features is regulated by a set of general principles that we will only state in prose here; we refer the reader to Bonami & Crysmann (2013) or Crysmann & Bonami (2016) for a more explicit formulation. Let us start with the relation-

$$
\begin{bmatrix}
word \\
\text{PHON} \quad /\text{byvjɔ̃}/ \\
\text{MPH} \quad \left\langle \begin{bmatrix} \text{PH} & /\text{byv}/ \\ \text{PC} & 0 \end{bmatrix}, \begin{bmatrix} \text{PH} & /j/ \\ \text{PC} & 2 \end{bmatrix}, \begin{bmatrix} \text{PH} & /ɔ̃/ \\ \text{PC} & 3 \end{bmatrix} \right\rangle \\
\text{RR} \quad \left\langle \begin{bmatrix} \text{MPH} \left\langle \begin{bmatrix} \text{PH} & \boxed{1}/\text{byv}/ \\ \text{PC} & 0 \end{bmatrix} \right\rangle \\ \text{MUD} \left\{ \begin{bmatrix} \text{LID} \begin{bmatrix} drink\text{-}lid \\ \text{STEMS} & \langle \boxed{1}, \ldots \rangle \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \begin{bmatrix} \text{MPH} \left\langle \begin{bmatrix} \text{PH} & /j/ \\ \text{PC} & 2 \end{bmatrix} \right\rangle \\ \text{MUD} \left\{ \begin{bmatrix} \text{TNS} & pst \\ \text{ASP} & ipfv \end{bmatrix} \right\} \end{bmatrix} \begin{bmatrix} \text{MPH} \left\langle \begin{bmatrix} \text{PH} & /ɔ̃/ \\ \text{PC} & 3 \end{bmatrix} \right\rangle \\ \text{MUD} \left\{ \begin{bmatrix} subj \\ \text{PER} & 1 \\ \text{NUM} & pl \end{bmatrix} \right\} \end{bmatrix} \right\rangle \\
\text{MS} \quad \left\{ \begin{bmatrix} \text{LID} \begin{bmatrix} drink\text{-}lid \\ \text{STEMS} & \langle /\text{byv}/, /\text{bwav}/, /\text{bwa}/ \rangle \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{TNS} & pst \\ \text{ASP} & ipfv \end{bmatrix} \begin{bmatrix} subj \\ \text{PER} & 1 \\ \text{NUM} & pl \end{bmatrix} \right\} \\
\text{SYNSEM} \quad \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{LID} & drink\text{-}lid \\ \text{TNS} & pst \\ \text{ASP} & ipfv \end{bmatrix} \\ \text{ARG-ST} \quad \langle \text{NP}_{1PL}, \text{NP} \rangle \end{bmatrix}
\end{bmatrix}
$$

Figure 9: A sample IbM analysis: the French ɪᴘꜰv.2ᴘʟ word *buvions* 'we drank'

ship between the sʏɴsᴇᴍ and ᴍs values. This is regulated by a set of language-specific constraints, since which aspects of syntax and semantics are realised by inflection is a highly parochial matter. Two features of this interface are worth noting. First, lexeme-specific information on inflection class and stem alternants is included in ᴍs inside the *lid* value. In particular, a list-valued feature sᴛᴇᴍs provides an indexed set of stem alternants, also known as a stem space (Bonami & Boyé 2006).[12] The choice of a particular stem is then effected by a realisation rule of *stem selection* (Stump 2001), picking out the appropriate value in this list, depending on the morphosyntactic context; In the present instance, the default of picking the first stem applies. In other words, in IbM, even the stem is taken to be the realisation of some word-level information, namely lexical identity. Second, ᴍs values are relatively flat in comparison to sʏɴsᴇᴍ values, consisting of a set of small feature structures, rather than one large, deeply recursive feature structure. This is necessitated by the different demands of morphological and syntactic combination.[13]

---

[12]Bonami & Boyé (2006) argue that the French stem space has 12 coordinates. for simplicity we show only 3 in the example in Figure 9.

[13]The distinction between sʏɴsᴇᴍ and ᴍᴏʀsʏɴ may also be used to account for mismatches between content and form at the morphology-syntax interface, as variously captured in the literature by distinguishing syntactic and morphological features (Sadler & Spencer 2001, Corbett & Baerman 2006, Bonami 2015) or content and form paradigms (Stump 2006, 2016).

We may now turn to the relationship between MS and RR. This is regulated by a principle of morphological wellformedness: the MS value of a word must be identical to the disjoint union of the MUDs of the realisation rules. In other words, each morphosyntactic property must be realised by exactly one rule, although a single rule may realise multiple properties at once.[14]

Finally, the relation between RR, MPH and PHON is rather straightforward. First, the MPH value of a word is the union of the MPH values of its realisation rules: in other words, every morph must be licensed by at least one realisation rule, although a realisation rule may license more than one morph (extended or overlapping exponence), or even no morph at all (zero exponence). Second, a word's phonology is determined by appending the phonology of its morphs in accordance with the linear sequence of position class indices. Note that, although the system of position class indices encodes the notion of a morphotactic template, it does so with appropriate flexibility. There is no notion of an 'empty position' in the template: position class indices regulate the relative order of morphs, but morph ordering is not effected by putting bits of phonology in slots, just by appending bits of phonology in order. More importantly, realisation rules may partially underspecify the position they assign morphs to, allowing one to capture an unprecedented set of situations of variable morphotactics. Note also that, although a realisation rule may encode zero exponence, it is not equivalent to a zero morpheme: having no morph as one's exponent is not the same thing as having a morph with no phonological realisation. In particular, since no empty morphs are postulated, no sybilline decisions need to be taken as to the positioning of inaudible elements.

2.2 The role of the lexeme in IbM

Now that we have outlined the main features of IbM, let us consider the role of the lexeme in such a framework. Remember that in classical HPSG, inflection rules take the form of unary rules relating an abstract sign, the lexeme, to a surface sign, the inflected word. IbM has no use for such a notion of inflection rule, since inflection is stated directly as a relation between content and form at the word level. On the other hand, IbM makes crucial use of the notion of a lexeme identifier to state lexeme-specific phonological and morphological information; and the word/lexeme opposition is still a useful way of capturing the relationship between lexical entries and inflected words, and making a clear distinction between lexeme formation and inflection.

We thus assume that, while there are no inflectional lexical rules, there is a general constraint on objects of type *word* to the effect that they are the realisation of a lexeme, as indicated in (2). This constraint enforces the monotonic character of inflection: unlike derivation, inflection does not modify syntax or semantics but merely realises whatever features are made available by paradigm structure and compatible with the syntactic context. This is enforced by the identity of SYNSEM values at the *lexeme* and *word* levels.

---

[14]Implicit here are two assumptions familiar from Paradigm Function Morphology: (i) if two realisation rules are appropriate in some context, only the rule realising more content may apply (Panini's Principle); and (ii) there exists a universal rule of default non-realisation, ensuring that a property set remains unrealised if and only if the inflection system provides no other rule for its realisation.

$$(2) \quad word \rightarrow \begin{bmatrix} \text{SYNSEM} & \boxed{1} \\ \text{M-DTRS} & \left\langle \begin{bmatrix} lexeme \\ \text{SYNSEM} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

As a consequence of (2), an inflected word will inherit any constraint imposed by the lexeme's lexical entry within SYNSEM, including, crucially, lexical identity and stem alternants as specified through the LID feature. Note that we assume the PHON attribute to be appropriate only for *syn-sign* objects (that is, words and phrases): lexemes constrain the phonology of their inflected forms through the STEMS feature instead (Bonami & Boyé 2006). The inflection-specific features MPH, RR and SC are appropriate for *word*s only. The format of lexical entries and lexeme formation rules is essentially unchanged.

3 Lexemes and flexemes

In this section we build on the general architecture just presented and argue that a distinction between two notions of lexical identity needs to be made.

3.1 Introducing the flexeme

Up to now, we have assumed a simple relationship between lexemes and inflectional paradigms: the value of the same feature LID is used for purposes of lexeme selection and for purposes of individuating inflectional paradigms. In doing so we have been following standard practice in realisational morphology, where paradigm functions take 'lexemes' (Stump 2001, 2016) or equivalently a 'lexemic index' (Spencer 2013) as an argument.

In an important but rarely cited paper, Fradin & Kerleroux (2003) note that matters are not so simple, for reasons having to do with lexical ambiguity and the division of labour between inflection and lexeme formation.[15] Rules of inflection are not generally concerned with matters of lexical ambiguity: from the point of view of inflection, the two French verbs DEVOIR$_1$ 'must' and DEVOIR$_2$ 'owe' are indistiguishable, as they have the same (highly irregular) inflectional paradigm. From the point of view of derivation, however, things are different. Derived lexemes normally relate to one sense of their base: for instance, while the French noun FILLE is ambiguous between two readings FILLE$_1$ 'girl' and FILLE$_2$ 'daughter', the diminutive FILLETTE 'small girl' only relates to the first.[16] Fradin & Kerleroux (2003) argue that this warrants a distinction between two kinds of abstract lexical objects: *lexemes* and *flexemes*. Inflection is about flexemes, while derivation is about lexemes. Because of the pervasive nature of lexical ambiguity, a single flexeme often corresponds to multiple lexemes.

---

[15]We purposefully use the general term 'lexical ambiguity' because whether the relevant examples are instances of polysemy or homonymy does not affect the argument.

[16]This very short summary does not do justice to Fradin and Kerleroux's insights, which build on an examination of the compatibility of various lexeme formation rules in French (Fradin & Kerleroux 2003) with various families of meanings. See also Fradin & Kerleroux (2009) for more discussion.

In the remainder we follow Walther (2013) in assuming that inflection is strictly concerned with flexemes, and propose an implementation of the lexeme-flexeme distinction in IbM.

3.2 LID and PID

Within an HPSG view of the world, it is tempting to capture the relationship between lexemes and flexemes in terms of underspecification in an inheritance hierarchy: flexemes would then be abstract groupings of lexemes. Suppose for concreteness a hierarchical organisation of LID values such as that indicated in Figure 10. Rules of inflection can then be stated in terms of the supertype *fille*, while lexemes are properly individuated in terms of the subtypes; and hence FILLETTE can be uniquely related to the lexeme whose LID value is *fille*$_1$.

$$
\begin{array}{c}
\textit{lid} \\
| \\
\begin{bmatrix} \textit{fille-lid} \\ \text{STEMS} \quad \text{/fij/} \end{bmatrix} \\
\overbrace{\textit{fille-1-lid} \qquad \textit{fille-2-lid}}
\end{array}
$$
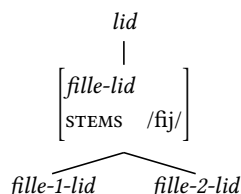
Figure 10: A first pass at flexemes in HPSG: flexemes as underspecified LID values

While this is technically feasible, such an approach only obscures the orthogonal roles played by the two notions. As illustrated above, IbM LID values are structured objects, which encompass all lexically-specified information relevant to inflection, including most notably stem alternants and inflection class. Such information is clearly irrelevant to syntax, although it is an indispensable component of inflection. On the other hand, studies that use LID for purposes of syntactic selection presuppose a tight correspondence between LID values and lexical semantic identity, and have no use for purely morphological information on stem alternants or inflection classes. In particular, Sag 2012 argues that LID values are to be identified with the main semantic predicate associated with a lexeme. One clear advantage of this convention is avoidance of redundancy in lexical entries: it is not necessary to postulate a new symbol as the LID value of each lexeme, since such a symbol is already present in the lexical entry as the constant designating the lexeme's main semantic predicate.

We now propose to clarify the situation by adopting Sag's view of LID. This entails that, for purposes of inflection, a separate index must be posited that individuates words according to which flexeme they instantiate. We call this index PID, standing for 'paradigm identifier'. While LID resides in HEAD and is thus available for selection in idioms, complex predicate constructions, or periphrastic constructions, PID is a top-level feature carried by signs of type *lexeme* only. As such it can be specified by lexical entries or manipulated by lexeme formation rules. In addition, it is universally constrained to be present among the features realised by inflection through inclusion in MS, as indicated in (3). This

is crucial to ensuring that inflection is always concerned with the realisation of lexical identity.

$$
(3) \quad word \rightarrow \begin{bmatrix} \text{MS} & \left\{ \boxed{1}, \ldots \right\} \\[1em] \text{M-DTRS} & \left\langle \begin{bmatrix} lexeme \\ \text{PID} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}
$$

In this architecture then, lexical entries need to specify both an LID and a PID value. To elaborate on the same example, an appropriate analysis of FILLE would posit two lexical entries sharing the same PID object while having different LID values, as indicated in Figure 11.

$$
\begin{bmatrix} lexeme \\[0.5em] \text{SS} \begin{bmatrix} \text{CAT|HD|LID} & \boxed{1}\ girl\text{-}rel \\ \text{SEM|RESTR} & \left\{ \boxed{1} \right\} \end{bmatrix} \\[1em] \text{PID} \quad fille\text{-}pid \end{bmatrix}
\qquad
\begin{bmatrix} lexeme \\[0.5em] \text{SS} \begin{bmatrix} \text{CAT|HD|LID} & \boxed{1}\ daughter\text{-}rel \\ \text{SEM|RESTR} & \left\{ \boxed{1} \right\} \end{bmatrix} \\[1em] \text{PID} \quad fille\text{-}pid \end{bmatrix}
$$

Figure 11: Proposed lexical entries for the two lexemes FILLE.

Under this analysis, the two lexemes FILLE are related by virtue of having indistinguishable PIDs, but they are still distinguishable in terms of LID. Hence, as indicated in the lexical entry in Figure 12, the derived noun FILLETTE adds diminutive semantics (*dim-rel*) to the semantics of its base which is constrained to be that lexeme with LID *girl-rel*, i.e., the left-hand lexeme in Figure 11. This captures the notion of formal lexical identity at the level of PID while implementing Fradin and Kerleroux's insight that derivational morphology operates on fully specific rather than underspecified lexemes.

$$
\begin{bmatrix} lexeme \\[1em] \text{SS} \begin{bmatrix} \text{CAT|HD|LID} & \boxed{1} \begin{bmatrix} dim\text{-}rel \\ \text{INST} & \boxed{x} \end{bmatrix} \\[1em] \text{SEM|RESTR} & \left\{ \boxed{1}, \boxed{2} \right\} \end{bmatrix} \\[1em] \text{PID} \quad fillette\text{-}pid \\[1em] \text{M-DTRS} \left\langle \begin{bmatrix} lexeme \\[1em] \text{SS} \begin{bmatrix} \text{CAT|HD|LID} & \boxed{2} \begin{bmatrix} girl\text{-}rel \\ \text{INST} & \boxed{x} \end{bmatrix} \\[1em] \text{SEM|RESTR} & \left\{ \boxed{2} \right\} \end{bmatrix} \\[1em] \text{PID} \quad fille\text{-}pid \end{bmatrix} \right\rangle \end{bmatrix}
$$

Figure 12: Proposed lexical entry for the lexeme FILLETTE 'small girl'.

3.3 Individuating flexemes: stem spaces

We now turn to the nature of *pid* objects. Evidently, there should be enough distinct PID values to be able to distinguish each flexeme from one another; that is necessary and sufficient to capture the notion of a flexeme. In the context of a typed-feature structure ontology, however, it is very natural to use PID to capture all aspects of inflectional identity. We thus take PIDs to be structured objects providing enough phonological and inflectional information to deduce a whole paradigm with minimal redundancy: Hence, at the very least, for the simplest inflectional systems, a basic stem. For systems of any complexity, this basic information needs to be supplemented with inflection class information (if there is more than one inflectional strategy) and information on stem alternants (if there are unpredictable stem alternations).

We illustrate a simple approach to the encoding of stem alternations by adapting the HPSG analysis of French conjugation presented in Bonami & Boyé (2006). French verbs exhibit pervasive stem alternations, illustrated in Table 1 in the indicative present subparadigms. Regular verbs from the first conjugation use a uniform stem in the present, and regular verbs from the second conjugation use an augmented stem in /-s/ in the plural. In addition to these two patterns, however, there are hundreds of irregular verbs instantiating others, which can be grouped into three types: either there is one stem for the singular and one for the plural, or the same stem is used for the singular and for the third plural, or three different stems are used following the pattern illustrated by BOIRE.[17]

Table 1: Sample French present indicative paradigms illustrating recurrent stem alternation patterns

	1SG	2SG	3SG	1PL	2PL	3PL	
LAVER 'wash'	lav	lav	lav	lav-ɔ̃	lav-e	lav	(1st conjugation)
FINIR 'finish'	fini	fini	fini	finis-ɔ̃	finis-e	finis	(2nd conjugation)
ENVOYER 'send'	ɑ̃vwa	ɑ̃vwa	ɑ̃vwa	ɑ̃vwaj-ɔ̃	ɑ̃vwaj-e	ɑ̃vwa	
JOINDRE 'join'	ʒwɛ̃	ʒwɛ̃	ʒwɛ̃	ʒwaɲ-ɔ̃	ʒwaɲ-e	ʒwaɲ	(other patterns)
BOIRE 'drink'	bwa	bwa	bwa	byv-ɔ̃	byv-e	bwav	

Given the pervasive nature of these alternations and the general unpredictability of the shapes of the alternants, Bonami & Boyé (2003a) build on previous work by Aronoff (1994), Brown (1998), Hippisley (1998), and Stump (2001), and posit that each lexeme is associated with a STEM SPACE, a vector of phonological shapes indicating the shape of the stem used in some zone of the paradigm. Limiting attention again to the stems found in the indicative present, the stem space of the verbs under consideration is indicated in Table 2: Stem 1 the default stem, Stem 2 is used in the 3PL, and Stem 3 is used in the singular.

---

[17] Bonami & Boyé (2006) deliberately set apart a handful of highly irregular and very frequent verbs instantiating an unpredictable form in the 1SG, 1PL or 2PL.

Table 2: Stem spaces for a sample of French verbs in the present indicative

	Stem 1	Stem 2	Stem 3
LAVER 'wash'	lav	lav	lav
FINIR 'finish'	finis	finis	fini
ENVOYER 'send'	ɑ̃vwaj	ɑ̃vwa	ɑ̃vwa
JOINDRE 'join'	ʒwaɲ	ʒwaɲ	ʒwɛ̃
BOIRE 'drink'	byv	bwav	bwa

In the context of an Item-and-Process view of inflection, Bonami & Boyé (2006) propose to encode stem spaces as the value of a feature carried by lexemes, and posit a hierarchy of stem space types capturing different patterns of identity among coordinates in the stem space. This analysis can be readily adapted to the current framework by assuming that stem spaces are represented inside *pid* objects using a list-valued feature STEMS. Let us first consider the lexical entry of BOIRE 'drink'. This needs to list three unpredictable stems, as indicated in Figure 13.

$$
\begin{bmatrix}
\textit{lexeme} \\
\text{SS|CAT|HD} \begin{bmatrix} \textit{verb} \\ \text{LID} \quad \textit{drink-rel} \end{bmatrix} \\
\text{PID} \begin{bmatrix} \textit{boire-pid} \\ \text{STEMS} \quad \langle /byv/,/bwav/,/bwa/\rangle \end{bmatrix}
\end{bmatrix}
$$
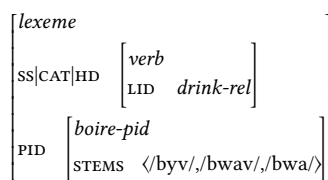
Figure 13: Lexical entry for BOIRE 'drink'

The grammar then needs to specify in which context each element in STEMS is to be used. Following insights from Stump (2001: chap. 6), we assume that this is effected by STEM SELECTION RULES, a special kind of realisation rule that selects a stem alternant for insertion. The relevant rules are presented in Figure 14.[18]

The first rule states that, by default, lexical identity (i.e. PID) is realised by inserting the first element on the STEMS list as a morph in position 0.[19] The two other rules add some allomorphic conditioning: the second element is only used if the morphosyntactic context is that of a 3PL subject, while the third is used when it is that of a SG subject.

Note that the stem selection rules are in no way sensitive to inflection class. This is in keeping with Bonami and Boyé's (2003b, 2006) analysis, which starts from the assumption that all variation in French conjugation originates in differential distributions of

---

[18] We use the em dash ('—') to denote an unconstrained string of segments. '—' in a STEMS value thus indicates that the shape of that stem is not constrained by the rule, type, or lexical entry under consideration.

[19] This rule can be thought of as capturing an inflectional universal, as it simply states that some stem must be provided for every word. In systems without unpredictable stem allomorphy, this will be the sole element on the STEMS list. In systems with stem allomorphy, by convention, we place the default stem alternant first.

$$
\begin{bmatrix} \text{MPH} & \left\langle \begin{bmatrix} \text{PH} & \boxed{1} \\ \text{PC} & 0 \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} pid \\ \text{STEMS} & \langle \boxed{1}, \ldots \rangle \end{bmatrix} \right\} \end{bmatrix}
\quad
\begin{bmatrix} \text{MPH} & \left\langle \begin{bmatrix} \text{PH} & \boxed{1} \\ \text{PC} & 0 \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} pid \\ \text{STEMS} & \langle -, \boxed{1}, - \rangle \end{bmatrix} \right\} \\ \text{MS} & \left\langle \begin{bmatrix} \text{PER} & 3 \\ \text{NUM} & pl \end{bmatrix}, \ldots \right\rangle \end{bmatrix}
\quad
\begin{bmatrix} \text{MPH} & \left\langle \begin{bmatrix} \text{PH} & \boxed{1} \\ \text{PC} & 0 \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} pid \\ \text{STEMS} & \langle -, -, \boxed{1} \rangle \end{bmatrix} \right\} \\ \text{MS} & \left\langle \begin{bmatrix} \text{NUM} & sg \end{bmatrix}, \ldots \right\rangle \end{bmatrix}
$$
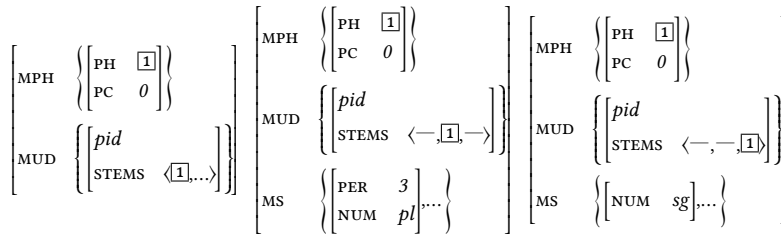
Figure 14: Stem selection rules for French present indicative

alternants in the stem space. That being said, it is useful to characterise classes of flexemes in terms of the patterns of identity they instantiate. In the present context, such a classification can be stated in the form of a type hierarchy of *pid* objects, as indicated in Figure 15.
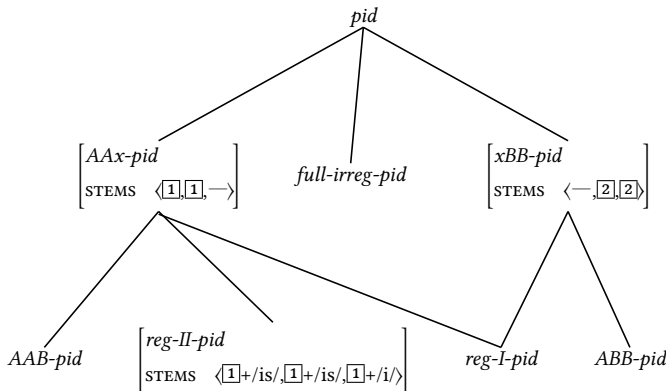
Figure 15: Hierarchy of *pid* subtypes capturing aspects of the French verbal stem space

The hierarchy of *pid* objects highlights the structure of the system, and allows the grammar writer to minimise redundancy in the statement of lexical entries. In particular, all regular verbs can be described with mention of the first stem only, while different types of irregulars necessitate information on two or more stems in different coordinates of the stem space. More sample lexical entries are provided in Figure 16 for illustration. Note that the lexical entry for BOIRE of Figure 13 does not need to mention a subtype of *pid* explicitly, since *full-irreg-pid* is the only subtype compatible with the listing of three distinct stems.

Finally, the distinction between PID types and stem inventories provides a simple account of situations where two verbs belonging to different stem alternation types have the same basic stem, as is the case e.g. with TAPIR 'hide' and TAPISSER 'paper', wich have both have a basic stem /tapis/, witness the ambiguous PRS.1PL /tapisɔ̃/ 'we hide'/'we paper'. Figure 17 shows the relevant lexical entries.

$$
\begin{bmatrix}
lexeme \\
\text{SS|CAT|HD} \begin{bmatrix} verb \\ \text{LID} \quad wash\text{-}rel \end{bmatrix} \\
\text{PID} \begin{bmatrix} reg\text{-}I\text{-}pid \\ \text{STEMS} \quad \langle /\text{lav}/, —, — \rangle \end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
lexeme \\
\text{SS|CAT|HD} \begin{bmatrix} verb \\ \text{LID} \quad finish\text{-}rel \end{bmatrix} \\
\text{PID} \begin{bmatrix} reg\text{-}II\text{-}pid \\ \text{STEMS} \quad \langle /\text{finis}/, —, — \rangle \end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
lexeme \\
\text{SS|CAT|HD} \begin{bmatrix} verb \\ \text{LID} \quad send\text{-}rel \end{bmatrix} \\
\text{PID} \begin{bmatrix} ABB\text{-}pid \\ \text{STEMS} \quad \langle /\tilde{a}\text{vwaj}/, /\tilde{a}\text{vwa}/, — \rangle \end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
lexeme \\
\text{SS|CAT|HD} \begin{bmatrix} verb \\ \text{LID} \quad join\text{-}rel \end{bmatrix} \\
\text{PID} \begin{bmatrix} AAB\text{-}pid \\ \text{STEMS} \quad \langle /\text{ʒwaɲ}/, —, /\text{ʒwɛ̃}/ \rangle \end{bmatrix}
\end{bmatrix}
$$

Figure 16: Lexical entries for a sample of French verbs

$$
\begin{bmatrix}
lexeme \\
\text{SS|CAT|HD} \begin{bmatrix} verb \\ \text{LID} \quad paper\text{-}rel \end{bmatrix} \\
\text{PID} \begin{bmatrix} reg\text{-}I\text{-}pid \\ \text{STEMS} \quad \langle /\text{tapis}/, —, — \rangle \end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
lexeme \\
\text{SS|CAT|HD} \begin{bmatrix} verb \\ \text{LID} \quad hide\text{-}rel \end{bmatrix} \\
\text{PID} \begin{bmatrix} reg\text{-}II\text{-}pid \\ \text{STEMS} \quad \langle /\text{tapis}/, —, — \rangle \end{bmatrix}
\end{bmatrix}
$$
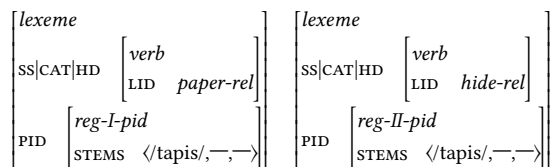
Figure 17: Lexical entries for two French verbs with homophonous basic stems

To sum up then, PID provides a natural locus for the representation of lexical information on stem alternations, and allows for a natural encoding of Bonami and Boyé's notion of a stem space. In addition, in a system where (by hypothesis) all variation in inflection is located in the stems, the indication of a specific vector of stem alternants is sufficient to fully individuate flexemes. In such a system, the hierarchy of *pid* values is merely used to limit the statement of redundant information in lexical entries.

3.4 Individuating flexemes: affixal inflection classes

We now turn to the role of PID in a system with nontrivial affixal inflection classes. As an illustration, let us examine a subset of the Czech nominal declension system. Table 3 provides partial paradigms for four nouns belonging to four of the major inflection classes of masculine inanimate and neuter nouns.

The distinction between hard and soft declension is correlated with the phonological properties of the stem-final consonant; however, it is not in general possible to categorically predict whether a noun will belong to a hard or soft declension on the basis of the phonological shape of its stem. Groups of declensions do share characteristics of exponence; in particular, it is evident from the table that some exponent strategies are common to the soft declensions (e.g. *-e* marking the GEN.SG), to the masculine declensions (e.g. *-ů* in the GEN.PL), or to larger groups of declensions (e.g. *-ům* is used in the DAT.PL across the declensions shown here, except in the soft neuter). These observations

Table 3: Partial declension for the four inflection classes of Czech inanimate nouns

		MASCULINE		NEUTER	
		hard	soft	hard	soft
SG	NOM	most	pokoj	měst-o	moř-e
	GEN	most-u	pokoj-e	měst-a	moř-e
	DAT	most-u	pokoj-i	měst-u	moř-i
	ACC	most	pokoj	měst-o	moř-e
	VOC	most-e	pokoj-i	měst-o	moř-e
	LOC	most-ě	pokoj-i	měst-ě	moř-i
	INS	most-em	pokoj-em	měst-em	moř-em
PL	NOM	most-y	pokoj-e	měst-a	moř-e
	GEN	most-ů	pokoj-ů	měst	moř-í
	DAT	most-ům	pokoj-ům	měst-ům	moř-ím
	ACC	most-y	pokoj-e	měst-a	moř-e
	VOC	most-y	pokoj-e	měst-a	moř-e
	LOC	most-ech	pokoj-ích	měst-ech	moř-ích
	INS	most-y	pokoji	měst-y	moř-i
		'bridge'	'room'	'town'	'sea'

motivate arranging flexemes in a hierarchy of classes, so that the application of rules of exponence can be restricted to arbitrary collections of declension classes. We thus propose a simpler hierarchy of *pid* objects reflecting the distinction between hard and soft declensions, as indicated in Figure 18.
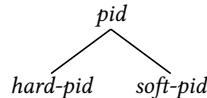
Figure 18: Premiminary hierarchy of *pid* subtypes for Czech declension

In addition, we propose that, since gender is inherent for nouns (in contrast to agreement gender) yet still conditions inflectional realisation, it should be represented as part of PID. Hence the lexical entries of the 4 nouns under consideration are as indicated in Figure 19. Note that traditional declensions correspond to a combination of a *pid* subtype and a gender value.[20]

---

[20]This bidimensional representation of declension classes is possible because gender is a strict predictor of inflection class in Czech: all members of each declension class belong to the same gender. Some declension classes corresponding to different genders are very similar, but always differ in at least one paradigm cell: e.g. masculine TÁTA 'dad' inflects like a feminine hard noun in only about half of its paradigm cells. Also note that a full description of the system would require more subtypes of *pid*, as there are more than two classes per gender, and hence organizing the *pid* hierarchy as a dense semi-lattice of inflection class groupings (Beniamine & Bonami 2016-09).

$$
\begin{bmatrix}
\textit{lexeme} \\[4pt]
\text{SS}|\text{CAT}|\text{HD} \quad
\begin{bmatrix}
\textit{noun} \\
\text{LID} \quad \textit{bridge-rel}
\end{bmatrix} \\[10pt]
\text{PID} \quad
\begin{bmatrix}
\textit{hard-pid} \\
\text{GEN} \quad \textit{mas} \\
\text{STEMS} \quad \langle /\text{most}/\rangle
\end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
\textit{lexeme} \\[4pt]
\text{SS}|\text{CAT}|\text{HD} \quad
\begin{bmatrix}
\textit{noun} \\
\text{LID} \quad \textit{room-rel}
\end{bmatrix} \\[10pt]
\text{PID} \quad
\begin{bmatrix}
\textit{soft-pid} \\
\text{GEN} \quad \textit{mas} \\
\text{STEMS} \quad \langle /\text{pokoj}/\rangle
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\textit{lexeme} \\[4pt]
\text{SS}|\text{CAT}|\text{HD} \quad
\begin{bmatrix}
\textit{noun} \\
\text{LID} \quad \textit{town-rel}
\end{bmatrix} \\[10pt]
\text{PID} \quad
\begin{bmatrix}
\textit{hard-pid} \\
\text{GEN} \quad \textit{neu} \\
\text{STEMS} \quad \langle /\text{měst}/\rangle
\end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
\textit{lexeme} \\[4pt]
\text{SS}|\text{CAT}|\text{HD} \quad
\begin{bmatrix}
\textit{noun} \\
\text{LID} \quad \textit{sea-rel}
\end{bmatrix} \\[10pt]
\text{PID} \quad
\begin{bmatrix}
\textit{soft-pid} \\
\text{GEN} \quad \textit{neu} \\
\text{STEMS} \quad \langle /\text{moř}/\rangle
\end{bmatrix}
\end{bmatrix}
$$

Figure 19: Preliminary lexical entries for a sample of Czech nouns

$$
\textit{rln-rule}
$$

$$
\begin{bmatrix}
\textit{gs-rule} \\[4pt]
\text{MPH} \quad \left\{ \begin{bmatrix} \text{PC} \quad 2 \end{bmatrix} \right\} \\[10pt]
\text{MUD} \quad \left\{ \begin{bmatrix} \text{CASE} \quad \textit{gen} \\ \text{NUM} \quad \textit{sg} \end{bmatrix} \right\}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{MPH} \quad \left\{ \begin{bmatrix} \text{PH} \quad \langle /\text{u}/\rangle \end{bmatrix} \right\} \\[10pt]
\text{MS} \quad \left\{ \begin{bmatrix} \textit{hard-pid} \\ \text{GEN} \quad \textit{mas} \end{bmatrix}, \ldots \right\}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{MPH} \quad \left\{ \begin{bmatrix} \text{PH} \quad \langle /\text{a}/\rangle \end{bmatrix} \right\} \\[10pt]
\text{MS} \quad \left\{ \begin{bmatrix} \textit{hard-pid} \\ \text{GEN} \quad \textit{neu} \end{bmatrix}, \ldots \right\}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{MPH} \quad \left\{ \begin{bmatrix} \text{PH} \quad \langle /\text{e}/\rangle \end{bmatrix} \right\} \\[10pt]
\text{MS} \quad \left\{ \textit{soft-pid}, \ldots \right\}
\end{bmatrix}
$$
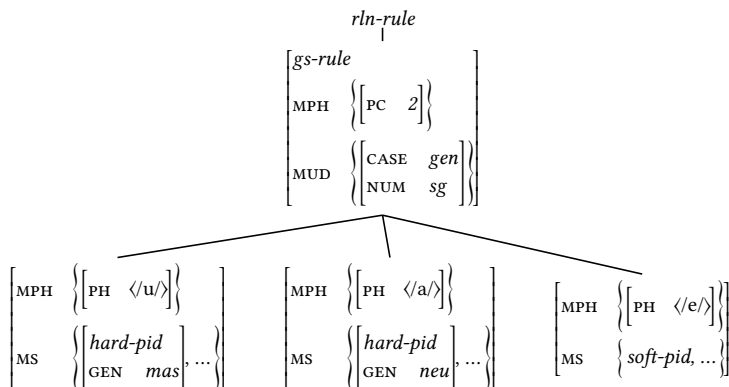
Figure 20: Preliminary realisation rules for Czech GEN.SG

To see how this hierarchy helps in capturing the distribution of exponents in Czech, consider the partial hierarchy of rules of exponence for the expression of GEN.SG in Figure 20. The three rules have the same general structure: they associate a specific phonological shape with the expression (through the MUD value) of the GEN.SG, but place a condition on that expression by restricting the MS value to contain specific information in its *pid* value. That is, they limit the use of an exponent to flexemes belonging to a particular inflection class or group of inflection classes. The first two rules express the conditioning in terms of both a type in the *pid* hierarchy and a gender value. The third one, however, does not mention gender, and hence can apply both in the case of masculine and neuter soft nouns.

This simple example illustrates how the typed feature structure architecture allows for a straightforward statement of generalisations on exponence across declension types by

locating inherent inflectional information in PID values and conditioning the application of rules of exponence to families of possible PID values.

We conclude this section by noting that the use of stem spaces, inherent features such as gender, and type of PID does not necessarily exhaust the inventory of relevant information that should be coded inside PID for the languages of the world. For instance, Bonami & Lacroix (2011) proposed that lexical information on thematic suffixes in the conjugation of the Kartvelian language Laz should be stored as the value of a dedicated feature inside the PID, since information on the shape of the thematic affix needs to be lexically stipulated but the affix is neither always present nor always contiguous to the root; and Crysmann & Bonami (2017) propose a concrete implementation of that idea in the context of Estonian declension. Our general claim is that PID should be the sole locus of lexically stipulated information on inflection.

4 Flexemes and overabundance

In previous sections we have justified the distinction between lexemes and flexemes by arguing that a single flexeme (characterised by a single inflectional paradigm) may correspond to multiple lexemes (characterised by different lexical semantic and/or syntactic properties). In this final section we explore situations where one may want to argue the opposite: multiple flexemes corresponding to a single lexeme.

Although we have not made use of it yet, the analytic scheme defined in the previous section certainly leaves room for such a possibility. Both for French verbs and Czech nouns, we have proposed that *pid* objects be organised in a hierarchy, capturing families of inflectional behavior. The lexical entries used thus far all introduce a PID value corresponding to a specific leaf type in the hierarchy: hence one flexeme for each lexeme. However, if some lexical entries were to refer to some *pid* supertype, this would authorise multiple inflectional behaviours for the same lexeme – hence, in a sense, multiple flexemes for one lexeme.

As a matter of fact, both French conjugation and Czech declension provide examples of phenomena that are insightfully analysed in this fashion. The phenomena at hand fall under the general heading of OVERABUNDANCE (Thornton 2011, 2012, to appear), that is, of situations where a single lexeme has multiple realisations for the same set of morphosyntactic properties.

First consider the French verb ASSEOIR. There is considerable variation in the realisation of different paradigm cells of this verb, leading to free variation at least for some paradigm cells in some varieties (Bonami & Boyé 2010). Limiting ourselves again to the indicative present, there seem to be two equally felicitous forms for each person-number combination in Standard French, as indicated in Table 4.

Although this situation could be described in terms of overabundance in individual paradigm cells, such an approach would not capture the fact that the forms seem to be organised in two distinct paradigms, each with two stem alternants, and each instantiating a different but familiar pattern of stem allomorphy: the /aswa/ /aswaj/ contrast follows an ABB pattern similar to that of ENVOYER (see Table 1), while the /asje/ /asɛj/

Table 4: The two main indicative present subparadigm of ASSEOIR 'sit'

1SG	2SG	3SG	1PL	2PL	3PL
aswa	aswa	aswa	aswaj-ɔ̃	aswaj-e	aswa
asje	asje	asje	asɛj-ɔ̃	asɛj-e	asɛj

contrast follows an AAB pattern similar to that of JOINDRE. It is thus more perspicuous to describe this case of overabundance as involving two different stem spaces, and hence two different PID values, rather than variation in individual paradigm cells. Figure 21 shows two appropriate lexical entries corresponding to the two paradigms of ASSEOIR that readily integrate with the analysis presented in Section 3 and account for overabundance directly.

$$
\begin{bmatrix}
\textit{lexeme} \\
\text{SS|CAT|HD} \begin{bmatrix} \textit{verb} \\ \text{LID} \quad \textit{sit-rel} \end{bmatrix} \\
\text{PID} \begin{bmatrix} \textit{AAB-pid} \\ \text{STEMS} \quad \langle /\text{asej}/, —, /\text{asje}/ \rangle \end{bmatrix}
\end{bmatrix}
\quad
\begin{bmatrix}
\textit{lexeme} \\
\text{SS|CAT|HD} \begin{bmatrix} \textit{verb} \\ \text{LID} \quad \textit{sit-rel} \end{bmatrix} \\
\text{PID} \begin{bmatrix} \textit{ABB-pid} \\ \text{STEMS} \quad \langle /\text{aswaj}/, /\text{aswa}/, — \rangle \end{bmatrix}
\end{bmatrix}
$$

Figure 21: Lexical entries for two variants of the verb ASSEOIR 'sit'

The French verb ASSEOIR exemplifies a case of stem-based overabundance, which is readily accommodated by having two stem spaces for a single lexeme. Let us now turn to Czech and discuss a situation of exponent-based overabundance.

In Section 3.4 we discussed the fact that the Czech inflection system distinguishes 'hard' and 'soft' declensions. As it happens, some lexemes follow a hybrid or 'mixed' pattern that does not clearly fall into one type or the other, but rather makes use of both hard and soft exponents. However, this has different manifestations for neuter and masculine inanimate nouns, as evidenced by the examples in Table 5.

The paradigm of the mixed neuter noun KUŘE 'chicken' exhibits HETEROCLISIS (Stump 2006): KUŘE inflects like a soft noun in the singular, but like a hard noun in the plural. By contrast, the paradigm of the mixed masculine noun PRAMEN 'spring' exhibits a combination of heteroclisis and partial overabundance. In the plural, PRAMEN inflects like a hard noun; in the singular, it may inflect either like a hard noun or like a soft noun. Correctly capturing the difference between these two types of mixed inflectional behaviour is a serious challenge for any theory of inflection.

Both behaviours are readily accomodated in the present framework, using a more refined hierarchy of PID values. The crucial insight is that overabundance amounts to ambiguity, i.e. disjunctive membership of two inflection classes, whereas heteroclisis involves simultaneous membership of two classes: while the former is modelled straightforwardly by means of underspecification, corresponding to the JOIN in the semi-lattice

Table 5: Overabundance and Heteroclisis in Czech declension

| | | MASCULINE | | | NEUTER | | |
		hard	mixed	soft	hard	mixed	soft
SG	NOM	most	pramen	pokoj	měst-o	kuř-e	moř-e
	GEN	most-u	pramen-u~pramen-e	pokoj-e	měst-a	kuř-et-e	moř-e
	DAT	most-u	pramen-u~pramen-i	pokoj-i	měst-u	kuř-et-i	moř-i
	ACC	most	pramen	pokoj	měst-o	kuř-e	moř-e
	VOC	most-e	pramen-e~pramen-i	pokoj-i	měst-o	kuř-e	moř-e
	LOC	most-ě	pramen-u~pramen-i	pokoj-i	měst-ě	kuř-et-i	moř-i
	INS	most-em	pramen-em	pokoj-em	měst-em	kuř-et-em	moř-em
PL	NOM	most-y	pramen-y	pokoj-e	měst-a	kuř-at-a	moř-e
	GEN	most-ů	pramen-ů	pokoj-ů	měst	kuř-at	moř-í
	DAT	most-ům	pramen-ům	pokoj-ům	měst-ům	kuř-at-ům	moř-ím
	ACC	most-y	pramen-y	pokoj-e	měst-a	kuř-at-a	moř-e
	VOC	most-y	pramen-y	pokoj-e	měst-a	kuř-at-a	moř-e
	LOC	most-ech	pramen-ech	pokoj-ích	měst-ech	kuř-at-ech	moř-ích
	INS	most-y	pramen-y	pokoji	měst-y	kuř-at-y	moř-i
		'bridge'	'spring'	'room'	'town'	'chicken'	'sea'
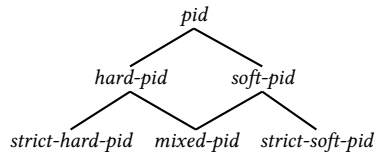
Figure 22: Improved hierarchy of *pid* subtypes capturing heteroclite Czech declension classes

of *pid* types, the latter can be captured by overspecification, i.e. the MEET, as shown by the type hierarchy in Figure 22.

Figure 23 shows schematically to which PID value each noun is assigned, and Figure 24 which PID value rules of exponence for the GEN.SG (left hand side) and NOM.PL (right hand side) are restricted to. More detailed lexical entries and rules of exponence are presented below in Figures 25 and 26. Any noun can be inflected using a realisation rule declared with a compatible PID value. That is, any point in the hierarchy that is identical to that of the noun, dominates it, or is dominated by it.

As shown in Figure 23, nouns belonging to non-mixed declensions are assigned to either of the two simple leaf types *strict-hard-pid* (MOST, MĚSTO) and *strict-soft-pid* (POKOJ, MOŘE). The heteroclite noun KUŘE is assigned to *mixed-pid*, and hence may inflect using either *hard* or *soft* exponents, but not *strict-hard* or *strict-soft* ones. The assignment of exponents to *pid* values (shown in Figure 24) ensures that it must use soft exponents in the singular, yet hard exponents in the plural. By contrast, the overabundant noun PRAMEN is assigned to an underspecified inflection class, namely *hard-pid*. As such it may

195
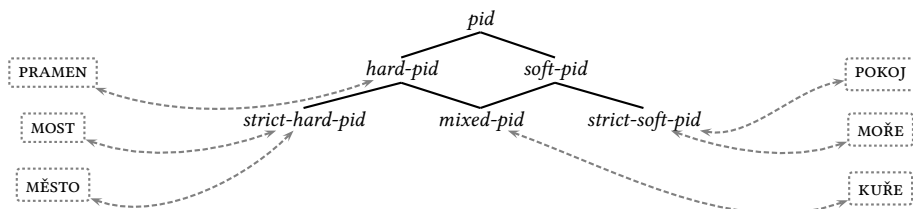
Figure 23: Schematic representation of inflection class assignment for Czech nouns
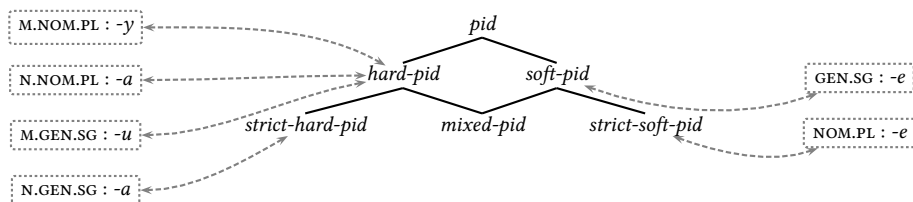
Figure 24: Schematic representation of the scope of rules of exponence for Czech nouns

Figure 25: Lexical entries for six Czech nouns
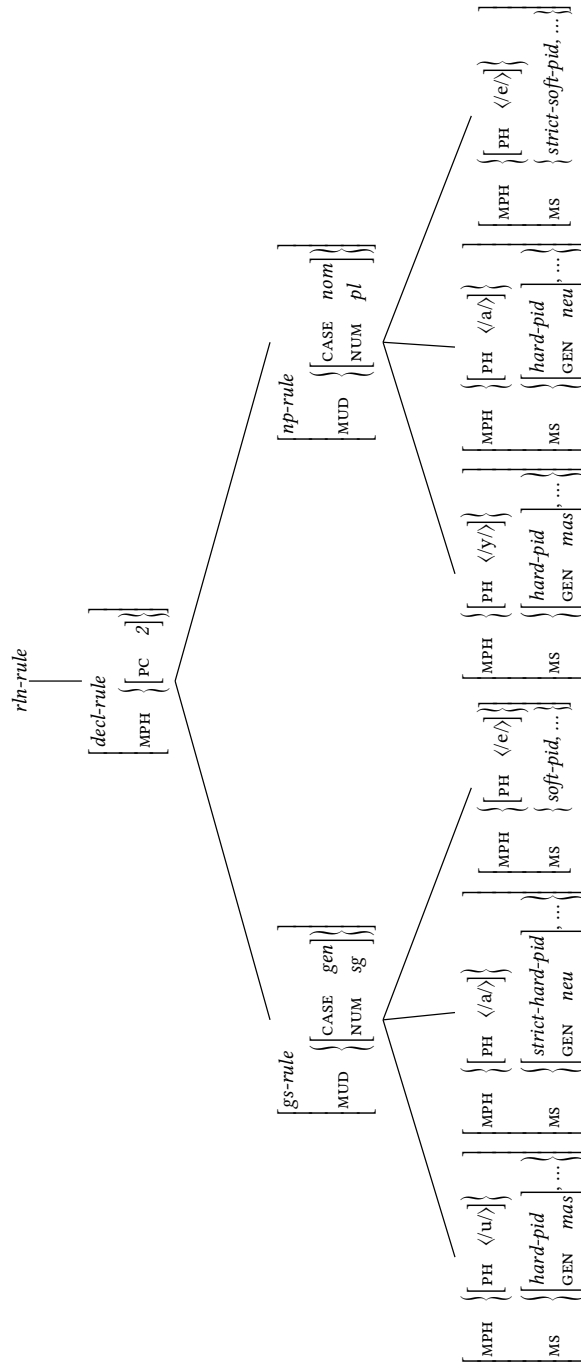
Figure 26: Realisation rules for Czech GEN.SG and NOM.PL nouns

use any one of *hard-pid*, *strict-hard-pid*, *mixed-pid*, or *soft-pid* exponents, but, crucially, not *strict-soft* exponents. This accounts pretty concisely for its contrasting behaviour in the singular and the plural: since the GEN.SG exponent *-e* is only *soft-pid*, there are two GEN.SG exponents available for PRAMEN, which is thus overabundant: inflection with *-e* by resolving *soft-pid* demanded by the rule and *hard-pid* demanded by PRAMEN to the hetoroclite type *mixed-pid*, or else with *-u*, by the sheer fact that this is the exponent available for all *hard-pid* words, whether strict or heteroclite. By contrast, the NOM.PL exponent *-e* is constrained to *strict-soft*. As such, it is inaccessible to PRAMEN, which hence behaves like a simple hard masculine noun in the plural.

We have thus established that mixed overabundant declensions can be accommodated by assigning a lexeme to a supertype in the *pid* hierarchy, while mixed heteroclite declensions can be accommodated by introducing a subtype intermediate between the hard and soft declensions.

The discussion in this section has exhibited the benefits of associating multiple *pid* objects with a single LID value to address some situations of overabundance; which amounts to positing that a single lexeme may correspond to multiple flexemes. We by no means claim that all overabundance phenomena are best thought of in such terms; See Thornton (this volume) for relevant discussion. Rather, we suggest that, where overabundance results from a lexeme being ambiguous between two classes of paradigms, lexically underspecified *pid*s make good sense of the situation.

5 Conclusions

In this paper we have addressed the representation of lexical identity in morphology. Following Fradin & Kerleroux (2003), we have argued that a distinction should be made between lexemes, individuated in terms of lexical semantics, and FLEXEMES, individuated in terms of inflectional paradigms. We have then shown that lexemes and flexemes stand in a many-to-many relation: in cases of lexical ambiguity, one flexeme realises multiple lexemes; in at least some situations of overabundance, multiple flexemes realise the same lexeme. We have shown how this distinction can be integrated into Information-based Morphology by providing words with two independent indices: LID and PID.

The distinction between LID and PID clarifies the role of lexical identity at the interface between inflectional morphology and syntax: syntax cares about lexemes, but not flexemes; inflectional morphology cares about flexemes, but not about lexemes. In the present framework, this is captured by the fact that LID is not represented in MS, the input to rules of inflection. Arguably, the distinction is also useful to clarify the role of lexical identity in lexeme formation. Recent work on French lexeme formation has highlighted the many-to-many nature of lexeme formation rules (see Bonami & Crysmann 2016: §3.1 and references cited therein): typically, a single formal process may be associated with multiple meanings, and the same type of meaning may be realised by multiple processes. Bonami & Tribout (2012) and Tribout & Bonami (2014-07) explore how the LID/*pid* can be used to make sense of that distinction. In their analytic scheme, lexeme formation rules are organised in a bidimensional multiple inheritance hierarchy, with

one dimension laying out formal strategies, and the other dimension describing a syntactic/semantic operation. Formal strategies determine a new *pid* from that of the base, while syntactic/semantic operations amount to constructing a new LID from that of the base.

More work is needed to integrate Bonami and Tribout's insights into IbM, but this integration paves the way towards a general, underspecification-based framework for morphological analysis.

Acknowledgments

References

Ackerman, Farrell & Gert Webelhuth. 1998. *A theory of predicates.* Stanford: CSLI Publications.

Anderson, Stephen R. 1992. *A-morphous morphology.* Cambridge: Cambridge University Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes.* Cambridge: MIT Press.

Bargmann, Sascha. forthcoming. *A semantic representation account of English VP-idioms.* Goethe-Universität Frankfurt am Main dissertation.

Beniamine, Sacha & Olivier Bonami. 2016-09. A comprehensive view on inflectional classification. In *Annual meeting of the linguistic association of Great Britain.* Paris.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.

Blevins, James P. 2016. *Word and paradigm morphology.* Oxford: Oxford University Press.

Boas, Hans & Ivan A. Sag (eds.). 2012. *Sign-based construction grammar.* Stanford: CSLI Publications.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25. 63–110.

Bonami, Olivier, Robert D. Borsley & Maggie Tallerman. 2016. On pseudo-non-finite clauses in Welsh. In Doug Arnold, Miriam Butt, Berthold Crysamnn, Tracy Holloway King & Stefan Muller (eds.), *Proceedings of the joint 2016 conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 104–124.

Bonami, Olivier & Gilles Boyé. 2003a. Supplétion et classes flexionnelles dans la conjugaison du français. *Langages* 152. 102–126.

Bonami, Olivier & Gilles Boyé. 2003b. Supplétion et classes flexionnelles dans la conjugaison du français. *Langages* 152. 102–126.

Bonami, Olivier & Gilles Boyé. 2006. Deriving inflectional irregularity. In *Proceedings of the 13th international conference on HPSG*, 39–59. Stanford: CSLI Publications.

Bonami, Olivier & Gilles Boyé. 2010. La morphologie flexionnelle est-elle une fonction ? In Injoo Choi-Jonin, Marc Duval & Olivier Soutet (eds.), *Typologie et comparatisme, hommage offert à Alain Lemaréchal*, 21–35. Leuven: Peeters.

Bonami, Olivier & Berthold Crysmann. 2013. Morphotactics in an information-based model of realisational morphology. In Stefan Müller (ed.), *Proceedings of HPSG 2013*, 27–47. Stanford: CSLI Publications.

Bonami, Olivier & Berthold Crysmann. 2016. The role of morphology in constraint-based lexicalist grammars. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge handbook of morphology*, 609–656. Cambridge: Cambridge University Press.

Bonami, Olivier & René Lacroix. 2011. Unblurring the inflection/derivation divide in Laz. Paper presented at the 8th Mediterranean Morphology Meeting, Cagliari.

Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2). 327–382.

Bonami, Olivier & Delphine Tribout. 2012. Underspecification and the semantics of lexeme formation. Vienna. http://www.llf.cnrs.fr/sites/llf.cnrs.fr/files/biblio//diapos-vienne-v5.pdf, accessed 2016-11-06. Paper presented at the 15th International Morphology Meeting.

Bonami, Olivier & Gert Webelhuth. 2012. The phrase-structural diversity of periphrasis: A lexicalist account. In Marina Chumakina & Greville G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms* (Proceedings of the British Academy 180), 141–167. Oxford: Oxford University Press/British Academy.

Brown, Dunstan. 1998. Stem lndexing and morphonological selection in the Russian verb: A network morphology account. In R. Fabri, A. Ortmann & T. Parodi (eds.), *Models of inflection*, 196–224. Tübingen: Niemeyer.

Brown, Dunstan & Roger Evans. 2012. Morphological complexity and unsupervised learning: Validating Russian inflectional classes using high frequency data. In F. Kiefer, M. Ladányi & P. Siptár (eds.), *Current issues in morphological theory: (Ir)regularity, analogy and frequency*, 135–162. Amsterdam: John Benjamins.

Corbett, Greville G. & Matthew Baerman. 2006. Prolegomena to a typology of morphological features. *Morphology* 16. 231–246.

Crysmann, Berthold. 2002. *Constraint-based coanalysis*. Universität des Saarlandes dissertation.

Crysmann, Berthold. 2017. Inferential-realisational morphology without rule blocks: An information-based approach. In Andrew Hippisley & Nik Gisborne (eds.), *Defaults in morphological theory*, 182–213. Oxford: Oxford University Press.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374.

Crysmann, Berthold & Olivier Bonami. 2017. Atomistic and holistic exponence in Underspecified Realisational Morphology. In *Proceedings of the 24th international conference on HPSG*.

Emerson, Guy & Ann Copestake. 2015. Lacking integrity: HPSG as a morphosyntactic theory. In Stefan Müller (ed.), *Proceedings of the 22nd international conference on Head-driven Phrase Structure Grammar*, 75–95. Stanford: CSLI Publications.

Erjavec, Tomaž. 1994. Formalizing realizational morphology in typed feature structures. In Gosse Bouma & Gertjan van Noord (eds.), *Proceedings of the fourth CLIN meeting*, 47–58. Groningen.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Fradin, Bernard & Françoise Kerleroux. 2009. L'identité lexémique. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 83–102. Saint-Denis: Presses Universitaires de Vincennes.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations. The form, meaning, and use of English interrogatives*. Stanford: CSLI Publications.

Hippisley, Andrew. 1998. Indexed stems and Russian word formation: A Network Morphology account of Russian personal nouns. *Linguistics* 36. 1039–1124.

Hockett, Charles F. 1967. The Yawelmani basic verb. *Language* 43. 208–222.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

Matthews, P. H. 1972. *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Matthews, P. H. 1974. *Morphology*. Cambridge: Cambridge University Press.

Miller, Philip & Ivan A. Sag. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15. 573–639.

Müller, Stefan. 2002. *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German*. Stanford: CSLI Publications.

Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46. 601–655.

Orgun, Cemil Orham. 1996. *Sign-based morphology and phonology*. University of California, Berkeley dissertation.

Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics*. Stanford: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Richter, Frank & Manfred Sailer. 2010. Phraseological clauses in constructional HPSG. In *Proceedings of HPSG 2009*, 297–317. Stanford: CSLI Publications.

Riehemann, Susanne Z. 1998. Type-based derivational morphology. *Journal of Comparative Germanic Linguistics* 2. 49–77.

Robins, R. H. 1959. In defense of WP. *Transactions of the Philological Society*. 116–144.

Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2000*, 71–96. Dordrecht: Kluwer Academic Publishers.

Sag, Ivan A. 2007. Remarks on locality. In Stefan Müller (ed.), *Proceedings of HPSG 2007*, 394–414. Stanford: CSLI Publications.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 69–202. Stanford: CSLI Publications.

Sag, Ivan A., Thomas Wasow & Emily Bender. 2003. *Syntactic theory: A formal introduction*. 2nd edition. Stanford: CSLI Publications.

Soehn, Jan-Philipp. 2006. *Über Bärendienste und erstaunte Bauklötze – Idiome ohne freie Lesart in der HPSG*. Frankfurt: Peter Lang.

Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2006. Heteroclisis and paradigm linkage. *Language* 82. 279–322.

Stump, Gregory T. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.

Thornton, Anna M. 2011. Overabundance (multiple forms realizing the same cell): A non-canonical phenomenon in Italian verb morphology. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 358–381. Oxford: Oxford University Press.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5(2). 183–207.

Thornton, Anna M. to appear. Overabundance: A canonical typology. In Francesco Gardani, Hans Christian Luschützky & Franz Rainer (eds.), *Competition in morphology*. Berlin: Springer. U. L'Acquila.

Tribout, Delphine & Olivier Bonami. 2014-07. The polysemy of lexeme formation rules. In *Semantics of derivational morphology: Empirical evidence and theoretical modeling*. Düsseldorf.

Walther, Géraldine. 2013. *De la canonicité en morphologie: perspective empirique, théorique et computationnelle*. Université Paris Diderot dissertation.

Zwicky, Arnold M. 1985. How to describe inflection. In *Proceedings of the 11th meeting of the Berkeley Linguistics Society*, 372–386.

Chapter 9

The morphology of essence predicates in Chatino

Hilaria Cruz
Dartmouth College

Gregory Stump
Professor emeritus, University of Kentucky

In the Chatino language [Oto-Manguean; Mexico], essence predicates are a class of predicative lexemes exhibiting a special complex of properties that distinguishes them from other kinds of predicates. We characterize this complex of properties with evidence from the San Juan Quiahije (SJQ) variety of Chatino. After examining the principal morphosyntactic characteristics of essence predicates, we focus particular attention on their patterns of person/number marking, on which basis we distinguish two possible hypotheses about the grammatical status of essence predicates: the possessed-subject hypothesis and the compound predicate hypothesis. We then assess these hypotheses in light of four kinds of evidence: the structural variety of essence predicates, their external syntax, their general lack of semantic compositionality, and their relation to the distributional flexibility of subject-agreement marking in Chatino. On the basis of this evidence, we conclude that neither the possessed-subject hypothesis nor the compound predicate hypothesis is fully adequate; we therefore propose an alternative way of situating essence predicates in the wider context of Chatino morphosyntax.

1 Introduction

Our intention here is to characterize a distinctive class of predicates in Chatino; we call this the class of ESSENCE PREDICATES. As we show, the members of this class share certain distinctive morphosyntactic characteristics; at the same time, they are also heterogeneous with respect to various criteria. Their interest here resides in the superficial ambiguity of their structure: in some ways, this resembles the syntactic combination

of a verb and its subject, while in other ways, it resembles the morphological structure of a compound predicate. In section 1, we examine the fundamental features of essence predicates. Their patterns of person/number marking (section 2) suggest two alternative analyses of their structure, one syntactic, the other morphological. In section 3, we examine four characteristics of essence predicates as a way of gauging the relative adequacy of the two competing analyses. In view of the equivocal outcome of this examination, we conclude (section 4) that essence predicates are, in fact, neither verb-subject combinations nor ordinary compound predicates, but lexemes whose realization is invariably periphrastic and whose content stems from the special function of a handful of grammaticalized nouns.

2 Basic characteristics of essence predicates

One of the defining features of essence predicates is their structure, which comprises a predicative base followed by a nominal component. For example, the essence predicate $ndi^4 \ riq^2$'s/he was thirsty'[1] comprises the predicative base ndi^4 'be thirsty' and the noun riq^2 'essence'; its inflectional paradigm is given in Table 1. Essence predicates exhibit a wide range of predicative bases, but there is only a handful of choices for the nominal component, the most common being riq^2.

Table 1: Paradigm of the essence predicate $ndi^4 \ riq^2$'s/he was thirsty' [thirsty essence] in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	$ndi^4 \ renq^{20}$	$ndi^{32} \ renq^{20}$	$ndyi^4 \ renq^{20}$	$tyi^{32} \ renq^{20}$
2SG	$ndi^4 \ riq^1$	$ndi^{32} \ riq^1$	$ndyi^4 \ riq^1$	$tyi^{32} \ riq^1$
3SG	$ndi^4 \ riq^2$	$ndi^{32} \ riq^2$	$ndyi^4 \ riq^2$	$tyi^{32} \ riq^2$
1INCL	$ndi^4 \ renq^2 \ en^1$	$ndi^{32} \ renq^2 \ en^1$	$ndyi^4 \ renq^2 \ en^1$	$tyi^{32} \ renq^2 \ en^1$
1EXCL	$ndi^4 \ riq^2 \ wa^{42}$	$ndi^{32} \ riq^2 \ wa^{42}$	$ndyi^4 \ riq^2 \ wa^{42}$	$tyi^{32} \ riq^2 \ wa^{42}$
2PL	$ndi^4 \ riq^2 \ wan^1$	$ndi^{32} \ riq^2 \ wan^1$	$ndyi^4 \ riq^2 \ wan^1$	$tyi^{32} \ riq^2 \ wan^1$
3PL	$ndi^4 \ riq^2 \ renq^1$	$ndi^{32} \ riq^2 \ renq^1$	$ndyi^4 \ riq^2 \ renq^1$	$tyi^{32} \ riq^2 \ renq^1$

In view of its structure, the inflectional morphology of essence predicates differs from that of simple verbal lexemes. These differences can be seen by comparing the inflectional

---

[1]Here and throughout, we generally use a verb's third-person singular completive form as its citation form; deviations from this practice are duly noted. We employ the following abbreviations: CPL 'completive aspect', PROG 'progressive aspect', HAB 'habitual aspect', POT 'potential mood'; DEM 'demonstrative'; ABS signifies that a referring expression's referent is absent; ESS = riq^2, tye^{32} or qin^4; EV.MOD = event modifier; and CBM = cranberry morpheme. A superscript 0 represents a floating super high tone, 1 a high tone, 2 a mid high tone, 3 a low mid tone, and 4 a low tone. Contour tones are represented as combinations of these numerals. For details concerning the SJQ Chatino tone system, see Cruz (2011), Woodbury (to appear).

paradigm of the essence predicate *ndi*[4] *riq*[2] 's/he was thirsty' in Table 1 with that of the simple verbal lexeme *yqan*[42] 's/he washed' in Table 2.[2]

Table 2: Paradigm of the simple verbal lexeme *yqan*[42] 's/he washed' in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	*yqan*[24]	*ntyqan*[1]	*ntyqan*[24]	*yqan*[24]
2SG	*yqan*[32]	*ntyqan*[32]	*ntyqan*[42]	*yqan*[42]
3SG	*yqan*[42]	*ntyqan*[32]	*ntyqan*[24]	*yqan*[24]
1INCL	*yqan*[42] *an*[42]	*ntyqan*[1] *an*[1]	*ntyqan*[14]	*yqan*[14]
1EXCL	*yqan*[42] *wa*[42]	*ntyqan*[32] *wa*[42]	*ntyqan*[14] *wa*[42]	*yqan*[14] *wa*[42]
2PL	*yqan*[42] *wan*[4]	*ntyqan*[32] *wan*[4]	*ntyqan*[24] *wan*[32]	*yqan*[24] *wan*[32]
3PL	*yqan*[42] *renq*[4]	*ntyqan*[32] *renq*[4]	*ntyqan*[24] *renq*[32]	*yqan*[24] *renq*[32]

As Table 2 shows, the singular forms of a simple verbal lexeme are single, synthetic word forms inflected both for aspect/mood and for subject person and number. The corresponding plural forms consist of a verb form inflected for aspect/mood and an enclitic pronominal element marking person and number; in general, this pronominal element appears only in the absence of an overt subject constituent, in the presence of which the verb simply appears in its default third-person singular form. As Table 1 shows, essence predicates differ from simple verbal lexemes in satisfying what Rasch (2002) calls the Compound Inflection Criterion, according to which an essence predicate exhibits aspect/mood marking on its predicative base but person and number marking on its nominal component, where, again, the marking of plural persons takes the form of an enclitic that only appears in the absence of an overt subject constituent. The one complication is that in the first-person plural inclusive, subject agreement is marked twice, not only by the clitic *en*[1], but also by ablauting of the nominal component, which appears as *renq*[2] rather than as *riq*[2] in Table 1.

Tables 1 and 2 show that the essence predicate *ndi*[4] *riq*[2] is like a verb in inflecting for aspectual and modal properties; but not all essence predicates are similarly verb-like. We take this as evidence that essence predicates are heterogeneous with respect to their syntactic category membership. In SJQ Chatino, the criteria in (1) are diagnostics of the distinction between verbs and adjectives. By criterion (1a), the predicate *yqan*[42] 's/he washed' in Table 2 is a verb because it exhibits distinct completive, progressive, habitual

---

[2]The 1INCL clitic appearing as *en*[1] in Table 1 and as *an*[42] ~ *an*[1] in Table 2 gets its vowel quality from its host and is manifested as a lengthening of the preceding vowel mora. (Note, however, that verbs with tone 14 do not undergo mora lengthening in the first person inclusive, so that superficially, they appear to lack the 1INCL enclitic, as in Table 2.) Its tone is generally determined by a process of progressive tone sandhi (Chen 2004); but verbs whose basic tone is 4 instead exhibit a regressive process by which their tone becomes 24. It is evidently the historical reflex of a clitic that was once constant in form. This constant form survives as the clitic *na*[4] in Zenzontepec Chatino (Campbell 2011). For details of the idiosyncratic sandhi exhibited by the 1INCL enclitic, see Cruz (2011).

and potential subparadigms. By contrast, the predicate tqi^4 'sick' in Table 3 does not, and is therefore an adjective according to criterion (1a). Similarly, $yqan^{42}$ and tqi^4 may both be used predicatively (as in (2)), but only tqi^4 is used attributively (e.g. (3a)); in order to modify a noun as part of a noun phrase, $yqan^{42}$ must appear as part of a relative clause introduced by the pronominal no^4 'one', as in (3b). Thus, criterion (1b) also leads to the conclusion that $yqan^{42}$ is a verb and tqi^4, an adjective.

(1)   a.   Verbs inflect for aspect and mood, but adjectives do not.

    b.   Adjectives may be used attributively, but verbs cannot (except as part of a relative clause).

(2)   a.   tqi^4 no^4 kiqyu^1.
        sick one(s) male
        'The men are sick.'

    b.   ntyqan^{32} no^4 kiqyu^1.
        wash.PROG one(s) male
        'The men are washing.'

(3)   a.   no^4 kiqyu^1 tqi^4
        one(s) male   sick
        'the sick men'

    b.   no^4 kiqyu^1 *(no^4)   ntyqan^{32}
        one(s) male   one(s) wash.PROG
        'the men that are washing'

Table 3: Paradigm of the adjective tqi^4 'sick' in SJQ Chatino

1SG	$tqen^{20}$
2SG	tqi^{32}
3SG	tqi^4
1INCL	$tqen^{24}$ en^{32}
1EXCL	tqi^4 wa^{42}
2PL	tqi^4 wan^4
3PL	tqi^4 $renq^4$

By these diagnostics, it appears that some essence predicates are verbs and others, adjectives. Unlike the essence predicate ndi^4 riq^2 's/he was thirsty' but like the adjective tqi^4 'sick', the essence predicate tqi^4 riq^2 [sick essence] 's/he was scornful' in Table 4 does not inflect for aspect and mood. Moreover, a comparison of (4) and (5) reveals that while tqi^4 riq^2 readily appears in attributive position, the attributive use of ndi^4 riq^2 requires a relative clause construction. Thus, although ndi^4 riq^2 and tqi^4 riq^2 are both

essence predicates, the diagnostics in (1) suggest that the former is a verb[3] and the latter, an adjective.[4]

Table 4: Paradigm of the essence predicate tqi^4 riq^2 [sick essence] 's/he is scornful' in SJQ Chatino

1SG	tqi^4 $renq^{20}$
2SG	tqi^4 riq^1
3SG	tqi^4 riq^2
1INCL	tqi^4 $renq^2$ en^1
1EXCL	tqi^4 riq^2 wa^{42}
2PL	tqi^4 riq^2 wan^1
3PL	tqi^4 riq^2 $renq^1$

(4) Ntqan^{24} qin^{32} no^4 kiqyu^1 tqi^4 riq^2 qnya^1 kanq^{42}.
see.CPL.1SG ACC one male sick essence.3SG me DEM
'I saw the guy scornful of me.'

(5) Ntqan^{24} qin^{32} no^4 kiqyu^1 *(no^4) ndi^{32} riq^{2}kanq^{42}.
see.CPL.1SG ACC one male *(one) thirsty.PROG essence.3SGDEM
'I saw the guy who is thirsty.'

Most essence predicates denote a particular psychological state or disposition, as the representative examples in Table 5 reveal. Some essence predicates, however, denote a physical state, as in Table 6; and there are also occasional examples that have an active rather than a stative or dispositional meaning, as in Table 7.

In nearly all cases, riq^2 'essence' seems to be interpretable as 'X's self', making the essence predicate similar to a lexically reflexive verb in a language like French; $skeq^1$ riq^0 'il se méprend', $sqwi^4$ riq^2 'elle se souvient', $ndwe^4$ riq^2 'il s'inquiète', tno^4 nga^{24} tye^{32} 'elle se sent courageuse'. Note, however, that argument reflexives are expressed by means of a reflexive pronoun in Chatino, as in (6) and (7). We return to the semantic issues raised by essence predicates in Section 3.3.

(6) Ti^2 kwenq^{42} en^{42} qnyi^4 qnya^4.
EV.MOD:only myself hit.CPL.OBJ.PRON.1SG
'I flagellated myself.'

---

[3]This conclusion further implies that ndi^4 is itself a verb, but its status as a verb is not independently demonstrable, given that it is a kind of cranberry morpheme, appearing as part of the essence predicate ndi^4 riq^2 and nowhere else.

[4]The question naturally arises whether an essence predicate's predicative base is ever a noun. There are occasional instances in which this superficially appears to be the case, but closer scrutiny leaves room for doubt. For example, the essence predicate $tnya^3$ riq^2 's/he is hardworking' seems to have the noun $tnya^3$ 'work' as its predicative base, but $tnya^3$ also seems to have adjectival uses, as in

No^4 nga^{24} $tnya^4$ [one be.PROG working] 'the ones who are authorities'.

Table 5: Some representative essence predicates in SJQ Chatino

Essence predicate	Gloss	Component parts
$nkqan^4\ riq^2$	's/he remembered'	[sit essence]
$sa^4\ riq^2$	's/he was smart, fast, agile'	[light essence]
$ndon^{42}\ riq^2$	's/he was happy'	[standing essence]
$sqwi^4\ riq^2$	's/he remembered'	[exist essence]
$senq^{14}\ riq^0$	's/he was upset'	[CBM essence]
$qna^3\ riq^2$	's/he pitied'	[pity essence]
$xkuq^{42}\ riq^2$	's/he was sad'	[turn.around essence]
$ndwe^4\ riq^2$	's/he worried'	[minced essence]
$skwa^3\ riq^2$	's/he was fed up'	[lying essence]
$tqi^4\ riq^2$	's/he hated'	[sick essence]
$sqwe^3\ riq^2/tye^{32}$	's/he was generous/happy'	[good essence/chest]
$liqa^{14}\ riq^0$	's/he was taciturn'	[slow essence]
$chin^4\ nga^{24}\ tye^{32}$	's/he was scared/queasy'	[ugly feel chest]
$ndya^{32}\ riq^2\ tye^{32}$	's/he liked'	[like essence chest]
$xqan^1\ nga^{04}\ tye^{32}$	's/he felt angry'	[mean feel chest]

Table 6: Some essence predicates denoting physical states in SJQ Chatino

	Essence predicate	Gloss of component parts
's/he is fair-skinned'	$lwi^3\ riq^2 \sim lwi^3\ tye^{32}$	[clean essence ~ chest]
's/he was thirsty, wheezing'	$ndyi^4\ riq^2$	[CBM essence]
's/he is dark-skinned'	$nta^{14}\ riq^0$	[dark essence]
's/he is hungry'	$nteq^{32}\ riq^2$	[hungry essence]
's/he is skinny'	$ti^4\ riq^2$	[skinny essence]
's/he is sturdy'	$tjoq^4\ riq^2$	[strong essence]
's/he is cold'	$tlyaq^4\ riq^2$	[cold essence]
's/he is skinny'	$tyjyan^{20}\ riq^2$	[skinny essence]
's/he is hot'	$tykeq^{14}\ riq^0$	[hot essence]

Table 7: Some essence predicates with an active denotation in SJQ Chatino

	Essence predicate	Glosses of component parts
's/he mocks'	$lyeq^3\ riq^2 \sim lyeq^3\ tye^{32}$	[bully essence ~ chest]
's/he placates'	$tlaq^{14}\ riq^0 \sim tlaq^{14}\ tye^{32}$	[cool essence ~ chest]
's/he takes a liking to'	$skwi^1\ riq^2$	[CBM essence]

(7) ti^2 kwiq^{42} ti^4 Tyu^{14} kwa^0 qnyi^1 qin^{24}.
     EV.MOD:still himself EV.MOD:only Peter DET hit.CPL.OBJ.PRON.3SG

     'Peter flagellated himself.'

3 Person/number marking in essence predicates

An essence predicate exhibits person/number marking on its nominal component. Person/number marking has a complex distributional pattern in Chatino; in this section, we propose to situate essence predicates within this complex pattern by comparing them with simplex verbs, inalienably possessed nouns, and compound verbs. These comparisons lead us to entertain two competing hypotheses about the morphosyntax of essence predicates: the possessed-subject hypothesis (according to which essence predicates embody a verb-subject construction, defined by the syntax of Chatino) and the compound predicate hypothesis (according to which essence predicates belong to a larger class of predicative—mainly verbal—compounds, defined by the morphology of the language).

3.1 Comparison to person/number marking in simplex verbs

A prominent feature of Chatino grammar is the heavy use of tone contrasts in its inflectional system (Cruz 2011, Cruz & Woodbury 2013). Consider, for example, the paradigm of the simple verb *sqi*^{2} 's/he bought' in Table 8. In this paradigm, contrasts in aspect/ mood are marked in three ways: (i) a nasal prefix distinguishes the progressive and the habitual from the completive and the potential, (ii) a stem-initial consonant alternation distinguishes the completive and the progressive (both with stem-initial *s*) from the habitual (stem-initial *ch*) and the potential (stem-initial *x*), and (iii) the completive and the progressive share one pattern of tone alternation, while the habitual and the potential share another. Within a particular aspect/mood subparadigm, contrasts in person and number are marked both tonally and—in the plural forms—by the use of personal clitics (in the absence of an overt subject constituent); in first-person singular and first-person plural inclusive forms, the verb stem also exhibits nasalization, sometimes in combination with ablaut. Verbs fall into a number of different conjugation classes that are distinguished mainly by their paradigms' patterns of tone alternation. Thus, despite some similarities, the pattern of tone alternation in the paradigm of *sqi*^{2} 's/he bought' contrasts with the pattern of *yqan*^{42} 's/he washed' observed earlier in Table 2; these contrasting tone patterns are given in Table 9. For extensive details on conjugation-class distinctions in Chatino, see Cruz & Woodbury (2013), Woodbury (to appear).

Essence predicates participate in this system of tone contrasts, but in a different manner from simplex verbs. In the inflection of a simplex verb, a verb form's tone exhibits a kind of cumulative exponence, serving to distinguish (or to help distinguish) both the form's aspect/mood and its person/number. In the inflection of an essence predicate, by contrast, forms do not exhibit this sort of cumulation, but conform to the Compound Inflection Criterion, with the predicative base carrying the tone relevant to identifying

Hilaria Cruz & Gregory Stump

Table 8: Paradigm of the verbal lexeme sqi^2 's/he bought' in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	$sqen^{40}$	$nsqen^{40}$	$nchqin^{40}$	$xqin^{40}$
2SG	sqi^1	$nsqi^1$	$nchqi^{20}$	xqi^{20}
3SG	sqi^2	$nsqi^2$	$nchqi^{14}$	xqi^{14}
1INCL	$sqen^2\ en^1$	$nsqen^2\ en^1$	$nchqin^{14}$	$xqin^{14}$
1EXCL	$sqi^2\ wa^{42}$	$nsqi^2\ wa^{42}$	$nchqi1^{40}\ wa^{42}$	$xqi1^{40}\ wa^{42}$
2PL	$sqi^2\ wan^1$	$nsqi^2\ wan^1$	$nchqi^{14}\ wan^0$	$xqi^{14}\ wan^0$
3PL	$sqi^2\ renq^1$	$nsqi^2\ renq^1$	$nchq^{14}\ renq^0$	$xqi^{14}\ renq^0$

Table 9: Tone patterns of two verbal lexemes in SJQ Chatino

		COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
sqi^2	1SG	40	40	40	40
's/he bought'	2SG	1	1	20	20
	3SG	2	32	14	14
	1INCL	2-1	2-1	14	14
	1EXCL	2-42	2-42	140-42	140-42
	2PL	2-1	2-1	14-0	14-0
	3PL	2-1	2-1	14-0	14-0
$yqan^{42}$	1SG	24	24	24	
's/he washed'	2SG	32	32	42	42
	3SG	42	32	24	24
	1INCL	42-42	1-1	14	14
	1EXCL	42-42	32-42	14-42	14-42
	2PL	42-4	32-4	24-32	24-32
	3PL	42-4	32-4	24-32	24-32

its aspect or mood and its nominal component carrying the tone that helps distinguish its person and number. (See again the inflection of $ndi^4 riq^2$'s/he was thirsty' in Table 1.)

3.2 Comparison to person/number marking in inalienably possessed nouns

The exponents of person and number employed in verb inflection also appear in the inflection of nouns, where they serve to express the properties of an inalienable possessor. Thus, in the paradigm of the noun $skon^2$ 'arm' in Table 10, the person and number of an inalienable possessor are expressed by tone and—in the plural (in the absence of an overt possessor constituent)—by a clitic. Different nouns exhibit different patterns of tone alternation in their inflection for an inalienable possessor; thus, the tone pattern in the paradigm of $yqan^1$ 'mother' (Table 11) is different from that of $skon^2$ 'arm'. Cruz (2016) distinguishes seven classes of nouns according to their patterns of tone alternation.

Table 10: Inflection of the noun $skon^2$ 'arm' for an inalienable possessor's person and number in SJQ Chatino (E. Cruz)

Possessor	Possessum	
1SG	$skon^{40}$	'my arm'
2SG	$skon^1$	'your (sg) arm'
3SG	$skon^2$	'his/her arm'
1pl incl	$skon^2 on^1$	'our arm'
1pl excl	$skon^2 wa^{42}$	'our arm'
2PL	$skon^2 wan^1$	'your (pl) arm'
3PL	$skon^2 renq^1$	'their arm'

Table 11: Inflection of the noun $yqan^1$ 'mother' for an inalienable possessor's person and number in SJQ Chatino (E. Cruz)

Possessor	Possessum	
1SG	$yqan^{20}$	'my mother'
2SG	$yqan^{42}$	'your (sg) mother'
3SG	$yqan^1$	'his/her mother'
1INCL	$yqan^1 an^1$	'our mother'
1EXCL	$yqan^1 wa^{42}$	'our mother'
2PL	$yqan^1 wan^{24}$	'your (pl) mother'
3PL	$yqan^1 renq^{24}$	'their mother'

In view of this correspondence of form between a verb's subject-agreement marking and a noun's inalienable possessor marking, one might hypothesize that an essence

predicate's nominal component is in fact a subject denoting an individual's inalienably possessed essence, and that its person-number marking therefore marks the person and number of the possessor of this essence. Indeed, riq^2 belongs to an inflection class differing minimally from that of $skon^2$ 'arm', exhibiting the same pattern of tone alternation as in Table 10 except in the first-person singular (where riq^2 exhibits tone 20 instead of tone 40). Accordingly, given the additional fact that Chatino is verb-initial, one might be drawn to conclude that the literal sense of the form ndi^4 $renq^{20}$ (analyzed in Table 1 as 'I was thirsty') is 'my essence is thirsty'—that of a verb-subject combination whose subject is the noun riq^2 'essence' inflected for a first-person singular inalienable possessor and whose predicate is, appropriately, the third-person singular progressive form of ndi^{32} 'be thirsty'. On this POSSESSED-SUBJECT HYPOTHESIS, an overt noun phrase apparently serving as the subject of an essence predicate is instead seen as a possessor, so that (i) no^4 $kyqyu^1$ kwa^3 'that guy' is a possessor in (8) exactly as in (9), and (ii) the head of the subject constituent in (8) is riq^2 '(his) essence', paralleling $tqwa^4$ '(his) mouth' in (9).

(8) La¹ riq²   no¹ kyqyu¹ kwa³.
     open essence one male   that

     'That guy is friendly, talkative.' [= 'That guy's essence is open', according to the possessed-subject hypothesis.]

(9) La¹ tqwa⁴ no⁴ kyqyu¹ kwa³.
     open mouth one male   that

     'The guy's mouth is open.'

This is a tempting analysis, but there is also an alternative possibility—the COMPOUND PREDICATE HYPOTHESIS, according to which essence predicates are a class of compound predicates taking mostly experiencer subjects. In order to evaluate this hypothesis, we now consider person/number marking in compound predicates in SJQ Chatino.

3.3 Comparison to person/number marking in compound verbs

Consider the compound verbs yku^4 $jyaq^3$'s/he tasted' [eat amount] and $ykwiq^4$ sla^3 [speak tiredness] 's/he dreamed', whose paradigms are given in Tables 12 and 13. Each compound consists of a verbal element (yku^4 's/he ate', $ykwiq^4$'s/he spoke') and a nominal element ($jyaq^3$ 'amount', sla^3 'tiredness'). The verbal element is like an essence predicate's predicative base, inflecting for aspect/mood but not ordinarily for person and number (though the verbal element sometimes exhibits agreement in the first person singular, as in Table 12); likewise, the nominal element is like an essence predicate's nominal component, since it carries the person/number inflection. In other words, the inflectional pattern again tends to conform to Rasch's Compound Inflection Criterion.[5]

---

[5]Compound predicates are nevertheless somewhat varied in their properties in SJQ Chatino. Compound verbs whose inflection deviates from the Compound Inflection Criterion may do so in more than one way. In the inflection of some compound verbs, person and number, like aspect and mood, are marked on the first, verbal element rather than on the following nominal element (e.g. $snyi^4$ $chaq^3$'s/he dealt, negotiated'

Table 12: Paradigm of the compound predicate yku^4 $jyaq^3$'s/he tasted' [eat amount] in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	$ykon^1$ $jyanq^3$	$ntykon^1$ $jyanq^3$	$ntykon^{20}$ $jyanq^3$	kon^{20} $jyanq^3$
2SG	yku^4 $jyaq^1$	$ntyku^{32}$ $jyaq^1$	$ntyku^4$ $jyaq^1$	ku^4 $jyaq^1$
3SG	yku^4 $jyaq^3$	$ntyku^{32}$ $jyaq^3$	$ntyku^4$ $jyaq^3$	ku^4 $jyaq^3$
1INCL	yku^4 $jyan^2$ anq^1	$ntyku^{32}$ $jyanq^2$ an^1	$ntyku^4$ $jyanq^2$ an^1	ku^4 $jyanq^2$ an^1
1EXCL	yku^4 $jyaq^3$ wa^{42}	$ntyku^{32}$ $jyaq^3$ wa^{42}	$ntyku^4$ $jyaq^3$ wa^{42}	ku^4 $jyaq^3$ wa^{42}
2PL	yku^4 $jyaq^3$ wan^{24}	$ntyku^{32}$ $jyaq^3$ wan^{24}	$ntyku^4$ $jyaq^3$ wan^{24}	ku^4 $jyaq^3$ wan^{24}
3PL	yku^4 $jyaq^3$ $renq^{24}$	$ntyku^{32}$ $jyaq^3$ $renq^{24}$	$ntyku^4$ $jyaq^3$ $renq^{24}$	ku^4 $jyaq^3$ $renq^{24}$

Table 13: Paradigm of the compound verb $ykwiq^4$ sla^3 [speak tiredness] 's/he dreamed' in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	$ykwiq^4$ $slan^{40}$	$ntykwiq^{32}$ $slan^{40}$	$ntykwiq^4$ $slan^{40}$	$tykwiq^4$ $slan^{40}$
2SG	$ykwiq^4$ sla^1	$ntykwiq^{32}$ sla^1	$ntykwiq^4$ sla^1	$tykwiq^4$ sla^1
3SG	$ykwiq^4$ sla^3	$ntykwiq^{32}$ sla^3	$ntykwiq^4$ sla^3	$tykwiq^4$ sla^3
1INCL	$ykwiq^4$ $slan^2$ an^1	$ntykwiq^{32}$ $slan^2$ an^1	$ntykwiq^4$ $slan^2$ an^1	$tykwiq^4$ $slan^2$ an^1
1EXCL	$ykwiq^4$ sla^3 wa^{42}	$ntykwiq^{32}$ sla^3 wa^{42}	$ntykwiq^4$ sla^3 wa^{42}	$tykwiq^4$ sla^3 wa^{42}
2PL	$ykwiq^4$ sla^3 wan^{14}	$ntykwiq^{32}$ sla^3 wan^{14}	$ntykwiq^4$ sla^3 wan^{14}	$tykwiq^4$ sla^3 wan^{14}
3PL	$ykwiq^4$ sla^3 $renq^{24}$	$ntykwiq^{32}$ sla^3 $renq^{24}$	$ntykwiq^4$ sla^3 $renq^{24}$	$tykwiq^4$ sla^3 $renq^{24}$

As Rasch (2002) and Cruz & Woodbury (2013) observe, compound verbs in Chatino are quite varied in their structure, consisting of a verb paired with a stem of any of a range of categories to form either a head-complement structure (as in (10a)) or a head-modifier structure (as in (10b)), but not, in general, to form a verb-subject structure.[6]

[grab word]); in the inflection of other compound verbs, aspect and mood, like person and number, are marked on the second, nominal element rather than on the preceding verbal element (e.g. xi^{42} $skwa^3$'s/he turned (s.o.) over' [cause be.in.elevated.position]); still others sporadically exhibit person/number marking on both the verbal and the nominal elements (as with $ykon^1$ $jyang^3$ 'I tasted' in Table 12); and yet others exhibit marking of aspect and mood on both the verbal and the nominal elements (e.g. sti^1 qo^{20} 's/he made fun of' [laugh with]). See Cruz & Woodbury (2013) for details concerning these deviations from the Compound Inflection Criterion in SJQ Chatino.

[6]Despite initial resemblances, a compound verb such as $ykwiq^4$ sla^3 's/he dreamed' cannot be seen as the phrasal combination of a verb with an independent postverbal constituent. As a VSO language, Chatino ordinarily positions a verb's subject between the verb and a following complement or modifier, as in (i); but a compound verb is followed by its subject, as in (ii). Moreover, the nominal component of a compound verb carries the verb's person/number inflection, as in (iii), but a verb's object does not, as (iv) shows.

(10) a. nchu^1 yaq^2

's/he clapped' [hit hand]

   b. yku^4 na^2

's/he ate in secret' [eat hidden]

Whether as a verb-complement structure or a verb-modifier structure, the compound verb tends to conform to the Compound Inflection Criterion. This similarity between an essence predicate such as *ndi^4 riq^2* 's/he was thirsty' and a compound verb such as *yku^4 jyaq^3* 's/he tasted' raises the possibility that essence predicates are in fact simply a subclass of compound predicates. If this is so, then an essence predicate's nominal component does not obviously function as an argument of its predicative base. Instead, it seems to serve as a quasi-adverbial modifier: *ndi^4 renq^{20}* 'I was thirsty inside'. On this analysis, the person/number marking on an essence predicate's nominal component is not an expression of possession, but (as in the compound verb *yku^4 jyaq^3* 's/he tasted') an ordinary expression of subject agreement.

In the following section, we assess the relative adequacy of the possessed-subject and compound predicate hypotheses in light of four kinds of evidence.

4 Assessing the possessed-subject and compound predicate hypotheses

We now consider four important characteristics of essence predicates in SJQ Chatino: their structural variety, their external syntax, their general lack of semantic compositionality, and their relation to the distributional flexibility of subject-agreement marking. As we show, this evidence reveals that neither the possessed-subject hypothesis nor the compound predicate hypothesis accounts for the full range of characteristics exhibited by essence predicates.

(i) Ykwiq^4 no^4 qan^1 kwa^3 ska^4 poema^{24}.
speak.CPL one female that one poem

'That woman spoke a poem.'

(ii) Ykwiq^4 sla^3 no^4 qan^1 kwa^3.
speak.CPL tiredness one female that

'That woman dreamt.'

(iii) Ykwiq^4 slan^{40} nka^3.
speak.CPL tiredness.1SG yesterday

'I dreamed yesterday.'

(iv) Ykwenq^1 chaq^3-xlya^{10} nka^3
speak.CPL.1SG word-Castilian yesterday.

'I spoke Spanish yesterday.'

4.1 Structural variety

Essence predicates vary in their structure in at least three ways. First, there is variation with respect to the identity of the nominal component, which we have so far exemplified mainly with *riq²* 'essence'. Second, there is variation with respect to the possibility of employing more than one nominal component within the same essence predicate. And third, essence predicates vary with respect to their predicative base—specifically, with respect to whether the predicative base has independent uses apart from its use in an essence predicate. Consider each of these areas of variation.

4.1.1 Choice of nominal component

The examples of essence predicates cited so far have nearly all had the noun *riq²* 'essence' as their nominal component. This is, indeed, the most usual nominal component for essence predicates. There is, however, a sizeable class of essence predicates whose nominal component is instead *tye³²* 'chest'; one such predicate is *nqne⁴² tye³²* 's/he dared', whose paradigm is given in Table 14. Still another class of essence predicates has the nominal component *qin⁴* (whose low tone makes it frequently susceptible to tone sandhi; an example is the predicate *skeq¹ qin²⁴* 'he (wrongly) thought or believed' [imagine essence] in Table 15.

Table 14: Paradigm of the essence predicate $nqne^{42}$ tye^{32} 's/he dared' [do chest] in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	qne^{42} $tyin^{20}$	$nqne^{32}$ $tyin^{20}$	$nqne^{24}$ $tyin^{20}$	qne^{24} $tyin^{20}$
2SG	qne^{42} tye^{32}	$nqne^{32}$ tye^{32}	$nqne^{24}$ tye^{32}	qne^{24} tye^{32}
3SG	qne^{42} tye^{32}	$nqne^{32}$ tye^{32}	$nqne^{24}$ tye^{32}	qne^{24} tye^{32}
1INCL	qne^{42} $tyin^1 in^1$	$nqne^{32}$ $tyin^1 in^1$	$nqne^{24}$ $tyin^1 in^1$	qne^{24} $tyin^1 in^1$
1EXCL	qne^{42} tye^{32} wa^{42}	$nqne^{32}$ tye^{32} wa^{42}	$nqne^{24}$ tye^{32} wa^{42}	qne^{24} tye^{32} wa^{42}
2PL	qne^{42} tye^{32} wan^4	$nqne^{32}$ tye^{32} wan^4	$nqne^{24}$ tye^{32} wan^4	qne^{24} tye^{32} wan^4
3PL	qne^{42} tye^{32} $renq^4$	$nqne^{32}$ tye^{32} $renq^4$	$nqne^{24}$ tye^{32} $renq^4$	qne^{24} tye^{32} $renq^4$

The identity of *qin⁴* in *skeq¹ qin²⁴* 's/he wrongly thought or believed' is debatable, since *qin⁴* has a variety of functions in Chatino; for example, *qin⁴* functions (with tone sandhi) as a third-person singular pronoun in (11a), but arguably as an animal classifier in (11b).

(11) a. Ye⁴² qa²⁴ yku²⁴ tykwen¹ qin²⁴   sen³².
         very EMPH eat.CPL bedbug OBJ.PRON:3SG last.night

      'Bedbugs bit her last night.'

Table 15: Paradigm of *skeq*[1] *qin*[24] 's/he (wrongly) thought or believed' [imagine essence] in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	*skeq*^{1} *qnya*^{24}	*nskeq*^{1} *qnya*^{24}	*nxkeq*^{1} *qnya*^{24}	*xkeq*^{1} *qnya*^{24}
2SG	*skeq*^{1} *qin*^{42}	*nskeq*^{1} *qnya*^{24}	*nxkeq*^{1} *qnya*^{24}	*xkeq*^{1} *qnya*^{24}
3SG	*skeq*^{1} *qin*^{24}	*nskeq*^{1} *qin*^{24}	*nxkeq*^{1} *qin*^{24}	*xkeq*^{1} *qin*^{24}
1INCL	*skeq*^{1} *qin*^{24}	*nskeq*^{1} *qin*^{24}	*nxkeq*^{1} *qin*^{24}	*xkeq*^{1} *qin*^{24}
1EXCL	*skeq*^{1} *qwa*^{42}	*nskeq*^{1} *qwa*^{42}	*nxkeq*^{1} *qwa*^{42}	*xkeq*^{1} *qwa*^{42}
2PL	*skeq*^{1} *qwan*^{14}	*nskeq*^{1} *qwan*^{14}	*nxkeq*^{1} *qwan*^{14}	*xkeq*^{1} *qwan*^{14}
3PL	*skeq*^{1} *qin*^{24} *renq*^{24}	*nskeq*^{1} *qin*^{24} *renq*^{24}	*nxkeq*^{1} *qin*^{24} *renq*^{24}	*xkeq*^{1} *qin*^{24} *renq*^{24}

b. Yla^{42} qin^4 qo^1 snyiq^{24} qin^{24}.
arrive.CPL ANIMAL.CLF with offspring OBJ.PRON:3SG
'The (animal) returned home with his offspring.'

Although *riq*^{2}, *tye*^{32} and *qin*^{4} are not freely interchangeable as the nominal component of an essence predicate, they do exhibit a partial overlap in their distribution; in cases of overlap, the choice of nominal component may or may not serve to express a difference in meaning. The forms in (12) constitute a minimal triplet in which the predicative base *sqwe*^{3} 'good' combines with *riq*^{2} ('essence'), *tye*^{32} ('chest'), or *qin*^{4} ('his or her essence'), with each combination expressing a different meaning.

(12)   a. sqwe^3 riq^2
            's/he was in a good mood'
        b. sqwe^3 tye^{32}
            's/he was generous'
        c. sqwe^3 qin^{24}
            's/he was affable'

Several cases in which *riq*^{2}, *tye*^{32} and *qin*^{4} may be used more or less interchangeably are listed in Table 16a. The essence predicates in Table 16b involve *riq*^{2} and *tye*^{32} but have no alternative with *qin*^{4}; conversely, those in Table 16c involve *riq*^{2} and *qin*^{4} and have no alternative with *tye*^{32}. Those in Table 16d involve *riq*^{2} but not *tye*^{32} or *qin*^{4}; those in Table 16e involve *tye*^{32} but not *riq*^{2} or *qin*^{4}; and those in Table 16f involve *qin*^{4} but not *riq*^{2} or *tye*^{32}.[7]

Even where the choice of nominal component corresponds to a difference of meaning, it is not clear that the nature of this difference is predictable. For example, the general sense of pity may be expressed by an essence predicate consisting of *qna*^{3} and either

---

[7]It might appear that in Table 16d, *tqi*^{4} *riq*^{2} 's/he hates' has a counterpart with *tye*^{32}, but *tqi*^{4} *tye*^{32} only has the literal meaning 'her/his chest hurts', not that of an essence predicate.

Table 16: Some essence predicates in SJQ Chatino

	Based on riq^2	Based on tye^{32}	Based on qin^4	
a. 's/he understood'	$nkwa^2\ jyaq^3\ riq^2$	$nkwa^2\ jyaq^3\ tye^{24}$	[$nkwa^2\ jyaq^3\ qin^{24}$'s/he was tried']	[realized ESS]
's/he is naïve'	$ntu^1\ riq^0$	$ntu^1\ tye^0$	$ntu^1\ qin^0$	[stupid ESS]
's/he is getting angry'	$ntykwen^3\ riq^{24}$	$ntykwen^3\ tye^{24}$	[$ntykwen^3\ qin^{24}$'s/he choked on sth']	[choke ESS]
's/he misperceives'	$skeq^1\ riq^0$	$skeq^1\ tye^{32}$	$skeq^1\ qin^{24}$	[imagine ESS]
's/he is happy'	$stu^1\ riq^1$	$stu^1\ tye^{32}$	$stu^1\ qin^0$	[gusto ESS]
's/he is fuzzy'	$swaq^{24}\ riq^{32}$	$swaq^{24}\ tye^{32}$	$swaq^{24}\ qin^{32}$	[slovenly ESS]
's/he is cool (not hot)'	$tlaq^{14}\ riq^0$	$tlaq^{14}\ tye^{32}$	[$tlaq^{14}\ qin^0$'s/he is in peace']	[cool ESS]
's/he is hot'	$tykeq^{14}\ riq^0$	[$tykeq^{14}\ tye^0$'s/he is angry']	$tykeq^{14}\ qin^0$	[hot ESS]
b. 's/he pities'	$qna^3\ riq^2$	$qna^3\ tye^{32}$		[pity ESS]
's/he remembers'	$sqwi^4\ riq^2$	[$sqwi^4\ tye^{32}$'s/he holds a grudge']		[exist ESS]
's/he is greedy'	$tkonq^1\ riq^2$	$tkonq^1\ tye^{32}$		[greedy ESS]
's/he is sad'	$xkuq^{42}\ riq^2$	$xkuq^{42}\ tye^{32}$		[CBM ESS]
c. 's/he is fast or in a hurry'	$sa^4\ riq^2, ndla^2\ riq^2$		$ndla^2\ qin^0$	[hurry ESS]
's/he is satisfied/ satiated'	$ylaq^{42}\ riq^2$		$sa^4\ qin^4, ylaq^{42}\ qin^4$	[satiated ESS]
d. 's/he hates'	$tqi^4\ riq^2$			[sick ESS]
's/he knows/is aware'	$jlyo^{20}\ riq^2$			[CBM ESS]
's/he is worry'	$ndwe^{32}\ riq^2$			[minced ESS]
's/he remembers'	$nkya^{42}\ yqwi^{32}\ riq^1$			[remember ESS]
's/he is disgusted'	$stya^4\ riq^2$			[place ESS]
's/he is ecstatic'	$styi^1\ riq^2$			[laugh ESS]
e. 's/he is scared/ queasy'		$chin^4\ nga^{24}\ tye^{32}$		[ugly ESS]
's/he likes'		$ndya^{24}\ riq^2\ tye^{32}$		[like ESS]
's/he feels brave'		$tno^4\ nga^{24}\ tye^{32}$		[big ESS]
's/he feels angry'		$xqan^{10}\ nga^{24}\ tye^{32}$		[mean ESS]
's/he feels sad'		$tqwa^{14}\ nka^{24}\ tye^{32}$		[cool ESS]
f. 's/he is affable'			$sqwe^3\ qin^{24}$	[good ESS]
's/he is a thief or fast'		'	$sa^4\ qin^4$	[light ESS]

In the three central columns, bracketed essence predicates have a meaning different from that of the corresponding essence predicate with riq^2. Note that tone sandhi alters the expected tonality of third-person singular riq^2 in some of these forms.

riq^2 or tye^{32}, and the nuanced difference expressed by this choice in (13) is not obviously predictable from the semantic difference between riq^2 'essence' and tye^{32} 'chest'. Note, by way of contrast, that the meaning of disgust expressed by the essence predicate $stya^4\ riq^2$ has no counterpart with tye^{32}: *$stya^4\ tye^{32}$. Moreover, the meaning 's/he is sad' may be expressed by an essence predicate with either riq^2 or tye^{32} (as either $xkuq^{42}\ riq^2$ or $xkuq^{42}$ tye^{32}), but the meaning 's/he feels sad' is expressed by an essence predicate requiring tye^{32} rather than riq^2 (as $tqwa^{14}\ nka^{24}\ tye^{32}$ but not *$tqwa^{14}\ nka^{24}\ riq^2$).

(13)   a.   Qna^3 qa^{24} riq^2 La^{20}ya^{24} kwa^3 xneq^2 qin^2.
           pity very essence Hilaria   that   dog   POSS.3SG
           'Hilaria feels bad for her dog.'

       b.   Qna^3 qa^{24} tye^{32} La^{20}ya^{24} kwa^3, nkjwi^{42} xneq^2 qin^1.
           pity very essence Hilaria   that   die.CPL dog   POSS.3SG
           'Hilaria is pitiable, her dog died.'

These facts suggest that choices among the nominal components *riq^2*, *tye^{32}* and *qin^4* in essence predicates are often (perhaps always) determined by lexical stipulation.

4.1.2 Combinability of nominal components

It is often possible to use *riq^2* and *tye^{32}* in tandem, as in Table 17.[8] In such cases, it is *tye^{32}* rather than *riq^2* that exhibits the person-number agreement; for instance, the first-person singular completive form of *njlya^{32} riq^2 tye^{32}* 's/he forgot' is *njlya^{32} riq^2 tyin^{20}* 'I forgot'. It is not clear that *qin^4* appears in tandem with either *riq^2* and *tye^{32}* in its function as the nominal component of an essence predicate; in those cases in which it might appear to do so, it instead serves one of its other functions, e.g. that of an animal classifier (as in *tkonq^1 riq^2 tye^{32} qin^{24}* 'that animal is gluttonous').

4.1.3 Cranberry predicative bases

Essence predicates also vary with respect to the independence of their predicative base. On one hand, there are essence predicates whose predicative base also appears independently (though usually not with the same meaning as the essence predicate), as in (14). On the other hand, there are instances whose predicative base does not have an independent use as a predicate, as in (15)–(18); such predicative bases are in effect cranberry morphemes.

(14)   a.   Sqwe^3 riq^2 / tye^{32} no^4 kyqyu^1.
           good essence / chest.3SG the male (ones)
           'The men are in a good mood / generous.'

       b.   Sqwe^3 no^4 kyqyu^1.
           good   the male (ones)
           'The men are good.'

(15)   a.   Ndi^{32} riq^2 Xwa^3 kwa^3.
           thirsty.PROG essence.3SG Juan that
           'Juan is thirsty.'

       b. * Ndi^{32} Xwa^3 kwa^3.
           thirsty.PROG Juan that
           Sought interpretation: 'Juan is thirsty.'

---

[8]In Table 17 and some later tables, '#' marks forms that we have not encountered and that aren't clearly acceptable, but whose acceptability to at least some speakers we do not wish to rule out.

Table 17: Instances of riq^2 used in tandem with tye^{32} in SJQ Chatino

Gloss	$riq^2 + tye^{32}$	Gloss	$riq^2 + tye^{32}$
's/he forgives'	$chaq^3\ tlyu^2\ riq^2\ tye^{32}$	's/he likes sth'	$snyi^4\ riq^2\ tye^{32}$
's/he forgets'	$jlya^{32}\ riq^2\ tye^{32}$	's/he is generous/happy'	$sqwe^3\ riq^2\ tye^{32}$
's/he knows/is aware'	$jlyo^{20}\ riq^2\ tye^{32}$	's/he remembers'	$sqwi^4\ riq^2\ tye^{32}$
's/he mischievous, playful'	$jnya^{20}\ riq^2\ tye^{32}$	's/he is strong/sturdy'	$sqye^{14}\ riq^0\ tye^{32}$
's/he worries'	$ndwe^4\ riq^2\ tye^{32}$	's/he is happy'	$stu^1\ riq^0\ tye^{32}$
's/he is open, extroverted'	$la^1\ riq^2\ tye^{32}$	's/he is disgusted'	$stya^4\ riq^2\ tye^{32}$
's/he is taciturn'	$liqa^{24}\ riq^1\ tye^{32}$	's/he is ecstatic'	$styi^1\ riq^2\ tye^{32}$
's/he is fair-skinned'	$lwi^3\ riq^2\ tye^{32}$	's/he is hard-working'	$t(j)nya^3\ riq^2\ tye^{32}$
's/he mocks'	$lyeq^3\ riq^2\ tye^{32}$	's/he is frugal/takes care of sth'	$tjenq^3\ riq^2\ tye^{32}$
's/he realizes'	$ndi^{20}\ riq^2\ tye^{32}$	's/he is sturdy'	$tjoq^4\ riq^2\ tye^{32}$
's/he is happy'	$ndon^{42}\ riq^2\ tye^{32}$	's/he is greedy'	$tkonq^1\ riq^2\ tye^{32}$
's/he is worry'	$ndwe^{32}\ riq^2\ tye^{32}$	's/he placates, calms'	$tlaq^{14}\ riq^0\ tye^{32}$
's/he likes, loves'	$ndya^{24}\ riq^2\ tye^{32}$	's/he is cold'	$tlyaq^4\ riq^2\ tye^{32}$
's/he is thirsty, wheezing'	$ndyi^{32}\ riq^2\ tye^{32}$	's/he is tired'	$tnyaq^4\ riq^2\ tye^{32}$
's/he remembers'	$nkqan^4\ riq^2\ tye^{32}$	's/he is fully conscious'	$tqa^{24}\ riq^2\ tye^{32}$
's/he realizes'	$nkwa^2\ jyaq^3\ riq^2\ tye^{32}$	's/he scorns'	$tqi^4\ riq^2\ tye^{32}$
's/he remembers'	$nkya^{42}\ yqwi^{32}\ riq^1\ tye^{32}$	's/he is flirtatious'	$tsa^3\ riq^2\ tye^{32}$
's/he is dark-skinned'	$nta^{14}\ riq^0\ tye^{32}$	's/he is smart'	$tya^{20}\ riq^2\ tye^{32}$
's/he is hungry'	$nteq^{32}\ riq^2\ tye^{32}$	's/he is slow'	$tyaq^4\ riq^2\ tye^{32}$
's/he is feeling lazy'	$ntja^1\ riq^2\ tye^{32}$	's/he is hot'	$tykeq^{14}\ riq^0\ tye^{32}$
's/he is weak'	$ntqan^1\ riq^1\ tye^{32}$	's/he made up her/his mind'	$wa^2\ xtya^{20}\ riq^2\ tye^{32}$
's/he is stupid'	$ntu^{10}\ riq^1\ tye^{32}$	's/he is sad'	$xkuq^{42}\ riq^2\ tye^{32}$
's/he gets mad'	$ntykwen^3\ riq^2\ tye^{32}$	's/he is afraid'	$xqnyi^4\ riq^2\ tye^{32}$
's/he gets used to'	$ntyqan^1\ riq^2\ tye^{32}$	's/he is fed up with/tired of'	$xyaq^2\ riq^2\ tye^{32}$
's/he pities'	$qna^3\ riq^2\ tye^{32}$	's/he bullies'	$xyuq^1\ riq^2\ tye^{32}$
's/he is smart, fast, agile'	$sa^4\ riq^2\ tye^{32}$	's/he believes/is gullible'	$ya^{42}\ ntyqan^4\ riq^1\ qin^{24}\ tye^{32}$
's/he is upset'	$senq^{24}\ riq^1\ tye^{32}$	's/he is satisfied/satiated'	$ylaq^{42}/ndlaq^{42}\ riq^2\ tye^{32}$
's/he is standoffish'	$siyeq^3\ riq^2\ tye^{32}$	's/he is shy'	$yqu^{20}\ riq^2\ tye^{32}$
's/he misperceives'	$skeq^1\ riq^2\ tye^{32}$	's/he breaks a bad habit/learns a lesson'	$\#ksa^4\ riq^2\ tye^{32}$
's/he is fed up'	$skwa^3\ riq^2\ tye^{32}$	's/he is skinny'	$\#ti^4\ riq^2\ tye^{32}$
's/he takes a liking to'	$skwi^1\ riq^2\ tye^{32}$	's/he is skinny'	$\#tyjyan^{20}\ riq^2\ tye^{32}$
's/he is desirous'	$snya^1\ riq^2\ tye^{32}$		

(16)   a.   Ndi³²   riq²   sti¹   Xwa³ kwa³.
            thirsty.ᴘʀᴏɢ essence.3sɢ father.3sɢ Juan   that

            'Juan's father is thirsty.'

       b.   * Ndi³²   sti⁴   Xwa³ kwa³.
            thirsty.ᴘʀᴏɢ father.3sɢ Juan   that

            Sought interpretation: 'Juan's father is thirsty.'

(17)   a.   Ndi³²   renq²⁰.
            thirsty.ᴘʀᴏɢ essence.1sg

            'I am thirsty.'

       b.   * Ndi³².
            thirsty.ᴘʀᴏɢ

            Sought interpretation: 'I am thirsty.'

(18)   a.   Ndi³²   riq²   sten¹.
            thirsty.ᴘʀᴏɢ essence.3sɢ father.1sɢ

            'My father is thirsty.'

       b.   * ndi³²   sten⁴.
            thirsty.ᴘʀᴏɢ father.1sɢ

            Sought interpretation: 'My father is thirsty.'

Table 18 lists some essence predicates whose predicative bases have independent uses, and Table 19, some whose predicative bases are cranberry morphemes. As inspection of both tables reveals, the meaning expressed by an essence predicate L usually cannot be equivalently expressed by using L's predicative base by itself; either the predicative base of L differs in meaning from L (as in (14)) or it is simply unavailable for use as an independent predicate (as in (15)–(18)).

Summarizing, we have seen that essence predicates exhibit three sorts of structural variety: in their choice of nominal component; in whether they exhibit one nominal component or two; and in whether their predicative base has uses apart from the essence predicate. None of these sorts of structural variety is unexpected under the compound predicate hypothesis. Because a compound constitutes a lexeme, two compounds may differ in lexically idiosyncratic ways. Despite their closely related meanings, the English compound nouns *German shepherd* and *Shetland sheepdog* differ in their internal logic; while one can imagine alternative combinations such as *Germany sheepdog* and *Shetlander shepherd*, each breed has its own conventional name agreed upon on the occasion of its coinage. In the same way, the use of *riq²*, *tye³²*, *qin⁴* or the combination *riq² tye³²* as an essence predicate's nominal component is a matter of convention enforced by the lexicon of Chatino. The incidence of essence predicates whose predicative base is a cranberry morpheme is further testimony to their lexical status; in such cases, the predicative base, like the *were-* in English *werewolf*, persists long after losing its status as an independent lexeme. If one instead views essence predicates as predicates having inalienably possessed subjects, the structural variety examined here is somewhat unexpected. On

Table 18: Essence predicates whose predicative bases are also used independently in SJQ Chatino

Essence predicate	Gloss	Independent use of predicative base	
$chaq^3$ $tlyu^2$ riq^2	's/he forgives'	[strong essence]	$chaq^3$ $tlyu^2$ no^4 $kyqyu^1$ 'because the men are strong'
ksa^4 riq^2	's/he breaks a bad habit/learns a lesson'	[break essence]	ksa^4 yka^4 'to break a piece of wood'
$ndwe^4$ riq^2	's/he worries'	[minced essence]	$ndwe^4$ no^4 $kyqyu^1$ 'The men will be minced.'
la^1 riq^2	's/he is open, extroverted'	[open essence]	la^1 no^4 $kyqyu^1$ 'The men got open.'
lwi^3 riq^2	's/he is fair-skinned'	[clean essence]	lwi^3 no^4 $kyqyu^1$ 'The men are clean.'
ndo^{42} riq^2	's/he is happy'	[stand essence]	ndo^{42} no^4 $kyqyu^1$ 'The men are standing.'
$ndya^{24}$ riq^2	's/he likes, loves'	[arrive essence]	$ndya^{24}$ no^4 $kyqyu^1$ 'The men arrived.'
$nkqan^4$ riq^2	's/he remembers'	[sit essence]	$nkqan^4$ no^4 $kyqyu^1$ 'The men are sitting.'
nta^{14} riq^0	's/he is dark-skinned'	[black essence]	nta^{14} no^0 $kyqyu^1$ 'Men are black.'
$ntja^1$ riq^2	's/he is feeling lazy'	[lazy essence]	$ntja^1$ no^4 $kyqyu^1$ 'Men are lazy.'
ntu^{10} riq^0	's/he is stupid'	[stupid essence]	ntu^1 no^0 $kyqyu^1$ 'The men are stupid or slow.'
$ntykwen^3$ riq^2	's/he gets mad'	[climb essence]	$ntykwen^3$ no^4 $kyqyu^1$ 'The men climbed up.'
qna^3 riq^2	's/he pities'	[poor essence]	qna^3 no^4 $kyqyu^1$ 'the poor men'
sa^4 riq^2	's/he is smart, fast, agile'	[light/fast essence]	sa^4 no^4 $kyqyu^1$ 'The men are light or fast.'
$siyeq^3$ riq^2	's/he is standoffish'	[vain essence]	$siyeq^3$ no^4 kyqyu^1 'Men are vain.'
$skwa^3$ riq^2	's/he is fed up'	[lie.elevated essence]	$skwa^3$ no^4 $kyqyu^1$ 'The men are lying elevated.'
$snyi^4$ riq^2	's/he likes sth'	[grab essence]	$snyi^4$ no^4 $kyqyu^1$ 'The men grabbed.'
$sqwe^3$ riq^2/tye^{32}	's/he is generous/ happy'	[good essence]	$sqwe^3$ no^4 $kyqyu^1$ 'Men are good.'
$sqwi^4$ riq^2	's/he remembers'	[exist essence]	$sqwi^4$ no^4 $kyqyu^1$ 'The men exist.'
$sqye^{14}$ riq^0	's/he is strong/ sturdy'	[strong essence]	$sqye^{14}$ no^0 $kyqyu^1$ 'The men are strong.'
$styi^1$ riq^2	's/he is ecstatic'	[laugh essence]	$styi^1$ no^4 $kyqyu^1$ 'The men are laughing.'

$t(j)nya^3$ riq^2	's/he is hard-working'	[work essence]	no^4 $kyqyu^1$ $tnya^3$	'the men who are authorities'
ti^4 riq^2	's/he is skinny'	[skinny essence]	ti^4 no^4 $kyqyu^1$	'The men are skinny.'
$tjenq^3$ riq^2	's/he is frugal or to take care of sth'	[sticky essence]	$tjenq^3$ no^4 $kyqyu^1$	'The men are sticky.'
$tjoq^4$ riq^2	's/he is sturdy'	[strong essence]	$tjoq^4$ no^4 $kyqyu^1$	'The men are strong.'
$tkonq^1$ riq^2	's/he is greedy'	[ambitious essence]	$tkonq^1$ no^4 $kyqyu^1$	'The men are ambitious.'
$tlaq^{14}$ riq^0	's/he placates, calms'	[cool essence]	$tlaq^{14}$ no^4 $kyqyu^1$	'The men are cooled, calm.'
$tnyaq^4$ riq^2	's/he is tired'	[tired essence]	$tnyaq^4$ no^0 $kyqyu^1$	'The men are tired.'
tqa^{24} riq^2	's/he is fully conscious'	[complete essence]	tqa^{24} no^4 $kyqyu^1$	'all men'
tqi^4 riq^2	's/he hates'	[sick essence]	tqi^4 no^4 $kyqyu^1$	'The men are sick.'
tya^{20} riq^2	's/he is smart'	[smart essence]	tya^{20} no^4 $kyqyu^1$	'The men are smart.'
$tyaq^4$ riq^2	's/he is slow'	[slow essence]	$tyaq^4$ no^4 $kyqyu^1$	'The men are slow.'
$tyjyan^{20}$ riq^2	's/he is skinny'	[skinny essence]	$tyjyan^{20}$ no^4 $kyqyu^1$	'The men are skinny.'
$tykeq^{14}$ riq^0	's/he is hot'	[hot essence]	$tykeq^{14}$ no^0 $kyqyu^1$	'The men are hot (temp).'
$xyaq^2$ riq^2	's/he is fed up with/tired of'	[mix essence]	$xyaq^2$ no^4 $kyqyu^1$	'The men will be mixed.'

that conception, the choice among riq^2, tye^{32} and qin^4 as subjects should seemingly be independent of the choice of predicate, and they should not appear in tandem (any more than *you* and *they* should appear in tandem to produce sentences such as **You they left*).

4.2 External syntax

With only occasional exceptions, the components of an essence predicate can be interrupted by members of a small class of elements; their syntax relative to these elements is a revealing criterion for evaluating the possessed-subject and compound predicate hypotheses. The class of interruptors includes the elements in (19), some of which Rasch (2002: 10) labels EVENT MODIFIERS; we extend his terminology to the full class. These may intervene between a verb and its subject, as in examples (20)–(25) (where verb and subject are in boldface). Correspondingly, they may sometimes intervene between an essence predicate's predicative base and its nominal element, as in (26)–(33), in which the interrupted essence predicates are in boldface.

Table 19: Essence predicates whose predicative bases are not used independently in SJQ Chatino

Essence predicate	Gloss
jlya^{32} riq^2	's/he forgot'
jlyo^{20} riq^2	's/he knews/was aware'
jnya^{20} riq^2	's/he was mischievous, playful'
ngwi^3 riq^2	's/he realized'
ndwe^4 riq^2	's/he worried'
ndya^{32} riq^2	's/he liked, loved'
ndi^4 riq^2	's/he was thirsty, wheezing'
nteq^4 riq^2	's/he was hungry'
ntqan^1 riq^1	's/he was weak'
ntyqan^1 riq^2	's/he got used to'
qna^3 riq^2	's/he pitied'
senq^{24} riq^1	's/he was upset'
skeq^1 riq^2 / qin^4	's/he mistook, misperceived'
skwi^1 riq^2	's/he took a liking to'
snya^1 riq^2	's/he was desirous'
stu^1 riq^0	's/he was happy'
stya^4 riq^2	's/he was disgusted'
tlyaq^4 riq^2	's/he was cold'
tsa^3 riq^2	's/he was flirtatious'
tya^{20} riq^2	's/he was careful'
xkuq^{42} riq^2	's/he was sad'
xqnyi^4 riq^2	's/he was afraid'
xyaq^2 riq^2	's/he was fed up with/ tired of'
xyuq^1 riq^2	's/he bullied'
ylaq^{42}/ ndlaq^{42} riq^2	's/he was satisfied/satiated'

(19) a. sqwe^3
'good, well'

b. ka^{24}
'able to; expression of emphasis'

c. ye^{42}
'very'

d. la^{24}
'comparative'

e. qa^{24}
'very'

f. kcha^4
'crazy'

g. ti[2], ti[4]
    'very, still, just'

h. kcha[4] qa[1]
    'crazy'

(20) **Ntqan[42]** ti[4]   **La[20]ya[24]** kwa[3] kna[1], kwen[42] qa[24] ntqo[1]   xqya[24].
      see.CPL EV.MOD:just Hilaria   that   snake loud   very leave.CPL scream

      'As soon as Hilaria saw the snake, she screamed very loudly.'

(21) **Nkwa[2]** **tqi[4]** ka[3]   **nten[14]** **no[0]** **yku[24]** **kla[32]** **xi[1]** **kanq[42]** qa[24].
      be.CPL sick EV.MOD:EMPH people that eat.CPL fish   bad that.ABS very

      'People who ate the bad fish got really sick.'

(22) Ti[2]   **ykwiq[1]**   ye[42]   **silya[14]** qo[0]   chaq[3] tyqo[1] qo[1] ja[4]
      EV.MOD:very speak.CPL EV.MOD:very police   with.3SG to   leave and NEG
      slya[1]   qa[1].
      agree.CPL   NEG

      'The police pleaded with him to leave and he refused (to leave).'

(23) **Ykwiq[4]**   la[1]   **sti[4]-qo[2]**   **kwa[4]** ke[4] neq[4]-sya[10]   kwa[1].
      speak.CPL EV.MOD:more father-saint that   then type.people-justice that

      'The priest spoke more than the authorities.'

(24) **Ya[42]**   kcha[4]   **no[4]** qan[1]   **lyuq[04]** kwa[3].
      go.away.CPL EV.MOD:crazy one female little   that

      'That little girl went (somewhere) aimlessly.'

(25) **Qya[42]**   kcha[4]   qa[1]   **kyo[24]** nka[3].
      fall.CPL EV.MOD:crazy EV.MOD:very rain   yesterday

      'It rained crazy, unpredictably yesterday.'

(26) **Nkqan[4]**   sqwe[3]   **riq[2]**   no[4] qan[1]   chaq[3] tsa[24]
      sitting.ground.PROG EV.MOD:well essence the female (one) COMPL go.away.POT
      kwaq[3].
      that

      'That woman remembers well that she has to go to the party.'

(27) **sqwa[140]**   ye[42]   **chaq[3]**   **tye[32]**.
      put CPL.3SG EV.MOD:very word/thing chest

      'He really encouraged him.'

(28) **Ndon[42]**   qa[24]   **riq[2]**   Xwa[3] kwa[3] ndon[42]   tqwa[4]-tqan[4]   qin[4].
      stand.PROG EV.MOD:very essence Juan that   stand.PROG mouth-house his

      'Juan was very happy standing in is front porch.'

(29) **Ndon[42]**   ti[4]   **riq[2]**   Xwa[3] kwa[3] ndon[42]   tqwa[4]-tqan[4]   qin[4].
      stand.PROG EV.MOD:only essence Juan that   stand.PROG mouth-house his

      'Juan was just happy standing in his front porch.'

(30) **Ndon⁴²** ka²⁴ **riq²** Xwa³ kwa³ ndon⁴² tqwa⁴-tqan⁴ qin⁴.
stand.PROG EV.MOD:very essence Juan that stand.PROG mouth-house his
'Juan was sure very happy standing in his front porch.'

(31) **Ndon⁴²** la²⁴ **riq²** Xwa³ kwa³ ndon⁴² tqwa⁴-tqan⁴ qin⁴.
stand.PROG EV.MOD:more essence Juan that stand.PROG mouth-house his
'Juan was happier standing in his front porch.'

(32) **Ndo⁴²** kcha⁴ qa¹ **riq²**.
happy EV.MOD:crazy EV.MOD:very essence
'He is crazy happy.'

(33) **Stu¹⁰** kcha⁰ *(qa²⁴) **riq¹** Xwa³ kwa³ nkya²⁴.
gusto EV.MOD:crazy EV.MOD:very essence Juan that go.away.CPL
'Juan left awfully happy.'

Strikingly, compound predicates generally resist the intrusion of an event modifier, a fact reflected by the unacceptability of (34). When an event modifier combines with a compound predicate, it generally follows it, as in (35). Yet, event modifiers in general do not follow essence predicates, as the evidence in (36) and (37) attests. Similarly, event modifiers do not typically follow the subject of a clause. Thus, in (38), the event modifier may intrude between the verb *ylu²* 'it grew' and its subject *yka²⁴-knyi²⁴ kwa³* 'that tree graft' (as in (38a)) but cannot follow the subject (*(38b)). The overarching generalization is that an event modifier typically follows the head of a predicate phrase, whether this head be simplex or compound. This generalization suggests that because an event modifier typically follows an essence predicate's predicative base, the essence predicate itself is phrasal.

(34) * **Ykon¹** ten²⁴ **jyanq²⁴** skwa¹ qin²⁴.
eat.CPL.1SG EV.MOD:only.1SG measure.1SG soup his
Sought interpretation: 'I only tasted his soup.'

(35) **Ykon¹** **jyanq²⁴** ten²⁴ skwa¹ qin²⁴.
eat.CPL.1SG measure.1SG EV.MOD:only.1SG soup his
'I only tasted his soup.'

(36) a. **Ndon⁴²** qa²⁴ **riq²**.
stand.PROG EV.MOD:very essence
'S/he is very happy.'

b. * **Ndon⁴²** **riq²** qa¹.
stand.PROG essence EV.MOD:very
Sought interpretation: 'S/he is very happy.'

(37) a. **Qne⁴²** sqwe³ **tye³**.
do.CPL EV.MOD:good chest.3SG
'S/he dared do something.'

b. * **Qne^{42} tye^{32}** sqwe^3.
do.CPL chest.3SG EV.MOD:good

Sought interpretation: 'S/he dared do something.'

(38) a. **Ylu^2** sqwe^3 **yka^{24}-knyi^{24} kwa^3**.
grow.CPL EV.MOD:well tree-graft that

'That grafted tree grew really well.'

b. * ***Ylu^2** **yka^{24}-knyi^{24}** sqwe^3.
grow.CPL tree-graft EV.MOD:good

Sought interpretation: 'That grafted tree grew really well.'

This distributional generalization about event modifiers is, however, deceptively broad, because event modifiers exhibit a number of idiosyncrasies in their interaction with essence predicates. On one hand, the event modifiers *ti^2 / ti^4* 'very, still, just', *ka^{24}* 'able to', *la^{24}* 'comparative', *kcha^4* 'crazy', and *kcha^4 qa^1* 'crazy' intervene quite freely between the parts of an essence predicate with two components; thus, all of these event modifiers may appear in the contexts in (39). On the other hand, if an essence predicate has three or more components, these event modifiers exhibit a much more variable pattern of distribution, as the examples in (40) suggest.

(39) 's/he worries' ndwe^4 — riq^2
's/he remembers' nkqan^4 — riq^2
's/he is standoffish' siyeq^3 — riq^2
's/he is daring' tno^4 — tye^{32}
's/he is afraid' xqnyi^4 — riq^2

(40) 's/he forgives' chaq^3 * tlyu^2 * riq^2 (very idiomatic)
's/he realizes' nkwa^2 * jyaq^3 ✓ riq^2
's/he made up her/his mind' wa^2 * xtya^{20} ✓ riq^2/tye^{32}
's/he feels sad' tqwa^{14} ✓ nka^{24} * tye^{32}
's/he believes/is gullible' ya^{42} * ntyqan^4 ✓ riq^1 * qin^{24}

Moreover, the event modifiers *sqwe^3* 'good', *ye^{42}* 'very' and *qa^{24}* 'very' exhibit a much higher degree of idiosyncrasy in their capacity to intervene between the parts of an essence predicate, as the examples in Table 20 show. This irregularity very likely has more than one cause. Some interventions are semantically improbable, e.g. **senq^{24} sqwe^3 riq^1* 's/he is well upset'. But it also appears that essence predicates are simply more fully grammaticalized as tightly bound units, more strongly resisting intrusive formatives.

We conclude that although the distribution of event modifiers exhibits a number of idiosyncrasies, essence predicates resemble verb + subject combinations more closely than they resemble compound predicates as regards their interaction with event modifiers. Thus, this evidence militates in favor of the possessed-subject hypothesis and against the compound predicate hypothesis.

Table 20: Intervention of the event modifiers $sqwe^3$ 'good', ye^{42} 'very' and qa^{24} 'very' into essence predicates in SJQ Chatino

Gloss	Essence predicate	$sqwe^3$ 'good' ye^{42} 'very'	qa^{24} 'very'
's/he remembered'	$nkqan^4 - riq^2$	✓	✓
's/he was smart, fast, agile'	$sa^4 - riq^2$	✓	✓
's/he was happy'	$ndon^{42} - riq^2$	✓	*
's/he remembered'	$sqwi^4 - riq^2$	✓	*
's/he was upset'	$senq^{14} - riq^0$	*	✓
's/he pitied'	$qna^3 - riq^2$	*	✓
's/he was sad'	$xkuq^{42} - riq^2$	*	*
's/he worried'	$ndwe^4 - riq^2$	*	*
's/he was fed up'	$skwa^3 - riq^2$	#	#
's/he hated'	$tqi^4 - riq^2$	#	#
's/he was generous/happy'	$sqwe^3 - riq^2/tye^{32}$	#	*
's/he was taciturn'	$liqa^{24} - riq^1$	#	*
's/he was scared/queasy'	$chin^4 - nga^{24}\ tye^{32}$	*	*
	$chin^4\ nga^{24} - tye^{32}$	*	*
's/he liked'	$ndya^{32} - riq^2\ tye^{32}$	*	✓
	$ndya^{24}\ riq^2 - tye^{32}$	#	#
's/he felt angry'	$xqan^{10} - nga^{24}\ tye^{32}$	*	#
	$xqan^{10}\ nga^{24} - tye^{32}$	*	*

4.3 Lack of compositionality

As we have seen, essence predicates tend to refer psychological states, with some exceptions. In a large proportion of cases, essence predicates are not transparently compositional. There are, to be sure, those whose semantics is directly deducible from their parts; examples are the essence predicates in Table 21. But a substantial number of essence predicates exhibit various degrees of departure from compositionality; the examples in Table 22 illustrate. The analogy of essence predicates to lexically reflexive verbs (noted in section 1) is again apt, since reflexive predicates are often idiosyncratic in their semantics; compare *attendre* 'wait for' to *s'attendre* (*à*) 'expect', *douter* 'doubt' to *se douter* 'suspect', *rendre* 'return' to *se rendre* (*à*) 'go to'. In the case of essence predicates whose predicative base is a cranberry morpheme appearing in no context other than the essence predicate itself (see again Table 19), there is no real question of compositionality. Here, too, the analogy to lexically reflexive verbs holds, since they also may be based on cranberry morphemes, as in the case of French *s'évanouir* 'faint' (whose verbal base *évanouir* has no independent use).

Table 21: Semantically transparent essence predicates in SJQ Chatino

	Essence predicate	Glosses of component parts
's/he was hard-working'	$t(j)nya^3\ riq^2$	[work essence]
's/he was open, extrovert'	$la^1\ riq^2$	[open essence]
's/he realized'	$ngwi^3\ riq^2$	[awake essence]
's/he got used to'	$nt(y)qan^1\ riq^2$	[used.to essence]
's/he was hungry'	$ntenq^{32}\ riq^2$	[hungry essence]
's/he was feeling lazy'	$ntja^1\ riq^2$	[lazy essence]
's/he was stupid'	$ntu^1\ riq^0$	[stupid essence]
's/he misperceived'	$skeq^1\ riq^0$	[imagine essence]
's/he was strong/ sturdy'	$sqye^{14}\ riq^0$	[strong essence]
's/he was skinny'	$ti^4\ riq^2$	[skinny essence]
's/he was sturdy'	$tjoq^4\ riq^2$	[sturdy essence]
's/he was greedy'	$tkonq^1\ riq^2$	[greedy essence]
's/he was cold'	$tlyaq^4\ riq^2$	[cold essence]
's/he was tired'	$tnyaq^4\ riq^2$	[tired essence]
's/he was slow'	$tyaq^4\ riq^2$	[slow essence]
's/he was skinny'	$tyjyan^{20}\ riq^2$	[skinny essence]
's/he was hot'	$tykeq^{14}\ riq^0$	[hot essence]
's/he was shy'	$yqu^{20}\ riq^2$	[embarrassed essence]

These facts about the semantics of essence predicates might be seen as favoring the compound predicate hypothesis; the observed variability in semantic transparency is, of course, typical of compounds. But the semantic noncompositionality of many essence predicates might be reconciled with the possessed-subject hypothesis by regarding them as idioms; even the incidence of essence predicates with cranberry morphemes might be likened to the fact that idioms sometimes involve words that have no use outside the idiom (e.g. *jiffy* in the idiom *in a jiffy*, *dint* in *by dint of*, *fro* in *to and fro*). Neverthe-less, recurring commonalities of form and content among essence predicates might be argued to make them different from idioms, which tend not to possess this high degree of systematicity.

4.4 Distributional flexibility of subject–agreement marking

An important feature of Chatino subject-agreement marking is its flexibility: in the in-flection of a simplex verb, subject-agreement marking is expressed cumulatively with aspect/mood marking (as in the case of sqi^2 's/he bought'—Table 8); but in the inflection of a compound predicate, aspect/mood is marked on the first member, and subject agree-ment is marked separately, on the second member (as in the case of $yku^4\ jyaq^3$'s/he tasted'—Table 12). This flexibility extends even farther: If a simplex verb is followed by

Table 22: Semantically opaque essence predicates in SJQ Chatino

	Essence predicate	Glosses of component parts
's/he was mischievous, playful'	$jnya^{20}\ riq^2$	[borrow essence]
's/he broke a bad habit/learned a lesson'	$ksa^4\ riq^2$	[break essence]
's/he worried'	$ndwe^4\ riq^2$	[minced essence]
's/he was fair-skinned'	$lwi^3\ riq^2$	[clean essence]
's/he was standoffish'	$lyaq^{14}\ riq^0$	[quiet essence]
's/he mocked'	$lyeq^3\ riq^2$	[fun essence]
's/he was satisfied/satiated'	$ndla^2\ riq^2$	[fast essence]
's/he was happy'	$ndon^{42}\ riq^2$	[stand essence]
's/he remembered'	$nkqan^4\ riq^2$	[sit essence]
's/he realized'	$nkwa^2\ jyaq^3\ riq^2$	[be.able measure essence]
's/he was dark-skinned'	$nta^{14}\ riq^0$	[dark essence]
's/he pitied'	$qna^3\ riq^2$	[poor essence]
's/he was smart/fast agile'	$sa^4\ riq^2$	[airy essence]
's/he was fed up'	$skwa^3\ riq^2$	[lie.elevated essence]
's/he took a liking to'	$skwi^1\ riq^2$	[round essence]
's/he liked sth'	$snyi^4\ riq^2$	[grab essence]
's/he was excited'	$sti^1\ riq^2$	[laugh essence]
's/he was standoffish'	$syeq^3\ riq^2$	[happy essence]
's/he was frugal or took care of sth'	$tjenq^3\ riq^2$	[sticky essence]
's/he placated'	$tlaq^{14}\ riq^0$	[cool essence]
's/he was fully conscious'	$tqa^{24}\ riq^2$	[complete essence]
's/he was envious'	$tqi^4\ riq^2$	[sick essence]
's/he was afraid'	$xqnyi^4\ riq^2$	[sad essence]
's/he was fed up with/tired of'	$xyaq^2\ riq^2$	[mix essence]
's/he bullied'	$xyuq^1\ riq^2$	[naughty essence]

an event modifier, the event modifier may carry the verb's subject-agreement morphology; thus, compare the inflection of $ykwiq^4$'s/he spoke' in Table 23 with that of $ykwiq^4$ ti^4 's/he just spoke' [speak EVENT.MODIFIER] in Table 24.[9]

The compound predicate hypothesis entails that in the inflection of an essence predicate, the nominal component (riq^2, tye^{32} or qin^4, alone or in combination) functions very much like the event modifier ti^4 in the inflection of $ykwiq^4$ ti^4 's/he just spoke': not as a subject, but as an adverbial or quasi-adverbial modifier of the predicate's head; in either instance, the modifier's adjacency to the preceding head makes it available to carry the head's agreement morphology. On this view, the literal meaning of an essence predicate's nominal component does not combine in a compositional way with the lit-

---

[9]Note that as in the inflection of the compound verb yku^4 $jyaq^3$'s/he tasted' [eat amount] in Table 12, the inflection of the verb + event modifier combination $ykwiq^4$ ti^4 's/he just spoke' [speak EVENT.MODIFIER] exhibits ablaut of its verbal element in the first person singular.

Table 23: Paradigm of the verb *ykwiq*^{4} 's/he spoke' in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	*ykwenq*^{1}	*ntykwenq*^{1}	*ntykwenq*^{20}	*tykwenq*^{20}
2SG	*ykwiq*^{42}	*ntykwiq*^{42}	*ntykwiq*^{42}	*tykwiq*^{42}
3SG	*ykwiq*^{4}	*ntykwiq*^{32}	*ntykwiq*^{4}	*tykwiq*^{4}
1INCL	*ykwenq*^{24} *en*^{32}	*ntykwenq*^{1} *en*^{32}	*ntykwenq*^{24} *en*^{32}	*tykwenq*^{24} *en*^{32}
1EXCL	*ykwiq*^{4} *wa*^{42}	*ntykwiq*^{32} *wa*^{42}	*ntykwiq*^{4} *wa*^{42}	*tykwiq*^{4} *wa*^{42}
2PL	*ykwiq*^{4} *wan*^{4}	*ntykwiq*^{32} *wan*^{4}	*ntykwiq*^{4} *wan*^{4}	*tykwiq*^{4} *wan*^{4}
3PL	*ykwiq*^{4} *renq*^{4}	*ntykwiq*^{32} *renq*^{4}	*ntykwiq*^{4} *renq*^{4}	*tykwiq*^{4} *renq*^{4}

Table 24: Paradigm of *ykwiq*^{4} *ti*^{4} 's/he just spoke' [speak EVENT.MODIFIER] in SJQ Chatino

	COMPLETIVE	PROGRESSIVE	HABITUAL	POTENTIAL
1SG	*ykwenq*^{1} *ten*^{24}	*ntykwenq*^{1} *ten*^{24}	*ntykwenq*^{20} *ten*^{24}	*tykwenq*^{20} *ten*^{24}
2SG	*ykwiq*^{4} *ti*^{42}	*ntykwiq*^{32} *ti*^{42}	*ntykwiq*^{4} *ti*^{42}	*tykwiq*^{4} *ti*^{42}
3SG	*ykwiq*^{4} *ti*^{4}	*ntykwiq*^{32} *ti*^{4}	*ntykwiq*^{4} *ti*^{4}	*tykwiq*^{4} *ti*^{4}
1INCL	*ykwiq*^{4} *ten*^{24} *en*^{32}	*ntykwiq*^{32} *ten*^{24} *en*^{32}	*ntykwenq*^{24} *ten*^{24} *en*^{32}	*tykwenq*^{24} *ten*^{24} *en*^{32}
1EXCL	*ykwiq*^{4} *ti*^{4} *wa*^{42}	*ntykwiq*^{32} *ti*^{4} *wa*^{42}	*ntykwiq*^{4} *ti*^{4} *wa*^{42}	*tykwiq*^{4} *ti*^{4} *wa*^{42}
2PL	*ykwiq*^{4} *ti*^{4} *wan*^{4}	*ntykwiq*^{32} *ti*^{4} *wan*^{4}	*ntykwiq*^{4} *wan*^{4}	*tykwiq*^{4} *wan*^{4}
3PL	*ykwiq*^{4} *ti*^{4} *renq*^{4}	*ntykwiq*^{32} *ti*^{4} *renq*^{4}	*ntykwiq*^{4} *renq*^{4}	*tykwiq*^{4} *renq*^{4}

eral meaning of the predicative base; instead, the nominal component has been grammaticalized with a meaning something like that of English *inside* in experiencer-based expressions such as *ntykwen*^{3} *riq*^{24} 's/he got angry inside'; note again that reflexive pronouns have been grammaticalized with much the same function in expressions such as *elle s'est fâchée* 'she got angry inside'. Thus, the compound predicate hypothesis situates the expression of subject agreement in essence predicates within a larger, independently motivated system in which other compound predicates and verb + event modifier combinations also participate in parallel fashion. The distributional flexibility of subject agreement therefore yields equivocal results. Both the possessed-subject hypothesis and the compound predicate hypothesis relate the person/number marker on an essence predicate's nominal component to an independent phenomenon in Chatino: according to the possessed-subject hypothesis, the person/number marking on an essence predicate's nominal component can be identified with a noun's inflection for the person and number of an inalienable possessor; by contrast, the compound predicate hypothesis entails that an essence predicate's nominal component reflects a more general pattern in which the person and number of a predicate's subject are marked on a nonsubject constituent—on the second member of a compound predicate, on an event modifier, or on a quasi-adverbial essence word. Given that both of these patterns of person/number

marking must in any event be countenanced in an adequate grammar of Chatino, it is not clear that the present criterion provides compelling evidence for choosing either of the two hypotheses over the other.

5 Essence predicates: A formal interpretation

Superficially, the properties of essence predicates seem ambiguous in their implications for a formal analysis. The essence predicate in (41) on the one hand resembles the verb-subject construction in (42): in both cases, the predicative word (in boldface) is inflected for aspect/mood and the nominal element (in italics) is inflected for person and number. At the same time, the essence predicate in (41) resembles the compound verb in (43): here, too, the boldface predicative word is inflected for aspect/mood and the nominal element is inflected for person and number. Finally, the essence predicate in (41) resembles the verb + event modifier combination in (44), where the predicative word is again inflected for aspect/mood and the event modifier, for person and number.

(41) **Ndi⁴** *riq²* no⁴ kyqyu¹ kwa³.
 thirsty.CPL essence.3SG one male that

 'That guy was thirsty.'

(42) **Nkya²⁴** *sti¹* Xwa³ kwa³.
 go.baseward.CPL father.3SG Juan that

 'Juan's father left.'

(43) **Ykwiq⁴** *sla³* no⁴ qan¹ kwa³.
 speak.CPL tiredness.3SG one female that

 'That woman dreamt.'

(44) **Ykwiq⁴** *ti⁴* no⁴ kyqyu¹ kwa³.
 speak.CPL EV.MOD.3SG one male that

 'That guy just spoke.'

According to the possessed-subject hypothesis, an essence predicate is a predicate-subject construction comparable to that of (42): its nominal element (*riq²* 'essence' in (41)) is a subject, and as in (42), the inflectional marking on the subject expresses the person and number of an inalienable possessor; this entails that *no⁴ kyqyu¹ kwa³* 'that guy' is not the subject of (41), but instead denotes an inalienable possessor, like *Xwa³* 'Juan' in (42).

According to the compound predicate hypothesis, an essence predicate is a compound predicate comparable to those of (43) and (44). In a compound predicate, the second element is not a subject, but is either a complement or a modifier of the predicate (as in (43) and (44) respectively), so that its inflection encodes the person and number of the predicate's subject rather than that of an inalienable possessor. This suggests that through grammaticalization, an essence predicate's nominal component has come to

serve a quasi-adverbial function, ordinarily causing the predicate to refer to the psychological or physical state of its subject's referent.

In section 3, we examined four characteristics of essence predicates: their structural variety, their external syntax relative to event modifiers, their general lack of semantic compositionality, and their possible relation to the distributional flexibility of Chatino subject-agreement marking. As we have seen, these four criteria do not decisively favor either of the two hypotheses under consideration. The criterion of external syntax seems to favor the possessed-subject hypothesis; the criteria of structural variety and lack of compositionality seem to favor the compound predicate hypothesis; and the criterion of the distributional flexibility of subject agreement marking does not clearly favor either hypothesis.

It is clear from this impasse that a third hypothesis is necessary to account for the properties of essence predicates. We therefore suggest the following account.

- We regard an essence predicate as a lexeme whose predicative base and nominal component act as separate constituents in syntax.[10] They are different from compound predicates in Chatino: their parts may be interrupted by event modifiers, while those of a compound predicate in general cannot.

- We propose that every Chatino predicate has an "inflectional domain" to which its agreement morphology is confined. A predicate is ordinarily its own domain; this is true whether the predicate is simplex or compound. In the former case, aspect/mood and agreement are expressed cumulatively. In the latter case, inflection is regulated by a principle of distributed exponence which we here equate with the Compound Inflection Criterion; according to this principle, a compound predicate's inflection is ordinarily bipartite, with aspect/mood marked on its head and agreement marked on nonhead component. (The details of this principle are complicated by deviations from this ordinarily bipartite pattern, as e.g. in the first-person singular forms in Tables 12 and 24; see Cruz & Woodbury (2013) for discussion of the range of such deviations.)

- Certain kinds of syntactic combinations also constitute inflectional domains. If a simplex verb is modified by an adjacent event modifier, these two words compose an inflectional domain, whose inflection again involves the distributed exponence

---

[10]There is abundant evidence that lexemes may inflect periphrastically; for discussion, see Börjars et al. (1997), Sadler & Spencer (2001), Ackerman & Stump (2004), Ackerman et al. (2011), Chumakina & Corbett (2013), Bonami & Samvelian (2009), and Bonami (2015). In many languages, a lexeme's paradigm may include both synthetic and periphrastic realizations; that is, periphrasis is used for the realization of particular morphosyntactic property sets (as in Latin, where periphrastic realizations occupy the perfective passive cells in paradigms whose other cells are realized synthetically). An essence predicate, however, is uniformly periphrastic in its realization; that is, the incidence of periphrasis is not restricted to the realization of particular morphosyntactic property sets, but is characteristic of all of an essence predicate's realizations. This view of essence predicates as lexemes whose realization is invariably periphrastic recalls the similar conception of Persian complex predicates proposed by Bonami & Samvelian (2010).

prescribed by the Compound Inflection Criterion.[11] In addition, an essence predicate is a lexeme whose periphrastic realization functions as an inflectional domain, exhibiting the same pattern of distributed exponence. In particular, its person/number marking is situated on its nominal component and is an expression of subject agreement rather than inalienable possession.

- An essence predicate's nominal component is not a subject, but has been grammaticalized as a quasi-adverbial formative ordinarily serving to express the psychological or physical state of the referent of the essence predicate's subject.

- The structural variety of essence predicates and their semantic idiosyncrasy reflect their status as lexemes listed in the lexicon.

- In most instances, the parts of an essence predicate are recognizably associated with independent lexemes, but this is not invariably the case. In English, the derivational suffix *-ize* may transparently relate a verb with a causative or inchoative meaning to a nominal or adjectival stem (*magnet→ magnetize*, *popular → popularize*) but may also simply mark a causative or inchoative verb that is not synchronically related to any nominal or adjectival base (*baptize, ostracize, recognize*). Analogously, a Chatino essence predicate denoting a psychological or physical state may be transparently related (in form if not in content) to an independent predicate (as in (45)) but there are also "intrinsic" essence predicates that are synchronically unrelated to any independent predicate (as in (46)). The observed parallelism of reflexive verbs is again telling: *demander* 'ask' → *se demander* 'wonder', but *se moquer* 'mock' (**moquer*).

(45) skwa^3 riq^2 (← *skwa^3* 's/he lay elevated')
      's/he was fed up'

(46) ndi^{32} riq^2 (**ndi^{32}* without *riq^2*)
      's/he was thirsty'

Other Oto-Manguean languages possess essence predicates exhibiting both similarities to and differences from those of SJQ Chatino; future work on these similarities and differences will likely shed additional light on the properties of this distinctive class of predicates.

Acknowledgements

---

[11]There are also cases in which the combination of a compound predicate with an adjacent event modifier constitutes an inflectional domain in which subject agreement is marked both on the compound predicate's non-head component and on the event modifier; sentence (35) is an example of this sort.

References

Ackerman, Farrell & Gregory T. Stump. 2004. Paradigms and periphrastic expression: A study in realization-based lexicalism. In Louisa Sadler & Andrew Spencer (eds.), *Projecting morphology*, 111–157. CSLI Publications.

Ackerman, Farrell, Gregory T. Stump & Gert Webelhuth. 2011. Lexicalism, periphrasis and implicative morphology. In Robert D. Borsley & Kersti Börjars (eds.), *Non-transformational theories of grammar*, 325–358. Oxford University Press.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25. 63–110.

Bonami, Olivier & Pollet Samvelian. 2009. Inflectional periphrasis in Persian. In Stefan Müller (ed.), *Proceedings of the HPSG 2009 conference*, 26–46. CSLI Publications.

Bonami, Olivier & Pollet Samvelian. 2010. Persian complex predicates: Lexeme formation by itself. Presented at the Septièmes Décembrettes, December 2010.

Börjars, Kersti, Nigel Vincent & Carol Chapman. 1997. Paradigms, periphrases and pronominal inflection: A feature-based account. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1996*, 155–180. Dordrecht: Kluwer Academic Publishers.

Campbell, Eric. 2011. Zenzontepec Chatino aspect morphology and Zapotecan verb classes. *International Journal of American Linguistics* 77(2). 219–246.

Chen, Matthew. 2004. *Tone sandhi: Patterns across Chinese dialects*. Cambridge University Press.

Chumakina, Marina & Greville G. Corbett (eds.). 2013. *Periphrasis: The role of syntax and morphology in paradigms*. Oxford University Press.

Cruz, Emiliana. 2011. *Phonology, tone and the functions of tone in San Juan Quiahije Chatino*. Austin TX dissertation.

Cruz, Emiliana. 2016. San Juan Quiahije noun possession. Unpublished manuscript.

Cruz, Emiliana & Anthony C. Woodbury. 2013. Tonal complexity in San Juan Quiahije Eastern Chatino compound verb inflection. Paper presented at the LSA & SSILA special session: Inflectional classes in the languages of the Americas, Boston MA.

Rasch, Jeffrey. 2002. *The basic morpho-syntax of yaitepec chatino*. Houston TX dissertation.

Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2000*, 71–96. Dordrecht: Kluwer Academic Publishers.

Woodbury, Anthony C. to appear. Verb inflection in the Chatino languages: The separate life cycles of prefixal vs. tonal conjugational classes. In Enrique L. Palancar, Matthew Baerman & Timothy Feist (eds.), *New lights on inflectional classes: Endangered complexity in the Oto-Manguean languages of Mexico*, 129–158. Loc: Surrey ePubs.

Chapter 10

Why traces of the feminine survive where they do, in Oslo and Istria: How to circumvent some "troubles with lexemes"

Hans-Olav Enger

The paper examines a surprising parallel in the development of the feminine gender in Oslo Norwegian on the one hand and Istro-Romanian (spoken in Croatia) on the other. In both cases, the feminine gender is lost on all 'normal' gender markers, but a trace of the feminine remains on the definite suffix, which is the 'last redoubt' of the feminine gender. An attempt is made to link this development to a slightly modified version of the Agreement Hierarchy. It is suggested that the Hierarchy may be linked to grammaticalisation, and that we should not draw too strict lines between different kinds of agreement.

1 The main point

The starting-point for what follows is a parallel between Norwegian as spoken in Oslo, Norway, and Istro-Romanian, as spoken on the Istrian peninsula in Croatia. In both cases, feminine agreement is reduced, diachronically, and in both cases, traces of the feminine remain longer in one specific place, namely word-internally, than elsewhere. Why would there be such a parallel? I suggest an account which involves a modified version of Corbett's (1979, 2006) Agreement Hierarchy. In brief, the 'definite article', when it is a suffix, has a different status than other elements that signal gender. Furthermore, Furthermore, an examination of the hierarchy reveals that it may be 'anchored' in the workings of diachrony and psycholinguistics.

2 The empirical background

2.1 Oslo

In the Oslo dialect of Norwegian, a change has taken place. A century ago, this dialect had three genders (in the singular, like German).[1] Compare (1):

(1) Three genders in Oslo dialect ca. 1900; examples in Norwegian Bokmål orthography.

   a. en liten gutt, en fin gutt, denne gutten, ikke noen gutt
     a.M little.M boy, a.M fine.MF boy, this.MF boy.DEF.SG.{M}, not any.M boy

   b. en liten stol, en fin stol, denne stolen, ikke noen stol
     a.M little.M chair, a.M fine.MF chair, this.MF chair.DEF.SG.{M}, not any.M chair

   c. ei lita jente, ei fin jente, denne jenta, ikke noa jente
     a.F little.F girl, a.F. fine.MF girl, this.MF girl.DEF.SG.{F}, not any.F girl

   d. ei lita jakke, ei fin jakke, denne jakka, ikke noa jakke
     a.F little.F jacket, a.F fine.MF jacket, this.MF jacket.DEF.SG.{F}, not any.F jacket

   e. et lite barn, et fint barn, dette barnet, ikke noe barn
     a.N little.N child, a.N fine.N child, this.N child.DEF.SG{N}, not any.N child

   f. et lite hus, et fint hus, dette huset, ikke noe hus
     a.N small.N house, a.N fine.N house, this.N house.DEF.SG{N}, not any.N house

There is clear evidence for three genders, masculine (1a,1b), feminine (1c,1d) and neuter (1e,1f). The formal differentiation between the masculine and the feminine is not so clearly marked as that of both of them in opposition to the neuter. The masculine–feminine distinction is not realised on all associated words, but it is realised on some very central determiners and a few highly frequent adjectives, such as the adjective *liten* 'small', which is overdifferentiated; showing 'too many' contrasts (cf. Corbett 2007). By contrast, the adjective *fin* 'fine' is 'regular', showing only the opposition neuter vs. non-neuter, in the same way as the proximal determiner *denne*.[2] In such cases, I have assigned the value 'mf'.

The status of the suffix in the definite singular of nouns is intriguing (see e.g. Enger & Corbett 2012 and Section 3.2.3 below). Genders are defined as classes of nouns reflected in the behaviour of associated words (Corbett 1991). Suffixes do not count as 'associated words'; and yet, in the nouns in (1), the suffixes are in a strict 1:1 relation with the gender exponents. If a noun takes *-a* in the definite singular (e.g. *jente* 'girl'), it will invariably also take *ei* 'a.F', *lita* 'little.F', *noa* 'any.F' and other 'associated words' expected from a feminine: if it takes *-en* in the definite singular, it will also take *en* 'a.M', *liten* 'small.M',

---

[1] The following draws on Larsen (1907) and Lødrup (2011) in particular; but cf. also Enger (2004a,c) and Opsahl (2009).

[2] There are also adjectives in which the gender distinction does not show at all, e.g. *rosa* 'pink', *gammaldags* 'old-fashioned'.

noen 'any.M', as expected from a masculine. This is the background for the use of curly brackets in (1).

In Oslo these days, there is no longer any evidence from 'associated words' in favour of a separate feminine gender. In other words, the feminine agreement has been ousted by the old masculine. The old suffix *-a*, by contrast, is retained. The system, at least for most of the speakers, is as described in (2):

(2) Two genders in recent Oslo dialect (compare example 1; again, examples given in Bokmål)

    a. en liten   gutt, en fin   gutt, denne gutten,   ikke noen gutt
       a.M small.M boy, a.M fine.M boy, this.M boy.DEF.SG.{M} not any.M boy

    b. en liten   stol, en fin   stol, denne stolen,   ikke noen stol
       a.M small.M chair, a.M fine.M chair, this.M chair.DEF.SG.{M} not any.M chair

    c. en liten   jente, en fin   jente, denne jenta,   ikke noen jente
       a.M small.M girl, a.M fine.M girl, this.M girl.DEF.SG.{?} not any.M girl

    d. en liten   jakke, en fin jakke, denne jakka,   ikke noen jakke
       a.M small.M jacket, a.M fine jacket, this.M jacket.DEF.SG.{?} not any.M jacket

    e. et lite   barn, et fint   barn, dette barnet,   ikke noe barn
       a.N small.N child, a.N fine.N. child, this.N child.DEF.SG.{N} not any.N child

    f. et lite   hus, et fint hus, dette   huset,   ikke noe
       a.N small.N house, a.N fine.N house, this.NEUT house.DEF.SG.{N} not any.N
       hus
       house

The usual interpretation of the data in (2), as indicated by the glossing, is that the old feminine is no longer a separate gender in the Oslo dialect, 'merely' an inflection class (Lødrup 2011, cf. also Enger 2004a,c and many others).[3] The definite singular suffix *-a* might seem 'the last redoubt' of the old feminine, cf. (2c-d), and some would like to analyse it is a gender marker (cf. Section 3.2.3 below); that is the reason for using "{?}".

A development from gender to inflection class is far from unique; such developments have been referred to as grammaticalisation (cf. Lehmann 1982, 2016, Wurzel 1986). The old feminine is changing into an inflection class also in some other Norwegian dialects, such as Tromsø (Westergaard & Rodina 2015, 2016), and it is absent also in some contact varieties in the North (Conzett et al. 2011). Essentially the same development is found in the Jämtland dialect in Sweden (Van Epps & Carling 2017).[4,5]

---

[3]There is considerable discussion about whether to take pronouns into consideration for the purposes of gender agreement. At this stage, they are left out, for expository reasons (but cf. Section 4.2 below).

[4]On the whole, it is pointless to debate whether dialects in Scandinavia are dialects of one or the other language, since Scandinavia generally counts as one dialect continuum. The point of interest is the parallel between Jämtland and Oslo.

[5]A next step after the system shown in (2) is that also the old *-a* suffix is lost. In that way, old masculines and old feminines become indistinguishable. This is found with some Oslo speakers, who will say *en liten jakke, jakken,* just like *en liten gutt, gutten.* (Essentially the same system is found in "standard" Swedish and Danish.)

2.2 Istro-Romanian

We now turn to Istro-Romanian, which is "spoken in some localities in north-eastern Istria (Croatia) to the south of Mt Učka, and in the town of Žejane to its north. Its speakers probably descend from pastoral communities originally resident in Bosnia, Serbia, and Croatia in the late Middle Ages, who settled in Istria from about the fifteenth century. The language's place of origin, and whether it originally broke away from varieties spoken in the Romanian lands, or from those spoken in the Balkans, or represents dialect mixing, remain controversial. There are today perhaps 200-250 speakers in Croatia, mainly elderly and all bilingual in Croatian" (Maiden 2016b: 91).

The number of genders in Istro-Romanian might be disputed. The system used to be essentially the same as that of Romanian, and the number of genders in Romanian has been much disputed (cf. Corbett 1991, Maiden 2016a,d, Loporcaro 2016). Besides the masculine and the feminine, which are uncontroversial, there is also, at least according to Corbett (1991) and Loporcaro (2016), a third gender. This gender has been referred to as 'neuter' and as 'genus alternans'. This gender has practically no morphology of its own, as Table 1 shows.

Table 1: Romanian gender.

Singular	Plural	
trandafir **frumos**	*trandafiri frumoși*	(beautiful rose, M)
casa frumoasă	*case frumo**ase***	(beautiful house, F)
palton **frumos**	*paltoane frumo**ase***	(beautiful coat, N)

Singular with definite article		Plural with definite article	
*pom - pom**ul***	(tree - the tree, M)	*pomii*	(the trees)
cutie - cutia	(box - the box, F)	*cutii**le***	(the boxes)
*loc - loc**ul***	(place - the place, N)	*locuri**le***	(the places)

The 'neuter' patterns with the masculine in the singular, with the feminine in the plural. Thus, it alternates between the two, hence the label *genus alternans*. In Table 1, some endings have been boldfaced so as to show this. According to Martin Maiden (personal communication, and 2016c), in Istro-Romanian, while the masculine and the feminine happily persist,

> The plural endings which originally selected feminine gender (alternating with masculine singulars) have lost the alternating gender and the relevant nouns have become masculine in singular and plural alike, ***except*** that they may continue to have a ***distinctively feminine*** definite article (suffixed, as in Norwegian) … this could indicate that the definite article is in a rather different category from other agreeing elements, at least when it is enclitic to the noun (Martin Maiden, e-mail).

The different status of the 'definite article', when it is 'inside' the noun (word-internal), is indeed a central theme of this paper.

2.3 Clitic or suffix?

It is necessary to address the status of the 'definite article', in both Istro-Romanian and Norwegian. Traditional wisdom has it that the Romanian 'definite article' is a clitic, but Ledgeway (2016a,b) has argued that it is not a syntactic 'head' at all, but rather a piece of inflectional morphology, expressing definiteness. Apparently, the Romanian definite article shows many of the characteristics of inflection, such as fusion, obligatoriness, defectiveness and erratic allomorphy. This conclusion carries over to Istro-Romanian.

The Norwegian 'definite article' has traditionally been analysed as a suffix, but some would analyse it as a clitic (e.g. Lahiri et al. 2005). However, Lødrup (2016) presents good arguments for the traditional suffix analysis (cf. also Faarlund 2009): There are unexpected 'gaps' in the inflection in the indefinite singular. Nouns that do not have to take a definiteness suffix, even when they quite clearly occur in the definite, and these nouns do not form a natural class. Consider first (4a,b):

(3) Gutten   er i byen   og sjekker kneet
     Boy.DEF.SG.{M} is in town-DEF.SG.{M} and checks knee-DEF.SG.{N}

     'The boy is in town getting his knee checked'

A corresponding sentence without the definiteness suffixes, as in (4), would be strange:

(4)   * Gutt er i by og sjekker kne

Intriguingly, if the words for 'boy', 'town' and 'knee' are replaced with the words for 'dean [of a faculty at a university]', 'city centre' and 'larynx', grammaticality judgments would be the opposite, as (4c,d) show:

(5) a. Dekanus_ er i   sentrum_ og sjekker larynks_
         Dean   is in centre   checking larynx

         'The dean is in the [city] centre getting his larynx checked'

     b. * Dekanusen   er i sentrumet   og sjekker larynksen
           Dean.DEF.SG.{M} is in centre.DEF.SG.{N} checking larynx.DEF.SG.{M}

Thus, there are 'gaps' in the marking of definiteness, and that does not square with clitic status. Some (mainly learned) nouns denoting (mainly) people and body parts do not take the definite article – but these nouns do not make up a natural class, as Lødrup (2016) shows. In other words, not all learned nouns behave like *dekanus, sentrum, larynx*, and not all nouns that can behave like *dekanus* are learned, Latinate nouns. Compare (6):

(6) a. Dekanus har foreslått at …
         'Dean has suggested that …'

b. * Diakon/ Diakonen har foreslått
   'Deacon has suggested'

c. * Leder/ Lederen har foreslått
   'Chief has suggested'

d. Avdelingsleder/Avdelingslederen har foreslått
   'Head of section has suggested'

The noun *diakon* 'deacon' is a clear loan, but it behaves like *gutt* 'boy' and not like *dekanus* 'dean', cf. (6b). Conversely, there is nothing Latinate over the word *avdelingsleder* 'head of section', which still can behave like *dekanus*, cf. (6d) (and contrasts intriguingly with the simplex *leder*, cf. (6c)).

One might add other arguments for taking the 'article' as a suffix, including the observation that the 'definite article' is restricted to one word-class, and that it cannot be skipped on co-ordinated nouns, cf. (7a), thus differing from the 'possessive' *-s*, usually considered a clitic, cf. (7b):

(7) a. gutten og faren – not *gutt og faren
       'the boy and the father'

    b. fars og mors – far og mors
       'father's and mother's

Also, at least for some Oslo speakers, the stem vowel of the one noun 'mother', *mor* is changed from the indefinite /mu:r/ to the definite /mura/, and that is unexpected under a clitic analysis, whereas inflectional suffixes can induce irregularity.[6]

2.4 Parallels in support

The diachronic parallel between Oslo and Istria is interesting. In both cases, a 'word-internal' element is where traces of the feminine stay on the longest. In Oslo, *-a* lingers on as a suffix long after agreeing words such as *lita* 'little.F', *noa* 'some.F' and even *ei* 'a.F' have been lost. In Istria, the suffix is the last relic of the old genus alternans. The parallel is close enough to warrant further examination, and the reason is probably structural; contact can safely be ruled out. Some other innovations in Scandinavian may be noted in support.

2.4.1 Danish

For a couple of centuries, Standard Danish has had a two-gender system, with an opposition between masculine (or common gender, a merger of the former feminine and masculine) and neuter (cf. Section 2 and Footnote 5). Historically speaking, the Danish

---

[6]Some readers may wonder if the change in stem vowel quantity for 'mother' might be some kind of compensatory lengthening, which might be analysed as phonologically rather than morphologically triggered. This seems unlikely, as the example is isolated.

system has influenced the Oslo development, although the change in Oslo is probably not due to contact only (Enger 2004c).

In current Danish, the mass nouns *vodka* 'vodka', *cement* 'cement' are usually masculine (as are their cognates in Norwegian). However, alongside the expected masculine determiner *den*, as in *den vodka* 'the.ᴍ vodka', *den cement* 'the.ᴍ concrete', Danish also allows for *det vodka* 'the.ɴ vodka', *det cement* 'the.ɴ concrete' with neuter agreement on the attributive determiner. These nouns thus allow for alternative agreement patterns; they have become hybrids, in Corbett's (1991, 2006) terminology. The neuter agreement in *det vodka, det cement* has been called semantic agreement (Hansen & Heltoft 2011: 232, Enger 2013).[7]

On this point, Danish goes further than its Scandinavian sister languages/dialects (cf. also Josefsson 2014b). Danish, Norwegian and Swedish allow 'pancake sentences', in which there is neuter agreement on the predicative adjective, even if the subject appears to have another feature. Consider example (8):

(8)   Vodka   (det)   er godt
       Vodka(m) (it.ɴᴇᴜᴛ) is good.ɴᴇᴜᴛ.sɢ

At least according to one analysis (e.g. Enger 2004b, Wechsler 2013, Haugen & Enger forthcoming), pancake sentences can be considered semantic (or 'referential') agreement.[8]

The same nouns, e.g. *vodka, sement* (Norwegian spelling)/ *cement* (Swedish and Danish spelling) can take a neuter pronoun in Swedish, Norwegian and Danish, and they can take a predicative adjective in the neuter, as in (8). However, Swedish and Norwegian do not allow *det vodka; in other words, they do not allow semantic agreement inside the NP in such examples. Danish allows *det vodka* 'that.ɴᴇᴜᴛ vodka', *det cement* 'that.ɴᴇᴜᴛ concrete' with semantic agreement, but even in Danish, only *cementen* 'concrete.ᴅᴇꜰ.sɢ{ᴍ}', *vodkaen* 'vodka.ᴅᴇꜰ.sɢ{ᴍ}' with the suffix associated with the masculine is accepted. In other words, also in Danish, **cementet,* *vodkaet is ruled out; the possibility of semantic agreement (neuter) found on the attributive determiner has not (yet?) spread to the suffix. Thus, the suffix is again more resistant against diachronic change than other, more word-like elements.

At this stage, a caveat is in order. I have used the terms 'pronoun' and 'determiner', but words that can be used pronominally in Norwegian can typically also be used as determiners, compare, for example the two uses of *det* in (9):

(9)   a. Hva   synes   du om det   huset?
           What think.ᴘʀs you of   that.ɴᴇᴜᴛ house.ᴅᴇꜰ.sɢ{ɴᴇᴜᴛ}?

           'What do you think of that house?'

---

[7]The terms 'hybrid noun' and 'semantic agreement' and 'referential agreement' have been debated (cf. Dahl 1999, Corbett 2006), but for present purposes, we may set this aside.

[8]For further discussion of pancake sentences, see e.g. Corbett & Fedden (2016), Enger (2013), Josefsson (2009, 2014a), Haugen & Enger (2014).

b. Det   er   fint
   It.NEUT be.PRS fine.NEUT
   'It is fine'

Thus, it is far from obvious that there is a categorical split between pronouns and determiners (Kristoffersen 2000, Halmøy 2016: 162-3 *et passim*, see also Hansen & Heltoft 2011: 183 for Danish), and in this paper, the terms 'pronoun' and 'determiner' refer to use only.

2.4.2 A peripheral change in (some) Norwegian Bokmål

Norwegian Bokmål presents many examples of a slightly different, but related kind (see also Enger & Corbett 2012, Enger 2015). Here, a new semantically motivated feminine gender agreement is found, formerly not available, as in the examples in (10a, 10b) (from the web):

(10)   a. Ei god venn som alltid   er der
          a.F good friend who always is there.

          'a good friend who is always there'

       b. B. har fått ei lærer   som …og hun …
          B. has got a.F teacher who …and she …

          'B. has got a teacher who … and she …'

The nouns *venn* 'friend', *lærer* 'teacher' are masculines in traditional three-gender systems, so one would expect the determiner *en*. Since the masculine is ousting the feminine, in many dialects (cf. Section 2 above), one would not expect the opposite to happen as well; it is strange to see the feminine *ei* spread. So a natural reaction may be to dismiss examples such as (10a, 10b) as wrong.

However, data like these do occur, if not terribly frequently (even in the speech of some, although I have only anecdotal evidence on this point), and the examples are not random. They relate to nouns denoting humans, and whenever the feminine is employed, it refers to females. The data therefore deserve to be taken seriously, and their immediate interest is that while the article/determiner can be changed, from *en venn* to *ei venn*, from *en lærer* to *ei lærer*, the suffix is not changed accordingly. The same two authors that produced *ei venn* and *ei lærer*, write *vennen* 'friend.DEF.SG.{M}', *læreren* 'teacher.DEF.SG.{M}' (and not *venna, *lærera) respectively, even if reference clearly is made to a woman. (See further Section 4.1 below.)

So even if these nouns change the attributive determiner *en* to *ei*, they do not change the suffix *-en* to *-a*. Again, the suffix is more resistant towards change than the other elements, which, unlike the suffix, are independent words.

3 Suggested analysis

3.1 The original Agreement Hierarchy

The similarities surveyed in Section 2 are probably not accidental, and one way ahead is to relate them to the Agreement Hierarchy (Corbett 1979, 2006). This hierarchy involves four 'pegs' for four different kinds of agreement controllers, as shown in Figure 1.

Attributive > Predicative > Relative > Personal Pronoun
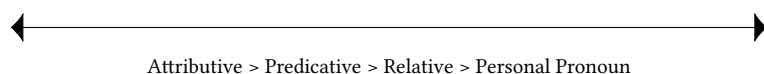
Figure 1: The Agreement Hierarchy.

Corbett (2006: 207) says that for "any controller that permits alternative agreements, as we move rightwards along the Agreement Hierarchy, the likelihood of agreement with greater semantic justification will increase monotonically". In other words: The possibility for semantic agreement will increase towards the right; if possible on the predicative, it will be possible on the personal pronoun too, but not necessarily the other way around. A case in point is the agreement patterns noted for some Scandinavian mass nouns (Section 2.4). Given that Danish allows semantic agreement on the attributive determiner (*det vodka*), semantic agreement is expected also on the predicative. In standard Swedish, semantic agreement is possible on the predicative; so, semantic agreement is expected also on personal pronouns, but it is no problem that semantic agreement is outlawed on the determiner.

While Corbett's hierarchy was originally formulated as a synchronic constraint, it "can easily be adapted to the diachronic perspective, predicting gender exponents to begin and/or complete the transition from lexical [syntactic] to referential [semantic] assignment the earlier, the further they are located on the right of the implicational hierarchy", as noted by Dolberg (2014: 55).

3.2 The revised Agreement Hierarchy

3.2.1 Suggestion and background

The suggestion now is to modify the hierarchy, at least for some purposes, by expanding it with an additional position or 'peg', which is 'word-internal', cf. Figure 2.

'Word-Internal' > Attributive > Predicative > Relative > Personal Pronoun
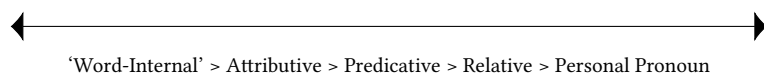
Figure 2: Modified Agreement Hierarchy.

The idea is that the Agreement Hierarchy has to do with 'tightness' of grammatical relations, and thus with grammaticalisation, and that grammatical relations generally are tighter inside the word than inside the phrase, and tighter inside the phrase than

outside it, – and across clauses weaker still. The idea that the Agreement Hierarchy may have to do with grammaticalisation is far from original (cf. Lehmann 1982, 2016), but it has not received quite the attention it merits (though see Jobin 2004).

When suggesting the hierarchy, Corbett (1979: 217) noted that it did not match then-current syntactic frameworks too well, and suggested that it was an "independent feature of natural languages". Nearly forty years later, this suggestion seems less appealing. As Dolberg (2014: 58) notes, from a diachronic perspective, Corbett's Agreement Hierarchy "is to be credited with being of remarkable predictive accuracy, yet it does not yield much in the way of explanatory power: even though it reliably tells us what to expect to happen in the exponents of changing gender systems, it provides little information regarding why this is so."

It would if the Agreement Hierarchy could be grounded in something else. In recent years, many linguists have come to see constraints "not so much as constraints on possible synchronic grammars [than, HOE] as constraints on diachronic developments" (Timberlake 2003: 194, cf. also e.g. Evans & Levinson 2009). On such a view, at least some of the explanatory burden is shifted from synchrony towards diachrony.

According to Lehmann (1982, 2015 and elsewhere), there is a unidirectional movement from semantic agreement towards syntactic agreement, but not vice versa. In other words, what starts out as semantic agreement may become 'syntacticised' and less meaningful; changes in the other direction should not occur. Becoming somehow 'semantically reduced' is a standard criterion for grammaticalisation, another is becoming more obligatory. Both criteria would seem to hold for 'syntactic' agreement compared to semantic; Wechsler (2009) even prefers the term 'grammatical' agreement. This fits with the broad picture of grammaticalisation; it is largely unidirectional. On the assumption that diachronic tendencies motivate the Agreement Hierarchy, the hierarchy can be related to a larger framework, viz. that of grammaticalisation.

3.2.2 Objection I: motivating the fifth peg

The fifth peg may seem like cheating, for two reasons. Firstly, 'word-internal (or noun-internal) agreement' is a controversial notion.[9] The other 'pegs' are syntactic heads; the suffix in Norwegian is morphology (cf. Section 2.3), and the idea of 'morphology-free syntax' is well-established (Zwicky 1992, Corbett 2014). Secondly, merely positing a fifth peg does not automatically solve the problem; the new peg does require some kind of motivation. As the Agreement Hierarchy has already been linked to grammaticalisation (Section 3.2.1), the latter problem will be discussed first.

There are different versions around of the Agreement Hierarchy. Köpcke et al. (2010) try to make their version less system-internal and more functional. In the words of Dolberg (2014: 18), they "assign pragmatic functions to the syntactic categories identified by Corbett, resulting in this altered agreement hierarchy: specifying – modifying – predicating – referent-tracking". Dolberg (2014: 58) argues that it makes sense to consider this version of the hierarchy together with Corbett's original:

---

[9]While Stolz (2007) argues at length in favour of the notion of word-internal agreement, the point I am trying to make here is orthogonal to his.

> [M]otivating this expected pathway of referential agreement encroaching into (predominantly) lexical gender systems is comparably straightforward in the functional version of the Agreement Hierarchy [Köpcke et al. 2010], simply by taking recourse to the basic surmise that changes will occur generally first in those areas, in which the change is most conducive and/or least detrimental to language use. Thus, the underlying assumption of the functional version of the Agreement Hierarchy is that personal pronouns changing to referential gender yield the largest gain in freeing cognitive capacity, as their lexical gender needs no longer be remembered over comparably long stretches of discourse, because the appropriate pronoun form is now simply being derived from attributes of the referent, or, more precisely, the interlocutor's mental representation thereof, which needs to be kept in working memory anyway. This putative gain then gradually diminishes the further one moves to the left in the Hierarchy. (Dolberg 2014: 58)

Relating the Agreement Hierarchy to grammaticalisation (cf. Section 3.2.1) means relating it to the 'tightness' of grammatical relations; one of Lehmann's (2015: 131) 'parameters' of grammaticalisation is bondedness or 'tightness': "The cohesion of a sign with other signs in a syntagm will be called its bondedness; this is the degree to which it depends on, or attaches to, such other signs." Lehmann (2015: 157) says the syntagmatic cohesion or bondedness of a sign "is the intimacy with which it is connected with another sign to which it bears a syntagmatic relation".

The relation between a noun and an attributive adjective is tighter, more "intimate", than that between a noun and a predicative adjective, which is in turn tighter than that between a noun and a pronoun. Elements in attributive position are inside the noun phrase, and the syntax of the phrase is, as a rule, tighter than that of the clause and sentence. The relation between a pronoun and its antecedent is typically 'loose', compared with that of determiner to noun, hence, semantic agreement is more characteristic of pronouns. A related 'parameter' for Lehmann (2015: 131) is that of syntagmatic variability; the possibility of 'shifting around' a sign in its construction. This also fits with the Agreement Hierarchy, and the relation between noun and suffix is tighter than any of the relations in Corbett's original hierarchy. The suffix has to occur immediately to the right of the noun stem; nothing else can intervene.

This fits with the suggestions made by Köpcke et al. (2010) and Dolberg (2014). Pronouns are unlikely to be 'stored' in the mental lexicon together with their controlling noun, and this opens for semantic agreement. By contrast, it seems likely that suffixes are stored with their controller, as some idioms show. Two set phrases in Norwegian are *få sparken* 'get the sack, be fired' and *gi sparken* 'sack, fire'. The verbs *få* and *gi* mean 'get, receive' and 'give' respectively, and they are both very general and frequent, but the noun *sparken* only rarely occurs outside these two idioms; it is difficult to ascribe a meaning to *sparken* in isolation. There is no indefinite singular; there are no plurals. Even if the suffix indicates a masculine noun, there is no noun phrase **en spark*.[10] If the whole *få sparken* were stored, that would weaken the case for saying that only stem and

---

[10]Strictly speaking, there is a noun *en spark* 'kicksled, spark', but it is a homonym, synchronically.

suffix are stored together, but *sparken* can marginally be found on its own, cf. examples from the web in (11):

(11) Examples of *sparken* without *få*:

    a. *Facebook betyr ikke sparken* 'Facebook does not [have to] mean the sack'

    b. *dermed ble det sparken* 'lit. thereby became it sack; so I was sacked'

Similar examples include *snurten*, which it hardly makes sense to translate in isolation; it is mostly known from the idiom *se ikke snurten av* 'not see anything/the least bit of'. This noun does occur marginally in some other contexts, though, even without negation, cf. (12), again, examples are taken from the web:

(12) Examples of *snurten* without *ikke* (and without *av*):

    a. *aldri sett snurten av* 'never seen anything of'

    b. *uten å se snurten til* 'without seeing anything of'

    c. *… kan man skimte snurten av peisen* '… can one spot a little of the fireplace'

Scandinavian diachrony presents at least one example where the definite singular suffix has become part of the stem. This is the noun meaning 'world'. Swedish has *värld*, Danish has *verden* (cf. def. sg. *världen* vs. *verdenen*). The Danish cognate is an innovation; the old def.sg. suffix has become part of the stem. Pragmatically, this makes sense; for most speakers, there is only one world (at least most of the time). Istro-Romanian also presents examples where the plural 'definite article' has become lexicalised (Maiden 2016c). It is difficult to think of an example where the pronoun would merge with the stem in the same way, also because pronouns do not typically occur next to a noun (as they occur 'instead of a noun').

It is more difficult to come up with examples in which the determiner must be stored than where the suffix must, but there are some. The phrase *ikke det spøtt* means 'not the least', and one might expect the noun *spøtt* to inflect as a regular neuter would. Yet at least in my Norwegian, there is no definite singular form, nor any plurals. For *spøtt*, then, it seems the determiner is stored with the noun.[11] An obvious question is if *ikke* 'not' also has to be stored, but *aldri sett det spøtt* 'never seen no nothing' shows it does not have to.

It probably does not happen often that the pronoun is stored together with the noun; this probably happens more often with the determiner. It seems even more likely that suffixes be stored with the corresponding noun (also because suffixes are 'salient', cf. Section 3.2.3 below).[12]

In Section 3.2.1, we considered an argument in favour of seeing the Agreement Hierarchy in terms of grammaticalisation having to do with 'semantic reduction'. According to

---

[11]Admittedly, dictionaries also mention *et spøtt*. But that is unknown to many speakers, and dictionaries tend to strive for completeness, sometimes at the expense of actual usage.

[12]The suggestion that determiner or affix may be stored together with the noun does not exclude the idea that generalisations may be made over the gender or inflection class of a noun (cf. e.g. Conzett 2006).

Heine (2003: 583), semantic reduction is the central factor behind grammaticalisation. It is helpful to think of semantic reduction in terms of reduction of uncertainty (entropy). The less surprising X is, the less is its information value. Consider now the examples in (13):

(13)   Pronoun and determiner in use

    a. Bilen   står   framfor   huset.   Den er faktisk
        Car.DEF.SG.{M} is (lit. stands) in front of house.DEF.SG.{N}. It.M is actually
        rosa.
        pink.
        'The car is in front of the house. It – i.e. the car – is actually pink.'

    b. Bilen   står   framfor   huset.   Det er faktisk
        Car.DEF.SG.{M} is (lit. stands) in front of house.DEF.SG.{N}. It.N is actually
        rosa.
        pink.
        'The car is in front of the house. It – i.e. the house – is actually pink.'

    c. Den   bilen   som står   framfor   huset,   er
        The.{M} car.DEF.SG.M that is (lit. stands) in front of house.DEF.SG.{N} is
        faktisk   rosa.
        actually pink

Recall from Section 2.4.1 that Norwegian pronouns can typically also be used as determiners. In (13a, 13b), *den* contrasts with *det*. In (13c), *den* does not contrast with *det*, since *det bilen is ungrammatical. In other words, the first *den* tells us the speaker is talking about the car, the last *den* merely tells us that a masculine or feminine will follow (and that it is a definite, specific example). Thus, the information value of *den* is higher when used pronominally than when used determinatively. Another argument in the same direction would be that the first (personal pronoun) *den* can be stressed, but the last (determiner) *den* cannot. This indicates that in general, the attributive determiner has a lower information value than the personal pronouns. The suffix has an even lower information value than the determiner (cf. Dahl 2015: 123). (Recall that the suffix is also even more 'bonded', which is one of Lehmann's 2015: 131 parameters for grammaticalisation.)

3.2.3 Objection II: Agreement between parts of words?

Patching suffixes on to the Agreement Hierarchy may seem a bad idea on theoretical grounds; this might at first glance seem tantamount to denying the claim that syntax is morphology-free (Zwicky 1992, Corbett 2014: 38f). This is a large issue which cannot be discussed in detail here, but the lexeme, the line between syntax and morphology, has not been handed down on tablets of stone; there are 'troubles with lexemes', as argued by Fradin & Kerleroux (2003), Haspelmath (2011) and many others. A very influential

adherent of lexeme-based models, Matthews (1991: 100), even says "it is often the mark of a genuine unit, like the lexeme, that we have trouble with it!"[13]

There has been some debate over whether the Norwegian definite singular suffix should be taken as a marker of gender or of inflection class (cf. 2.1), and this also relates to the problem of the delimitation morphology–syntax. Åfarli & Lohndal (2015) argue that the suffix -*a* should count as a marker of gender (and not 'only' of inflection class), also in the recent Oslo system described in example 2. Åfarli & Lohndal are not worried about violating lexicalist doctrines, and that is surely fair enough, given their theoretical stand; yet it remains too open, in my view, what the consequences will be: many things normally not included as 'gender' will then have to fall under that label (many inflection classes, for instance). From the opposite side of the spectrum, Lødrup (2011) squarely rejects analysing -*a* as a gender marker, as it is not an 'associated word'. An in-between course is suggested by Enger (2004a), who discusses a system like that in example (1):

> If genders are defined only on the basis of word-external agreement, it seems dubious to treat the definite singular suffix as an exponent of gender. However, one may wonder if there is any reason for speakers not to consider the definite singular suffix a gender marker, given that the correlation with gender is perfect. In other words, it seems perverse to deny that the definite singular suffix is an exponent of gender, **when there is one and only one definite singular suffix associated with each gender** [emphasis added here]. [...] even if what determines gender contrasts is what patterns show up on the target (and not on the controller), affix contrasts that show up on the controller and that correspond to gender contrasts on targets have to be considered markers of gender as well. (Enger 2004a: 65)

This means taking the definite sg. suffix as an exponent of gender in the classical Oslo dialect (1), but not in the present-day one (2), since the suffix did correlate with gender then, but does not do so now. A possible defence of taking *some* suffixes into consideration is that agreement evidence is less salient; considering agreement evidence requires more subtle reasoning (cf. also Carstairs-McCarthy 1994: 766).[14] There is interesting psycholinguistic evidence that Norwegian children acquire the suffixes for the definite singular much earlier than the gender in agreeing words (e.g. Westergaard & Rodina 2015, 2016) .

However, once the Agreement Hierarchy is seen as a product of other factors, it may become a bit less pressing whether, say, in an example such as *gutten min* 'boy.DEF.SG{M} my.M', the relation between *gutt* 'boy' and *min* 'my' and that between *gutt* and -*en* should both be subsumed under 'agreement'. Corbett (e.g. 2006) has presented strong arguments

---

[13]Maiden (2016d) argues, on the basis of an impressive set of data taken from dialects and diachrony, that Romanian "nouns showing *genus alternans* are not a class defined by the agreement behaviour of associated words, but **a class the agreement behaviour of whose associated words is dictated by inflexional morphology** [boldface mine, HOE]". The implications are intriguing. Yet Maiden's analysis has also been criticised (by Loporcaro 2016). Anyway, the subject of 'morphology-free syntax' is too large for this paper.

[14]Wurzel (1986) even suggested that, in general, exponents on the word itself should count.

in favour of including pronouns under the label of agreement: There are important similarities between pronouns and other elements in the hierarchy, so that drawing a line at any one specific point at the hierarchy will entail an arbitrary choice and the loss of worthwhile generalisations. By the same token, I suggest there are some worthwhile generalisations to be made by including *some* suffixes under the scope of the Agreement Hierarchy. Theories should be about opening doors, not about closing them. The only reason not to include these suffixes would be substantial empirical evidence showing that they behave very differently from the predictions of the hierarchy.[15]

In *gutten min*, both *min* and *-en* convey information about *gutt*. The notion of 'intra-morphological meaning' can be useful and productive here (e.g. Carstairs-McCarthy 1994, Maiden 2005, Enger 2004a); the notion that an element of a word may 'signal' say, a particular property of the stem. In (1), *-a* has intra-morphological meaning, signalling the noun's inflection class and its gender. This does not mean that *-a* is an 'associated word', only that it gives information about gender. In (2), *-a* also carries intra-morphological meaning, but now signalling inflection class only, because there is now no gender agreement related to it.

4   The danger of drawing too sharp lines

4.1   Automatisation

Lehmann (1982) drew a sharp line between NP-internal and NP-external agreement. One of Corbett's (2006) arguments against this is that there can be referential/semantic agreement also inside the NP, and Danish *det vodka* and Norwegian *ei lærer* (cf. Section 2.4) support Corbett's view. Perhaps paradoxically, if Lehmann is right in arguing that agreement has to do with grammaticalisation (cf. Section 3.2.1), then it is to be expected that Corbett should be right in not drawing a sharp line. Grammaticalisation tends to be a gradual affair; I see no reason why it should come to a complete halt exactly at the NP.

As noted, a development from (feminine) gender to inflection class may be described as grammaticalisation (cf. Section 2). Grammaticalisation may in turn be related to automatisation, according to Lehmann (2016).[16] He sees inflectional classes as more 'automatised' than genders, and he says one almost has to be a linguist to wilfully produce the wrong allophone of a phoneme or to choose the wrong inflectional suffix. Pronominal gender is at the other end of the spectrum. It is for pronouns that there is most 'leeway'. They are the least 'automatised'. This perspective fits the one adopted here.

However, under certain circumstances, even inflection class suffixes can be manipulated consciously, and not only by linguists. When looking for examples like *ei lærer* (Section 2.4.2, Enger 2015), I found (in a net forum for 'nurse jokes') *ei søt sykepleier* 'a.ꜰ

---

[15]Thanks to Florian Dolberg for pointing this out to me.

[16]There are many suggestions in the literature that are similar to that of Lehmann. Boye & Harder (2012) relate grammaticalisation to 'backgrounding'; automatisation and backgrounding are related. Bybee (2003) relates grammaticalisation to 'chunking'; her explanation of this concept makes it quite clear that automation is relevant here too. Haiman (1994) links grammaticalisation to ritualization and repetition. Lehmann (2016) does not address the relation between his suggestion and these others.

cute.MF nurse'. Now, in Norwegian Bokmål, *en søt sykepleier* 'a.M cute.MF nurse', with masculine determiner *en*, is the only conventional choice. In writing *ei søt sykepleier*, the author emphasises that the nurse is a woman. Another author on the same net forum reacted to the wording in an interesting way. Rather than criticise the choice of *ei* directly, he lists a part of the paradigm, the way it is taught to school-children, and then comments (my translation and editing) in (14):

(14) ei sykepleier, sykepleiera?
       'Where did you learn your Norwegian?'

  This is an argument *ad absurdum*: if you say A (*ei sykepleier*), then B (*sykepleiera*) follows, and given that B (*sykepleiera*) is absurd, A (*ei sykepleier*) must be rejected. For present purposes, the point of interest is B: Using the old feminine suffix is apparently even worse than the use of feminine determiner. In short, even if the suffix is extremely automatised, it can be manipulated and changed.

4.2 Pronouns

4.2.1 A problem for the present approach?

Lehmann (1982, 2016) is not the only linguist who has wished to draw a sharp line between NP-internal agreement and pronominal agreement. So far, pronouns have been kept out of the picture, but they are worth including. In the Oslo dialect today, there are four pronouns. Consider (15).

(15) Pronouns in the current Oslo dialect
       a. *gutten*.M (the boy) – *han* 'he'
       b. *jenta*.{?} (the girl) – *hun* 'she'
       c. *låven*.M (the barn) / *jakka*.{?} (the jacket) – *den* 'it.NON-NEUT'
       d. *barnet*.N (the child) – *det* 'it.NEUT'

  The choice of pronoun relates to animacy. The pronouns *han, hun* are used with animates (males and females respectively), *den, det* with non-animates (*den* with non-neuters, *det* with neuters). Animacy does not generally play a role for gender agreement inside the NP in Scandinavian (though cf. Enger 2013: 286–289). Pronoun agreement and noun-phrase-internal agreement thus follow partly different rules in this system, as in Danish and Swedish. Therefore, some conclude that pronouns are not subject to gender agreement (e.g. Josefsson 2009, 2014a). An alternative view is that pronouns should be included under gender (e.g. Corbett 2006, Enger 2013, Dolberg 2014, Haugen & Enger 2014, Van Epps & Carling 2017).

  Once pronouns are taken into account, it may seem that the modified Agreement Hierarchy gets into trouble: It might seem as if the feminine in Oslo now is retained in the very extremes of the hierarchy, viz. the pronominal peg and the suffix peg, and not

in-between. On closer inspection, however, this is not so. As noted, the Agreement Hierarchy predicts that a new gender system, if semantically based, will start from the right end of the hierarchy and the old system will stay on the longest at the very left end. The word *hun* in (13) indicates a human – or a higher animal – of female sex. That is not the intra-morphological meaning of -*a* (cf. Section 3.2.3). While the intra-morphological meaning of -*a* can be roughly given as 'the stem to my left belongs to a particular inflection class, including words as *jakke* 'jacket' and many others', the meaning of *hun* is roughly 'the noun to my left denotes a person of female sex'.[17]

4.2.2 A problem for another approach

In their Swedish grammar, Holmes & Hinchliffe (2013: 4) say that "Nouns ending in -a [in the indefinite sg., thus ending in -an in the definite sg., HOE] which denote animals are often treated as feminine irrespective of their true gender [i.e. biological sex, HOE]: *råttan – hon* the rat – she, *åsnan – hon* the donkey – she".

This observation is interesting, as it represents a problem for an important approach to Scandinavian gender. According to Josefsson (2009: 40, 2014a), lexical gender, which is found within the DP, does not carry any meaning. By contrast, gender is a meaningful category in the pronominal domain. Thus, Josefsson's approach implies a sharp boundary between pronominal agreement, which is meaningful, and DP-internal agreement, which is not. However, if we wish to explain why Swedish *råttan* 'the rat' and *åsnan* 'the donkey' are more often referred to with *hon* than, say, *musen* 'the mouse' and *hästen* 'the horse', we are stuck with the fact that the former end in −*a* in the indefinite singular [*råtta, åsna*], the latter do not [*mus, häst*]. Yet 'ending in an -*a* in the indefinite singular' is hardly a meaningful property. (See Haugen & Enger forthcoming, for a summary of other arguments against Josefsson's approach, and further references.)

5 Conclusions

I have pointed out a parallel between Oslo Norwegian and Istro-Romanian. In both cases, the 'last redoubt' of the old feminine is a suffix on the noun. The parallel is not coincidental; there are other Scandinavian examples (cf. Section 2.4) indicating that the noun's suffix is more 'resistant' towards change than are 'associated words'. The difference can relate to a somewhat modified version of the Agreement Hierarchy (Corbett 1979, 2006, Köpcke et al. 2010), in which an extra 'peg' is added for the suffix. This modification is in line with the spirit of Fradin & Kerleroux (2003); they also note 'troubles with lexemes', but they do not use those problems as arguments against the lexeme as such. Rather than getting stuck in such problems, we may, for example, utilise the handy concept of intra-morphological meaning (Section 3.2.3). Following Lehmann (1982), I have argued that relating the Agreement Hierarchy to grammaticalisation may be useful, at least for some purposes.

---

[17]The example also illustrates 'semantic reduction', cf. Section 3.2.2.

Acknowledgments

References

Åfarli, Tor A. & Terje Lohndal. 2015. Genus vs. Bøyingsklasse? Talk and ppt at MONS (Møte om norsk språk) 16, Kristiansand, November 2015.

Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88. 1–44.

Bybee, Joan. 2003. Mechanisms of change in grammaticization. In Richard D. Janda & Brian D. Joseph (eds.), *Handbook of historical linguistics*, 602–624. Oxford: Blackwell.

Carstairs-McCarthy, Andrew. 1994. Inflection classes, gender, and the principle of contrast. *Language* 70(4). 737–788.

Conzett, Philipp. 2006. Gender assignment and the structure of the lexicon. *Sprachtypologie und Universalienforschung* 59. 223–240.

Conzett, Philipp, Hilde Sollid & Åse Mette Johansen. 2011. Genus og substantivbøying i nordnorske språkkontaktområder. *Nordand: nordisk tidsskrift for andrespråksforskning* 6. 35–71.

Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15. 202–224.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words. *Language* 83(1). 8–42.

Corbett, Greville G. 2014. Features in inflection. In Matthew Baerman (ed.), *The Oxford handbook of inflection*, 35–47. Oxford: Oxford University Press.

Corbett, Greville G. & Sebastian Fedden. 2016. Canonical gender. *Journal of Linguistics* 52. 495–531.

Dahl, Östen. 1999. Animacy and the notion of gender. In Barbara Unterbeck & Matti Rissanen (eds.), 99–115. Berlin: Mouton de Gruyter.

Dahl, Östen. 2015. *Grammaticalization in the north : Noun phrase morphosyntax in Scandinavian vernaculars*. Language Science Press.

Dolberg, Florian. 2014. *Gender variation, change and loss in Medieval English: Evidence from the Anglo-Saxon chronicle*. University of Hamburg dissertation.

Enger, Hans-Olav. 2004a. On the relation between gender and declension. *Studies in Language* 28(1). 51–82.

Enger, Hans-Olav. 2004b. Scandinavian pancake sentences as semantic agreement. *Nordic Journal of Linguistics* 27(1). 5–34.

Enger, Hans-Olav. 2004c. Tre endringer i det skandinaviske genussystemet i lys av grammatikaliseringsteori. *Arkiv för nordisk filologi* 119. 125–147.

Enger, Hans-Olav. 2013. Scandinavian pancake sentences revisited. *Nordic Journal of Linguistics* 36. 275–301.

Enger, Hans-Olav. 2015. When friends and teachers become hybrids. In Jürg Fleischer, Elisabeth Rieken & Paul Widmer (eds.), *Agreement from a diachronic perspective*, 215–235. Berlin: Mouton de Gruyter.

Enger, Hans-Olav & Greville G. Corbett. 2012. Definiteness, gender, and hybrids: Evidence from Norwegian dialects. *Journal of Germanic Linguistics* 24. 287–324.

Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals. *Behavioral and Brain Sciences* 32. 429–448.

Faarlund, Jan Terje. 2009. On the history of definiteness marking in Scandinavian. *Journal of Linguistics* 45. 617–639.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Haiman, John. 1994. Ritualization and the development of language. In William Pagliuca (ed.), *Perspectives on grammaticalization*, 3–28. Amsterdam: John Benjamins.

Halmøy, Madeleine. 2016. *The Norwegian nominal system: A Neo-Saussurean perspective*. Berlin: Mouton de Gruyter.

Hansen, Erik & Lars Heltoft. 2011. *Grammatik over det Danske Sprog I*. Copenhagen: Det danske Sprog- og Litteraturselskab/Syddansk Universitetsforlag.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80.

Haugen, Tor Arne & Hans-Olav Enger. 2014. Scandinavian pancake constructions as a family of constructions. *Cognitive Linguistic Studies* 1. 171–196.

Haugen, Tor Arne & Hans-Olav Enger. Forthcoming. The semantics of Scandinavian pancake constructions. *Linguistics*.

Heine, Bernd. 2003. Grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 575–601. Oxford: Blackwell.

Holmes, Phil & Ian Hinchliffe. 2013. *Swedish: A comprehensive grammar*. 3rd edition. Abingdon: Routledge.

Jobin, Bettina. 2004. *Genus im Wandel. Studien zu Genus und Animatizität*. Stockholm: Almqvist & Wiksell.

Josefsson, Gunlög. 2009. Peas and pancakes. On apparent disagreement and (null) light verbs in Swedish. *Nordic Journal of Linguistics* 32. 35–72.

Josefsson, Gunlög. 2014a. Pancake sentences and the semanticization of formal gender in Mainland Scandinavian. *Language Sciences* 43. 62–76.

Josefsson, Gunlög. 2014b. Scandinavian gender and pancake sentences: A reply to Hans-Olav Enger. *Nordic Journal of Linguistics* 37. 431–449.

Köpcke, Klaus-Michael, Klaus-Uwe Panther & David Zubin. 2010. Motivating grammatical and conceptual gender agreement in German. In Hans-Jörg Schmid & Susanne Handl (eds.), *Cognitive foundations of linguistic usage patterns*, 171–194. Berlin: Mouton de Gruyter.

Kristoffersen, Kristian E. 2000. Ordklassane pronomen og determinativ – kor gode er argumenta for å skilje dei i norsk? *Maal og Minne* 2000. 181–194.

Lahiri, Aditi, Allison Wetterlin & Elisabet Jönsson-Steiner. 2005. Lexical specification of tone in North Germanic. *Nordic Journal of Linguistics* 28. 61–96.

Larsen, Amund B. 1907. *Kristiania bymaal*. Kristiania [=Oslo]: Cammermeyer.

Ledgeway, Adam. 2016a. Minunile limbii române: Lecţii pentru romanişti. Talk given at the conference "Studiile Româneşti din Lume – Perspective Actuale" 29/09 – 1/10 2016 Universitatea din Bucureşti.

Ledgeway, Adam. 2016b. The Romanian definite article in a comparative Romance perspective. Manuscript, Cambridge.

Lehmann, Christian. 1982. Universal and typological aspects of agreement. In Hansjakob Seiler & Franz Josef Stachowiak (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen, Teil II: Die Techniken und ihr Zusammenhang in Einzelsprachen*, 201–268. Tübingen: Gunter Narr Verlag.

Lehmann, Christian. 2015. *Thoughts on grammaticalization*. 3rd edition. Berlin: Language Science Press.

Lehmann, Christian. 2016. Grammaticalization and automation. Talk presented at the 23. LIPP-Symposium 'Grammatikalisierung in interdisziplinärer Perspektive'. Internationale Konferenz der Graduiertenschule Sprache & Literatur München, 6.–8. July.

Lødrup, Helge. 2011. Hvor mange genus er det i Oslo-dialekten? *Maal og Minne* 2011. 120–136.

Lødrup, Helge. 2016. Dekanus er i sentrum og sjekker larynks: Lærde substantiver med definitt tolking i norsk. *Maal og Minne* 2016. 31–50.

Loporcaro, Michele. 2016. Noun inflection and gender in Romanian. Talk at the 17th International Morphology Meeting, Vienna.

Maiden, Martin. 2005. Morphological autonomy and diachrony. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2004*, 137–175. Dordrecht: Kluwer Academic Publishers.

Maiden, Martin. 2016a. Inflexional morphology and gender in Romanian. Why Romanian only has two genders, and how it might yet acquire a third. Talk at the 17th International Morphology Meeting, Vienna.

Maiden, Martin. 2016b. Romanian, Istro-Romanian, Megleno-Romanian, and Aromanian. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 91–125. Oxford: Oxford University Press.

Maiden, Martin. 2016c. The death of the genus alternans in Istro-Romanian. Manuscript, Oxford.

Maiden, Martin. 2016d. The Romanian alternating gender in diachrony and synchrony. *Folia Linguistica Historica* 37. 111–144.

Matthews, P. H. 1991. *Morphology*. 2nd edition. Cambridge: Cambridge University Press.

Opsahl, Toril. 2009. *''egentlig alle kan bidra'': En samling sosiolingvistiske studier av strukturelle trekk ved norsk i multietniske ungdomsmiljøer i Oslo.* University of Oslo dissertation.

Stolz, Thomas. 2007. Word-internal agreement. *Sprachtypologie und Universalienforschung* 60. 219–251.

Timberlake, Alan. 2003. Review of Corbett: Number and Senft: Systems of nominal classification. *Journal of Linguistics* 39. 189–195.

Van Epps, Briana & Gerd Carling. 2017. From three genders to two: The sociolinguistics of gender shift in the Jämtlandic dialect of Sweden. *Acta Linguistica Hafniensia* 49. 53–84.

Wechsler, Stephen. 2009. Agreement features. *Language and Linguistics Compass* 3. 384–405.

Wechsler, Stephen. 2013. The structure of Swedish *Pancakes.* In Philip Hofmeister & Elisabeth Norcliffe (eds.), *The Core and the Periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 71–98. Stanford: CSLI Publications.

Westergaard, Marit & Yulia Rodina. 2015. Grammatical gender in Norwegian: Language acquisition and language change. *Journal of Germanic Linguistics* 27. 147–189.

Westergaard, Marit & Yulia Rodina. 2016. Hvor mange genus er det i Tromsø-dialekten? *Maal og Minne* 2016. 159–189.

Wurzel, Wolfgang Ullrich. 1986. Die wiederholte Klassifikation von Substantiven: Zur Entstehung von Deklinationsklassen. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39. 76–96.

Zwicky, Arnold M. 1992. Some choices in the theory of morphology. In Robert D. Levine (ed.), *Formal grammar. Theory and implementation*, 327–371. Oxford: Oxford University Press.

Chapter 11

The Haitian Creole copula and types of predication: A Word-and-Pattern account

Alain Kihm

CNRS, Université Paris-Diderot

Haitian Creole is a French-based creole language spoken by about 10 millions people in Haiti. In Haitian Creole the copula consists in the two forms *se* and *ye* and it may not be expressed. The present paper argues that, despite claims to the contrary, the Haitian Creole copula is a verbal lexeme realized through two overt suppletive stems and a phonologically null stem. Selecting one stem or the other does not depend on inherent and/or contextual inflectional features as in English *am* vs. *is* vs. *was* vs. *were*, but on the syntax and semantics of the predicate headed by the copula lexeme.

1 Introduction

In Haitian Creole (HC), a French-based creole spoken by about 10 millions people in Haiti, the copula is expressed via two overt forms *se* and *ye* and it may also not be expressed. Various studies, most of them couched in syntactic transformational terms, have been devoted to this variation (Valdman 1978, Damoiseau 1985, DeGraff 1992, Kihm 1993, Déprez & Vinet 1997, Déprez 2003). The main debate centred around the issue of whether the two overt forms are verbs (e.g. Valdman 1978, Kihm 1993) or pronouns (DeGraff 1992) or both (Déprez 2003).

Here I will try to support the four following assumptions: (i) the Haitian Creole copula is a verb throughout; (ii) the two overt forms are word forms in the sense of Matthews (1972), realizing alternative suppletive stems of the copular lexeme; (iii) the lexeme also includes a null stem, devoid of phonological substance; (iv) selecting one stem or the other (including the null stem) does not depend on inherent and contextual inflectional features as is often the case (cf. English *am* vs. *is* vs. *was* vs. *were*, *go* vs. *went*), but on the syntax and semantics of the predicate headed by a given form of the lexeme.

The Haitian Creole stem alternation thus differs not only from the English instances just mentioned, but also from cases where the phonological shape of an item merely

depends on the syntactic environment, i.e. on what the item appears next to. Zwicky (1985, 1990) gives several examples, such as the French singular possessive determiners which take on the masculine form when preceding a feminine item beginning with a vowel: e.g. *mon ombrelle* 'my sunshade', not **ma ombrelle* (cf. *une ombrelle* 'a sunshade'). Yet, as argued by Zwicky (1985), it wouldn't make sense to assume that the gender feature common to both components of the NP [*mon ombrelle*] is not the same as in e.g. *ma maison* 'my house'. What is in fact needed to account for such an apparent mismatch is a rule of referral stipulating that the shape — but not the content — of feminine singular possessive determiners is identical to that of masculine singular possessive determiners just in the case that the adjacent word begins with a vowel. (For rules of referral also see Stump 2001: 36–37) And note that the adjacent word need not be the head noun: cf. *mon ancienne maison* 'my old house'.

In Haitian Creole, in contrast, inserting *se* or *ye* or nothing audible depends not on the shape of what follows, but it is related to the lexical category of the complement to some extent and, more importantly, to the semantics of the predication type. The *ser*/*estar* alternation in Portuguese and Spanish may provide an analogue (Mateus et al. 1989: 98–102), except for the fact that *ser* and *estar* are likelier to represent two distinct lexemes than distinct stems of the same lexeme as in Haitian Creole. In the latter, as we shall see, the equivalent of the *ser*/*estar* contrast is the *se* vs. nothing contrast. Now, it is not detrimental to parsimony to assume a null stem of a given lexeme, provided it belongs to a paradigm whose other members are all overt forms, so that the content of the null form can be unambiguously retrieved thanks to contrast with the overt forms' contents (see Sag et al. 2003 on the copula in African-American Vernacular English). Lexemes devoid of phonological realization would be much harder to justify, in contrast. Moreover the conditions on *ye*'s insertion find no equivalent in the *ser*/*estar* alternation, while supporting the suppletive stem hypothesis.

What I am proposing, therefore, is a fully lexicalist account which accounts for most of the facts and avoids the unnecessary complexities and implausible assumptions of the previous syntactic treatments. First I review the facts. Then I show how these facts can be accounted for by assuming one copular lexeme, the lexical entry of which mentions several stems, each of which identifies a particular lexical entry of type *word*, whose valence and semantics are subsets of the valence and semantics of the lexeme. Collocations of these words with tense-mode-aspect (TMA) markers are realized via realization rules written in an Information-based Morphology (IbM) format (Crysmann & Bonami 2015).In the conclusion, I point out what remains, to my mind, in need of an account and I suggest some lines of research that might lead to a fuller understanding of the Haitian Creole copula, especially from a diachronic viewpoint.

2 The facts of the HC copula

Part of the Haitian Creole copula's paradigm can be retrieved from the following examples (Déprez 2003: 135, 136, 139; Fattier 2013: 201) :

(1) Jan se yon pwofesè.
    John COP INDF teacher

    'John is a teacher.'

(2) Jan chapantyè.
    John carpenter

    'John is a carpenter.'

(3) Jan malad.
    John sick

    'John is sick.'

(4) Jan nan lekol la.
    John in school DEF

    'John is at school.'

(5) Elifèt te anba tab la.
    E. PST under table DEF

    'Elifèt was under the table.'

(6) Se frè mwen Jan ye.
    COP brother 1SG John COP

    'It is my brother that John is.'

As mentioned above, three forms come out from these examples:, (i) *se* in (1) and (6), obviously from French *c'est* /sɛ/ 'it is'; (ii) the null form in (2)–(5); (iii) *ye* in (6), from French *est* /ɛ/ 'is' or *i(l) est* /jɛ/ 'he is'.

Let us first compare (1), where the copula is realized as *se*, with (2) where it is not realized at all. The difference seems to lie in the syntactic category of the complement, an NP in (1) and a NOM in (2) (Sag & Wasow 1999: 84). And note that *chapantyè* in (2) can be modified by an attributive adjective: e.g. *Jan bon chapantyè* 'John is a good carpenter'.

The crucial difference, however, actually resides in the individual-level (permanent, identificational) character of the property predicated by means of *se*, in the present case being a professor (Carlson 1977, Diesing 1988, Chierchia 1995, Kratzer 1995). *Se*'s complements need not be indefinite NPs involving the indefinite determiner *yon* 'a' as in (1). Whenever the complement denotes some obviously permanent quality of the subject, determination can be dispensed with. See for instance the following extract from a poem by Bonel Auguste (Chalmers et al. 2015: 20), where being man's limit is presented as a defining property of man's dream:

(7) Rèv lòm se limit lòm.
    dream man COP limit man

    'Man's dream is man's limit.' (*Le rêve de l'homme est la limite de l'homme*)

Despite the absence of the definite articles one sees in the French translation, *limit lòm* is a definite NP in (7) by virtue of being a genitive construction whose complement *lòm* is

itself definite as it refers to the maximal set of human beings (see Lyons 1999:181–184 on "class generics"; Huddleston & Pullum 2002:407; Kihm 2003). Bare nouns (i.e. NOMs) are also acceptable under the same conditions as in *Mari se fanm* 'Mary is a woman' (Glaude 2012), alternating with the almost synonymous *Mari se yon fanm*. In French as well, in a somewhat literary register, *Marie est femme* is an acceptable alternative to *Marie est une femme*.

Given this, (2) appears to be ambiguous, in the sense that being a carpenter may be viewed as a permanent, individual-level quality of John, or as just a stage-level description of what John is at the time the sentence is uttered. Nouns denoting professions or trades typically trigger that kind of ambiguity, always allowing for referentially equivalent predicates with or without *se*. (For similar facts in French, see Kupferman 1979, Boone 1987)

The individual- vs. stage-level contrast can also be made manifest in adjective predicates. Contrary to the received idea that Haitian Creole adjectives are in fact stative verbs that never need a copula, Damoiseau (1996) demonstrates on the basis of a corpus study that for more than half of the items (including *malad*) adjective predicates without an overt copula as in (3) imply a stage-level interpretation, while the same with *se* as in *Jan se malad* are understood as predicating an individual-level property of the subject (also see Pompilius 1976). This is patently shown by the distinct clefting strategies implied by either possibility. Clefting stage-level predications (no overt copula) is done by way of "doubling" as in (8) (Déprez 2003: 146):

(8) Se damou Jan damou.
     COP in.love John in.love

     'John ɪs in love.'

Compare *Se manje Jan manje* {COP eat J. eat} 'John did eat'. Clefted individual-level predications (involving *se*), in contrast, are like (6). See (9) (Damoiseau 1996: 157):

(9) Se grangou   li ye.
     COP unscrupulous 3SG COP

     'S/he ɪs unscrupulous.'

Interestingly *grangou* also has the stage-level meaning 'hungry', in which case clefting employs the same strategy as for *damou* 'in love' in (8): *Se grangou Jan grangou* 'John ɪs hungry'.

Example (4) shows the copula is not realized when the complement is a PP. However, not all PP complements behave alike: PP complements, locative or not, predicating a potentially permanent situation require *se* as shown in (10) and (11) (Déprez 2003: 141–142):

(10) Tout sa se pou ou.
     all this COP for 2SG

     'All this is for you.'

(11) M   pa   te   di   ou   vi   mwen se   nan navigasyon.
      1SG NEG PST tell 2SG life 1SG   COP in   navigation

      'I did not tell you my life is in navigation.'

The descriptive generalization therefore seems to be that the copula is realized as *se* before a noun, adjective or prepositional phrase denoting a potentially individual-level property of the subject, while it has no exponent when the denoted property is potentially stage-level. I hedge this statement with "potentially" because it seems to be rare that being viewed as a stage or individual-level property does not to some extent depend on the intentionality of the speaker rather than being entirely anchored in the ontology of the property itself.

In (5) one might wonder whether *te* is not actually the past form of the copula. Two considerations oppose this supposition. First, complementary data show *te* to be a past tense marker (a 'particlexeme' in Zwicky's 1990 terminology) that may combine with other undisputable TMA markers. See the following examples from Fattier (2013: 199, 201):

(12) Li   te   gen   twa   zoranj.
      3SG PST have three orange

      'S/he had three oranges.'

(13) Li   t(e) ap   boukanen mayi.
      3SG PST PROG roast   maize

      'S/he was roasting maize.'

Yet, there still might exist two homophonous *te*, one a past marker, the other the copula's past form. Actually, such an assumption would have history on its side, since *te* obviously comes from the French imperfect *était* 'was' and/or the past participle *été* 'been' and the TMA sequence in (13) can be traced back to the obsolete and/or dialectal French past progressive periphrase *était après* or *(a) été après*.

Synchronically, however, there is good reason not to regard *te* as the past copula, namely that transposing (6) into the past gives us *Se frè mwen Jan te ye* 'It's my brother that John was', not **Se frè mwen Jan te*, as we would expect if *te* was the past copula. I will therefore assume that the past tense marker *te* in (5) "precedes" (if one may say so) the same null form of the copula as is evidenced in (2)–(4).

Example (6) illustrates both the use of *se* in clefts and the copula's third form *ye*. Let us begin with the latter. Its peculiarity is to require a gap to its immediate right. The gap, the foot of a long distance dependency (LDD) (Sag et al. 2003), may be part of a cleft as in (6) or of a WH-construction as in (14) from a poem by André Fouad (Chalmers et al. 2015: 62):

(14) Di   m   kijan lavi te   ye.
      tell 1SG how   life PST COP

      'tell me how life was.' (*dis-moi comment était la vie.*)

Note it wouldn't do to simply state that *ye* must be followed by nothing (meaning an utterance-final pause). Something may indeed occur after it, provided it is not a complement, but rather dislocated material as in (15) (Tessonneau 1980: 18) or an adjunct as in (16) (Déprez 2003: 148):

(15) Sa   l' ye nèg la ki marye avè fi a?
     what 3SG COP man DEF REL marry with girl DEF

     'What is he, the man who married the girl?'

(16) Nonm nan te pi   gran m te ye lè   sa   a.
     man   DEF PST more big   1SG PST COP time DEM DEF

     'The man was bigger than I was at that time.'

Conceivably *ye*'s immediate follower in (16) is a gap whose filler is *gran* 'big'. Note that *ye* is neutral as to the stage vs. individual-level contrast. This is expected since *ye* only occurs in clauses involving LDDs, whose neutral, declarative or noncomparative counterparts may involve either type of predication: e.g. the answer to (15) might be *Nèg la ki marye avè fi a **se** yon pwofesè* 'The man who married the girl is a professor', while a possible non-comparative counterpart of (16) would be *Nonm nan gran* 'The man (is) big'.

As mentioned, the fact that initial *se* in (6) lacks a subject has led some authors to cast doubt on its verbal character (DeGraff 1992) or to define it as an "introducer" — whatever that may be — distinct from copular *se* (see discussion in Valdman 1978).Yet, null subjects do exist in Haitian Creole as shown by the following two examples (Déprez 1992a:24; Déprez 1992b:198):

(17) Rete   yon nèg nan kay   la.
     remain one man in   house DEF

     'There remains one man in the house.'

(18) Sanble Mari renmen Jan.
     seem   Mary love   John

     'It seems that Mary loves John.'

Such unrealized subjects correspond to expletive subjects in languages like English or French where nullity is disallowed: compare *Il reste un homme dans la maison*, *Il semble que Marie aime Jean*. But note that in 17th century French *sembler* and *rester* could be used without expletive *il* in sentences quite similar to (17) and (18) (Haase 1935: 15–16). The null subject of *se* in (6) and in such sentences as *Se vre* {COP true} 'It's true' (French *C'est vrai*) falls under this generalization. Although *se*'s initial /s/ obviously originates in the French neutral pronoun *ce* of *c'est* 'it is', this is highly unlikely ever to have had any relevance in the fully emerged Creole — that is since the end of the 18th century — where *se* has become an unanalysable item, contrary to what I argued in Kihm (1993). I therefore conclude that *se* is a verbal copula across the board, and it belongs to the small

set of verbs that allow expletive null subjects, a feature to be mentioned in its lexical definition.

Se presents still other properties. First, contrary to what the examples so far may suggest, it is not limited to third person. See (19) from a poem by Solèy (Chalmers et al. 2015: 22) where its subject is the clitic form *m* of *mwen* 'I, me', occurring with all verbs (cf. *m pati* 'I left'):

(19) M se espas nan mitan de pyebwa.
      1SG COP space in middle two tree

      'I am the space between two trees.' (*je suis l'espace entre deux arbres*)

And see (16), which shows that *ye*, like *se*, is compatible with all person-number values.

An intriguing property of *se* is its position vis-à-vis TMA markers and the negator, as illustrated in the three following examples (Glaude 2012: 39; Valdman 1978: 240; Cavé in Chalmers et al. 2015: 46):

(20) Jan se pa te papa w.
      John COP NEG PST father 2SG

      'John wasn't your father.'

(21) Sa se va yon gwo nouvèl.
      that COP FUT INDF great news

      'That will be great news.'

(22) Se tap   yon tan pèdi.
      COP PST.PROG INDF time lose

      'It would be time lost.' (*Ce serait une perte de temps*)

As shown by (20) the grammatical order is *se* < NEG < TMA, whereas it is NEG < TMA < V with all other verbs, including *ye* (cf. 14). Examples (20)–(22) suggest that all simple or complex TMA markers are admissible with *se*. However, not all native speakers accept *se va* and *se ap*.[1]

Another peculiarity of *se* is that the possibility of its being preceded by all subject pronouns gets drastically reduced whenever it combines with TMA markers and/or the negation. The pronoun is then obligatorily 3SG, it is left-dislocated and only the emphatic form *li-mèm* may be used. See the following contrast (Déprez 2003: 151):

(23) * Li se te zanmi mwen.
        3SG COP pst friend 1SG

      Intended: 'S/he was my friend.'

(24) Li-mèm, se te zanmi mwen.
      3SG-self COP PST friend 1SG

      'S/he was my friend.'

---

[1] I am grateful to Jean Noël Whig for these judgments.

The same ungrammaticality affects **Li se pa zanmi mwen* contrasting with *Li-mèm, se pa zanmi mwen* 'S/he isn't my friend' and **Ou(-mèm) se (pa) te zanmi mwen*, whose grammatical alternative is *Ou (pa) te zanmi mwen* 'You were (not) my friend', using the null form of the copula. In (24) the subject of *se* is therefore the null subject bearing 3SG as its only possible value.

Déprez (2003: 151) relates the ungrammaticality of **Ou(-mèm) se…* to that of French **Toi, c'est/c'était mon ami* next to *Elle/lui, c'est/c'était mon ami(e)*. There certainly is truth in this parallel. Yet it does not account for the well-formedness of *Ou se zanmi mwen* 'You are my friend' or *Jan se zanmi mwen* 'John is my friend'. In fact, it seems to be a true generalization that *se* modified by TMA markers and/or the negation only selects for the null subject, so that *Jan* in (20) is actually left-dislocated as is *li-mèm* in (24) and as is *Jean* in the French equivalent *Jean, c'est/c'était mon ami*. This — as it is not so obvious as with pronouns — has to be checked with careful prosodic analyses.

Another noteworthy fact is the neutralization of the stage- vs. individual-level contrast with non-third person subjects and inflected *se*, since *Ou (pa) te malad* 'You were (not) sick' is the only negative and/or past counterpart of the positive present contrasting pair *Ou malad* 'You're sick' and *Ou se malad* 'You're a sick person'.

Finally, it is worthwhile noting that *se* may be elided as *s'* before *yon* 'a' yielding the portmanteau /sɔ̃/. See the following lines by Solèy (Chalmers et al. 2015: 22):

(25) Labote / s'   on   zwazo benyen an san.
      beauty COP INDF bird bath   in blood
      'beauty / is a bird bathed in blood.' (*la beauté / est un oiseau ensanglanté*)

This confirms, if need be, that *se* is unanalysable as a single word despite its etymology.

As for the null form, it is compatible with all TMA markers and the negator, as shown by (5) as well as by (26) (Glaude 2012: 49) and (27) (DeGraff 2007: 114):

(26) Jan ap   doktè.
      John PROG doctor

      'John will be a doctor.'

(27) Duvalye pa   prezidan Ayiti.
      Duvallier NEG president Haiti

      'Duvallier isn't the president of Haiti.'

As Glaude points out, (26) cannot mean 'John is being a doctor', quite normally in fact: interpreting the progressive as a future is a general possibility, and the only one with stative verbs (Fattier 2013). The positive counterpart of (27) is *Duvalye prezidan Ayiti* 'Duvallier is the president of Haiti', whereas the negative of the also acceptable *Duvalye se prezidan Ayiti* is *Duvalye, se pa prezidan Ayiti* (see above).

3 A formal account of the Haitian Creole copula

In this section I will only try to account for the clearest facts as exemplified in (1)–(6). What I leave aside for future research will be set out in the conclusion.

As stated in the introduction, I assume the Haitian Creole copula to be one verbal lexeme realized as three stems, one null, selected according to predication type. This lexeme can be represented as the lexical entry below:

(28)

$$
\begin{bmatrix}
\textit{copv-lexm} \\
\text{LID} \quad \textit{cop} \\
\text{SYN} \begin{bmatrix}
\text{HEAD} \begin{bmatrix} \text{PRED} + \end{bmatrix} \\
\text{VAL} \begin{bmatrix}
\text{SPR} & \boxed{1}\langle \text{NP} \mid \textit{null} \rangle \\
\text{COMPS} & \boxed{2}\langle \text{NP} \mid \text{NOM} \mid \text{PP} \mid \text{ADJP} \mid \text{ADV} \mid \textit{gap} \rangle \\
\text{ARG-ST} & \boxed{1} + \boxed{2}
\end{bmatrix}
\end{bmatrix} \\
\text{FORM} \begin{bmatrix} \text{STEM} & \langle \boxed{A}, \boxed{B}, \boxed{C} \rangle \end{bmatrix} \\
\text{SEM} \begin{bmatrix}
\text{MODE} & \textit{prop} \\
\text{INDEX} & s \\
\text{RESTR} \begin{bmatrix}
\text{RLN} & \textit{cop} \\
\text{SIT} & s \\
\text{SBJ} & i \\
\text{PRED} & j \textit{ pred stlev} \mid \textit{indlev}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

That is to say, the Haitian Creole copula is a predicator whose valence includes (i) a specifier that is a possibly unrealized NP; (ii) a complement that may be an NP, a NOM, a PP, an adjective phrase, an adverb (e.g. *Se konsa* 'It's so'), or a gap. Recall that NOM is the label for noun phrases unspecified for (in)definiteness, such as *chapantyè* in (2).

Let me also point out that Haitian Creole personal pronouns are best analysed as members of the NP category. There seems to be no good reason, in particular, to view their reduced forms (see Table 1) as anything but phonological clitics, since (i) reduced and unreduced forms alternate without change of meaning; (ii) sequences of reduced forms and TMA markers or verbs do not give rise to any particular phonological phenomena as is the case with English contracted auxiliaries (Bender & Sag 2000). For instance, 3sg *li* may but need not reduce to *l* when preceding a vowel-initial verb or TMA marker, e.g. *l ap chante* ~ *li ap chante* 's/he/it is singing' (but *li /*l chante* 's/he/it sang'); similarly in object position following a vowel-final verb, e.g. *yo wè li* ~ *yo wè l* 'they saw her/him/it' (but *yo bat li/*l* 'they struck her/him/it'). The crucial factors seem to be register and speed of delivery.

Expressions headed by the copula are propositions about some situations and they are semantically restricted to predicating stage-level (*stlev*) or individual-level (*indlev*) properties of a given subject. This has to be specified, since it conditions the choice of

Table 1: Haitian Creole personal pronouns

	sg	pl
1	mwen / m	nou / n
2	ou / w	nou / n
3	li / l	yo / y

the proper stem among the three stems that realize the copula, tagged A (the null stem), B (*se*), and C (*ye*) according to degrees of nondefaultness.

The syntactic environment calling for the null stem (A) is summed up in (29):

(29) Jan (pa) (te) (bon) chapantyè / malad (anpil) / nan lekol la / konsa.
John (NEG) (PST) (good) carpenter sick (very) in school DEF so
'John is/was (not) a (good) carpenter/(very) sick/at school/so.'

That is to say, the copula's null stem is required if (i) the subject is an NP; (ii) the complement is a NOM, or an ADJP, or a PP, or an adverb; (iii) the denoted property is viewed as being transitory, that is of the stage-level sort. Whatever the complement, the copula may be negated and/or specified for some TMA value.

The question now is to relate the copula's stems to the syntactic and semantic properties calling for one or the other. Since (28) describes the lexeme labelled COP, each of the stems may be viewed as realizing a word-form of the lexeme, each word-form with its own lexical entry. The A stem is thus assigned the following lexical entry, where the phonological form is represented as the empty list, and the valence and semantics are subsets of the lexeme's valence and semantics:

(30)
$$
\begin{bmatrix}
\textit{verb word} \\
\text{LID} \quad \textit{cop} \\
\text{PHON} \quad \langle\,\rangle \\
\text{SYN} \begin{bmatrix}
\text{HEAD} \quad [\text{PRED } +] \\
\text{VAL} \begin{bmatrix}
\text{SPR} \quad \boxed{1}\langle\text{NP}\rangle \\
\text{COMPS} \quad \boxed{2}\langle\text{NOM} \mid \text{PP} \mid \text{ADJP} \mid \text{ADV}\rangle \\
\text{ARG-ST} \quad \boxed{1}+\boxed{2}
\end{bmatrix}
\end{bmatrix} \\
\text{SEM} \begin{bmatrix}
\text{MODE} \quad \textit{prop} \\
\text{INDEX} \quad s \\
\text{RESTR} \begin{bmatrix}
\text{RLN} \quad \textit{cop} \\
\text{SIT} \quad s \\
\text{SBJ} \quad i \\
\text{PRED} \quad \textit{j pred stlev}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

Suppose now we want to account for the predicate *te bon chapantyè* 'was a good carpenter' (French *était bon charpentier*). Following Bonami (2015), I assume Haitian Creole collocations such as *te chante* 'sang, used to sing' to be periphrases, that is multiword morphological units involving an ancillary and a main element, in which the former is a marker instead of a verb as in the English periphrase *has sung*. (See Van Eynde 1994 and Sag 2012 for the relevant notion of marker as a non-head element selecting a head and assigning it features.) The only difference between *te chante* and the case at hand is that the main verb's stem has no phonology associated with it. Hence the following realization rule for the collocation of the past marker *te* with the null stem of the copula, using Information-based Morphology formalism (Crysmann & Bonami 2015):

$$
(31) \quad
\begin{bmatrix}
\textit{mword} \\
\text{PHON} \quad \langle te \rangle \\
\text{MPH} \quad \langle \boxed{1}\begin{bmatrix} \text{PH} & \langle te \rangle \\ \text{PC} & 1 \end{bmatrix}, \boxed{2}\begin{bmatrix} \text{PH} & \langle \ \rangle \\ \text{PC} & 1 \end{bmatrix} \rangle \\
\text{MS} \quad \langle \boxed{3}[\text{TMA } \textit{pst}], \boxed{A}[\text{LID } \textit{cop}] \rangle \\
\text{RR1} \quad \begin{bmatrix} \text{MUD} & \boxed{3}[\text{TMA } \textit{pst}] \\ \text{MPH} & \boxed{1}\begin{bmatrix} \text{PH} & \langle te \rangle \\ \text{PC} & 1 \end{bmatrix} \\ \text{RS} & [\ ] \end{bmatrix} \\
\text{RR2} \quad \begin{bmatrix} \text{MUD} & \boxed{A}[\text{LID } \textit{cop}] \\ \text{MPH} & \boxed{2}\begin{bmatrix} \text{PH} & \langle \ \rangle \\ \text{PC} & 1 \end{bmatrix} \\ \text{RS} & [\ ] \end{bmatrix}
\end{bmatrix}
$$

Rule (31) realizes a multiword (*mword*) comprising the marker *te* and the null copula tagged A pointing to the relevant word-form and stem. Owing to this tagging we ensure that /te ⟨ ⟩/ will be inserted in the right syntactic and semantic contexts.

Note the reverse selection (RS) feature is given no value in (31). The function of this feature is to ensure that, in periphrases such as *has sung*, the main verb's form (e.g. the past participle) stands in the context of the ancillary item that requires it (e.g. *have*). In Haitian Creole, however, the form of the main verb never depends on the marker in collocation with which it assumes a given TMA value. Being a word, on the other hand, *te* includes a COMPS feature [VFORM *finite*] in its lexical entry.

In the morphophonological (MPH) tier of the rule, the phonological (PH) form ⟨te⟩ and the null stem are assigned the same position class (PC) 1. This is in order to avoid the awkward statement that *te* "precedes" something that is actually not there. From a morphophonological viewpoint, we may therefore consider *te* in *te bon chapantyè* a portmanteau word amalgamating the marker and the null stem, somewhat similar to French *du* for ⟨de le⟩.

Rule (31) will also account — *mutatis mutandis* — for the collocations *ap* ⟨ ⟩ and *pa* ⟨ ⟩ of (24) and (25).

Let us now tackle *se*. The syntactic environments calling for it are not so easy to sum up in one example. At least three are necessary, discounting for the moment the issue of the position of TMA markers and the negator:

(32) Mari se yon (bon) profesè / fanm   / sè   ou / malad.
      Mary COP INDF (good) teacher   woman   sister 2SG   sick

      'Mary is a (good) teacher / a woman/ your sister / a sick person.'

(33) Se vre / konsa / yon lòt   bagay.
      COP true so   INDF other thing

      'It's true / so / another thing.'

(34) Vi mwen se nan   navigasyon.
      life 1SG COP in navigation

      'My life is in navigation.'

Se is thus shown to be required when (i) the subject is an NP as in (32) and (34) or is null as in (33); (ii) the complement is an NP as in (32), or a NOM whose head clearly denotes some permanent quality such as being a woman, or an adjective phrase denoting an individual-level property as in (32) and (33), or a PP with the same type of denotation as in (34), or an adverb such as *konsa* in (33). Owing to questions about its valence, I leave aside *se* in clefts such as (6), although I'm confident it can be shown to represent the same lexeme as *se* in the other contexts. The lexical entry for the *se* word-form of the copula is therefore (35):

(35)

$$
\begin{bmatrix}
\textit{verb word} \\
\text{LID} \quad \boxed{B}\ \textit{cop} \\
\text{PHON} \quad \langle se \rangle \\
\text{SYN}
\begin{bmatrix}
\text{HEAD} \quad [\text{PRED} +] \\
\text{VAL}
\begin{bmatrix}
\text{SPR} \quad \boxed{1}\langle \text{NP} \mid \text{NOM} \rangle \\
\text{COMPS} \quad \boxed{2}\langle \text{NOM} \mid \text{PP} \mid \text{ADJP} \mid \text{ADV} \rangle \\
\text{ARG-ST} \quad \boxed{1}+\boxed{2}
\end{bmatrix}
\end{bmatrix} \\
\text{SEM}
\begin{bmatrix}
\text{MODE} \quad \textit{prop} \\
\text{INDEX} \quad s \\
\text{RESTR}
\begin{bmatrix}
\text{RLN} \quad \textit{cop} \\
\text{SIT} \quad s \\
\text{SBJ} \quad i \\
\text{PRED} \quad j\ \textit{pred indlev}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

I assume the present tense reference of *se* in examples (32)–(34) is a corollary of its not being modified by any TMA marker, so that there is no question of a "zero" marker. Hence the following realization rule for *se* in, for instance, (32) with *yon chapantyè* as a complement:

$$
(36)
\begin{bmatrix}
\textit{mword} \\
\text{PHON} \quad \langle se \rangle \\
\text{MPH} \quad \left\langle \boxed{1} \begin{bmatrix} \text{PH} & \langle se \rangle \\ \text{PC} & 1 \end{bmatrix} \right\rangle \\
\text{MS} \quad \langle \boxed{2}\, [\text{TMA } \textit{prs}], \boxed{B}\, [\text{LID } \textit{cop}] \rangle \\
\text{RR1} \quad \begin{bmatrix} \text{MUD} & \boxed{2}\, [\text{TMA } \textit{prs}] \\ \text{MPH} & \boxed{1} \begin{bmatrix} \text{PH} & \langle se \rangle \\ \text{PC} & 1 \end{bmatrix} \\ \text{RS} & [\,] \end{bmatrix} \\
\text{RR2} \quad \begin{bmatrix} \text{MUD} & \boxed{B}\, [\text{LID } \textit{cop}] \\ \text{MPH} & \boxed{2} \begin{bmatrix} \text{PH} & \langle se \rangle \\ \text{PC} & 1 \end{bmatrix} \\ \text{RS} & [\,] \end{bmatrix}
\end{bmatrix}
$$

In accordance with the "paradigmatic" view of TMA retrieval, [TMA *prs*] and the stem's realization are assigned the same phonology and position class.

What about the position of TMA markers and the negator as illustrated in (20)–(22)? Considering only the sequence ⟨se te⟩, one would be tempted to see it as one word *sete* meaning 'was/were', which would then have to count as a fourth stem of the copula or as an exceptionally synthetic inflection of the second stem. There are several hitches to that solution. First, one would have to deal with the fact that this putative word could be broken up by the negator *pa*, as one sees in (20). Infixes do exist, yet assuming *pa* to behave as an infix just in this case will certainly be felt to be too costly. The only solution coherent with the *sete* hypothesis would then be to view as one word not only it, but also the sequences ⟨se pa te⟩ 'was/were not' and ⟨se pa⟩ 'am/is/are not'.

It seems to me to be simpler and less offensive to Occam's razor to posit special realization rules such that TMA markers and the negator — a natural class as exponents of analytic inflection including polarity — exceptionally follow rather than precede the main verb when it is *se*. As usual, the explanation for such a crazy behaviour is bound to be diachronic to some extent: cf. French *c'est pas* /sɛ_pa/ 'it isn't' — but *c'était pas* /sɛtɛ_pa/ 'it wasn't', which confirms *te*'s identity as a TMA marker and shows the COP < NEG < TMA ordering to be a Haitian Creole innovation consequent to *te*'s emergence.

Rule (37) accounts for the sequence ⟨se pa te⟩ of *se pa te yon bon chapantye* 'wasn't a good carpenter':

$$
(37)\quad
\begin{bmatrix}
\textit{mword} \\
\textsc{phon} \quad \langle\textit{sepate}\rangle \\[4pt]
\textsc{mph} \quad \left\langle \boxed{1}\begin{bmatrix}\textsc{ph} & \langle\textit{se}\rangle \\ \textsc{pc} & 1\end{bmatrix},\; \boxed{2}\begin{bmatrix}\textsc{ph} & \langle\textit{pa}\rangle \\ \textsc{pc} & 2\end{bmatrix},\; \boxed{3}\begin{bmatrix}\textsc{ph} & \langle\textit{te}\rangle \\ \textsc{pc} & 3\end{bmatrix} \right\rangle \\[6pt]
\textsc{ms} \quad \langle \boxed{4}\,[\textsc{pol}\;\textit{neg}],\; \boxed{5}\,[\textsc{tma}\;\textit{pst}],\; \boxed{B}\,[\textsc{lid}\;\textit{cop}] \rangle \\[6pt]
\textsc{rr1} \quad \begin{bmatrix}\textsc{mud} & \boxed{B}\,[\textsc{lid}\;\textit{cop}] \\ \textsc{mph} & \boxed{1}\begin{bmatrix}\textsc{ph} & \langle\textit{se}\rangle \\ \textsc{pc} & 1\end{bmatrix} \\ \textsc{rs} & [\,] \end{bmatrix} \\[6pt]
\textsc{rr2} \quad \begin{bmatrix}\textsc{mud} & \boxed{4}\,[\textsc{pol}\;\textit{neg}] \\ \textsc{mph} & \boxed{2}\begin{bmatrix}\textsc{ph} & \langle\textit{pa}\rangle \\ \textsc{pc} & 2\end{bmatrix} \\ \textsc{rs} & [\,] \end{bmatrix} \\[6pt]
\textsc{rr3} \quad \begin{bmatrix}\textsc{mud} & \boxed{5}\,[\textsc{tma}\;\textit{pst}] \\ \textsc{mph} & \boxed{3}\begin{bmatrix}\textsc{ph} & \langle\textit{te}\rangle \\ \textsc{pc} & 3\end{bmatrix} \\ \textsc{rs} & [\,] \end{bmatrix}
\end{bmatrix}
$$
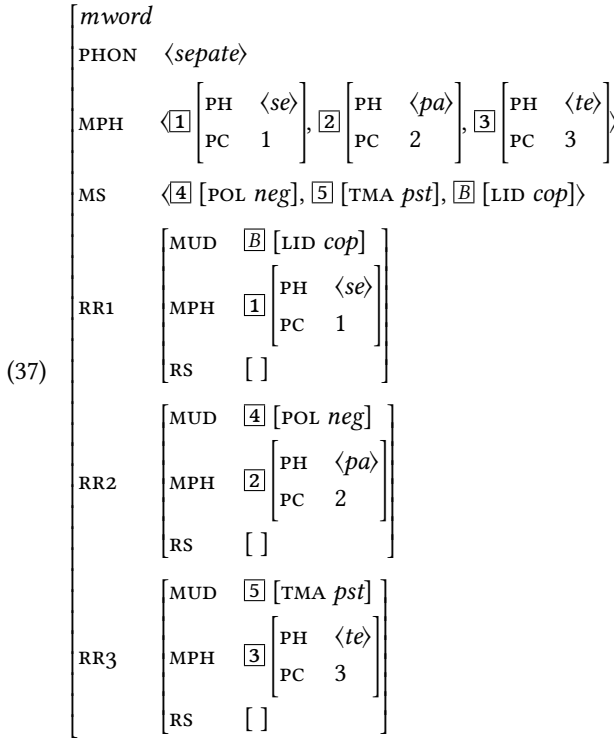
This rule should be contrasted with the rule accounting for the "normal" order /pa te V/ of, e.g., *pa te chante* 'didn't sing':

$$
(38) \quad
\begin{bmatrix}
\textit{mword} \\
\text{PHON} \quad \langle \textit{patechante} \rangle \\[4pt]
\text{MPH} \quad \left\langle \boxed{1}\begin{bmatrix} \text{PH} & \langle \textit{pa} \rangle \\ \text{PC} & 1 \end{bmatrix}, \boxed{2}\begin{bmatrix} \text{PH} & \langle \textit{te} \rangle \\ \text{PC} & 2 \end{bmatrix}, \boxed{3}\begin{bmatrix} \text{PH} & \langle \textit{chante} \rangle \\ \text{PC} & 3 \end{bmatrix} \right\rangle \\[8pt]
\text{MS} \quad \langle \boxed{4}\,[\text{POL } \textit{neg}], \boxed{5}\,[\text{TMA } \textit{pst}], \boxed{C}\,[\text{LID } \textit{chante}] \rangle \\[6pt]
\text{RR1} \quad
\begin{bmatrix}
\text{MUD} & \boxed{4}\,[\text{POL } \textit{neg}] \\
\text{MPH} & \boxed{1}\begin{bmatrix} \text{PH} & \langle \textit{pa} \rangle \\ \text{PC} & 1 \end{bmatrix} \\
\text{RS} & [\,]
\end{bmatrix} \\[10pt]
\text{RR32} \quad
\begin{bmatrix}
\text{MUD} & \boxed{5}\,[\text{TMA } \textit{pst}] \\
\text{MPH} & \boxed{2}\begin{bmatrix} \text{PH} & \langle \textit{te} \rangle \\ \text{PC} & 2 \end{bmatrix} \\
\text{RS} & [\,]
\end{bmatrix} \\[10pt]
\text{RR1} \quad
\begin{bmatrix}
\text{MUD} & \boxed{C}\,[\text{LID } \textit{chante}] \\
\text{MPH} & \boxed{3}\begin{bmatrix} \text{PH} & \langle \textit{chante} \rangle \\ \text{PC} & 3 \end{bmatrix} \\
\text{RS} & [\,]
\end{bmatrix}
\end{bmatrix}
$$

The main difference — apart from the fact that *chante*, like all verbs but *se* and raising verbs (see above), does not accept null subjects — lies in the respective position classes. It is particularly noteworthy that the mutual ordering of the negator and the TMA marker is fixed: *pa* ≺ TMA. It is this sequence that appears as a block on the "wrong" side when the verb is *se*.

Examples (6) *Se frè mwen Jan ye* 'It's my brother that John is' and (14) *kijan lavi te ye* 'how was life' suffice to illustrate the third stem's environment: its subject must be an NP and its complement a gap related to clefting as in (6) or questioning as in (14). Hence the following lexical entry:

(39)

$$
\begin{bmatrix}
\textit{verb word} \\
\text{LID} \quad \boxed{C}\ \textit{cop} \\
\text{PHON} \quad \langle \textit{ye} \rangle \\
\text{SYN} \quad
\begin{bmatrix}
\text{HEAD} \quad [\text{PRED} +] \\
\text{VAL} \quad
\begin{bmatrix}
\text{SPR} \quad \boxed{1}\langle \text{NP} \rangle \\
\text{COMPS} \quad \boxed{2}\langle \textit{gap} \rangle \\
\text{ARG-ST} \quad \boxed{1}+\boxed{2}
\end{bmatrix}
\end{bmatrix} \\
\text{SEM} \quad
\begin{bmatrix}
\text{MODE} \quad \textit{prop} \\
\text{INDEX} \quad s \\
\text{RESTR} \quad
\begin{bmatrix}
\text{RLN} \quad \textit{cop} \\
\text{SIT} \quad s \\
\text{SBJ} \quad i \\
\text{PRED} \quad j\ \textit{pred}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

As mentioned above, *ye* is neutral as to whether the predicated property is a stage- or individual-level one. Its occurrence in just one environment justifies my ranking it as the most non-default stem. On the other hand, the mutual ranking of the null stem and *se* in terms of defaultness may be judged moot. The numbers of triggering contexts are the same, and I can't see any good reason why stage-level properties should be deemed more default than individual-level properties. Be it as it may, since stems must be tagged in any event and nothing much hangs on the relative ordering of *se* and the null stem, I maintain the ranking of (28).

4 Conclusion: What has been done and what remains to do

Haitian Creole facts lie precisely at the interface of morphology and syntax, and it has been the aim of the present article to show how a word-based morphological model is especially fit to do justice to such an inherently morphosyntactic character.

Formalizing the data as I just have done is a necessary step in understanding how things work. It doesn't tell us, however, why things work the way they do, it doesn't explain why things are as they are. Explanation in the real sense of the term has to come from outside formal grammar. In the case at hand, the likeliest source is diachrony, that is the sociolinguistic conditions under which Haitian Creole emerged and the nature of the linguistic input at the origin of this emergence.

As to the first point, our best hypothesis is that Haitian Creole emerged between the 1680's and the end of the 18th century as a consequence of the massive importation of African slaves into Haiti, officially a French possession from 1697 to 1804 (see Holm 1989:382–387; Faraclas et al. 2007), and that it was mainly the product of a process of

second language acquisition (SLA) by adults in adverse conditions, where the target language French could only be acquired in an unguided fashion, "on the job", and was not actually acquired, but only a basic variety of it (Klein & Perdue 1997), which later expanded into a full-fledged language. The Africans' knowledge of their first languages (the substrate) played a role in this process, although apparently no direct one in the copula issue.

Where it may have proved influential is in the fact that the stage- vs. individual-level contrast is active in what seems to have been Haitian Creole's main substrate language, namely Fongbe (Lefebvre 1998). In Fongbe according to Ndayiragidje (1993: 63) "only predicates whose argument structure includes an event position — *Stage-Level Predicates...* may be clefted, contrary to those that do not include that position — *Individual-Level Predicates*" (my translation). This is what makes the difference between e.g. *gbà* 'to destroy' and *sè* 'to know'. In Haitian Creole as well the same difference obtains between *kraze* 'to destroy' and *konnen* 'to know' so that (40) is grammatical, whereas (41) — possibly meaning 'John does know that language' — is not (Lefebvre 1990 — and see (8)–(9):

(40) Se   kraze Bouki kraze   kay   la
      COP destroy B.   destroy house DEF

      'What Bouki did to the house was destroy it.'

(41) * Se   konnen Jan konnen lang   sa   a.
        COP know   J.   know   language DEM DEF

      Intended: 'John does know that language.'

The *se* vs. null form contrast therefore appears to be a special case of this overarching contrast permeating the whole verbal lexicon, which seems to be more central in Fongbe than it is in French, though it is present in the latter as well.

Concerning the French input, on the other hand, we unsurprisingly hold no recording of the sort of 17th century French in which the arriving slaves were addressed or could pick up from the native French speakers they were in generally unpleasant contact with. That it was a colonial koinè not too different from the central Parisian dialect, we can be reasonably sure of (Chaudenson 2004). Whether it was the full language or a foreigner talk reduction of it, we don't know, though there is evidence that the full lexifier languages were used in the Caribbean plantations where creole languages emerged (Alleyne 1980).

What we can and must do then, is first try to account for the facts that have been pushed under the rug in the present work, in particular the strange behaviour of *se* according to whether it is or is not modified by TMA markers and/or the negation, and why is then the stage- vs. individual-level contrast neutralized. Secondly, we should look up 17th century French grammar, using such ressources as Haase (1935), in order to determine as much as possible to what extent the Haitian Creole system inherits from its lexifier's system. For instance, although the substrate is likely to have been influential as suggested above, there probably is a relation between the distribution of *se* and the

null stem — requiring individual and stage-level complements respectively — and the distribution of *c'est* and *il/elle est* preceding a nominal complement in 17th century as well as contemporary French (Kupferman 1979, Boone 1987, Zribi-Hertz to appear). All this, however, belongs to the to-do tray. Let's hope it won't linger there too long.

References

Alleyne, Mervyn C. 1980. *Comparative Afro-American: An historical-comparative study of English-based Afro-American dialects of the New World*. Vol. 11. Karoma Publishers, Incorporated.

Bender, Emily & Ivan A. Sag. 2000. Incorporating contracted auxiliaries in English. In Ronnie Cann, Claire Grover & Philip Miller (eds.), *Grammatical interfaces in HPSG*, 17–32. Stanford: CSLI Publications.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25. 63–110.

Boone, Annie. 1987. Les constructions «il est linguiste»/«c'est un linguiste». *Langue française* (75). 94–106.

Carlson, Gregory N. 1977. A unified analysis of the English bare plural. *Linguistics & Philosophy* 1(3). 413–456.

Chalmers, Mehdi, Chantal Kénol, Jean-Laurent Lhérisson & Lyonel Trouillot (eds.). 2015. *Anthologie bilingue de la poésie creole haïtienne de 1986 à nos jours*. Arles: Actes Sud / Atelier Jeudi Soir.

Chaudenson, Robert. 2004. *La créolisation: théorie, applications, implications*. Editions L'Harmattan.

Chierchia, Gennaro. 1995. Individual-level predicates as inherent generics. In Gregory N. Carlson & Francis J. Pelletier (eds.), *The generic book*, 176–223. Chicago: The University of Chicago Press.

Crysmann, Berthold & Olivier Bonami. 2015. Variable morphotactics in information-based morphology. *Journal of Linguistics* 51(1). 1–64.

Damoiseau, Robert. 1985. *Nature et fonction des monèmes 'se' en créole haïtien.* Tech. rep. Port-au-Prince: Centre de Linguistique, Université d'État d'Haïti.

Damoiseau, Robert. 1996. Les adjectivaux en créole haïtien. In Daniel Véronique (ed.), *Matériaux pour l'étude des classes grammaticales dans les langues créoles,* 151–161. Aix en Provence: Publications de l'Université de Provence.

DeGraff, Michel. 1992. *Creole grammars and acquisition of syntax: The case of Haitian.* University of Pennsylvania dissertation.

DeGraff, Michel. 2007. Kreyòl Ayisyen, or Haitian Creole (Creole French). In John Holm & Peter L. Patrick (eds.), *Comparative Creole syntax: Parallel outlines of 18 Creole grammars*, 101–126. London: Battlebridge.

Déprez, Viviane. 1992a. Is Haitian Creole really a pro-drop language. *Travaux de recherche sur le créole haïtien* 11. 23–40.

Déprez, Viviane. 1992b. Raising constructions in Haitian Creole. *Natural Language & Linguistic Theory* 10(2). 191–231.

Déprez, Viviane. 2003. Haitian Creole *se*: A copula, a pronoun, both or neither? On the double life of a functional head. In Dany Adone (ed.), *Recent developments in creole studies*, 135–173. Tübingen: Max Niemeyer Verlag.

Déprez, Viviane & Marie-Thérèse Vinet. 1997. Predicative constructions and functional categories in Haitian Creole. *Journal of Pidgin and Creole Languages* 12(2). 203–235.

Diesing, Molly. 1988. Bare plural subjects and the stage/individual contrast. In *Genericity in natural language: Proceedings of the 1988 tübingen conference*, 107–154.

Faraclas, Nicholas, Don E. Walicek, Mervyn C. Alleyne, Wilfredo Geigel & Luis Ortiz. 2007. The complexity that really matters: The role of political economy in creole genesis. In Umberto Ansaldo, Stephen Matthews & Lisa Lim (eds.), *Deconstructing Creole*, 227–264. Amsterdam: John Benjamins.

Fattier, Dominique. 2013. Haitian Creole. In Susanne Michaelis, Peter Maurer, Martin Haspelmath & Maria Huber (eds.), *The survey of Pidgin and Creole languages*, vol. II, 195–204. Oxford: Oxford University Press.

Glaude, Herby. 2012. *Aspects de la syntaxe de l'haïtien*. Paris: Anibwé.

Haase, A. 1935. *Syntaxe française du XVIIe siècle*. Nouvelle édition traduite et remaniée par M. Obert avec l'autorisation de l'auteur. Paris: Delagrave.

Holm, John. 1989. *Pidgins and Creoles*. Cambridge: Cambridge University Press.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Kihm, Alain. 1993. What is it that you said? A study of obligatory focalization in two creoles and beyond. In Francis Byrne & Donald Winford (eds.), *Focus and grammatical relations in Creole languages*, 141–162. Amsterdam: John Benjamins.

Kihm, Alain. 2003. Haitian construct state nominals: A creole contribution to the theory of genitive phrases. In Dany Adone (ed.), *Recent developments in creole studies*, 203–222. Tübingen: Max Niemeyer Verlag.

Klein, W. & Clive Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?) *Second Language Research* 13. 301–347.

Kratzer, Angelika. 1995. Stage level and individual-level predicates. In Gregory N. Carlson & Francis Jeffry Pelletier (eds.), *The generic book*, 125–175. Chicago: The University of Chicago Press.

Kupferman, Lucien. 1979. Les constructions *Il est médecin / C'est un médecin*. Essai de solution. *Cahiers de linguistique* 9. 131–164.

Lefebvre, Claire. 1990. Stage and individual predicates: Additional evidence from Haitian and French. Université du Québec à Montréal, ms.

Lefebvre, Claire. 1998. *Creole genesis and the acquisition of grammar: The case of Haitian Creole*. Cambridge: Cambridge University Press.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Mateus, Maria Helena Mira, Ana Maria Brito, Inês Duarte, Isabel Hub Faria, Sónia Frota, Gabriela Matos, Fátima Oliveira, Marina Vigário & Alina Villalva. 1989. *Gramática de língua portuguesa*. 3a edição. Lisboa: Editorial Caminho.

Matthews, P. H. 1972. *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Ndayiragidje, Juvénal. 1993. Clivage du prédicat en fongbe. In Alain Kihm & Claire Lefebvre (eds.), *Aspects de la grammaire du fɔngbè. Etudes de phonologie, de syntaxe et de sémantique.* Paris: Peeters.

Pompilius, Pradel. 1976. *Contribution à l'étude comparée du créole et du français : morphologie et syntaxe.* Port-au-Prince: Editions Caraïbe.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 69–202. Stanford: CSLI Publications.

Sag, Ivan A. & Thomas Wasow. 1999. *Syntactic theory: A formal introduction.* Stanford: CSLI Publications.

Sag, Ivan A., Thomas Wasow & Emily Bender. 2003. *Syntactic theory: A formal introduction.* 2nd edition. Stanford: CSLI Publications.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure.* Cambridge: Cambridge University Press.

Tessonneau, Louise. 1980. *Contes créoles d'Haïti.* Paris: Conseil international de la langue française.

Valdman, Albert. 1978. *Le créole : structure, statut et origine.* Paris: Klincksieck.

Van Eynde, Frank. 1994. *Auxiliaries and verbal affixes. A monostratal cross-linguistic analysis.* Leuven dissertation.

Zribi-Hertz, Anne. to appear. Les pronoms démonstratifs. In *La Grande grammaire du français*, Actes Sud.

Zwicky, Arnold M. 1985. Rules of allomorphy and phonology-syntax interactions. *Journal of Linguistics* 21. 431–436.

Zwicky, Arnold M. 1990. Syntactic representations and phonological shapes. In Sharon Inkelas & Draga Zec (eds.), *The phonology-syntax connection*, 379–397. Chicago: The University of Chicago Press.

Chapter 12

On lexical entries and lexical representations

Andrew Spencer

University of Essex

Lexicalist models of syntax share with lexeme-and-paradigm models of morphology the assumption that the primary unit of the lexicon is the lexeme, an abstract representation of properties unifying a set of inflected word forms. Lexicalist syntactic models (such as Head-driven Phrase Structure Grammar, henceforth HPSG, and Sign-Based Construction Grammar, henceforth SBCG) distinguish modelled linguistic objects from descriptions of objects. A description, but not an object, can be a partial (underspecified) representation. However, a lexeme is by definition only partially specified, being underspecified for all those morphosyntactic properties that its word forms realize (the lexeme DOG realizes neither singular nor plural, unlike the word forms *dog, dogs*). This implies that lexemes are descriptions, not objects, which is incompatible with assumptions about the type hierarchy for signs and the lexicon in HPSG/SBCG. If we relax the definition of full specification to admit lexemes as objects then the question arises as to how many properties can be left unspecified. I argue for a maximally underspecified model. Even the declaration of properties for which the given class of lexemes inflects (the 'morpholexical signature', MORSIG) is underspecified to the extent that its contents are predictable. This entails that an inflected word form of a lexeme can be defined only after the MORSIG attribute is specified. Derivation, a lexeme-to-lexeme mapping, can therefore be defined over the same maximally underspecified lexical representations, whose inflection is then typically governed by a different morpholexical signature (e.g. when the derivation changes word class). All such specifications are given by default statements, which are overridden for irregular items. Verb-to-adjective transpositions (participles) are members of the verb's paradigm yet inflect according to the adjectival paradigm (the 'adjectival representation' of a verb). This gives the effect of a 'lexeme-within-a-lexeme', posing a challenge for lexeme-and-paradigm models. I present an analysis in which the definition of the participle is driven by a feature REPRESENTATION. This (re-)defines the MORSIG attribute, creating a representation which is identical to that of an adjective, while remaining part of the verb's paradigm. I discuss some of the implications of this analysis for lexical relatedness, the lexical type hierarchy of SBCG and the morphology-syntax interface.

Andrew Spencer

1 Introduction

The notion of word is by definition central to lexicalist models of syntax, so one would imagine that morphology, too, would occupy a central place in the construction of such models. However, there is as yet surprisingly little consensus between morphologists and syntacticians over fundamental aspects of word structure and the relations between words and syntax or semantics. In addition, I will argue that there is a systematic unclarity in conceptualizations of wordhood even amongst those of us who accept the primacy of the lexeme notion and its role in morphosyntax ('lexeme-and-paradigm' models). One central ontological question is 'what kind of a thing is a word?' The problem is that, whereas inflected word forms can be regarded as 'concrete' linguistic objects which combine with each other to form phrases (another type of object), lexemes are by their nature more abstract: they are ultimately representations which unite a set of related inflected word forms without themselves being a form. They are therefore underspecified representations, in the sense that they are not specified for the various morphosyntactic properties their word forms realize. The dictionary is a set of lexemes, so it, too, is an abstract construct.

The question of what lexemes are is made more acute when we examine a somewhat neglected, but theoretically and conceptually important, type of lexical relatedness, the (true) transposition, illustrated in this paper by the Russian deverbal participle. A participle is the adjectival 'representation' (Haspelmath 1996) of a verb. As such, it is part of the paradigm of a verb and yet it inflects exactly like an adjective and demonstrates much of the external syntax of an adjective (a true participle is used principally as an attributive modifier to a noun). Shifting morphosyntactic category in this fashion is characteristic of derivation, i.e. lexeme formation, yet a true participle (that is, a participle that has not undergone lexicalization, or some other process of grammaticalization) is not an autonomous lexeme, independent of its verb base, any more than the past tense or the infinitive form of a verb is an autonomous lexeme. The participle thus gives the appearance of being a 'lexeme-within-a-lexeme', posing obvious difficulties for any simple characterization of lexeme-and-paradigm inflectional morphology, and especially to the inferential-realizational (I-R) class of models in Stump's (2001) typology.

In this paper I investigate some of these questions against the backdrop of the class of I-R models called Paradigm Function Morphology (PFM: Stump 2001, Bonami & Stump 2016). Specifically, I will assume the overall architecture of a model of lexical relatedness proposed in Spencer (2013), Generalized Paradigm Function Morphology (GPFM). I confront the proposals about lexical representations and lexical relatedness made in GPFM with influential proposals put forward within the variant of HPSG developed by Sag (2012), Sign-Based Construction Grammar (SBCG). I argue that the HPSG/SBCG conception of the lexeme conceals important conceptual inconsistencies. In particular, a lexeme can only be described by a feature structure (FS) that is partially specified. However, this means that technically a lexeme is just a description and not an object. Yet the architecture of the HPSG lexicon demands that lexemes be bona fide linguistic objects, not descriptions of objects.

If we simply declare the lexemes as objects then the question arises as to how much the lexeme can be underspecified. Building on the defaults-based GPFM model I argue that a lexeme is best regarded as a *maximally* underspecified object, bearing all and only those properties which are not predictable from default specifications.[1] I show how the maximally underspecified lexemic representation can help solve the question of the status of transpositions such as participles.

I make a number of background assumptions.

- A dictionary is a list of lexemes.

- Inflectional morphology operates according to I-R principles and defines a paradigm for each class of lexemes, each cell of which is occupied by a pair ⟨ω,σ⟩ for the set σ of morphosyntactic properties realized by the word form ω.

- A fully specified representation of a lexeme includes a specification of a set of syntactic properties, a semantic representation (which for convenience I take to be a simplified form of Lexical Conceptual Structure, Jackendoff 1990) and a unique identifier, variously called the Lexical Identifier (LID), the Lexical Index, or the Lexemic Index (LI). (This is comparable in function to the lexicographer's lemma.)

- The syntactic properties of a lexeme include a specification of its argument structure (ARG-ST).

- The ARG-ST attribute of a lexeme includes a semantic function role (SF role, Spencer 2013), canonically R for nouns, E for verbs and A for adjectives

The chapter is structured as follows. I open by outlining four possible ways of representing lexemes, the fourth of which relies heavily on the device of defaults and overrides operating over a maximally underspecified entry. The next section addresses the question of whether a lexeme can be regarded as an object or not, and how many of its properties can be underspecified.

In §4 I turn briefly to the model of lexical representation proposed in Spencer (2013), and specifically to the way in which an inflectional feature declaration (MORSIG, 'morpholexical signature') can be defined and deployed in a defaults-based model of lexical representation. Against this background §5 addresses the architecturally important question of the place of transpositions such as deverbal participles. These are an important test case because they raise questions of lexemic identity and category membership: the participle behaves as a 'quasi-lexeme', without being the output of derivational lexeme formation proper. I deploy an attribute REPRESENTATION to define transpositions. I discuss the way that the adjectival inflectional paradigm can be incorporated into the paradigm of a verb by appropriate use of the MORSIG attribute. I illustrate with a description of the Russian participial system. I contrast the behaviour of true participles

---

[1]This corresponds to Sag's 2012 notion of listeme. The listeme has a somewhat unclear status in SBCG, but Sag explicitly describes it as a description and not an object, so it is not a perfect correspondent to the conception of lexeme proposed here.

with that of transpositional lexemes (Spencer 2013, 2016), which are derived autonomous lexemes formed from transpositions such as participles.

In §6 I ask how transpositions might be incorporated into a multiple inheritance hierarchy but note two problems. First, multiple inheritance hierarchies are not straightforwardly capable of distinguishing, say, the adjectival representation of a verb (participle) from the verbal representation of an adjective (inflecting predicative adjective). Second, there is in any case virtually no discussion in the morphological literature of transpositions and hence no consensus on how their morphological properties should be accounted for. I conclude with a tentative list of questions which arise from the discussion.

I will close this introduction with a terminological note. I shall simplify discussion wherever possible by assuming the correctness of my approach and taking the lexeme to be effectively identical to its description. That is, a lexeme is a dictionary entry, an abstract underspecified representation, which we can think of as a meta-representation, unifying the concrete representations in the complete set of its word forms. The obvious synonym for 'dictionary entry' is 'lexical entry'. However, in constraints-based syntactic models the notion of 'lexeme' is rather poorly developed, and the term 'lexical entry' is often (though not invariably!) used to refer not to the abstract object listed in a dictionary but rather to a concretely instantiated inflected word form of a lexeme. This terminological ploy is confusing, but is now ingrained. Following Dalrymple et al. (2015), I shall therefore adopt the term 'lexemic entry' for the standard lexicographic notion of dictionary entry. I will avoid the term 'lexical entry' and refer to the representation (fully or partially specified) of an inflected form as the lexical representation of that word form. This is more than a question of mere terminology, especially in HPSG, but proper evaluation of the issues would require a separate study.

2 The nature of the lexeme

In principle there are a good many ways in which dictionary entries can be represented. It will be useful to consider four of these. The first possibility is to list every inflected form separately with a complete specification of all its properties, whether idiosyncratic or predictable. This will include (i) all the morphological properties, such as inflection class, (ii) syntactic properties such as argument structure, including the SF (semantic function) roles, valence, selection, collocation, lexicosyntactic class features and others, together with (iii) contextual properties or properties relating to usage such as register, connotations, and other, not strictly linguistic, properties that a competent user would be expected to know about the word (what is sometimes called 'encyclopaedic information', though this term is difficult to pin down). I shall call this mode of representation the **unindexed full word form listing** model. Some psycholinguistic models of the mental lexicon appear to have essentially this structure. It does not define a dictionary entry in any direct sense because every word form of every lexeme has the same representational status as any other: *dog* and *dogs* are only marginally more related to each other on this model than are *dog* and *dig* or *dogs* and *geese*.

The unindexed full word form listing model effectively excludes any standard understanding of the notion of dictionary entry, therefore. However, it would be possible to reconstruct the traditional notion of dictionary entry by providing all the forms that unite under a given lexeme with a unique **lexemic index**. Thus, *dog, dogs* would both have the index DOG, distinct from that of *dig* (DIG) or *geese* (GOOSE). This would then define our second model of lexical representation, which I will call the **indexed full word form listing** model. The LI would have to be a secondary property associated with each component of a lexemic entry, FORM, SYN, SEM. At the level of FORM this would mean indexing the lexeme's root, its various stem forms and all its inflected forms (unless these were able to inherit the LI of their stems). At the level of SYN, SEM each individual sub-attribute (syntactic class, argument structure or whatever, depending on one's syntactic assumptions) would be furnished with the same LI, as would the basic meaning or lexical conceptual structure and any other aspects of meaning. This use of a lexemic index is very similar to that proposed by Jackendoff (1997) and integrated into the Simpler Syntax model (Culicover & Jackendoff 2005), though their model makes rather different assumptions about the structure of inflected words because it retains the morphemic concept and therefore is not strictly speaking lexeme-based.

These first two models share the property that all inflected word forms are fully listed. In such models there is effectively no morphology defining the lexical relatedness that holds between word forms of the same lexeme. In order to capture formal similarity/identity between word forms it would therefore be necessary to postulate lexical redundancy rules (Jackendoff 1975, Bochner 1993) or inflectional templates (Ackerman et al. 2009).

The third model I shall call the **fully specified lexemic entry** model. The term 'fully specified' refers to the fact that on this model (along with the previous two models) the lexemic entry includes fully predictable information about the FORM, SYN, SEM representations as well as unpredictable, idiosyncratic information. For instance, if all syntactic nouns in the language are also morphological nouns (i.e. if the language lacks category mixing with respect to the noun class) then the property of inflecting as a noun, that is, being a morphological noun, can be deduced from the SYNCAT label. However, under the fully specified lexemic entry model such a word would still be given the attribute [MORCAT *noun*] or the equivalent as part of its FORM representation. Where this third model differs from the previous two is in the important assumption that (regularly) inflected word forms are not included as part of the lexicon as such. Rather, such a model follows lexicographic tradition in abstracting away from inflected word forms, instead, defining them by means of a separate 'inflectional engine', such as PFM. On the fully specified lexemic entry model, the lexeme-as-dictionary-entry is accorded a special ontological status, that of a linguistic object. Depending on how such a model is implemented formally it may or may not be necessary to individuate dictionary entries by means of the arbitrary LI attribute. However, traditional lexicography certainly makes use of something very close to an LI in the form of a lemma or headword. An arbitrary label of this sort appears to be the most natural way of individuating entries.

The fourth model of lexical representation is the **underspecified lexemic entry** model, argued for in Spencer (2013). This model deploys the logic of default inheritance to ab-

stract away fully predictable lexical information. The lexemic representation in this case includes just the information that cannot be inferred by default from other aspects of the representation or from other facts in the grammar of the language. Thus, in our previous example, if the specification [MORCAT *noun*] is fully predictable from the specification [SYNCAT *noun*] then the MORCAT specification need not be stated in the lexemic entry itself (indeed, there need be no mention of the attribute MORCAT at all).

To see how the underspecified lexical entry model might define dictionary entries, consider a word such as TREE. This minimally has to specify a phonological form for the basic stem form (root), $STEM_0$ = /tri:/, as well as minimal information about the kind of meaning the word has. As far as morphosyntax and especially inflection is concerned it hardly matters, of course, what kind of a thing a tree is (much less where to draw the line between trees and bushes). Also, the difference between abstract and concrete denotations seems to have little grammatical import, in English. However, it is important to know that TREE denotes some type of *Thing* and that it is countable, in contrast to words such as VEGETATION, or WOOD (in the sense of 'material coming from a tree'). Informally, we can distinguish count *Things* and mass *Things* with a subscript: $Thing_c$/$Thing_m$. However, for English we should also have some way of representing the fact that TREE (and IDEA) denotes something which is not a sexed higher animal, such as a person or a horse and which therefore can only be referred to as *it*, not as *s/he*. In languages which distinguish a 'vegetable' gender (e.g. Bininj-Gunwok) we might need to indicate the fact that TREE (and perhaps VEGETATION but not IDEA) denotes a kind of plant. In other languages with semantically-driven gender other distinctions would have to be made. These observations hold for the determination of inflectional properties. However, for a specification of derivational morphology it is often necessary to appeal to very subtle nuances of meaning (Fradin & Kerleroux 2003).

The point of this discussion of lexical semantics is that once the right semantic properties are fixed much of the rest of the lexemic representation can be deduced by default. Thus, if an English lexeme belongs to the *Thing* ontological category (as opposed to the category *Event* or *Property*) then by default it will be a noun, with an argument structure that includes the SF role R. A syntactic noun will also be a noun morphologically, and if it is of subcategory $Thing_c$ it will have a singular and plural form. This is more than just a modern version of the notional parts-of-speech theory, however. Being defaults, all these inferences can, of course, be overridden by more specific lexical stipulations. Thus, a noun such as JOURNEY is ontologically an *Event* but grammatically it is a noun, so that the inference from *Event* to SF role E to [SYNCAT/MORCAT *verb*] is overridden in the lexemic entry (for instance, by stipulating that its SF role is a simplex R). Moreover, in many languages there will be non-default morphological information to stipulate in addition to the phonology of the root. For instance, the Russian noun STOLOVAJA 'canteen; dining room' is a noun syntactically, but it has the morphology of a (feminine gender) adjective, thus its [MORCAT *adjective*] value cannot be inferred from its [SYNCAT *noun*] value and has to be stipulated in the lexemic entry in some way. In some cases, not all argument structure or complementation properties can be deduced from the semantic representation so those would need to be specified lexically. Some of the contextual properties of a

lexeme such as special register, connotations, or other details of usage may also diverge from the default and will therefore have to be recorded in the lexeme's entry. But the limiting case of a lexical representation in the underspecified lexemic entry model is a pure pairing of basic meaning with the form of the root (what Sag 2012 refers to as a 'listeme'; see §3) .

3 Lexemes as objects or descriptions

The principal question to be addressed in this paper is: what kind of a representation is a dictionary (lexemic) entry? Specifically, is it a linguistic object in its own right? In this section I discuss the answers proposed in Sag's (2012) summary of SBCG.

In SBCG, as in HPSG generally, a distinction is drawn between linguistic objects and the representational technology used to describe those objects, notably feature structures (FSs) or attribute-value matrices (AVMs). An inflected word form, for example, is a linguistic object, but it can be described in various ways, including partial feature descriptions which underspecify certain aspects of the representation. A linguistic object proper, however, cannot be thus underspecified. This means, for instance, that Sag's listeme, the barest possible representation of a lexemic entry, must be a description, not an object in its own right.

Sag (p. 98) introduces the notion of the lexeme into the model, giving it a special place in the type hierarchy of signs shown in Figure 1. This hierarchy defines the lexeme as a lexical sign, just like a word form. However, word forms appear as parts of syntactic phrases which can ultimately be pronounced, and so they count as linguistic expressions. A lexeme cannot be pronounced. This is not because it is some kind of 'covert expression', however (like *gap* and *pro* in Sag's type hierarchy). A lexeme is an altogether different kind of sign, in fact, a unique type given the hierarchy in Figure 1.
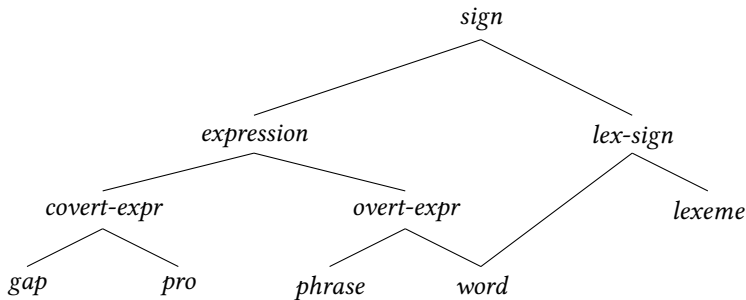
Figure 1: Sag's (2012) type hierarchy

Sag provides examples of representations of word forms from English (plurals, past tense forms) and in his Fig. 6 (p. 101), here reproduced as Figure 2, he gives the example of the lexeme LAUGH. Notice that this representation actually seems to specify the word

$$
\begin{bmatrix}
\textit{sintrans-v-lxm} \\
\text{PHON} \quad \textit{/læf/} \\
\text{FORM} \quad \langle \textit{laugh} \rangle \\
\text{ARG-ST} \quad \langle \text{NP}_i \rangle \\[2pt]
\text{SYN} \quad
\begin{bmatrix}
\textit{syn-obj} \\
\text{CAT}
\begin{bmatrix}
\textit{verb} \\
\text{SELECT} \quad \textit{none} \\
\text{VF} \quad \textit{psp} \\
\text{XARG} \quad NP_i \\
\text{LID} \quad \left\langle
\begin{bmatrix}
\textit{laughing-fr} \\
\text{LABEL} \quad l_1 \\
\text{SIT} \quad s \\
\text{S-SRCE} \quad i
\end{bmatrix} \right\rangle \\
\text{MRKG} \quad \textit{unmk} \\
\text{VAL} \quad \langle NP_i \rangle
\end{bmatrix}
\end{bmatrix} \\[2pt]
\text{SEM} \quad
\begin{bmatrix}
\textit{sem-obj} \\
\text{IND} \quad s \\
\text{LTOP} \quad l_1 \\
\text{FRAMES} \quad \left\langle
\begin{bmatrix}
\textit{laughing-fr} \\
\text{LABEL} \quad l_1 \\
\text{SIT} \quad s \\
\text{S-SRCE} \quad i
\end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$
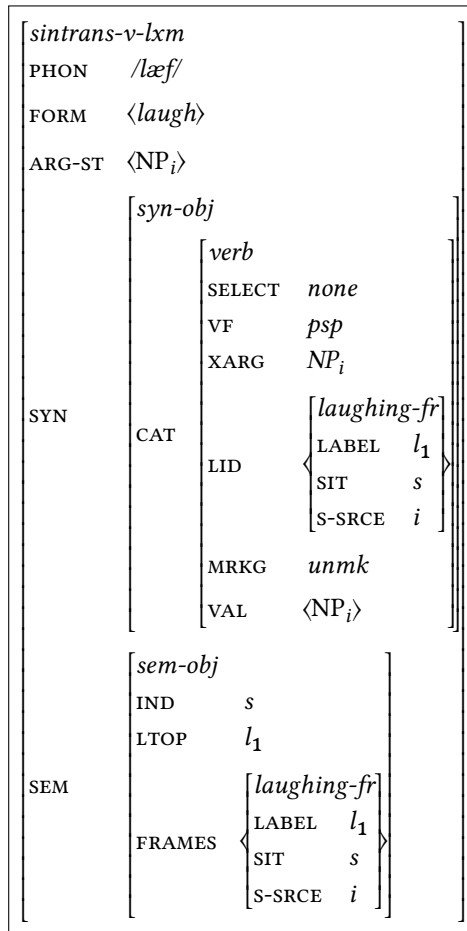
Figure 2: Sag's (2012: 111) representation of the lexeme LAUGH

form *laughed*, in that it bears the feature [VFORM *psp*]. It is worth citing Sag's justification for this choice of representation:

> [T]he value *psp* illustrated here […] represents an arbitrary expositional choice — any value of VFORM would satisfy the requirements imposed by the *laugh* listeme. And each such choice gives rise to a family of well-formed FSs licensed by that listeme. (Sag 2012: 99)

Sag here appeals to the LAUGH listeme. In SBCG a listeme *licenses* modelled linguistic objects. This means that it places restrictions on what properties a modelled object or sign may have (p. 105). Another way of characterizing the listeme is as "a lexeme description in the lexicon" (p. 107).

The type *lexeme* plays a central role in SBCG, in that it is the starting point for all morphology (Sag is here following PFM and related models). Inflection and derivation are modelled by means of morphological functions. An inflectional rule such as the English preterite (past tense) is modelled by a *preterite-cxt*, whose mother is the past tense form and whose daughter is the lexeme whose past tense form is being defined. A derivational rule is given by a construction whose mother is the derived lexeme and whose daughter is the base lexeme.

Sag summarizes the morphological functions by saying (p. 113) that they express "<…> the relation between the forms of two lexemes or the relation between the form of a lexeme and the form of a word that realizes that lexeme." This sounds like an expression of conventional wisdom in lexeme-based morphology, but it hides a serious conceptual flaw. This centres around the way that Sag's formulation uses the term 'form'. The problem is apparent from Sag's description of the lexeme LAUGH. He is obliged to provide this representation with an arbitrary inflectional feature specification, in effect defining not the lexeme as such but one of its inflected forms. This is because a lexeme is meant to be a modelled object, a subtype of *sign*, and a linguistic object must be fully specified. But the whole point of defining a lexemic level of representation is to abstract away from actual (concrete) word forms. This means that the lexeme is effectively a description, in fact a partial description, of the full set of word forms. But that is completely incompatible with Sag's type hierarchy and, indeed, with any coherent interpretation of the HPSG lexicon.

Given this reasoning we seem to have two logical courses of action. Either we can reconstruct the HPSG lexicon without recourse to the type *lexeme*, or we can redefine the notion of linguistic object in such a way as to make a dictionary entry a kind of modelled object, even though it appears to be underspecified. I shall adopt the second approach.

I propose to treat the lexicon as more than just a convenient descriptive fiction, as would be implied by a strict application of the object~description distinction. Rather, I take the lexicon to be a network of mentally represented (or representable) objects which can be defined and described by FSs just like (utterable and unutterable) linguistic expressions.

By simply declaring a dictionary (lexemic) entry to be a kind of object we solve the immediate problem: the lexeme can remain a type of sign, and can be a supertype of other

signs. Its unusual position in being partially underspecified is now reflected in the type hierarchy: only the *expression* type has to be fully specified, a lexical sign may be only partially specified (*lexeme*), though when a lexical sign is also a subtype of *expression* (*word*) it, too, can, and must, be fully specified.

Now, once we admit the possibility of an underspecified entity as an object in the linguistic ontology we are immediately faced with two sets of questions. The most general of these is 'are there other linguistic objects which can be less than fully specified? Can *any* partially specified representation be interpreted as a modelled object? If so, then what is the content of the original object~description distinction?" It seems that we should not be allowed to postulate such objects except in very special circumstances. But if we admit lexemes as less than fully specified objects what prevents us from postulating entirely arbitrary types? The simplest answer is to say that it is an architectural (i.e. stipulated) property of linguistic expressions that they be fully specified. However, whether this is really true may depend on how we perceive linguistic specification. Presumably, an object of type *word* such as *dogs* is to be regarded as a fully specified object and not a description, even when, for instance, its intonation and other prosodic characteristics are not specified. But in the strictest sense a word form remains partially underspecified until its full phonetic realization is given. Indeed, the same is true of sentences, which can be uttered with a very wide variety of affective intonation contours even when realizing one and the same set of discourse or information-structure functions.

The second question is more immediately relevant: if we are to admit as an object a lexeme underspecified for its inflection properties, how much further can we go with the underspecification? For instance, we might want to say that our lexeme LAUGH is underspecified for its inflectional properties by virtue of bearing the attribute values [TENSE u, VFORM u, SUBJAGR u, …] or whatever, where 'u' means 'not yet specified value', or we may wish to make the more radical proposal that LAUGH lacks the actual attributes [TENSE, VFORM, SUBJAGR, …]. This may turn out to be little more than a matter of notational convention, but in a more radical vein we can ask why we can't regard Sag's maximally underspecified listeme as a default lexeme object. In other words, can we not adopt the underspecified lexemic entry model for dictionary entries, as proposed in Spencer (2013)? We will see that the question assumes particular importance in defaults-based models of morphology such as PFM, where the lexeme concept finds its most elaborated implementation, and especially GPFM, where defaults define all aspects of lexical representation. Before turning to a consideration of the lexeme concept in such models I first discuss an important but generally neglected aspect of lexical representation and its relation to inflectional morphosyntax.

4 The morpholexical signature (MORSIG)

A lexeme of a given morpholexical class, such as 'noun', will (typically!) inflect for properties particular to that class (say, NUMBER, CASE, DEFINITENESS, POSSESSOR AGREEMENT) and may have intrinsic properties which determine its morphosyntax, such as GENDER. The actual set of properties is stipulated for each language, so a grammar has to include a

declaration of that set. In the Generalized Paradigm Function Morphology(GPFM) model of Spencer (2013) I refer to this declaration as the **morpholexical signature** (MORSIG). In GPFM the MORSIG attribute is itself treated as a default property with respect to lexemic entries/representations. By this I mean that the properties which make up the MORSIG are true of every regular lexeme of the given class, so it would be redundant to specify that information in the lexemic entry itself.

In Spencer (2013) I treat the MORSIG as a value of the FORM attribute, i.e. as a morphological property of a lexeme, but this is an oversimplification. It is well-known that the set of features needed to define a lexeme's syntactic distribution, and the set of grammatical meanings expressed by inflected word forms, are often at variance with the set of features needed to define the inflected morphological forms themselves. The most obvious mismatches are found in periphrases. We often find that the morphological form of one of the elements of the construction bears properties which contradict the feature content expressed by the periphrasis as a whole. Elsewhere, the morphological element may be morphomic and therefore not associated with any meaning, or the periphrasis may express a meaning in the manner of an idiom, so that no part of it can sensibly be associated with the meaning of the periphrasis as a whole (Brown et al. 2012). Periphrasis therefore motivates a distinction between m-features and s-features (mnemonically, morphological/syntactic features, Sadler & Spencer 2001). Similarly, Stump has argued for a modification of the original Paradigm Function Morphology (PFM) model, 'PFM1' (Stump 2001), in favour of a model, 'PFM2', which draws a distinction between FORM and CONTENT paradigms, on the basis of mismatches such as syncretisms, deponency and a variety of others (Stump 2002, 2006, 2016a,b). The obvious way to capture such distinctions in lexical representations is to assume that there is a SYN|MORSIG attribute which is mapped to a FORM|MORSIG attribute by means of a function, Stump's 'paradigm linkage'. By default, paradigm linkage is the identity function, in the sense that the FORM paradigm or m-feature set is identical to the CONTENT paradigm/s-feature set.

In GPFM the relation between the most highly underspecified lexical representation and a fully specified word form is mediated by two sets of functions. The second of these is effectively identical to the paradigm function of PFM2. It maps a pairing of ⟨*L*,σ⟩, for LI *L*, feature set σ, to a pair ⟨ω,σ⟩, where ω is the corresponding inflected word form. This function is, however, only defined for a complete and coherent feature set. In other words the function cannot be defined for a representation which lacks a specification of those features for which the lexeme inflects, that is, the MORSIG. Therefore, to be inflectable the lexeme's MORSIG attribute needs first to be specified (*Inflectional Specifiability Principle*, Spencer 2013: 199). This is achieved by the first of the two functions, the default specification of MORSIG for a given morphosyntactic lexical category.

An illustration of how this works can be given by (a simplified version of) the Turkish noun (following the discussion in Stump 2016a: 175–179). The minimal lexical information needed for, say, the word EV 'house' is shown in Figure 3 (using English as a metalanguage). Turkish grammar stipulates that a count noun inflects for the properties shown in Figure 4. The FORM|MORSIG attribute is almost identical except for a well-known syncretism between the 3sg possessed form of 'houses', and the 3pl possessed forms of

'house/houses' and the ordinary unpossessed plural. We would expect these to take the forms *evler, evlerler, evler* respectively, but the form *evlerler* is reduced by haplology to *evler*. Clearly, the FORM paradigm makes fewer distinctions than the CONTENT paradigm.

$$
\begin{bmatrix}
\text{FORM} & \begin{bmatrix} \text{STEM}_0 & \begin{bmatrix} \text{PHON} & /\text{ev}/ \end{bmatrix} \end{bmatrix} \\
\text{SEM} & \begin{bmatrix} \text{Thing}_c & \lambda x.\text{house}(x) \end{bmatrix} \\
\text{LI} & \textit{HOUSE}
\end{bmatrix}
$$

Figure 3: Lexemic entry for Turkish EV 'house'

$$
\begin{bmatrix}
\text{SYN}|\text{MORSIG} & \begin{bmatrix}
\text{NUMBER} & \{\, sg,pl \,\} \\
\text{CASE} & \{\, nom,acc,gen,dat,loc \,\} \\
\text{POSS} & \begin{bmatrix} \text{PERSON} & \{\, 1,2,3 \,\} \\ \text{NUMBER} & \{\, sg,pl \,\} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

Figure 4: MORSIG for Turkish count noun lexeme

In PFM2 this mismatch is defined via a Correspondence function, **Corr**, which specifies the distinct FORM features and CONTENT features and which defines the mismatches giving rise to syncretism, deponency and so on. The details are not relevant here so I simply assume the existence of the **Corr** mapping.

5 Lexical relatedness and the role of the Lexemic Index

The notion of lexemic representation (lexeme, lexical entry) plays an important role in the I-R class of models. This is especially true of GPFM, because that model attempts to unify inflection with (regular, productive, paradigmatic) derivational morphology. If we say, for the sake of argument, that English Subject Nominal (SubjNom) formation is paradigmatic then we can define it by recourse to a derivational feature (cf. Stump 2001: 257) SN, such that the generalized paradigm function, GPF, will map a verb lexeme to its subject nominal: GPF($\langle L, \text{SN} \rangle$) = $\langle L', \text{SN} \rangle$, where L' is the LI of the subject nominal of the verb L. However, the GPF cannot apply in exactly the way that the PF applies in PFM2. In PFM2 the PF maps a pairing of $\langle$LI,features$\rangle$ to a word form (via the **Corr** function). But the output of a derivational function has to be some representation of an independent lexeme. This means that when a derivational feature is in the domain of the GPF it must map to a representation of that derived lexeme, not to a word form. But the standard architecture of PFM2 (including the **Corr** function) does not permit this. The problem is at heart very familiar: while inflectional morphology specifies word forms that realize the particular morphosyntactic property set of a lexeme, derivational morphology effects

wholesale changes in syntactic and semantic representations, undermining the basic I-R assumptions under which morphology simply serves to realize property sets.

In the GPFM model of Spencer (2013), derivational morphology requires the GPF to perform a kind of 'deletion' of the base lexeme's properties, followed by respecification by means of defaults driven by the enriched SEM representation of the derived lexeme. However, a more parsimonious way to represent derivational morphology is to map the maximally underspecified base lexeme's entry to a maximally underspecified derived entry. This obviates the need to delete most of an entry's specifications, in that they are lacking in any case. Thus, for the lexeme DRIVE and its SubjNom DRIVER a schematic application of the GPF would be as in Figure 5 (where SN(DRIVE) is a function from LIS to LIS governed by the derivational feature, defining the LI of the derived lexeme, DRIVER). This type of application can be thought of as an elaborated, feature-driven word formation rule (*wfr*) in the sense of Aronoff 1976.

$$
\begin{bmatrix}
\text{FORM} & \begin{bmatrix} \text{STEM}_0|\text{PHON} & /\text{draɪv}/ \\ \text{STEM}_{\text{PST}}|\text{PHON} & /\text{droʊv}/ \\ \text{STEM}_{\text{PSTPTCP}}|\text{PHON} & /\text{drɪv}/ \end{bmatrix} \\
\text{SEM} & \begin{bmatrix} _{\text{Event}} \lambda x, y.\text{drive}(x, y) \end{bmatrix} \\
\text{LI} & \text{DRIVE}
\end{bmatrix}
$$

$$\Downarrow$$

$$
\begin{bmatrix}
\text{FORM} & \begin{bmatrix} \text{STEM}_0|\text{PHON} & /\text{draɪv}/\oplus/\text{ə}/ \end{bmatrix} \\
\text{SEM} & \begin{bmatrix} _{\text{Thing}} \lambda x[\text{person}(x) \wedge \exists y.\text{drive}(x, y)] \end{bmatrix} \\
\text{LI} & \text{SN}(\text{DRIVE})
\end{bmatrix}
$$
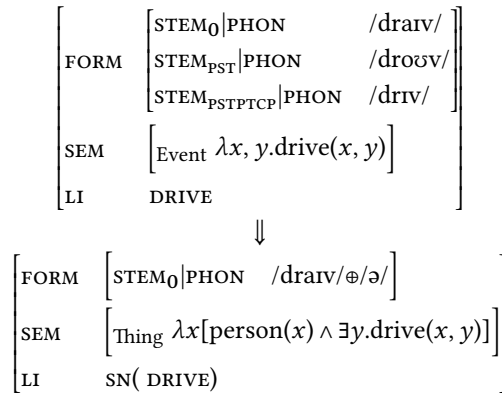
Figure 5: Derivation of DRIVER from DRIVE

Now, the output of the GPF is the representation of a *Thing*, so by default it will have all the morphosyntactic properties of a noun.[2] In languages with nominal inflectional classes the GPF may additionally have to specify which inflectional class the derived noun belongs to, as a FORM property overriding whatever the default specification for noun inflection class is, just as would be the case with a simplex (underived) lexemic entry belonging to a non-default inflectional class. The function in Figure 5 fails to transfer the non-default (stipulated) specification of the past tense and past participle stems from the base verb to the subject nominal, giving rise to a kind of despecification. There is an important rationale behind the despecification of lexemic entries in Spencer (2013): derivation, unlike inflection, leads to lexical opacity. Thus, the derived lexeme DRIVER lacks any specification which would identify it as having a base with past tense or past participle forms, irregular or otherwise, or, indeed, any of the morphosyntactic proper-

---

[2]DRIVER is a count noun, of course. I assume that this can be made to follow from the fact that a driver is a subtype of person.

ties associated with a finite verb. In this case the failure of the past and past participle forms to be inherited by the derived noun is the consequence of the definition of the MORSIG attribute for nouns as opposed to that for verbs. The GPF for SubjNom specifies exactly one STEM$_0$ form (for regular lexemes). This can be unified with the default MORSIG specification associated with *Thing* lexemes. Since the *Thing* ontological category does not license inflectional (s-feature) paradigm properties other than NUMBER in English, there would be no way for any tense or participle features to unify with the MORSIG attribute once it is specified. The only additional assumption that we need to make here is that SubjNom derivation is the kind of lexical relatedness which defines an entirely new MORSIG (i.e. one which 'deletes' the MORSIG of the base entry). I return later in this section to the question of how we characterize the class of relatedness functions which fail to preserve the base lexeme's MORSIG attribute in this way.

In true derivational morphology the LI of the output lexeme is always distinct from that of the base. This reflects the most significant difference between derivational types of lexical relatedness, on the one hand, and types of lexical relatedness broadly thought of as inflectional, on the other hand: derivation defines new lexemes while inflection defines forms of lexemes. However, in GPFM, preservation or alteration of the LI is just one parameter of relatedness, almost entirely independent of other parameters (this is the *Principle of Representational Independence*, Spencer 2013: 139). In particular, we systematically encounter two types of situation in which the crucial feature of the relatedness is the preservation or change of the base lexeme's LI.

The first of these is the class of relatedness types called **transpositions**, in which the morphosyntactic class of a word changes, as in typical derivation, but in which there is no creation of a novel lexeme with a distinct LI. In a canonical transposition the SEM value, that is, the conceptual content of the representation, does not change either.

The second type of case is very similar. Here the lexical relation defines a distinct lexeme but does not alter the conceptual content of the base. These are what I have called **transpositional lexemes** (Spencer 2013: 275; 359–60; Spencer 2016). Simple examples are adjectives derived from participles such as *interesting, bored* or so-called relational adjectives (in English and other European languages) such as *prepositional, ferrous*. These contrast with superficially similar cases in which the derived adjective differs semantically from its (etymological) base: *budding (linguist), harrowing (experience), gaping (hole); outspoken, unspoken, incensed, poised; popular (= 'well-liked'), spectacular*. Distinguishing true transpositions from transpositional lexemes and transpositional lexemes from other, often homophonous, adjectives is important for understanding the nature of lexical representations and types of lexical relatedness. In some cases, the only difference between the lexical representation of a true transposition and that of the homophonous transpositional lexeme is the difference in LI. However, in many cases the transpositional lexeme has different syntactic privileges from the homophonous transposition by virtue of being an independent lexeme. For instance, the adjective *interesting* has the complementation properties of an adjective, not of a verb or a true participle, as seen by comparing the true participle in (1) with the true adjective in (2).

(1) *interesting* = participle

    a.   the book (*very) interesting the children

    b.   * The book seems interesting the children.

(2) *interesting* = adjective

    a.   the book most interesting to the children

    b.   The book seems interesting to the children.

Comparable examples can be found with Russian participles and participial lexemes.

A clear instance of a true transposition is the (deverbal) participle, familiar from many languages, including almost all Indo-European languages. In Russian, for instance, we find four participles, realizing the properties [VOICE {act, pass}], [ASPECT {*pfv*, *ipfv*}] (Spencer 2017). These inflect exactly like adjectives and their principal function is that of attributive modifier to a noun. However, in addition to expressing the verbal properties of voice and aspect the participles also retain the argument structure/complementation of the base verb, including quirky case assignment. They are thus prototypical examples of mixed categories.

In Spencer (2013, 2017) I argue that participles belong to the base verb's paradigm in the broadest sense, and that this means their LI is that of the base verb. In an I-R model this means that the participles are defined by a ⟨feature, value⟩ pair, just like tense or number forms, and I propose the feature REPR(ESENTATION), following Russian descriptive tradition (see, for instance, Kuznecova et al. 1980, Helimski 1998 for the Samoyedic language Selkup, which is particularly rich in transpositions; see also Haspelmath 1996).

Following Spencer (2017) I notate the feature as REPR⟨K,Λ⟩, denoting a transposition from category K to category Λ. For example, a participle would be defined by the feature REPR⟨V,A⟩.[3] The GPF(⟨*V*,{REPR⟨V,A⟩,σ}⟩) applies to a verb lexeme *V* and defines a participle realizing features σ. For instance, the Russian perfective passive participle *udár'onn-* from UDARIT′ 'hit, strike' is defined by (3).

(3)   GPF(⟨UDARIT′,{REPR⟨V,A⟩,{[ASPECT *pfv*],[VOICE *pass*]}}⟩).

The GPF (3), however, only defines the stem of the participle. In order to inflect it as an adjective it must be given an appropriate MORSIG, inheriting CONCORD (agreement) features from the adjective class, permitting the participle to agree with the head noun. This addition to the MORSIG is an automatic consequence of redefining the morphosyntactic class as *adjective*. The technical details of exactly how this is achieved are provided in Spencer (2017). The GPF which defines the stem of the participle defines a lexical representation which is thus very similar to that of a (maximally underspecified) simplex adjective before it receives the default MORSIG specification. In this way the participle resembles an automonous adjectival lexeme, whilst remaining a form (better, the adjectival representation) of the verb, what we could call a 'quasi-lexeme'.

---

[3]The labels 'V, A' are for convenience. In fact, it is likely that all 'capital letter' lexical/phrasal ('c-structure') category labels (N, V, A, P) can be dispensed with, in favour of appeal to more fine-grained properties, especially the SF roles (Spencer 1998, 1999, 2013: 322–23; see also Chaves 2014 for similar remarks) .

Here is, in broad outline, how the GPF would deliver the quasi-lexeme form *udar´onn-*. A (partial) FS for the morsig of a typical transitive verb is shown in Figure 6. The FS in Figure 6 shows those morphosyntactic properties that are reflected in the grammatical system of Russian. It does not, however, tell us what the inflected forms are. This is because that FS defines the content paradigm feature set, not the form paradigm set. For instance, [tense *fut*] is only expressed morphologically in [aspect *pfv*] verb forms; in imperfective verb forms future tense is expressed periphrastically. Similarly, [voice *pass*] is only expressed synthetically in imperfective verb forms (where it actually borrows forms marked [reflexive *yes*]); in perfective verb forms it is expressed again periphrastically.

$$
\begin{bmatrix}
\text{SYN} \mid \text{MORSIG} &
\begin{bmatrix}
\text{ASPECT} & \{\textit{pfv,ipfv}\} \\
\text{VOICE} & \{\textit{act,pass}\} \\
\text{TENSE} & \{\textit{prs,pst,fut}\} \\
\ldots \\
\text{REPR}\langle\text{V,A}\rangle & \begin{bmatrix} \text{ASPECT} & \{\textit{pfv,ipfv}\} \\ \text{VOICE} & \{\textit{act,pass}\} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

Figure 6: Partial morsig for a Russian transitive verb

The somewhat complex mapping between content and form paradigms in Russian verbs is explored in greater detail in Spencer (2017). The precise characterization of the form or m-features for Russian verbs is controversial (as it is for most languages, including English). In Spencer (2017), for instance, I argue that the form paradigm has a single-valued m-[tense *prs-fut*] feature, accounting for both the present tense inflections of imperfective verb forms and the (identical) future tense inflections of perfective verb forms. Likewise, the content paradigm feature s-[tense *pst*] is expressed by a morphomic l-participle form ([vform *lptcp*]), which has no semantic interpretation of its own but which co-realizes s-[mood *conditional*] in conjunction with the particle *by*. Elsewhere, by default the l-participle realizes the content paradigm s-[tense *pst*] feature value. The specification [tense *pst*] has no form/m-feature counterpart.

The partial specification in Figure 6 also shows us that a transitive verb in Russian has four participial forms, listed in Table 1, where the parenthesized suffixes (-ij, …), (-yj, …) indicate the agreement inflections.

Table 1: Participles of Russian udar´it´ 'hit'

udar´-aju-šč(-ij, …)	imperfective active
udar´-aje-m(-yj, …)	imperfective passive
udar´-i-vš(-ij, …)	perfective active
udar´-on-n(-yj, …)	perfective passive

Given the MORSIG in Figure 6 the GPF can apply to a pairing ⟨*U*,π⟩, where *U* is the LI of UDAR′IT′ and π is a mnemonic shorthand for the set of participial features {[REPR⟨V,A⟩],{[ASPECT *pfv*],[VOICE *pass*]}}. In the original PFM models (PFM1 and PFM2) the paradigm function serves solely to define inflected forms (and periphrastic realizations of certain inflectional features). In terms of the lexical representational schemas discussed so far this means that the PF operates solely at the level of the FORM attribute. In GPFM the PF is generalized to four functions, operating over the FORM, SYN, SEM, LI attributes. The first of these, f_{form}, is the classical PF. For ordinary inflectional morphology the f_{syn}, f_{sem}, f_{li} functions have no material effect and behave like identity functions. Thus, the GPF for pure inflection collapses with the classical PF. However, for paradigmatic derivational morphology all four functions can introduce non-trivial changes as we saw earlier in the case of the derivation of DRIVER from DRIVE.

The case of transpositions such as participles is midway between that of pure or canonical inflection and derivation. The LI and SEM attributes remain unchanged but both FORM and SYN attributes have to be (re-)specified. Following Spencer (1999, 2013), in Spencer (2017) I assume that the category of a transposition is defined in terms of a complex SF role. A simplex verb has the SF role [ARG-ST|SF E] and an adjective the SF role [ARG-ST|SF A]. A participle is the adjectival representation of a lexeme with SF role E. The notion 'adjectival representation' is captured by defining a complex SF role ⟨A⟨E⟩⟩. To simplify the exposition I shall assume that the complex SF role is cashed out as a complex category label, [A [v]] (at the SYN level SYNCAT|[A [v]], at the FORM level MORCAT|[A [v]]).[4] The GPF for a participle, as defined by the attribute REPR⟨V,A⟩ will define a form with this new category, as shown in (4).

(4) $f_{syn}(⟨U,π⟩) = …$
     [SYN|SYNCAT V] ⟹ [SYN|SYNCAT [A [v]]]

The transpositional feature specification π will also define a restatement of the MORSIG attribute for the participle, as shown in (5).

(5) [ASPECT], [VOICE] ⊂ [SYN|MORSIG]

The statement in (5) is more specific than the default specification and hence it will override that default. However, the participles in Russian (unlike some languages) are actually adjectival forms. Therefore, their lexical representations must include a feature defining their agreement properties, which for convenience I will label CONCORD. This feature must be included there, in the participle's MORSIG. However, that fact, together with the definition of [CONCORD], is inherited from elsewhere in the grammar in the definition of adjectival inflection, as shown in (6).

(6) a. SF ⟨⟨A … ⟹ [CONCORD] ⊂ [SYN|MORSIG]

     b. [NUMBER], [GENDER], [CASE] ⊂ [CONCORD]

---

[4]In fact, it seems that the device of complex SF roles allows us to dispense entirely with traditional syntactic category labels (see also footnote 3).

Declaration (6) is so formulated that it applies to any word type whose 'outermost' category label is defined by the complex SF ⟨⟨A .... This will trivially include simplex adjectives, of course, but it also includes (true transpositional) participles (SF ⟨⟨A⟨E⟩⟩) and true relational adjectives (SF ⟨⟨A⟨R⟩⟩). Russian participles are well-behaved morphologically and so they will inherit very nearly all the FORM|MORSIG properties implied by the SYN|MORSIG specification.[5]

We are now in a position to state the full GPF defining the perfective passive participle, an extension of the GPF shown schematically in (3). This is shown in (7). It defines the object represented by the FS given in Figure 7.

(7) GPF for the perfective passive participle of UDARIT′ 'hit'
Where U is the Lexemic Index of the lexeme UDARIT′ 'hit' and π is the feature set [REPR⟨V,A⟩ [ASPECT *pfv*, VOICE *pass*]], the passive perfective participle stem form is defined by a generalized paradigm function, GPF(⟨U,π⟩) =

   (i) f_{form}(⟨U,π⟩) =
   [FORM STEM$_{ppp}$ = PHON STEM$_0$(U)⊕onn = /udar′onn/]

   (ii) f_{syn}(⟨U,π⟩) =

$$
\begin{bmatrix}
\text{SYN} & \begin{bmatrix}
\text{SYNCAT} & \begin{bmatrix} \text{A} & \begin{bmatrix} \text{V} \end{bmatrix} \end{bmatrix} \\
\text{ARG-ST} & \langle (x), y \rangle \\
\text{MORSIG} & \begin{bmatrix} \text{ASPECT} & pfv \\ \text{VOICE} & pass \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

   where (x) denotes the suppressed external argument of the passive.

   (iii) f_{sem}(⟨U,π⟩), f_{li}(⟨U,π⟩) are the 'identity function' (no change in representation).

The redefinition of the MORSIG attribute to include two attributes inherited from the verb base together with the new CONCORD attribute is part of the morphosyntactic definition of 'participle' in Russian. However, the subsequent inflection of the participle as an adjective follows entirely from the more general characterization of adjectives, independently of their origin. For instance, it is equally applicable to a purely derivational adjective such as *svet-l-yj* 'bright, light' from *svet* 'light', or *krov-av-yj (režim)* 'bloody (regime)' from *krov′* 'blood'. This means that the participle feature ensemble π defines an underspecified lexical representation which has exactly the same type of structure as an

---

[5]The main caveats here concern participles used as predicates, where there are a number of restrictions. The participle also retains crucial verb properties such as complementation and even quirky case assignment, so we need to ensure that those properties are inherited by the participle when the GPF is applied to π. This would require a much more detailed discussion of the lexical representation of verbs, so I refer the reader to Spencer (2017) where some of those details are worked out.

$$
\begin{bmatrix}
\text{FORM} & \begin{bmatrix} \text{STEM}_{PPP}|\text{PHON} & /\text{udar}'\text{onn}/ \end{bmatrix} \\[2ex]
\text{SYN} & \begin{bmatrix}
\text{SYNCAT} & \begin{bmatrix} \text{A} \begin{bmatrix} \text{V} \end{bmatrix} \end{bmatrix} \\[1ex]
\text{ARG-ST} & \langle (x), y \rangle \\[1ex]
\text{MORSIG} & \begin{bmatrix} \text{ASPECT PFV} \\ \text{VOICE PASS} \end{bmatrix}
\end{bmatrix} \\[4ex]
\text{SEM} & \text{'hit'} \\
\text{LI} & \text{UDARIT}'
\end{bmatrix}
$$

Figure 7: "Quasi-lexemic" feature structure for Russian passive perfective participle *udar̃onn*

$$
\begin{bmatrix}
\text{FORM} & \begin{bmatrix}
\text{STEM}_{PPP} |\text{PHON} & /\text{udar}'\text{onn}/ \\
\text{MORCAT} & \boxed{1}\begin{bmatrix} \text{DECL} & adj \end{bmatrix} \\
\text{MORSIG} & \boxed{2}
\end{bmatrix} \\[4ex]
\text{SYN} & \begin{bmatrix}
\text{SYNCAT} & \boxed{1}\begin{bmatrix} \text{A} \begin{bmatrix} \text{V} \end{bmatrix} \end{bmatrix} \\[1ex]
\text{ARG-ST} & \langle (\text{X}), \text{Y} \rangle \\[1ex]
\text{MORSIG} & \boxed{2}\begin{bmatrix}
\text{ASPECT} & pfv \\
\text{VOICE} & pass \\
\text{CONCORD} & \begin{bmatrix} \text{NUM} \\ \text{GEND} \\ \text{CASE} \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\[4ex]
\text{SEM} & \text{'hit'} \\
\text{LI} & \text{UDARIT}'
\end{bmatrix}
$$

Figure 8: Feature structure for passive perfective participle *udar̃onn* after default specification of MORSIG

independent simplex or derived adjectival lexeme. It is in this respect that the participle behaves as a quasi-lexeme, having the inflectional and morphosyntactic potential of an adjective but remaining a 'form' (more precisely, representation) of the base verb.

The analysis now brings us back to one of the questions posed earlier — is the representation in Figure 7 an object or a description?

If we regard Figure 7 as a description (vs. object) then it would presumably have to describe an object of type word. But this would entail that it describes some particular inflected form, say, the feminine instrumental plural. But the participle is not specified for those or any other CONCORD features, just as Sag's FS for LAUGH is underspecified for any inflectional feature set. This makes the participle FS look exactly like a lexemic entry, which *ex hypothesi* is an object not a description. It is this object that I have informally referred to as a quasi-lexeme. However, from the perspective of the grammatical system, it *is* a lexeme, albeit not one which is independent of its verb base.

The participle shares its Lexemic Index with the base verb in all its inflected forms. However, it is easy to imagine such a representation undergoing the simplest type of lexicalization, namely, to acquire its own unique LI. This would happen if the participle were recategorized as a simplex adjective, that is a member of the morphosyntactic category [A] rather than [A [V]]. This is then the representation of a transpositional lexeme of the type *interesting*. Russian, too, has such converted participial lexemes, though they often do not correspond to English transpositional lexemes. Examples are *potrˊasájuščij* 'amazing' from *potrˊasátˊ* 'to amaze', *izmúčonnyj* 'exhausted' from *izmúčitˊ* 'to exhaust' and many others (see Spencer 2017 for further discussion) . The crucial point is that these derived adjectival lexemes do not seem to differ from their verb bases in their semantics, just like true transpositions, yet they behave syntactically like independent lexemes.

6 Lexemes and types

We have arrived at the conclusion that the lexical representation of a participle is nondistinct in crucial ways from the representation of a lexeme, and for this reason the grammar will treat it as a linguistic object, akin to a lexeme. This invites the conclusion that the participle is, in fact, a subtype of the type *lexeme* in the hierarchy proposed by Sag (2012), say, *ptcp-lxm*. The problem would then be to define where *ptcp-lxm* fits in the type hierarchy. A participle inherits from both adjectives and verbs, as illustrated in Figure 9, adapting Sag's hierarchy for English (with obvious modifications for Russian).

This would be in keeping with Malouf's (2000) approach to deverbal nominalizations. However, there are a number of problems with this solution. One of these relates to the 'directionality' or 'headedness' of transpositions: a transposition is a representation of its base lexeme. In that respect a participial quasi-lexeme bears the same relationship to a verb that, say, the past tense form bears. But this is not captured in a hierarchy such as that sketched in Figure 9, where the relation between *verb-lxm, adj-lxm*, the two mothers of the participle *ptcp-lxm*, is equal. As a result, there will be no way of distinguishing between the adjectival representation of a verb and the verbal representation of an adjective (that is, a transpositional predicative adjective heading a finite clause and bearing inflections for verb features such as tense-mood-aspect-polarity or subject agreement).
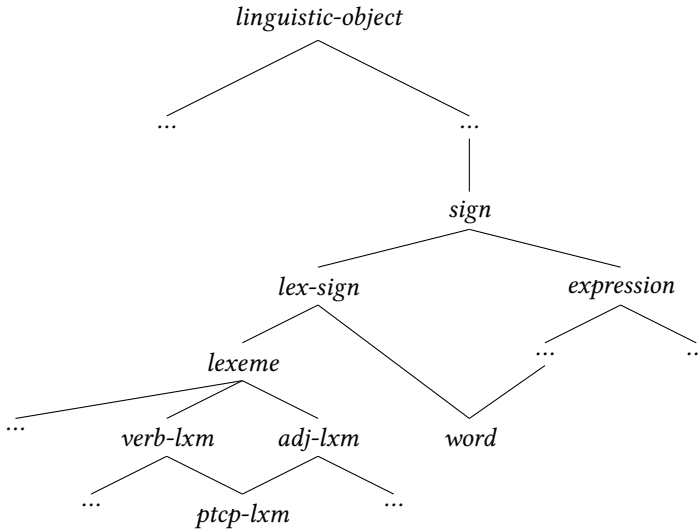
Figure 9: Revised partial type hierarchy

Perhaps, then we should adopt a different approach. Since participles are morphologically derived we can set up a construction type in SBCG (or a lexical rule in standard HPSG) which would perform the same role as the GPF applied to the REPR feature in GPFM. Sag defines two sorts of morphological construction relevant to us in this context, the *infl-cxt* and the *deriv-cxt*.

$$\text{(8)} \quad \textit{infl-cxt}: \begin{bmatrix} \text{MTR} & \textit{word} \\ \text{DTRS} & \textit{list(lexeme)} \end{bmatrix} \qquad \text{(Sag 2012: 115)}$$

$$\text{(9)} \quad \textit{deriv-cxt}: \begin{bmatrix} \text{MTR} & \textit{lexeme} \\ \text{DTRS} & \textit{list(lex-sign)} \end{bmatrix} \qquad \text{(Sag 2012: 119)}$$

The formulation in (9) additionally permits derivation from word forms, but in general derivation is defined over lexemes and to simplify the discussion I will assume that this is always the case. If we take a participle to be a subtype of *lexeme*, then participle formation will be a subtype of the derivational construction shown in (9).

One issue that has to be resolved when incorporating morphological models into lexicalist syntactic models arises from the fact that I-R models of morphology are generally based on default inheritance logic, while the syntactic models generally avoid the use of defaults and overrides. An important proposal for marrying the two systems is given by Bonami & Samvelian (2015) in the context of analysing periphrastic constructions in Persian (see also Bonami & Webelhuth 2012). The details depend on the specifics of their

analysis, but the overall import of their proposal is a 'meta-constraint' on signs of type *word*, such that a word is licenced in the (HPSG) syntax only if a corresponding representation of it is also licensed in the (PFM) morphology (Bonami & Samvelian 2015: 32). In effect, they treat the PFM morphology as a 'black box' whose outputs bear properties that can be recognized by the syntax.

The interface for canonical inflection works well. However, the proposals do not touch directly on other types of morphology, notably derivation and transpositions. Presumably, the interface principle could be extended so as to apply between a morphological engine and the HPSG lexicon. A major problem here is the lack of consensus over how to handle derivational morphology in I-R models. In PFM there has been very little discussion of derivation and no discussion of transpositions.[6] Concrete proposals for derivation and transpositions can be found in the Network Morphology model of Brown & Hippisley (2012) but it is not clear how that model would interface with syntax. Moreover, it is not clear how the Network Morphology model distinguishes between transpositions and canonical derivation, and between these and the (non-canonical) phenomenon of transpositional lexemes.

A detailed set of proposals for defining lexical relatedness is given in Spencer (2013), where I show that there are many other types of relatedness between words in addition to canonical inflection, canonical derivation and true (canonical) transposition. Any model of the lexicon has to be able to account for all these types. They include meaning-changing inflection, meaning-changing transposition, derivation which involves no change at all in FORM properties (morphologically inert derivation) and others. The conceptual problem here is that any of these types of relatedness might be part of the paradigmatic grammatical system in a given language, in which case the morphological means by which they are all expressed cannot be distinguished. Therefore, the same kind of machinery has to be deployed for paradigmatic derivation as for inflection. Given our current assumptions this means some form of paradigm function, defined in terms of defaults and overrides, and the challenge is therefore to ensure that the lexical representations so defined are compatible with the kinds of representations deployed in the syntax.

7 An agenda for lexical representation

The foregoing discussion raises more question than it answers, but the questions are important for lexicalist, constraints-based models generally, and for theories of lexical representation and morphology generally. Here, by way of a conclusion I summarize the main issues that have emerged.

- Are lexemes partially specified linguistic objects?

- What is the relationship between transpositional quasi-lexemes and canonical lexemes?

---

[6]This includes Stump (2016a,b), which are concerned exclusively with FORM/CONTENT mismatches.

- How do we ensure that I-R morphological models can interface with constraints-based syntactic models, including all aspects of paradigmatically organized morphology?

- To what extent can the morphological functions/constructions proposed in Sag (2012) be retained in their current form? To what extent can such constructional types, or their more traditional incarnations in standard HPSG, be made compatible with I-R models?

Finally, the most difficult question of all is the oldest and the one with the widest significance: what kind of a thing is a dictionary entry? Is it a real, mentally represented linguistic construction or is it merely the convenient fiction of the lexicographer? We cannot address this question without providing very explicit answers to the representational and ontological questions raised in this paper, and so I present my discussion of those questions as a modest contribution towards answering the much bigger question.

Abbreviations

ARG-ST	Argument Structure (attribute)
FS	feature structure
GPFM	Generalized Paradigm Function Morphology
HPSG	Head-driven Phrase Structure Grammar
I-R	inferential-realizational (model)
LI	Lexical/Lexemic Index
LID	Lexical Identifier
PFM	Paradigm Function Morphology
SBCG	Sign-Based Construction Grammar
SF	semantic function (role)
wfr	Word Formation Rule

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. Oxford: Oxford University Press.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: MIT Press.

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.

Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2). 327–382.

Bonami, Olivier & Gregory T. Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge Handbook of Morphology*, 449–481. Cambridge: Cambridge University Press.

Bonami, Olivier & Gert Webelhuth. 2012. The phrase-structural diversity of periphrasis: A lexicalist account. In Marina Chumakina & Greville G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms* (Proceedings of the British Academy 180), 141–167. Oxford: Oxford University Press/British Academy.

Brown, Dunstan, Marina Chumakina, Greville G. Corbett, Gergana Popova & Andrew Spencer. 2012. Defining 'periphrasis': Key notions. *Morphology* 22. 233–275.

Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology: A defaults based theory of word structure.* Cambridge: Cambridge University Press.

Chaves, Rui P. 2014. Grammatical alignments and the gradience of lexical categories. In Philip Hoffmeister & Elisabeth Norcliffe (eds.), *The Core and the Periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 167–220. Stanford: CSLI publications.

Culicover, Peter & Ray S. Jackendoff. 2005. *Simpler Syntax.* Oxford: Blackwell Publishers.

Dalrymple, Mary, Louise Mycock & John Lowe. 2015. Morphology in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG15 conference*, 43–62. Stanford: CSLI Publications.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Haspelmath, Martin. 1996. Word-class-changing inflection and morphological theory. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1995*, 43–66. Dordrecht: Kluwer Academic Publishers.

Helimski, Eugene. 1998. Selkup. In Daniel Abondolo (ed.), *The Uralic Languages*, 548–579. London: Routledge.

Jackendoff, Ray S. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3). 639–671.

Jackendoff, Ray S. 1990. *Semantic structures.* Cambridge: The MIT Press.

Jackendoff, Ray S. 1997. *The architecture of the language faculty.* Cambridge: The MIT Press.

Kuznecova, Ariadna Ivanovna, Evgenij Arnol'dovič Xelimskij & Elena Vladislavovna Gruškina. 1980. *Očerki po sel'kupskomu jazyku.* Moscow: Izdatel'stvo moskovskogo universiteta.

Malouf, Robert. 2000. *Mixed categories in the hierarchical lexicon.* Stanford: CSLI Publications.

Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2000*, 71–96. Dordrecht: Kluwer Academic Publishers.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 69–202. Stanford: CSLI Publications.

Spencer, Andrew. 1998. Relational adjectives and the redundancy of lexical categories. In Geert Booij, Angela Ralli & Sergio Scalise (eds.), *Proceedings of the First Mediterranean Conference of Morphology*, 19–29.

Spencer, Andrew. 1999. Transpositions and argument structure. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1998*, 73–102. Dordrecht: Kluwer Academic Publishers.

Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press.

Spencer, Andrew. 2016. How are words related? In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 1–26. Amsterdam: John Benjamins Publishing Co.

Spencer, Andrew. 2017. Split-morphology and lexicalist morphosyntax: The case of transpositions. In Claire Bowern, Laurence Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond)*, 385–421. Berlin: Language Science Press.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2002. Morphological and syntactic paradigms: Arguments for a theory of paradigm linkage. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2001*, 147–180. Dordrecht: Kluwer Academic Publishers.

Stump, Gregory T. 2006. A theory of heteroclite inflectional paradigms. *Language* 82. 279–322.

Stump, Gregory T. 2016a. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2016b. Paradigms at the interface of inflectional morphology. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 27–58. Amsterdam: John Benjamins Publishing Co.

Chapter 13

Troubles with flexemes

Anna M. Thornton

University of L'Aquila

This paper investigates an aspect of the notion *flexeme* (French *flexème*), introduced by Fradin & Kerleroux (2003), Fradin (2003). After a brief review of how this concept developed in these authors' work, and of how these authors conceive of lexemes (Section 2), the relation between flexemes and overabundance (Thornton 2011, 2012) is explored. Overabundance is introduced in Section 3, and Section 4 is devoted to some case studies, from Italian and other languages. It is shown that a single lexeme can map to more than one flexeme – and overabundance results from this mapping. Besides, it is shown that flexemes differing from each other in parallel ways can have various relations with lexemes: in some cases, mapping to different flexemes distinguishes two lexemes that are homophonous in their citation form (e.g., Italian SUCCEDERE[1] 'happen' with PST.PTCP *successo* and SUCCEDERE[2] 'succeed' with PST.PTCP *succeduto*), while in other cases flexemes that differ from each other in a way parallel to the previous one map to a single overabundant lexeme (e.g., Italian PERDERE 'lose' with PST.PTCP *perso* and *perduto*). I conclude that the distinction between lexemes and flexemes first proposed by Fradin & Kerleroux (2003) and Fradin (2003), as well as their definition of lexeme, based on semantic and constructional coherence rather than on inflectional coherence, is useful even beyond the area of lexeme formation for which it was originally proposed.

1 Introduction

In a paper titled "Troubles with lexemes", Bernard Fradin and Françoise Kerleroux (2003) laid the bases for a critique of the commonly held notion of lexeme, drawing data from the realm of word-formation. They observed at the beginning of their paper:

> the lexeme is supposed to constitute one lexical unit. **This unicity is guaranteed by inflection on the one hand** and by the semantic content of the lexeme, which is supposed to be unique, on the other (Fradin & Kerleroux 2003: 177, emphasis mine).

They proceeded then to show that the objects to which word-formation rules apply – which they propose to call lexemes, partially modifying the usual definition of this term – are semantically fully specified objects, that are, however, unspecified for inflection. In

the concluding section of that paper, they propose to distinguish three different theoretical entities: lexemes ("lexical individuals defined by the conjunction of three properties: category, underspecification for inflection, full specification for meaning", Fradin & Kerleroux 2003: 193), syntactic words (which are inflected, categorized, and fully specified for meaning), and a third entity, which they propose to call *inflecteme* in English and *flexème* in French (see also Fradin 2003: 259). Objects of this third type are categorized, uninflected and underspecified for meaning.

In this short contribution, I will discuss some aspects of these entities that have come to the fore of the debate in morphology after the publication of Fradin & Kerleroux (2003) and Fradin (2003). I prefer to refer to these units as *flexèmes*, because I think that the intentional and witty phonological and orthographic overlap with *lexème* 'lexeme' is too good to be lost, and as an *hommage* to the authors who first proposed this term. Following Fradin (forthcoming), in this paper I will use the English adaptation *flexeme*.

The paper is organized as follows: Section 2 reviews the development of the concept of flexeme; Section 3 introduces the concept of overabundance in inflectional paradigms; Section 4 presents several case studies from Italian and other languages, illustrating cases in which a single lexeme is overabundant in one or more cells, i.e., maps to two distinct flexemes; Section 5 concludes.

2 What are flexemes?

In different contributions by Bernard Fradin, sometimes in collaboration with Françoise Kerleroux, the concept of *flexème*/flexeme is presented differently: its coverage seems to have grown with time, probably in consequence of our growing understanding of the workings of inflectional morphology in the early years of the third millennium.

In Fradin & Kerleroux (2003: 193) the concept seems to be equivalent to that of *stem* (in the sense, e.g., of Aronoff 1994):

> This unit [i.e., the inflecteme/*flexème*] lacks semantic specification since it functions as the "inflectional stem".

However, the authors seem to have something more than just a single stem in mind, since immediately after this definition they observe: "This is correlated to the fact that "no semantic constraints hangs [sic] over the application of inflectional rules" (Corbin 1987: 6)". So the idea that flexemes have to do with instructions for building all the inflected forms that realize a lexeme seems to have been present already in Fradin & Kerleroux (2003).

Fradin (2003: 259) states that

> Les flexèmes […] comportent […] des informations relevant […] du syntactique interne (les différents thèmes flexionnels, sous forme hiérarchisée, s'il en existe plusieurs […]).

So the concept of flexeme seems to have developed from being used to refer to a stem to being used to refer to the whole stem-set of a lexeme. In Fradin (forthcoming) a new development appears.[1] The author, dealing with verbs, distinguishes between verbs as morphological units, called "morphological verbs", and verbs as lexical units, called "verbal lexemes". He states that "[m]orphologically, a V is defined by its inflectional paradigm", and maintains that the two French verbs RESSORTIR¹ ((de Y): *il ressort, il ressortait...*) 'go out again' and RESSORTIR² ((à Y): *il ressortit, il ressortissait...*) 'come under' "constitute distinct 'flexemes', see Fradin & Kerleroux (2003) [...] because the set of their word-forms is not identical" (Fradin forthcoming: 4).

In this passage, Fradin attributes to Fradin & Kerleroux (2003) a fully developed concept of flexeme, in which a flexeme contains all the information needed to generate all the inflectional forms in a paradigm: not only the information about which stem to select, but also inflectional class and realization rules for the different inflected forms. Roughly, it seems to me, a flexeme now corresponds to the entities called *form paradigm* and *realized paradigm* in paradigm-linkage theory (Stump 2016). Fradin (forthcoming) also equates the notion of flexeme with that of Paradigm Identifier adopted by Bonami & Tribout (2012). In turn, Bonami & Tribout (2012) state that their notion of Paradigm Identifier "[c]aptures Fradin & Kerleroux (2003)'s notion of a flexeme: a family of lexemes with the same inflectional paradigm" (Bonami & Tribout 2012: slide 16).[2]

Papers such as Fradin (forthcoming) and Bonami & Tribout (2012) address the question of how to deal with objects that are semantically different but morphologically identical, such as CIRAGE¹ 'polishing' and CIRAGE² 'shoe polish', or PERLER¹ 'sew beads on' and PERLER² 'form beads on', which share a flexeme (a form paradigm and a realized paradigm) but are different lexemes.[3]

In this paper, on the contrary, I will explore the issue of objects that are the same lexeme, in the sense of Fradin & Kerleroux (2003) and Fradin (2003, forthcoming), but can be realized, to variable degrees, by different flexemes.

3 Overabundance

In recent years, attention has been drawn to the phenomenon of overabundance in inflectional paradigms (Thornton 2011, Stump 2016: 147-151). Overabundance is defined as the situation in which two or more forms are available to realize the same cell in an inflectional paradigm; in terms of paradigm linkage theory, one content cell has more than one realization. Stump (2016: 148) gives an example from English. Consider the verbs SEEM, MEAN, and DREAM, and the realizations of their past tense: ⟨SEEM, {past}⟩ is realized by *seemed*, ⟨MEAN, {past}⟩ is realized by *meant*, and ⟨DREAM, {past}⟩ can be realized either by *dreamed* or by *dreamt*. The two (or more) forms that realize the same cell are sometimes called cell mates (Thornton 2011).

---

[1]The notion of flexeme is not mentioned in Fradin & Kerleroux (2009).

[2]The notion of Paradigm Identifier is clearly articulated by Bonami & Crysmann (this volume).

[3]This phenomenon is labelled "homomorphy" by Stump (2016: 65): "homomorphic lexemes are lexically and semantically distinct but alike in every detail of their morphology". English examples are WEAR¹ 'have on (an article of clothing)' and WEAR² 'erode'.

How does overabundance relate to the notion of flexemes? Does the existence of distinct but synonymous realizations for a given content cell force us to recognize distinct flexemes linked to a single lexeme?

Fradin (forthcoming) analyzes cases such as PERLER[1] 'sew beads on' and PERLER[2] 'form beads on' as distinct lexemes linked to the same flexeme. The case of *dreamed* 'dream.PST' and *dreamt* 'dream.PST' appears to be a mirror image of this case, with distinct flexemes linked to a single lexeme. The existence of such a state of affairs would be predicted in Fradin's theory, in which lexemes, defined as categorized and semantically fully specified but uninflected objects, are autonomous from the flexemes that provide instructions for the realization of their inflected forms. Recognizing the possibility that a single lexeme may be linked to two (or more) flexemes implies that a difference in inflectional realization cannot be invoked as one of the criteria that allow to distinguish between different lexemes vs. simply different senses/acceptations of a polysemous lexeme, as was sometimes done in traditional discussions of the homonymy/polysemy distinction (see e.g. Ullmann 1957: 127–132).[4] Indeed, flexemes that are distinct in parallel ways may map to a single lexeme or to distinct lexemes – where the criterion for recognizing distinct lexemes is semantic and constructional difference, as proposed by Fradin & Kerleroux (2003, 2009) and Fradin (2003, forthcoming).

In the following section, I will review some data that show that the mapping between flexemes and lexemes can be of several kinds.

4 Non-canonical mappings between lexemes and flexemes

In this section, I will present data, mostly from well-studied cases in familiar languages, that show how one and the same difference in inflectional realization may map either to distinct lexemes or to a single overabundant lexeme.

4.1 Case study 1: Noun plurals

Nouns in which apparently more than one plural form pairs with a single singular form are very easy to find in language descriptions. Usually authors assume, at least implicitly, the admittedly vaguely defined criterion of 'difference in meaning' to decide whether specific cases represent distinct lexical items with homophonous singular forms or a single lexical item which is overabundant in its plural cell(s). Besides, since data are usually found in works which aim at description rather than at theoretical analysis, often authors leave the matter undecided, because it is not necessary for descriptive purposes to establish whether a certain case is an instance of homonymy or polysemy; on the other hand, cases in which no semantic distinction is observable between two or more different plural forms are usually highlighted by authors of descriptions.

---

[4]Remember also the observation by Fradin & Kerleroux (2003: 177) quoted in Section 1, that unicity of a lexeme "is guaranteed by inflection" as well as by the semantic content.

Cases such as the English and Breton ones in (1) and (2) are typical:

(1) English (Aronoff 2000: 347)

    a. SG *brother* PL *brothers*
       'male sibling'

    b. SG *brother* PL *brethren*
       'fellow member of a profession, society or sect'

(2) Breton (Trépos 1980) [5]

    a. SG *eskob* PL *eskibien*
       'bishop'

    b. SG *eskob* PL *eskobou*
       'kingpin'[6]

In these cases most authors argue that the meanings of the two items are sufficiently distinct to allow us to consider them as distinct lexemes, which happen to be homophonous in their singular form.[7] In these cases, then, we have a 1:1 mapping between lexemes and flexemes, with the extra quirk represented by the fact that two distinct flexemes have homophonous singular forms.

However, by perusing the whole description of Breton noun plural offered by Trépos (1980), we discover that 'bishop' can have as many as three different plural forms (3a), and the same is true for 'coat' (3b):

(3) Breton (Trépos 1980: § 149)

    a. SG *eskob* PL *eskibien/eskobed / eskeb*
       'bishop'

    b. SG *mantell* PL *mentell/mentellou/mentilli*
       'coat'

A similar situation is common in Modern Standard Arabic, where nouns often have several plural forms; authors of descriptions usually comment on when they would prefer to assign the different plural forms to distinct lexemes, on the basis of a clear distinction in meaning, as in (4a vs. 4b, 4c vs. 4d), and when the different plural forms can be used interchangeably, and must be recognized as realizing the same lexeme, as in (5a-5b).

---

[5]Breton nouns inflect only for number.

[6]The French gloss given by Trépos (1980: 73) for *eskibien* is 'chevilles d'attelage'.

[7]Even if (1b) obviously derives from (1a) by means of a metaphorical extension.

(4) Modern Standard Arabic (Holes 2004, Kaye 2007)[8]

    a. SG *bayt* PL *buyu:t*
       'tent', 'house'

    b. SG *bayt* PL *ʔabya:t*
       'verse of poetry'

    c. SG *maktab* PL *maka:tib*
       'office'

    d. SG *maktab* PL *maktaba:t*
       'library', bookshop'

(5) Modern Standard Arabic (Kaye 2007)

    a. SG *ʕayn* PL *ʔaʕyun/ʕuyūn*
       'eye'

    b. SG *sāriq* PL *sāriqūn, saraqa, surrāq*
       'thief'

With respect to nouns such as those in (5), Kaye (2007: 234–235) observes that "[t]here are many nouns with two or more plural variants without any difference in meaning", while on the nouns in (4a-4b) he states that "[i]t is best to regard […] *bayt* as distinct lexemes" (Kaye 2007: 234).

Authors like Kaye rely on meaning distinction as the only criterion for distinguishing between lexemes, and (implicitly) accept the possibility that what they conceive of as single lexemes (like the ones in (5)) may have overabundant realizations in one or more cells, i.e., may map to more than one flexeme. Other authors, however, reject this possibility, and assume that a difference in inflectional realization (a difference in flexemes) must always correspond to a difference in lexemes. A champion of such a position is Paolo Acquaviva, who has articulated his point of view in his work on Italian double noun plurals (Acquaviva 2008).

Italian nouns have inherent gender (with two values: feminine and masculine) and inflect for number (with two values: singular and plural). About 20 Italian nouns are usually described as overabundant in the plural (e.g., in traditional reference grammars such as Battaglia & Pernicone 1954). These nouns have a singular form in *-o* which is masculine, a plural form in *-i* which is masculine, and a plural form in *-a* which is feminine. Some representative examples are given in (6):

(6) Italian (Acquaviva 2008, Thornton n.d.)

    a. SG *braccio* PL *braccia*/*bracci*
       'arm'

---

[8]MS Arabic nouns inflect for number (singular, dual, plural), case (nominative, genitive, accusative, with a syncretism of genitive and accusative (sometimes called oblique) in non-singular forms), and definiteness (definite, indefinite). In systems in which nouns inflect for other features besides number, if multiple forms with the same number value exist they are predicted to exist in all cells; e.g., in Arabic, multiple plural forms are predicted to exist in all case and definiteness values.

b. SG *corno* PL *corna*/*corni*
'horn'

c. SG *ginocchio* PL *ginocchia*/*ginocchi*
'knee'

d. SG *membro* PL *membra*/*membri*
'limb'/'member'

Acquaviva's position is that plurals in -*a*, independently of whether they differ in meaning from the plurals in -*i* with which they share a root, are distinct lexemes, *pluralia tantum*, derivationally related to the lexemes in -*o*/-*i* with which they share a root:

> plurals in -*a* [...] are lexical plurals: **distinct, inherently plural nouns**, related to the base noun by a word-formation process. (Acquaviva 2008: 123, emphasis mine)

> ***Braccia* 'arms' is not the plural of *braccio* 'arm'**; it is an inherently plural lexeme, derived from the same root as *braccio*/*bracci* (Acquaviva 2008: 157, emphasis mine)

He brings forward several arguments for his position, which are reviewed in Thornton (n.d.: 430–438), where it is shown that one of them (agreement with conjoined singular NPs) is based on a misunderstanding of the workings of Italian agreement resolution rules, and can be dismissed as irrelevant. His other arguments will be illustrated here.

The first argument is purely metatheoretical. Acquaviva states it as follows:

> The simple fact that a number of plurals in -*a* do not block their regular alternants in -*i* is enough to prove the point, **if we take seriously inflectional disjunctivity** (Acquaviva 2008: 145, emphasis mine).

This argument boils down to positing as a theoretical requirement the non-existence of overabundance, or the impossibility of a single lexeme to map to distinct flexemes. Such a choice eliminates the problem we are investigating by denying its existence, rather than by offering a solution. However, if we assume, as done in the canonical approach to morphological typology (Corbett 2005, 2006, 2007), that inflectional disjunctivity and lack of overabundance are only canonical properties of lexemes, rather than inviolable theoretical requirements, the problem reappears and requires to be investigated.

Another argument put forward by Acquaviva to establish that plurals in -*a* are distinct lexemes from their co-radicals in -*o*/-*i* is consonant with Fradin & Kerleroux's (2003) view of lexemes: Acquaviva observes that some plurals in -*a* appear to be the bases of word-formation processes. An example would be *cornificare* 'to make a cuckold of', which Acquaviva analyzes as derived from *corna* 'horns' (6b); *cornificare* is synonymous with the idiom *fare*/*mettere le corna* 'to make a cuckold of, lit. to make /put horns.F.PL', which is never realized by *fare/*mettere i corni,* with 'horns.M.PL'. On this basis, one can presume that *corna*, and not *corni*, is the base of *cornificare*. However, the idiom *fare*/*mettere un corno* 'to make a cuckold of, lit. to make/put a horn.M.SG' is also attested, so one cannot exclude that the base of *cornificare* is a non-defective lexeme *corno/corna*, rather than

a *plurale tantum* defective noun *corna*. In any case, this argument boils down to recognizing different lexemes when there is a difference in semantics and in the possibility of appearing in certain constructions, as proposed also by Fradin & Kerleroux (2003, 2009), Fradin (2003). This is orthogonal to the question whether a lexeme, defined on the basis of its semantics and distribution in constructions, can be overabundant in one or more cells. If we show that two plural forms appear in the same set of environments and constructions, they must be recognized as belonging to the same lexeme (unless, like Acquaviva, one wants to posit a difference of inflectional realization as sufficient for recognizing distinct lexemes, regardless of the equal semantics and distribution of the forms). Thornton (2010-2011) has shown, by means of corpus-based evidence, that in some cases two plurals in *-i* and *-a* are used interchangeably in the same context, and cannot therefore be considered as instances of distinct lexemes in Fradin & Kerleroux (2003)'s sense. This is the case for *ginocchi /ginocchia* 'knees' (6c), both of which appear interchangeably (as well as the singular form *ginocchio*) in a number of syntactic environments (Thornton n.d.: 465). In the case of *membra* and *membri* (6d), instead, there is evidence to posit two distinct lexemes, MEMBRO¹ 'limb (body part)', which is [–human], and MEMBRO² 'member (of a committee, organization, etc.)', which is [+human], and is obviously derived from MEMBRO¹ by means of a metaphoric extension. MEMBRO² is not overabundant: its plural is always *membri*, and it is the base of a derived feminine MEMBRA 'female member (of a committee, organization, etc.)', PL *membre* (Thornton 2014). MEMBRO¹ isn't overabundant either: its plural is *membra* 'limbs'; however, contrary to Acquaviva's analysis, it is not defective: the singular *membro* in the sense of 'limb, body part' is attested (cf. Thornton n.d.: 463, fn. 38). These examples show that each case in which we observe, in Italian, a feminine plural in *-a* and a masculine one in *-i* based on the same root, must be analyzed in its own right: the parallelism in the flexemes does not guarantee a parallelism in the lexemes. *Membri* and *membra* belong to different lexemes (defined according to Fradin & Kerleroux's (2003) and Fradin's (2003) semantic criteria), while *ginocchi* and *ginocchia* belong to the same lexeme – if we admit the possibility of overabundance, i.e. of a single lexeme mapping to more than one flexeme. The case of *bracci* and *braccia* is particularly complex: these very frequent forms, if submitted to Fradin & Kerleroux's (2003) and Fradin's (2003) criteria for the recognition of distinct lexemes, map to several semantically distinct lexemes, some of which are overabundant in the plural (e.g., 'arm (body part)'), while others select only one plural form (e.g., 'ell (measure of length)' selects *braccia*). Again, the mapping between lexemes and flexemes is not 1:1, as shown in Figure 1.
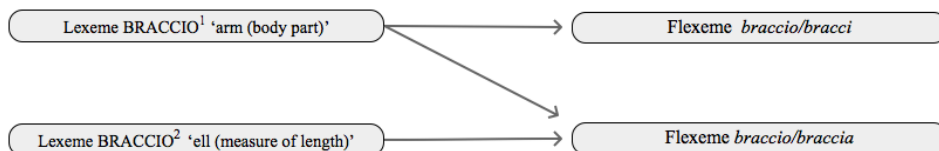
Figure 1: Mapping between two lexemes and two flexemes in Italian

4.2 Case study 2: Past participles

Another area in which mapping between semantically defined lexemes and flexemes is not always 1:1, and in which differences in flexemes do not invariably coincide with differences in lexemes, is verbal inflection.

In some cases, two semantically and constructionally distinct lexemes have quite different realized paradigms, even if the citation forms coincide. A case in point is that of Italian SUCCEDERE¹ 'happen' and SUCCEDERE² 'succeed'. SUCCEDERE¹ 'happen' is an impersonal verb, which is used only in 3ʳᵈ person forms; its PST.PTCP is *successo*. SUCCEDERE² 'succeed' is a bi-argumental verb; its second argument is introduced by the preposition *a* 'to'; it has a full set of realized forms, and its PST.PTCP is overabundant, according to various authoritative sources (Zingarelli 2016, Serianni 1988): it can be either *succeduto* or *successo*. The forms are shown in (7):

(7) Italian (Zingarelli 2016, Serianni 1988, personal knowledge)
lexeme	PST.PTCP
SUCCEDERE¹ 'happen'	*successo*
SUCCEDERE² 'succeed'	*successo/succeduto*

From (7) it would appear that SUCCEDERE¹ maps to a single flexeme, while SUCCEDERE² maps to two. However, for SUCCEDERE² the form *succeduto* is prescribed over *successo* by normatively oriented sources like Serianni (1988: § 316), and the most recent example of *successo* as a form of SUCCEDERE² cited by Serianni (1988) is from a novel published in 1960. Investigation of contemporary usage in corpora is difficult for practical reasons: *successo* has 87763 tokens in the corpus *la Repubblica 1985-2000* (380M tokens; I will consider data from this corpus as representative of contemporary Italian usage of *successo* and *succeduto*), making it impractical to examine each token to assign it to either SUCCEDERE¹ or SUCCEDERE². Besides, *successo* is a homonym of the SG form of the noun SUCCESSO 'success'. However, manual examination of the first 200 random tokens of the string *successo a*, which corresponds to both 'happened to' and 'succeeded to', suggests that in this context *successo* always realizes SUCCEDERE¹ 'happen', while, as expected, all the 374 tokens of *succeduto* in the corpus *la Repubblica 1985-2000* realize SUCCEDERE² 'succeed'. So, as far as the PST.PTCP is concerned, it appears that in contemporary Italian the two lexemes SUCCEDERE¹ 'happen' and SUCCEDERE² 'succeed' map to different flexemes.[9]

We can compare this situation with that of the verb PERDERE 'lose', which is genuinely overabundant in its PST.PTCP, as shown in (8):

---

[9]Things are more complicated with the simple past, which is (exemplifying with 3SG forms) *successe* for SUCCEDERE¹ and overabundant for SUCCEDERE² (*successe/ succedette;* a third form, *succedé*, is theoretically possible as 'succeed.PST.3SG', but it is not attested in the corpus *la Repubblica 1985-2000*). *Successe* has 1263 tokens and *succedette* 43 tokens in this corpus; all tokens of *succedette* realize SUCCEDERE² 'succeed'; manual examination of the 14 tokens of the string *successe a* 'happened to/succeeded to' reveals that in most cases it realizes SUCCEDERE¹ 'happen', but in 2 cases *successe* realizes SUCCEDERE² 'succeed', confirming that this verb is overabundant in its simple past. However, the simple past does not belong to the native grammar of many speakers of Italian, for whom it is a learned form; so it is unwise to draw strong conclusions from these data. Overabundance in the simple past in Italian shall be left for further research.

(8) Italian (personal knowledge)
     lexeme   PST.PTCP
     perdere 'lose'   *perso/perduto*

Speakers appear unaware of conditions regulating the selection of either one of the two forms, to the point that many speakers asked the *Accademia della Crusca*'s linguistic consulting service for advice on when to use each form (Thornton 2016). Speakers seem convinced that rules that govern a complementary distribution of the two forms should exist, but indeed the distribution of the two PST.PTCP forms is not complementary: they can be used interchangeably in many contexts, including idioms, as shown in (9a-b) and already shown by Thornton (2011: 369); the only case in which only one form is used is in titles of works of art (9c). Representative data, with frequencies from the corpus *la Repubblica 1985-2000* when relevant, are presented in (9).

(9) Italian (Thornton 2011: 369, Thornton 2016, personal knowledge)

     a. *occasione perduta* 291 / *occasione persa* 83
         'a chance lost'

     b. *perso la guerra* 109 / *perduto la guerra* 32
         'lost the war'

     c. *I predatori dell'arca perduta/*persa*
         'Raiders of the lost ark'

     d. *Alla ricerca del tempo perduto/*perso*
         À la recherche du temps perdu by Proust, literally 'In search of lost time';
         English translation's title 'Remembrance of things past'

     e. *Paradiso perduto/*perso*
         'Paradise lost'

This case study shows again a case in which similar differences in flexemes do not map in a parallel way to differences in lexemes: while SUCCEDERE[1] 'happen' and SUCCEDERE[2] 'succeed' map to distinct flexemes, in which the PST.PTCP forms are *successo* and *succeduto* respectively, PERDERE 'lose' maps to two flexemes, distinct from each other in a way parallel to the flexemes SUCCEDERE[1] and SUCCEDERE[2], and its PST.PTCP can be realized by both *perso* and *perduto*.

4.3 Systematic overabundance and overabundance in all cells

The two case studies illustrated above have shown examples in which there is an overabundant cell in the form paradigm and the realized paradigm of certain lexemes (such as Italian BRACCIO[1] 'arm' and PERDERE 'lose'). Technically, this should be enough to recognize that such lexemes map to distinct flexemes. However, if one wished to take into

account quantitative considerations, one might want to deal with these cases by recognizing a minor "exception", and still posit a single flexeme with a single exceptional, overabundant cell.

However, overabundance is not always confined to a single cell. In this section I will illustrate cases of "systematic overabundance" (Bonami & Stump 2016: 469), in which entire slabs or subparadigms are involved,[10] and cases of overabundance in all cells. These cases definitely deserve consideration in the context of exploring the possible deviations from a 1:1 mapping between lexemes and flexemes.

A particularly clear example of systematic overabundance is found in Spanish, where all verbs have two complete sets of forms, built by means of different endings, in the Imperfect Subjunctive, as shown in Table 1 for the verb *haber* 'have'.

Table 1: Imperfect Subjunctive of Spanish *haber* 'have'

	-ra set	*-se* set
1SG/3SG	*hubiera*	*hubiese*
2SG	*hubieras*	*hubieses*
1PL	*hubiéramos*	*hubiésemos*
2PL	*hubiérais*	*hubiéseis*
3PL	*hubieran*	*hubiesen*

Despite a suggestion by Bolinger (1956) that there is some subtle semantic difference between the two sets of forms, contemporary descriptions agree that "these two sets of forms are interchangeable" (Butt & Benjamin (2000: 167); see also Rojo & Veiga (1999: 2910): "las formas en *-ra* y *-se* son hoy por hoy perfectamente equivalentes"). Spanish verbal lexemes, then, appear to systematically map to two flexemes, which are distinct in the Imperfect Subjunctive forms – unless one wants to build overabundance within the definition of Spanish verbal flexemes, exactly because of its systematicity.

In other cases, however, we encounter overabundance in all cells of a given lexeme, but this is not systematic across all the lexemes within that part of speech in the language; therefore, the possibility of building overabundance in the definition of the flexemes to which these lexemes map is not viable, and we must recognize a 1:2 mapping between lexemes and flexemes.

A case in point is that of the Italian noun ORECCHIO 'ear'. This noun can be described as overabundant in all its cells: it has two SG forms and two PL forms, as shown in (10):

---

[10]The notion of slab has been introduced by Carstairs (1987: 81), who defines it as "a subset of the macroinflexions within one paradigm consisting of all the macroinflexions which are associated with some specified morphosyntactic property". His examples from Latin noun paradigms are the singular slab (all singular case-forms) or the genitive slab (GEN.SG and GEN.PL). The notion of sub-paradigm is used in a variety of senses, most commonly by scholars with a background in Slavonic languages. It aims at capturing subsets of cells in a paradigm which share more than just one feature value, such as verb tenses (the Present Indicative, the Present Subjunctive, etc.).

(10)   Italian (personal knowledge)
       lexeme   ORECCHIO
       SG forms   *orecchio* (M)   *orecchia* (F)
       PL forms   *orecchi* (M)   *orecchie* (F)

Of course, one could posit two distinct lexemes, ORECCHIO(M) and ORECCHIA(F), on the basis of the difference in gender, which is canonically an inherent fixed feature value in nouns. However, we already know from the cases discussed in Section 4.1 that Italian has nouns which change their gender value from the singular to the plural. Besides, according to Fradin's (2003) and Fradin & Kerleroux's (2003) definition of lexeme, which recognizes a single lexeme on the basis of identity of meaning and constructional distribution, the different forms in (10) appear to belong to the same lexeme, since they can be used interchangeably in the same contexts, even in idioms (11a-11b), as shown by the examples in (11):

(11)   Italian (personal knowledge; frequency data from the corpus *la Repubblica 1985-2000*)

   a. *fare orecchi da mercante* 18 / *orecchie da mercante* 139
      'to turn a deaf ear' lit., to do merchant's ears

   b. *dare una tirata d'orecchi* 122 / *tirata d'orecchie* 92
      'to give a dressing-down' lit., to give a tug of ears

   c. *occhi e orecchi* 19 / *occhi e orecchie* 68
      'eyes and ears'

   d. *da un'orecchia all'altra* 2 / *da un'orecchio all'altro* 13
      'from one ear to the other'

So Italian ORECCHIO can be analyzed as a single lexeme mapping to two flexemes, as shown in Figure 2.
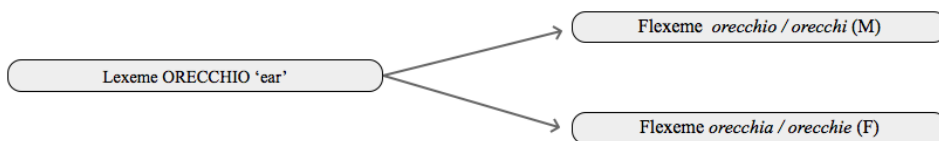
Figure 2: Mapping between one lexeme and two flexemes in Italian.

The flexemes are distinct; they instantiate nouns of different inflectional classes, while most Italian noun lexemes map to only one flexeme, belonging consistently to only one gender and one inflectional class, as shown by the examples in Table 2.

Lexemes such as BRACCIO[1], GINOCCHIO and ORECCHIO are non-canonical, in that they map to more than one flexeme, as seen above.

Table 2: Italian (personal knowledge).

lexeme	flexeme		gloss
	SG	PL	
OCCHIO (M)	*occhio*	*occhi*	'eye'
BOCCA(F)	*bocca*	*bocche*	'mouth'
MANO(F)	*mano*	*mani*	'hand'

The last case of non-canonical mapping between lexemes and flexemes that I will examine is that of certain Italian verbs, that are described as able to inflect according to two different conjugations; these are called "verbi sovrabbondanti" by Serianni (1988).

Grammars usually address together two kinds of such verbs: those in which the difference in conjugation does not bring along a difference in meaning (12a), and those in which the difference in inflectional class goes hand in hand with a difference in meaning (12b).

(12) Italian (Serianni 1988, personal knowledge)

    a.   i. *adempiere/adempire*
              'fulfil'

        ii. *compiere/compire*
              'complete'

      iii. *empiere/empire*
              'fill'

      iv. *riempiere/riempire*
              'fill'

    b.   i. *abbonare/abbonire*
               'subscribe'/'appease'

        ii. *arrossare/arrossire*
              'make red', 'dye red'/'redden', 'flush'

      iii. *fallare/fallire*
              'make a mistake'/'fail'

      iv. *imboscare/imboschire*
              'hide [in a wood]'/'afforest'

       v. *impazzare/impazzire*
              'be in full swing'/'go crazy'

      vi. *sfiorare/sfiorire*
              'brush', 'graze'/'wither', 'wilt'

Serianni (1988), from which the examples in (12) are taken, considers the cases in (12a) and (12b) as two groups of overabundant verbs, while Dardano & Trifone (1985) consider only cases (12a) as overabundant verbs, and propose that cases in (12b) are best analyzed as distinct lexemes; I concur with Dardano & Trifone, because of a clear difference in meaning between the two verbs in each pair in (12b); these verbs are different lexemes according to Fradin's (2003) and Fradin & Kerleroux's (2003) criteria, and will not be further discussed here.

Verbs in (12a) are claimed to have forms belonging to the two inflectional classes traditionally called 2nd conjugation (infinitive ending in -*ere*) and 3rd conjugation (infinitive ending in -*ire*); besides, the 3rd conjugation forms belong to the subclass of 3rd conjugation verbs which does not exhibit the element -*isc*- in the appropriate morphomic partition (so PRS.IND.1SG is *empio*, not *empisco, etc.). The 2nd conjugation and the -*isc*-less subclass of the 3rd conjugation have non-distinct inflection in several cells, listed in (13a), while they have distinct forms in other cells, listed in (13b), with examples from *riempiere* and *riempire*:[11]

(13)   Italian (personal knowledge)

    a. Cells with non-distinct realization for the verbs in (12a)
       Present Indicative: all person/number forms, except 2PL
       Present Subjunctive: all person/number forms
       Imperative 2SG
       Gerund
       (Present Participle)[12]

    b. Cells with distinct realization for the verbs in (12a)
       Present Indicative 2PL = Imperative 2PL (e.g., *riempiete* vs. *riempite*)
       Imperfective Past Indicative (*Imperfetto*): all person/number forms (e.g., 1SG *riempievo* vs. *riempivo,* etc.)
       Simple Perfective Past Indicative (*Passato Remoto*): all person/number forms (e.g., 1SG *riempietti* or *riempiei* vs. *riempii,* etc.)
       Future: all person/number forms (e.g., 1SG *riempierò* vs. *riempirò,* etc.)
       Imperfect Subjunctive: all person/number forms (e.g., 1SG *riempiessi* vs. *riempissi,* etc.)

---

[11] In (13) I consider only synthetic forms; periphrastic forms are formed by an inflected auxiliary followed by a Past Participle, so their distinctness is a function of the distinctness of the Past Participle form (therefore, they are always distinct for these two conjugations).

[12] A so-called Present Participle ending in -*nte* is normally listed as part of a verb's paradigm in Italian descriptive grammars, but it is extremely doubtful that such a cell should be recognized as a genuine part of verbal paradigms in Italian. Haspelmath (1996) contrasts these so-called present participles of Italian with those of other languages in terms of their syntactic properties (government of subject and non-subject arguments) and concludes that in Italian "active participles do not exist" (Haspelmath 1996: 61). Luraghi (1999) is less drastic, but shows that -*nte* forms have never been part of the spoken register in the history of the language, and that a verbal usage of -*nte* forms is only attested in some technical or bureaucratic registers, while adjectives and nouns in -*nte*, often unrelated to any verbal base, are common.

Present Conditional: all person/number forms (e.g., 1SG *riempierei* vs. *riempirei,* etc.)
Past Participle (e.g., *riempiuto* vs. *riempito*)
Infinitive (e.g., *riempiere* vs. *riempire*)

The verbs in (12a) are technically cases of single lexemes mapping to two distinct flexemes, but these flexemes are syncretic in all the cells listed in (13a).

As I am always wary of believing statements by grammars on the distribution of cell mates, I have checked the distribution in the corpus *la Repubblica 1985-2000* of the forms of the verbs in (12a) that are distinct in the two conjugations. Table 3 illustrates the results (figures for forms of the same Tense/Mood have been added together).

The data in Table 3 show the following picture: EMPIERE/EMPIRE 'fill' are almost extinct verbs in both conjugations, totaling only 13 forms overall; their meaning is normally expressed, in contemporary Italian, by RIEMPIRE; RIEMPIERE 'fill' is little used – there are a few tokens of the Infinitive and of the Imperfective Past Indicative (*Imperfetto*) in usage, but the ratio between forms of RIEMPIRE and forms of RIEMPIERE in the cells for which the two conjugations have distinct forms is so unbalanced (504:1) that the two verbs represent at best an extremely weak and non-canonical case of overabundance (or mapping from one lexeme to two flexemes) according to Thornton's (2012: 188–189) criteria for measuring the strength of overabundance on the basis of frequency ratios between two cell mates. ADEMPIERE and ADEMPIRE 'fulfil' have a less unbalanced frequency ratio (15.2:1) overall, but it must be observed that 99.5% of the forms of ADEMPIERE are realizations of the Infinitive and the Past Participle, while 93.3% of the forms of ADEMPIRE are realizations of tenses different from the Infinitive and the Past Participle. Indeed, all the Past Participle forms are 2nd conjugation forms (i.e., they are forms of ADEMPIERE, not possible forms of ADEMPIRE), so there is no overabundance in this cell; the only tenses in which the two verbs display some overabundance are the Future (with a ratio of 5.4:1 in favour of ADEMPIRE) and, very marginally, the Infinitive (with a very unbalanced ratio of 154:1 in favour of ADEMPIERE). The same picture, even more dramatically, is presented by COMPIERE/COMPIRE 'complete'. Assessment of overabundance in this case is made difficult by the fact that some Past Participle forms of COMPIRE are homographous with other forms in the paradigm, and/or with forms of the noun COMPITO 'task, homework', and/or of the adjective COMPITO 'corteous, polite', and/or of the verb COMPITARE 'spell out' (e.g., *compito* represents 'complete.PST.PTCP.M.SG', 'task(M).SG', 'courteous.M.SG' and 'spell_out.PRS.IND.1SG'; the noun for 'task' and the 1SG form of 'spell out' have antepenultimate stress, while the other forms have penultimate stress, but stresses on these syllables are not marked in the standard orthography of Italian, so all the forms are homographs even if they are not all homophonous); these homographies have been manually disambiguated for the forms ending in -*a* and -*e* (*compita* 'complete.PST.PTCP.F.SG', 'courteous.F.SG', 'spell_out.PRS.IND.3SG' and *compite* 'complete.PST.PTCP.F.PL', 'complete.PRS.IND.2PL', 'courteous.F.PL'), which have low frequency, thus making manual disambiguation practical; the lack of manual disambiguation for the high frequency forms in -*o* and -*i* explains why the exact frequency of these forms is

Table 3: Frequency in *la Repubblica 1985–2000* corpus of forms of the verbs in (12a) in cells that have distinct realizations for the two conjugations.

	EMPIERE	EMPIRE	RIEMPIERE	RIEMPIRE	ADEMPIERE	ADEMPIRE	COMPIERE	COMPIRE
2PL.PRS.IND = 2PL.IMP	--	1	--	368	--	2	--	2
Imperfetto	--	2	2	611	--	10	7	441
Passato Remoto	1	5	--	282	--	7	--	423
Future	1	--	--	294	5	27	23	1371
Imperfect Subjunctive	--	--	--	52	--	7	7	111
Present Conditional	--	--	--	52	--	3	3	76
Past Participle	--	--	--	2307 -o	233 -o	--	15907 -o	? -o
				537 -a	10 -a		4700 -a	3 -a
				450 -i	24 -i		5926 -i	? -i
				368 -e	24 -e		2397 -e	1 -e
Infinitive	--	3	15	3239	615	4	9267	8
Total	2	11	17	8560	911	60	38237	> 2400

not given in Table 3, and a question mark has been inserted instead.[13] The actual forms realizing 'complete.PST.PTCP.F.SG', and 'complete.PST.PTCP.F.PL' turned out to be a minority (3 over 48 (6%) for the F.SG form, 1 over 9 (11%) for the F.PL form). Therefore, it may be concluded with some safety that in the Past Participle cell the verb COMPIERE is favoured and COMPIRE is quite underrepresented. These two verbs show the same kind of "division of labour" already observed for ADEMPIERE and ADEMPIRE: COMPIERE specializes for the Infinitive and the Past Participle, and COMPIRE for all other tenses (among the ones that have distinct realizations for the two conjugations); however, in most tenses a few forms of COMPIERE are also attested, so COMPIERE/COMPIRE represent the best example of overabundance in all cells encountered so far among the Italian verbs commonly dubbed "sovrabbondanti" (although the frequency ratios render this case of overabundance not very canonical). It seems that ADEMPIERE/ADEMPIRE and COMPIERE/COMPIRE are on their way from overabundance to heteroclisis: at some point in the future, we might observe a lexeme with finite synthetic forms belonging to the 3rd conjugation and Infinitive and Past Participle (which carries with it all the periphrastic forms) belonging to the 2nd conjugation. RIEMPIERE/RIEMPIRE, instead, is just reducing overabundance in favour of the 3rd conjugation forms, and is quite advanced in this process.

If the process leading to heteroclisis is completed, we will have a single lexeme mapping to a single heteroclitic flexeme. At the moment, however, we have a number of Italian verbal lexemes that map to two flexemes, at least in parts of their paradigm.[14]

5 Conclusions

The data illustrated in this paper show that the distinction between lexemes and flexemes first proposed by Fradin & Kerleroux (2003) and Fradin (2003), as well as their definition of lexeme based on semantic and constructional coherence, is useful even beyond the area of lexeme formation, for which it was originally proposed. A separation between lexemes and flexemes, like the separation between content paradigms, form paradigms and realized paradigms adopted in paradigm-linkage theory, is a useful tool in models of morphological analysis that recognize a level of autonomous morphology.

References

Acquaviva, Paolo. 2008. *Lexical plurals: A morphosemantic approach.* Oxford: Oxford University Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes.* Cambridge: MIT Press.

---

[13]The raw frequency in *la Repubblica 1985-2000* of the form *compito* is 26450, that of *compiti* 9180.

[14]A reviewer observes that one could introduce overabundance in the very make-up of flexemes, rather than using it as the grounds for positing two (or more) distinct flexemes whenever a lexeme has two (or more) distinct realizations for a single cell, as done in this paper. This would involve allowing flexeme cells to comprise sets of forms rather than a single form. The relative merits of the two alternative approaches could be fully compared only in a formalized model, whose development exceeds the scope of this paper.

Aronoff, Mark. 2000. Morphology between lexicon and grammar. In Geert Booj, Christian Lehmann & Joachim Mugdan (eds.), *Morphology. An international handbook on inflection and word-formation*, vol. 1, 344–349. Berlin: Walter de Gruyter.

Battaglia, Salvatore & Vincenzo Pernicone. 1954. *La grammatica italiana. Seconda edizione migliorata.* Turin: Loescher Chiantore.

Bolinger, Dwight. 1956. Subjunctive *-ra* and *-se*: "Free Variation"? *Hispania* 39(3). 345–349.

Bonami, Olivier & Gregory T. Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge Handbook of Morphology*, 449–481. Cambridge: Cambridge University Press.

Bonami, Olivier & Delphine Tribout. 2012. Underspecification and the semantics of lexeme formation. Vienna. http://www.llf.cnrs.fr/sites/llf.cnrs.fr/files/biblio//diapos-vienne-v5.pdf, accessed 2016-11-06. Paper presented at the 15th International Morphology Meeting.

Butt, John & Carmen Benjamin. 2000. *A new reference grammar of modern Spanish.* 3rd edition. London: Arnold.

Carstairs, Andrew. 1987. *Allomorphy in inflexion.* London: Croom Helm.

Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic diversity and language theories*, 25–49. Amsterdam: John Benjamins.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words. *Language* 83(1). 8–42.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique.* Tübingen: Max Niemeyer Verlag.

Dardano, Maurizio & Pietro Trifone. 1985. *Grammatica italiana con nozioni di linguistica.* Bologna: Zanichelli.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie.* Paris: Presses Universitaires de France.

Fradin, Bernard. Forthcoming. Competition in derivation: What can we learn from duplicates? In Francesco Gardani, Hans Christian Luschützky, Franz Rainer & Wolfgang U. Dressler (eds.), *Competition in morphology.* Dordrecht: Springer.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Fradin, Bernard & Françoise Kerleroux. 2009. L'identité lexémique. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 83–102. Saint-Denis: Presses Universitaires de Vincennes.

Haspelmath, Martin. 1996. Word-class-changing inflection and morphological theory. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1995*, 43–66. Dordrecht: Kluwer Academic Publishers.

Holes, Clive. 2004. *Modern Arabic. Structures, functions and varieties*. Washington, D. C.: Georgetown University Press.

Kaye, Alan S. 2007. Arabic morphology. In Alan S. Kaye (ed.), *Morphologies of Asia and Africa*, 211–247. Winona Lake: Eisenbrauns.

la Repubblica 1985–2000 *corpus*. N.d. http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica. (Accessed 6 November 2016).

Luraghi, Silvia. 1999. Il suffisso -*ante/-ente* in italiano: fra flessione e derivazione. In Paola Benincà, Alberto M. Mioni & Laura Vanelli (eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia. Atti del XXXI Congresso della Società di linguistica italiana*, 539–550. Rome: Bulzoni.

Rojo, Guillermo & Alexandre Veiga. 1999. El tiempo verbal. Los tiempos simples. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, vol. 2, 2867–2934. Madrid: Espasa Calpe.

Serianni, Luca. 1988. *Grammatica italiana*. Turin: Utet. In cooperation with Alberto Castelvecchi.

Stump, Gregory T. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.

Thornton, Anna M. N.d. La non canonicità del tipo it. *braccio // braccia / bracci*: Sovrabbondanza, difettività o iperdifferenziazione? *Studi di grammatica italiana* 29–30. 419–477. Published in 2013.

Thornton, Anna M. 2011. Overabundance (multiple forms realizing the same cell): A non-canonical phenomenon in Italian verb morphology. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 358–381. Oxford: Oxford University Press.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5(2). 183–207.

Thornton, Anna M. 2014. Risposta n. 7 [le membre del comitato]. *La Crusca per voi* 49(II). 14–15.

Thornton, Anna M. 2016. *Perso* o *perduto? Visto* o *veduto?* http://www.accademiadellacrusca.it/it/lingua-italiana/consulenza-linguistica/domande-risposte/perso-perduto-visto-veduto. (Accessed 31 October 2016).

Trépos, Pierre. 1980. *Grammaire bretonne*. Rennes: Ouest France.

Ullmann, Stephen. 1957. *The principles of semantics*. Oxford: Basil Blackwell & Mott.

Zingarelli. 2016. *Lo Zingarelli 2016. Vocabolario della lingua italiana di Nicola Zingarelli*. Bologna: Zanichelli. Ed. by Mario Cannella and Beata Lazzarini.

Part IV

Troubles with Lexeme Formation Rules

Chapter 14

Reduplication across boundaries: The case of Mandarin

Chiara Melloni
University of Verona

Bianca Basciano
Ca' Foscari University of Venice

In this chapter, we shed new light on the reduplicative processes of Mandarin Chinese and assess the structural and interpretive properties of the input/base and output of these word formation phenomena. In particular, we focus on the categorial status of the base and address the issue of whether reduplication applies to category-free roots or full-fledged lexemes. Empirically, the privileged domain of research is *increasing* reduplication of disyllabic bases, or, as we dub it in the chapter, the AABB pattern, which is compared with *diminishing* reduplication, expressed by the template ABAB. The comparison between the two phenomena allows us to show that increasing and diminishing reduplication differ in the nature of the input units involved. On the grounds of a wide-ranging class of data, we argue that Mandarin reduplication takes base units of different 'size': word/lexeme-like units provided with category, namely verbs in the case of diminishing reduplication, and categoryless roots in the case of increasing reduplication. Throughout the chapter, we explore some category neutral properties of increasing reduplication and propose a unitary semantic operation capable to derive the various interpretive nuances of this phenomenon across lexical categories.

1 Introduction

1.1 Lexemes vs. words and reduplication phenomena

Lexemes are usually understood as sound/meaning pairs, i.e. linguistic signs provided with lexical category specification yet lacking inherent inflectional specification. Lexemes and words are thus considered as distinct entities in lexicalist approaches to word formation. As a matter of fact, while a word proper is a fully inflected entity functioning as a syntactic atom, a lexeme is the abstract version of the word-form lacking inflectional marking (Fradin & Kerleroux 2003). As put forward by Fradin & Kerleroux (2003), the

form of the lexeme can either be segmentally simple (*viz.* a root) or complex (*viz.* a stem), with affixal derivation, compounding and reduplication as phenomena possibly involved in lexeme formation.

Reduplication phenomena, however, are particularly challenging under this approach, since cross-linguistically the functions of reduplication are very varied and difficult to place categorically within the *derivational* domain of lexemes. In fact, whereas derivation typically forms new lexemes and can be category changing, reduplication often conveys values typically found in the *inflectional* domain. Although reduplication is attested with a variety of meanings (and forms) across languages, this phenomenon is consistently associated with its prototypical (iconic) function of intensification. In its increasing value, reduplication in the nominal domain gives as a result plural nouns, and in the domain of verbs it usually conveys aspectual meanings, i.e. pluractionality, iterative or progressive aspect, which are features prototypically expressed by inflection markings in most Indo-European languages. With adjectives, the prototypical value is intensification of the property/quality expressed by the base adjective. Nevertheless, independently of its semantic values, reduplication manifests several properties of word/lexeme formation and, formally, approaches derivational phenomena. First of all, (full) reduplication consists in the iteration of simple or complex roots (*viz.* stems), since it may also involve complex objects, such as compounds. Crucially, however, it typically applies to uninflected bases, with inflectional marking, if any, applying outside of/after reduplication. Moreover, reduplication shows many properties of compounding, since it often induces a reanalysis of the stress or tonal pattern of its base, or the insertion of epenthetic material between the two iterating units and/or some other phonological readjustment. Further, semantic drift and idiosyncrasy can characterize the outputs of reduplicative processes, while inflection phenomena are very transparent at the interpretive level (see Forza 2011, for an enlightening typological perspective) .

Therefore, under the lexeme/word distinction approach, we could argue that reduplication applies to roots or stems (traditionally understood as the phonological form of lexemes) and its domain of application is below the level of the word, or below X° in the standard X-bar approach.

1.2 Words, lexemes, and roots/stems in Mandarin Chinese

If the concept of lexeme appears empirically motivated in fusional or agglutinating languages whereby inflection markers modify the word form conveying relevant features in the syntactic contexts, its motivation is less grounded in isolating languages, where (concrete) words occur with none or a very low number of inflection markers, typically show invariable form and are virtually indistinguishable from the corresponding (abstract) lexemes. Mandarin Chinese is one of those languages where words have little or no inflection and where lexemes, expressing the abstract representation of a word, cannot be distinguished from word forms on a formal basis.

In Mandarin, the crucial distinction at the morphological level lies in the bound or free status of the root (a lexical morpheme), i.e. whether the root can 'stand alone' and occupy

a syntactic slot (1), equating thus free standing words in fusional languages, or whether it must be formally conjoined with another bound or free root, or with a derivational affix, to form an autonomous lexeme/word (2).

(1) free roots: 貓 *māo* 'cat', 走 *zǒu* 'walk, run away'

(2) bound roots: 衣 *yī* 'clothing, clothes' , 毆 *ōu* 'beat'

While the roots in (1) can be used by themselves in a sentence, those in (2) cannot stand alone but occur in complex words like e.g. 大衣 *dà-yī* 'big-clothes, overcoat, topcoat', 雨衣 *yǔ-yī* 'rain-clothes, raincoat', 衣櫃 *yī-guì* 'clothes-cupboard, wardrobe', 衣鉤 *yī-gōu* 'clothes-hook, clothes hook' (Arcodia & Basciano 2017: 105-106). Due to a strong tendency towards disyllabification attested in the evolution of the Chinese language over the centuries (see Shi 2002: 70-72), most roots are nowadays bound in Standard Mandarin (about 70% according to Packard 2000). Therefore, the majority of words or lexemes are compounds or other types of morphologically complex forms, typically ranging over all major lexical categories.

Another crucial aspect of Chinese morphology lies in the absence of strictly morphological criteria for the identification of the lexical category of roots (or stems, if morphologically complex), with some exceptions.[1] As a matter of fact, no category-specific morphology (such as declension/conjugation class markers in fusional languages) can be deployed to partition roots into lexical classes, with a verb like 走 *zǒu* 'walk, run away' being virtually indistinguishable at the morphological level from a noun like 書 *shū* 'book' (see Basciano 2017). Since there are no reliable morphological criteria to identify *lexeme*s as roots (or stems) endowed with lexical category features, the only reliable criterion is the distributional one. For instance, syntactic distribution only can discriminate among the adjectival, verbal or nominal use of a stem (namely, a combination of two roots) like 麻煩 *máfan* 'annoying, bother, trouble' (examples below from Basciano 2017: 561–562):

(3) a. 這件事很麻煩。
        zhè jiàn shì hěn máfan
        this CLF fact very troublesome

        'This fact is troublesome.'

     b. 他不願麻煩別人。
        tā   bù-yuàn   máfan biérén
        3SG.M not-willing trouble others

        'He is unwilling to trouble other people.'

     c. 你們在路上會遇到一些麻煩。
        nǐ-men zài lù-shang huì   yùdào yīxiē máfan
        2SG-PL in street-on may/will meet some trouble

        'You may/will run into some troubles on the road.'

---

[1]Examples are words containing suffixes such as 子 *-zi*, e.g. 刷子 *shuāzi* 'brush' (cf. 刷 *shuā* 'to brush'), and 頭 *-tou*, e.g. 想頭 *xiǎngtou* 'idea' (cf. 想 *xiǎng* 'to think'), which are always nouns (see Basciano 2017).

Thus, under the standard approach to lexemes proposed in 1.1, a relevant issue concerns the very existence of these units in the Chinese language where, at the lexical level, the very flexible distribution of lexical items seems to point in the direction of a lexicon whose base units (roots/stems) lack inherent category features. Moreover, the examples in (3) shed light on the need for a very loose semantics of roots/stems, arguably incompatible with the specific semantic meaning of lexemes, as proposed in Fradin & Kerleroux (2003). Under the hypothesis that roots bear no category specification, their meaning should be 'vague' enough to make it compatible with the adjectival, verbal or nominal meanings that might be instantiated in the syntax.[2] We may remark, however, that the great flexibility observed in previous stages of the language has been largely reduced over the centuries, first with a functional specialization of lexemes during the Han period (206 BCE-220 CE), and then with the proliferation of compound words, whose functional preference has been always much more rigid and stable (see Zádrapa 2017). Even though cases of 'regular ambiguity' like the one in (3) are found, in Modern Chinese lexemes tend to be more fixed as far as lexical category and distribution are concerned; many roots have a 'prototypical' distribution and cannot be easily coerced into other lexical categories. However, even very stable words may be occasionally placed in syntactic slots usually occupied by other word classes, creating "innovative ambiguities" (Kwong & Tsou 2003: 116; see also Basciano 2017) . As observed by Zádrapa (2017), although it is not possible to distinguish on a formal basis the prototypical from the non-prototypical use, it is still possible to perceive a functional "strain" (or "pragmatic markedness" in Bisang's 2008 terms), which always results in a semantic shift of varying dimension (see Croft 2001: 73).

1.3 Reduplication phenomena in Mandarin Chinese

Among word formation phenomena in Mandarin, reduplication is one of the most productive and, as we will see throughout this chapter, it is found across all major lexical categories with both increasing (iconic) and diminishing (countericonic) values. Whereas there is no perfect correspondence between lexical categories and reduplication functions (verbs, for instance, can be reduplicated along one or the other function), we will see there is instead a tight correspondence between the structural pattern of reduplication and its diminishing or increasing value, so that the two patterns are rigidly differentiated at the segmental and suprasegmental level.

In recent years there has been a growing attention to reduplication in Sinitic. In this chapter, we will try to shed new light on the reduplicative processes of Mandarin, and try to assess the structural and interpretive properties of the input (the bases of reduplication) and the output of reduplicative processes. In particular, we will focus on the question of the categorial status of the base of the reduplicative processes in Mandarin, i.e. what the base units are and, specifically, whether reduplication applies to category-less

---

[2]In syntactic approaches to word formation such as Distributed Morphology, the meaning of a word emerges constructionally once the root has been categorized by a selecting head (*n, v* or *a*) in the course of syntactic derivation, and cannot be determined lexically.

roots or to full-fledged lexemes/words. Empirically, the privileged domain of research will be the increasing reduplication of disyllabic bases, or, as we dub it here, the AABB pattern, which will be compared with the diminishing pattern, characterized by the disyllabic template ABAB.

The comparison between the two patterns will allow us to show that they differ in the type of units that constitute the basis of the reduplicative process. Mandarin reduplication, indeed, involves base units of different 'size', ranging from word/lexeme-like units provided with category and, namely, involving the verbal domain in the case of diminishing reduplication, to category-less roots in the case of increasing reduplication. Throughout the chapter, we will provide evidence for the latter claim, i.e. that reduplication phenomena involve roots, and we will explore some category neutral properties of increasing reduplication. We will conclude with some remarks on the semantic effects of this phenomenon, which we interpret as an *increased measure* function modifying the sortal type conveyed by the (combination of) roots.

1.4 Outline of the chapter

The chapter is organized as follows. Section 2 is dedicated to the presentation of the main patterns of full reduplication in Mandarin Chinese. Section 3 explores the characterizing features of increasing reduplication (AABB pattern) in some detail and discusses its formal and interpretive properties across lexical categories. Section 4 contains the structural analysis and some hypotheses about the semantics of AABB increasing reduplication, and section 5 draws the conclusions.

2 Data description

2.1 Reduplication in Mandarin: An overview

Reduplication in Mandarin Chinese is a widespread and productive phenomenon, virtually affecting all major lexical categories (V, Adj, N) and showing a tight relation between structural patterns (form) and semantic meanings (function). Semantically, Mandarin reduplications have augmentative/increasing and diminishing functions that are rigidly associated with different structural and/or suprasegmental patterns.

The diminishing function is only found in the verbal domain. Reduplicated verbs typically convey 'delimitative' or 'tentative' aspect (Chao 1968, Li & Thompson 1981, Tsao 2001), meaning to do something "a little bit/for a while" (Li & Thompson 1981: 29) or, by extension, to do something quickly, lightly, casually or just for a try.[3] Both monosyllabic (A → AA) and disyllabic (AB → ABAB) bases can reduplicate, but only in the case of monosyllabic reduplication the morpheme 一 *yi* (<*yī*) 'one' can occur between the base and the reduplicant:

---

[3]Further, it has the pragmatic function of marking a relaxed tone, casualness (Ding 2010), and thus reduplicated verbs are also used as mild imperatives (see Xiao & McEnery 2004).

(4) a. 教 (A)   →   教（一）教 (AA)
           jiāo   *jiāo (yi) jiao*
           teach   teach one teach
           'teach'   'teach a little'

    b. 休息 (AB)   →   休息休息 (ABAB)
           xiūxi   *xiūxi xiūxi*
           rest   rest   rest
           'rest'   'rest a little/for a while'

It has been argued that this reduplicative process is a syntactic phenomenon involving units in the *v*P domain (see Arcodia et al. 2014, Basciano & Melloni 2017). First of all, the reduplicated complex is not a syntactic atom, since it is possible to have intervening morphemes between the base and the reduplicant: beyond the numeral 一 *yi* (<*yī*) 'one' mentioned above, the perfective aspect marker 了 *le*[4] can intervene between the base and the reduplicant, as in (5):

(5) a. 走了走
       zǒu-le zou
       walk-PFV walk
       'walked a bit'

    b. 走了一走
       zǒu-le yi zǒu
       walk-PFV one walk
       'had a walk'

Moreover, diminishing reduplication is subject to event structure constraints (see Fradin & Kerleroux 2003, for similar constraints in French word formation): the base verb must be a process verb, typically controlled by an agent and crucially lacking a result, which captures the fact that achievements, accomplishments and resultative compounds are systematically excluded from reduplication. Aspectually, the reduplicated verb is incompatible with the progressive and durative aspectual markers while, as we have seen, it is perfectly compatible with the perfective aspect marker. Therefore, reduplication seems to modify the event structure of the base verb, providing a temporal boundary to the unbounded process expressed by the base (see Xiao & McEnery 2004). Other constraints, e.g. purely morphological constraints, are not observed.

In view of these facts and under the assumption that aspectual properties are *syntactically* encoded (see e.g. Travis 2000, 2010, Borer 1994, 2005, McClure 1995, Ramchand 2008), Arcodia et al. (2014) propose that diminishing reduplication is a syntactic phenomenon affecting the *v*P domain, and develop a syntactic analysis to account for it; the reader is referred to Arcodia et al. (2014) and Basciano & Melloni (2017) for further details of the analysis.

---

[4]Note that the perfective marker 了 *le* is generally placed after the second verb in resultatives and other kinds of compound verbs: 喝醉了 *hē-zuì-le* 'drink-drunk-PFV' *vs.* * 喝了醉 *hē-le zuì* 'drink-PFV drunk'.

Increasing reduplication exhibits several properties that make it a very different phenomenon from diminishing reduplication. First, increasing reduplication is found mainly among adjectives, but it can be found with verbs and nouns/classifiers too. Consider the following examples of adjectival reduplication:[5]

(6) a. 小 (A)   →   小小 (AA)
        xiǎo   *xiǎo~xiāo*
        small   small~small
        'small'   'very/really small'

     b. 高興 (AB)   →   高高興興 (AABB)
        gāoxìng   *gāo~gāo-xìng~xìng*
        'happy'   'very happy'

In the adjectival domain, the increasing function expressed by this kind of reduplication is not necessarily 'very Adj', but it rather makes the adjectives more descriptive, indicating a higher degree of liveliness and vividness.[6] As we will see in the next section, differently from diminishing reduplication, increasing reduplication requires that its base adjectives and verbs have specific structural properties.

Increasing reduplication applies to verbs too, but only if the base is bimorphemic and its constituents are in a relation of coordination.[7] In (7), for instance, the reduplicated verb portrays two interrelated actions which are performed alternately, repeatedly, or an action performed by a great number of people.

(7) 來往   →   來來往往
     lái-wǎng   *lái~lái-wǎng~wǎng*
     come-go   come~come-go~go
     'come and go'   'come and go repeatedly, come and
                                go in great numbers'

AABB verbs, beside expressing pluractionality or action in progress (see Hu 2006, Ding 2010), can also express vividness (8), or acquire an extended meaning, losing their verbal meaning and becoming more similar to adjectives in meaning and distribution

---

[5]According to Li & Thompson (1981: 33), in AABB reduplication of adjectives the second syllable is unstressed, and thus has a neutral tone. However, there is no clear consensus on tonal patterns in this kind of reduplication. For example, according to Tang (1988: 282), the second syllable is in the neutral tone, while the third and fourth syllables, or just the fourth syllable, are in the first tone. Further, Tang observes that in Taiwan most people use the original tones, i.e. there is no tonal modification in this reduplication pattern (see also the examples in Paul 2010).

[6]Xu (2012a: 6) states that, when adjectives are reduplicated, the degree of the adjective's quality is generally intensified. However, this does not seem to be always the case in the modern language: for example, she observes that colour perception can be subjective and variable, and thus adjectives indicating colours are prone to subjective interpretation.

[7]Reduplication of monosyllabic verbs (AA) in Modern Chinese does exist but has a diminishing meaning (see ex. (4a)). However, in previous stages of the language, before the appearance of the VV pattern with diminishing meaning, reduplication of monosyllabic verbs had an increasing function (repetition or action in progress); see e.g. Xu (2012a: 7).

(9),[8] depending on the linguistic context (on the meaning of AABB verbal reduplication, see Hu 2006).

(8) 跑跳   →   跑跑跳跳
　　　pǎo-tiào   *pǎo~pǎo-tiào~tiào*
　　　run-jump   run~run-jump~jump
　　　'run and jump'   'skip, run about, run and jump in a
　　　　　　　　　　　　　　　　　 vivacious way'

(9) 偷摸   →   偷偷摸摸
　　　tōu-mō   *tōu~tōu-mō~mō*
　　　steal-touch   steal~steal-touch~touch
　　　'pilfer'   'furtive, surreptitious, sneaky'

Finally, nouns can reduplicate too, conveying an overall increasing function, though AA reduplication no longer seems to be productive:

---

[8] See the following examples, where 偷偷摸摸 *tōu~tōu-mō~mō* is used as a nominal modifier (i) and as an adverbial, both with (ii.a) and without (ii.b) the adverbial marker 地 *-de* (examples from the Academia Sinica Balanced Corpus of Modern Chinese: http://lingcorpus.iis.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh?ukey=-78102521&qtype=1&ssl=7 [2017-08-25]).

(i) [...] 最機警與最偷偷摸摸的一種動物
　　zuì jījǐng yǔ zuì tōu~tōu-mō~mō de yī zhǒng dòngwù
　　most astute and most furtive   DET one CLF (type) animal

　　'[...] the most astute and furtive animal.'

(ii) a. 不要偷偷摸摸寫
　　　　bù yào tōu~tōu-mō~mō xiě
　　　　not have furtive   write

　　　　'You must not write furtively'

　　b. 也儘量不要躲在角落裡偷偷摸摸地拍攝
　　　　yě jǐnliàng bù yào duǒ-zài jiǎoluò lǐ tōu~tōu-mō~mō-de pāishè
　　　　also as.mush.as.possible not have hide-at corner in furtive-ADV   take.picture

　　　　'Also, as much as possible, you must not hide in a corner taking pictures furtively'

Generally speaking, adjectives may function as adverbs, modifying verbs. Adverbs are generally formed from adjectives (though sometimes they can be formed from abstract nouns) but not from verbs. Basically, an adjective may modify both a noun/NP or a verb/VP, while a verb may only modify a noun/NP (see Arcodia 2014).

It must be noted, though, that basically all reduplicated AABB verbs can have an adverbial use, and thus they all share an important property of adjectives:

(iii) 妻子和女兒説説笑笑地準備著晚飯。
　　qīzi hé nǚ'ér shuō~shuō-xiào~xiào-de zhǔnbèi-zhe wǎnfàn
　　wife and daughter talk~talk-laugh~laugh-ADV prepare-DUR dinner

　　'His wife and daughter were preparing dinner talking and laughing.'

(Center for Chinese Linguistics PKU corpus of Modern Chinese: http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xiandai [2017-07-24])

(10) a. 天 (A)   →   天天 (AA)
          tiān   *tiān~tiān*
          day   day~day
          'day'   'every day'

      b. 花草 (AB)   →   花花草草 (AABB)
          huā-căo   *huā~huā-căo~căo*
          flower-plant/grass   flower~flower-plant~plant
          'flowers and plants'   '(many) flowers and plants'

Reduplicated monosyllabic nouns are said to have a distributive (see e.g. Li & Thompson 1981, Hu 1994, Li 2009, Xu 2012b) or plural-collective (Paris 2007) meaning. Given the specific meaning of monosyllabic reduplications, their lack of productivity and the fact that many of the nouns that can reduplicate display classifier-like properties, it is disputable whether AA reduplication applies to actual nouns or nominal classifiers (functional elements in the *extended* NP domain); we will go back to this in section 3.3. As for disyllabic reduplicated nouns, the disyllabicity of the base (classifiers never are disyllabic) point to uncontroversially nominal bases. Semantically, Zhang (2015) argues that AABB reduplication is a plural marker, expressing 'greater plurality' (see Corbett 2000), but according to Xu (2012b) it indicates distributivity, as we will see in section 3.3.

2.2 Diminishing vs. increasing reduplication

From the brief overview provided above, a first interesting generalization arises. There is a correspondence between reduplicative pattern (with consistent structure and meaning) and lexical category, but limited to diminishing reduplication: AA or ABAB diminishing reduplication applies only to verbs, as input and output categories. Increasing reduplication is very different in this respect because it cross-cuts lexical categories rather than being firmly associated with a word class (although AA/monosyllabic reduplication is unproductive nowadays with nouns and classifiers).

Let us now focus on other differences between the two types of reduplication: it appears that the two functions of reduplication are associated with a set of different formal and selectional properties. A striking fact, especially in consideration of the great deal of unstable meaning-structure correspondences in reduplication cross-linguistically, is the tight correspondence between form and function observed in the reduplication of disyllabic bases.[9] While for monosyllabic bases the difference between increasing and diminishing reduplication is visible only at the suprasegmental level,[10] for disyllabic bases (AB), the difference arises at the segmental level.

---

[9]Many (if not most) languages do not exhibit such a clear correspondence between patterns and functions in reduplication (Mattes 2014).

[10]According to some, diminishing reduplicated verbs are toneless, whereas the reduplicated adjective always bears the first tone (Tang 1988: 282, Paul 2010: 120). However, according to Li & Thompson (1981: 33), the second syllable of reduplicated adjectives too is unstressed. As for the few monosyllabic nouns that reduplicate in Modern Chinese, it seems that the reduplicant keeps the same tone as the base noun.

In the diminishing function, the base is reduplicated as a whole (ABAB), as in the ex. (4b), while in the increasing function, each morpheme is reduplicated by itself (AABB), as seen in the examples (6b), (7)-(9) and (10b). Thus, it appears that there is a strong correlation between the function and the form of reduplication: as hinted at in section 2.1, the ABAB pattern always conveys diminishing meaning, whereas the AABB pattern is associated with increasing semantics, regardless of the word class of the input. Interestingly enough, the AABB pattern seems to be associated with increasing semantics also in other Sinitic languages (see Arcodia et al. 2015).

It is worth noting that some disyllabic words predominantly showing an adjectival distribution can not only occur in the (standard) increasing template AABB, but they may also appear in the diminishing ABAB template, so that the same base eventually enters two reduplication templates formally and functionally distinct:

(11)   a.   高興   →   高高興興 (AABB) (cf. 6b)
             gāoxìng   *gāo~gāo-xìng~xìng*
             'happy'   'very happy'

       b.   高興   →   高興高興 (ABAB)
             gāoxìng   *gāoxìng gāoxìng*
             'happy'   'have some fun'

Crucially, these minimal pairs are restricted to disyllabic bases amenable to a verbal/dynamic beyond an adjectival/stative interpretation, as we can see in the ABAB pattern in (11b). Therefore (11b) is not a counterexample to the generalization that only verbs can be reduplicated along the ABAB pattern.

Moreover, the difference between diminishing and increasing reduplication is not only semantic, but also concerns the restrictions on the input and on the output. As for diminishing reduplication, the selection restrictions, as we have seen, seem to be aspectual and allegedly dependent on event structure constraints, while for increasing reduplication these restrictions are (mostly) morphological, as we will see in the next section.

3 Increasing reduplication: input and output

Different from diminishing reduplication, increasing reduplication requires that its bases have specific morphotactic and semantic properties. In what follows we focus on the category-specific and category-neutral restrictions of increasing reduplication and describe the properties of the outputs of these reduplications across the major lexical categories.

3.1 Adjectives

In the adjectival domain increasing reduplication applies indifferently to monosyllabic and to disyllabic bases. In both cases, the base adjective must be gradable, thus absolute adjectives cannot reduplicate: e.g. 方 *fāng* 'square' cannot give rise to **方方 *fāng~fāng*

(see Paris 1979, cit. in Paul 2010: 139, fn. 18). [11] Therefore, adjectival reduplication only applies to bases that encode a degree/scalar value (see also Zhu 2003). At the morphotactic level, we find restrictions as far as disyllabic bases are concerned: as a matter of fact, the AABB pattern requires a disyllabic *and* bimorphemic base, whereas disyllabic monomorphemic words cannot be reduplicated (Paul 2010: 137):[12]

(12)   窈窕   →   *窈窈窕窕
        yǎotiǎo   **yǎo~yǎo-tiǎo~tiǎo*
        'graceful, gentle'

Also, the two morphemes must be lexical. For instance, adjectives formed with a prefix-like element cannot reduplicate (see Zhu 2003):

(13)   不安   →   *不不安安
        bù-ān   **bù~bù-ān~ān*
        not-peaceful
        'troubled/restless'

It thus appears that units are here handled strictly on a morphemic basis, rather than on a prosodic basis. Moreover, the possible bases for AABB reduplication are either lexicalized, non-transparent bases (14a), or adjectives formed by two morphemes with a similar meaning (14b) or in a logical coordination (14c):

(14)   a.   馬虎   →   馬馬虎虎
             mǎ-hu   *mǎ~ma-hū~hū*
             horse-tiger   horse~horse-tiger~tiger
             'careless, casual'   'careless, casual (stronger)'

---

[11]However, Tang (1988: 279-283) lists 方方 *fāng~fāng* 'square~square' among possible reduplicated adjectives. This could be possibly the result of a coerced interpretation (see e.g. English *very square face*). Indeed, Tang highlights that adjectives that express distinctive properties (e.g. appearance, size and colour) generally can reduplicate even when, as in the case of 方 *fāng* 'square', they are not used predicatively and cannot be modified by degree adverbs (examples from Tang 1988: 283):

(i)   ?? 他的臉很方
      tā   de liǎn hěn fāng
      3SG.M DET face very square
      'His face is very square.'

(ii) ?? (很) 方的臉
      (hěn) fāng de liǎn
      (very) square DET face
      'A (very) square face'

(iii) 方方的臉
      fāng~fāng   de liǎn
      square~square DET face
      'A (very/really) square face'

[12]窈窕 *yǎotiǎo* is an example of partial reduplication in Old Chinese, involving rhymes only, traditionally called 叠韵 *diéyùn* 'reduplicated rhymes': 窈窕 *ˀiwˀ-liwˀ > ewX-dewX > yǎotiǎo* (Sagart 1999: 137).

b. 快樂 → 快快樂樂
   kuài-lè   *kuài~kuài-lè~lè*
   pleased-happy   pleased~pleased-happy~happy
   'happy'   'very/really happy'

c. 高大 → 高高大大
   gāo-dà   *gāo~gāo-dà~dà*
   tall-big   tall~tall-big~big
   'tall and big'   'very/really tall and big'

These data show that the disyllabic AABB template applies to complex bases that are structurally and semantically symmetrical, i.e. exocentric or coordinative structures lacking a clearly identifiable head. Adjectival reduplication, thus, seems to be conditioned by morphosyntactic (word-internal) factors.

As for the output, the reduplicated adjective loses its gradability: while the base must be gradable, the reduplicated adjective is no longer gradable. As a matter of fact, whereas the (scalar) base adjective is compatible with degree modifiers such as 'very' and 'fairly', which indicate a high level on the scale of the (gradable) property expressed by the adjective they modify, the reduplicated adjective is not:

(15) a. 長 → 非常長
        cháng   *fēicháng cháng*
        'long'   'very long'

     b. 長長 → * 非常長長
        cháng~cháng   *fēicháng cháng~cháng*
        'long~long'   *'very long~long'

Moreover, whereas the base adjective can appear in the comparative construction, the reduplicated adjective cannot:

(16) a. 我的頭髮比他的長。
        wǒ de tóufa bǐ   tā   de cháng
        1SG DET hair COMP 3SG.M DET long

        'My hair is longer than his.'

     b. * 我的頭髮比他的長長。
        wǒ de tóufa bǐ   tā   de cháng~cháng
        1SG DET hair COMP 3SG.M DET long~long

However, there is a group of adjectives for which reduplication works differently. These are adjectives that typically involve a modifier-head structure, such as 雪白 *xuě-bái* 'snow-white', which reduplicates as ABAB (雪白雪白 *xuě-bái~xuě-bái*). The function is reportedly increasing, as in the case of AABB reduplicated adjectives. This might appear as an exception to the form-function identity between ABAB reduplication and diminish-

ing meaning in Mandarin.[13] It must be noted, though, that modifier-head adjectives like 雪白 *xuě-bái* 'snow-white' are not gradable and, indeed, they are not compatible with degree adverbs and cannot be used in the comparative construction. Therefore, reduplication does not result in a change in gradability of the base adjective, as it is the case with AA and AABB adjectival reduplication. Adjectival ABAB reduplication, thus, seems to be a phenomenon distinct from the other patterns of reduplications described in this section. We will go back to this issue in section 3.5., when discussing the word/lexeme status of the bases of increasing reduplication.

3.2 Verbs

As for verbs, increasing reduplication poses no aspectual requirements on the base unit since all kinds of verbs, including inherently telic verbs like 來 *lái* 'come', 進 *jìn* 'enter' or 出 *chū* 'exit', are allowed (see ex. (7), repeated here as (17c)). Nonetheless, increasing reduplication requires base units that possess specific structural properties. As a matter of fact, AABB increasing reduplication is generally possible only for coordinated complex verbs, the constituents of which may be either in a relation of logical coordination (17a), synonymy (17b) or antonymy (17c):

(17) a. 説笑 → 説説笑笑
shuō-xiào *shuō~shuō-xiào~xiào*
talk-laugh talk~talk-laugh~laugh
'talk and laugh' 'talk and laugh continuously'

b. 叫嚷 → 叫叫嚷嚷
jiào-rǎng *jiào~jiào-rǎng~rǎng*
call-shout call~call-shout~shout
'shout, howl' 'shout repeatedly'

c. 來往 → 來來往往
lái-wǎng *lái~lái-wǎng~wǎng*
come-go come~come-go~go
'come and go' 'come and go repeatedly, come and go in great numbers'

Note that in (17) the bases of reduplication are existing verbs, but this is not necessarily always the case, as e.g. 走走停停 *zǒu~zǒu-tíng~tíng* 'walk and stop' (there is no corresponding base verb 走停 *zǒu-tíng*).[14]

---

[13] According to Paul (2010: 137, fn. 15), "[the] reduplication pattern for 'modifier-adjectival head' compounds deriving an adjective of the form [$_{A^o}$ ABAB] is not to be confounded with the repetition of a disyllabic verb as a whole in syntax: [$_{V^o}$ AB] [$_{V^o}$ AB]".

[14] An alternative analysis might pose that verbal AABB reduplication is the result of the coordination of two reduplicated verbs, [A~A] [B~B]. However, note that since the reduplication of monosyllabic verbs expresses a delimitative meaning, the coordination of two monosyllabic reduplicated verbs should result in a delimitative semantics. Further, this analysis is not tenable because telic verbs like 來 *lái* 'come', as said above, cannot reduplicate by themselves, * 來來 *lái~lái*.

Also, it is worth remarking that the verbal reduplication pattern AABB may also be found with disyllabic monomorphemic verbs, such as (18a) or other kind of compound verbs (18b and 18c):

(18) a. 哆嗦
        duōsuo
        'tremble'
     b. 飄悠
        piāo-you
        'float-long/leisurely, wobble, stagger'
     c. 鬧騰
        nào-teng
        'noisy-jump, disturb/create confusion'

As for the prosodic properties of the pattern, the second morpheme/syllable of non-coordinate compound verbs that can undergo AABB reduplication generally has the neutral tone, suggesting that these are lexicalized forms.[15] Thus, similarly to adjectives, the AABB template in the verbal domain basically applies to structurally and semantically symmetrical bases, but it can also apply to unanalyzable morphemes or to lexicalized forms.[16] For some of these lexicalized forms, it is possible that they originate from coordinating structures whose relationship became opaque with time, but an in depth diachronic analysis is needed to substantiate this hypothesis.

As for the output, AABB reduplication of verbs seems to operate at the aspectual level, expressing repetition or action in progress. However, as we have seen, it can also express vividness (8), or other kinds of more abstract meanings (9), closely approaching *adjectival* reduplicative processes.

3.3 Nouns

As we have seen, reduplicated monosyllabic nouns are said to have a 'distributive' or 'plural collective' meaning:

(19) 人人都喜歡受人稱贊。
      rén~rén   dōu xǐhuan shòu   rén   chēngzàn
      person~person all   like   receive person praise
      'Everybody likes to be praised by people.'

---

[15]Toneless items in Chinese are typically grammatical morphemes, such as e.g. aspectual markers, (some) no longer productive derivational suffixes, and the second syllables of some reduplicated or compound words, as e.g. 爸爸 *bàba* 'father', 學生 *xuésheng* 'student'. Thus, lack of tone is a clue of either grammaticalization or lexicalization.

[16]The only constraint which does not seem to be morphological but rather aspectual concerns coordination of telic verbs: as we have seen, telic verbs may appear in the AABB pattern of reduplication, but if they do they must be antonyms (as in ex. 7/17c), i.e. reduplication of synonymic telic verbs does not seem to be possible (see Zhang 2016). This might be due to the fact that the coordination of two antonymic telic verbs (like *enter-exit*) results in the annulment of the *télos*, which seems to suggest that, actually, the bases of this kind of reduplication too must express an overall *atelic* event. This issue deserves further research.

Several authors (e.g. Hu 1994, Cai 2007, Li 2009) stress the fact that reduplication of monosyllabic nouns may be assimilated to classifier reduplication and that many of the nouns that can reduplicate show classifier-like properties. For example, Hu (1994: 103) observes that at least part of these (alleged) nominal bases can directly follow a numeral without an intervening classifier, as e.g. 一年 *yī nián* 'one year', 三戶 *sān hù* 'three households', and they can themselves work as classifiers, as e.g. 三戶人家 *sān hù rénjiā* 'three household (CLF) family, three families', thus exhibiting properties of (nominal) classifiers.

Reduplication of classifiers – how it is generally reported in reference grammars – seems to convey a distributive meaning:

(20)　看書的時候，書上的字不可能個個都認識。
　　　kàn shū de shíhou, shū shàng de zì　　 bù kěnéng gè~gè　dōu rènshi
　　　read book DET time　 book on　　 DET character not can　　 CLF~CLF all　 know
　　　'You cannot know all the characters/each character of the books you read.'

According to Paris (2007: 68), however, reduplicated classifiers get a (plural) distributive meaning when they appear in pre-verbal position (21a), while they get a plural collective interpretation when they occupy the post-verbal position (21b):[17]

(21)　a.　他個個學生都認得。
　　　　　tā　　 gè~gè　xuésheng dōu rènde
　　　　　3SG.M CLF~CLF student　 all　 be.acquainted.with
　　　　　'He knows all the students (individually).'

　　　b.　在分析上遇見種種困難
　　　　　zài fēnxī　　shàng yùjiàn zhǒng~zhǒng kùnnan
　　　　　at　 analysis on　　 meet　 CLF~CLF　　 difficulty
　　　　　'Come across all kinds of difficulties during the analysis.'

According to Zhang (2014), reduplication of classifiers in Mandarin is a type of plural marking; it denotes plurality of *units* (groups/collectives) rather than of individuals. Units and individuals can overlap, like in (22a), but it is not always the case, like in (22b), where 'lotus' is the individual, while 'lotus pile' is the unit that reduplicates (examples from Zhang 2014: 6):

(22)　a.　河裏漂著（一）多多蓮花。
　　　　　hé　　 lǐ piāo-zhe（ yī ）duō~duō liánhuā
　　　　　river in float-DUR (one) CLF~CLF lotus
　　　　　'There are many lotuses floating on the river.'

---

[17]Paris notes that it is not possible to have the noun preceded by the reduplicated classifier in post-verbal position with the same meaning as (21a), so that the following sentence is ungrammatical:

(i)　　* 他認得個個學生。
　　　tā　 rènde　　　　gè~gè　xuésheng
　　　3SG.M be.acquainted.with CLF~CLF student

b. 地上有一堆堆蓮花。

 dì   *shàng* *yǒu* *yī*   *duī~duī*   *liánhuā*

 earth on   have one CLF(pile)~CLF lotus

 'There are piles of lotuses on the ground.'

Zhang (2014: 12) argues that the distributive meaning emerges when reduplicated classifiers occur with the adverb 都 *dōu* 'all' (even when it is allowed but does not show up; see e.g. Guo 1999) or other kinds of adverbials:

(23)   個個學生都有自己的網頁。

  gè~gè   *xuésheng* *dōu* *yǒu* *zìjǐ* *de*   *wǎngyè*

  CLF~CLF student   all   have own DET webpage

  'All of the students have their own webpage.'

In contrast, according to Zhang, in (24), where no 都 *dōu* 'all' is allowed, the distributive meaning is not possible (example from Zhang 2014: 12):

(24)   雙雙情人步入會場。

  shuāng~shuāng *qíngrén* *bù-rù*   *huì-chǎng*

  CLF (pair)~CLF lover   step-enter meet-place

  'Many pairs of lovers stepped into the meeting place.'

According to Zhang (2014: 12), the fact that reduplicated classifiers do not have an intrinsic distributive reading is proven by the compatibility with collective verbs.

Going back to reduplication of monosyllabic nouns proper, Paris (2007) argues that it expresses a 'plural collective' meaning, more specifically it denotes a collectivity of elements sharing the same properties, which can function either as an argument or as an adverbial. According to Paris (2007: 69-70), reduplication of monosyllabic units does not have a distributive meaning, as shown by the contrast between (25a) and (25b), where the first one contains a reduplicated noun (天天 *tiān~tiān* 'day~day, every day'), while the second contains the quantifier 每 *měi* 'each'. In (25b) the object is necessarily distributed, i.e. it must be a different poem every day, while this is not necessarily the case in (25a).[18]

(25)   a. 他天天都讀一首詩。

    tā   *tiān~tiān* *dōu* *dú*   *yī*   *shǒu* *shī*

    3SG.M day~day all   read one CLF   poem

    'He reads a poem every day.'

    b. 他每一天都讀一首詩。

    tā   *měi*   *yī*   *tiān* *dōu* *dú*   *yī*   *shǒu* *shī*

    3SG.M each one day all   read one CLF   poem

    'Every day he reads a (different) poem.'

---

[18]Note that in (25a) 都 *dōu* 'all' is used but, according to Paris, we do not get the distributive reading. This contrasts with what Zhang argues about classifiers, where the presence of this adverb would lead to a distributive reading (see above).

Providing a detailed picture of the kind of plural readings expressed by reduplicated classifiers is beyond the scope of this chapter; however, what we want to stress here is that it is not easy to trace a clear boundary between different kinds of plural readings and that arguably different readings can be related to distributional/syntactic rather than solely lexical factors.

As for reduplication of disyllabic nouns, a first element is the undisputable categorial nature of the input, since classifiers are all monosyllabic. Structurally, nominal bases seem to be subject to the same morphological constraints observed for AABB adjectives and verbs. The AB base nouns usually entail a relation of coordination between their constituents: either logical coordination (see 26a), or synonyms or antonyms (26b) (see Tang 1979: 114; Zhang 2015):[19]

(26) a. 家戶 → 家家戶戶
         jiā-hù   *jiā~jiā-hù~hù*
         family-household   family~family-
                             household~household
         'household/family' 'every family/each house-
                             hold/many families'

      b. 老少 → 老老少少
         lǎo-shào   *lǎo~lǎo-shào~shào*
         old-young   old~old-young~young
         'the old and
          the young'   'old people and young people'

As we have seen with adjectives (14), we can also find more lexicalized forms like:

(27) a. 風雨 → 風風雨雨
         fēng-yǔ   *fēng~fēng-yǔ~yǔ*
         wind-rain   wind~wind-rain~rain
         'wind and rain/
          trials and
          hardships'   'trials and hardships/storms'

      b. 點滴 → 點點滴滴
         diǎn-dī   *diǎn~diǎn-dī~dī*
         dot-drip/drop   dot~dot-drop~drop
         'droplet'   'dribs and drabs/bit by bit'

The nominal AABB pattern of reduplication seems to be well-established in the Chinese lexicon (see e.g. Hu 1994, Wu & Shao 2001), and can be extended to disyllabic nouns that usually do not reduplicate (28a, Hu 1994: 106). Also, two monosyllabic nouns A and B

---

[19]Note that some AABB lexicalized nouns do not have a AB compound counterpart (see Wu & Shao 2001: 12): e.g. 生生世世 *shēng~shēng-shì~shì* 'life~life-generation~generation, generation after generation'(*生世 *shēng-shì*). Generally speaking, it is possible to form AABB nouns from the coordination of two items that do not form an AB compound (see (28b) and the related discussion).

that do not form a AB compound word, but satisfy the coordination requirements seen above, can reduplicate along the AABB pattern forming novel combinations (28b, see Wu & Shao 2001: 12):

(28) a. 情景   →   情情景景
          qíng-jǐng   *qíng~qíng-jǐng~jǐng*
          feeling-scene   feeling~feeling-scene~scene
          'scene, sight,   'every scene, all scenes'
           circumstances'

      b. 盆罐   →   盆盆罐罐
          pén   guàn   *pén~pén-guàn~guàn*
          'basin/pot' 'jar'   'pots and jars'

According to Zhang (2015: 7), though, the AABB nominal pattern is not productive, since many acceptable compound nouns formed by parallel constituents do not reduplicate (she argues the same for verbs too). This is however questionable since e.g. one of the example she mentions, i.e. 桌椅 *zhuō-yī* 'table-chair, tables and chairs' → 桌桌椅椅 *zhuō~zhuō-yī~yī* 'table~table-chair~chair', is listed as an example of reduplicated AABB noun by Wu & Shao (2001: 12-13), who put it among AABB 'temporary' combinations with low frequency. Even though it is not easy to establish the productivity of a pattern, we believe that 'occasional' usages and the possibility to coin new AABB nouns are hints of its productivity.

As for its function, as we have mentioned, Zhang (2015) argues that AABB expresses 'greater plurality' (see also Wu & Shao 2001), though it sometimes seems to have a distributive meaning, like in the case of reduplicated monosyllabic nouns; and, indeed, as we have seen, according to Xu (2012a), reduplicated AABB nouns indicate distributivity. See the examples below:[20]

(29) a. 家家戶戶的門前都掛著青天白日滿地紅的國旗 [...]
          jiā~jiā-hù~hù   *de mén-qián dōu guà-zhe*
          family~family-household~household DET door-front all hang-DUR
          qíng-tiān-bái-rì   mǎn-dì   hóng de guó-qí
          blue-sky-white-sun full-ground red   DET country-flag

          'In front of the door of each household hung the red national flag with the white sun in the blue sky [...]'

      b. 海水浴場裡，男男女女、老老少少，都穿著各種不同款式的泳裝 [...]
          hǎi-shuǐ   yù-chǎng lǐ, nán~nán-nǔ~nǔ,   lǎo~lǎo-shào~shào,   dōu
          sea-water bath-site in man~man-woman~woman old~old-young~young all
          chuān-zhe gè   zhǒng   bùtóng kuǎnshì de yǒng-zhuāng
          wear-DUR each CLF(kind) different style   DET swim-suit

          'Every man, woman, old and young bathing in the sea was wearing all different styles of swimming suits'

---

[20]Examples from Academia Sinica Balanced Corpus of Modern Chinese: http://app.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh [2016-11-24].

In any case, it is possible to argue that this reduplication pattern expresses a kind of plural and, indeed, Xu (2012a) argues that reduplication, like plural marking, is one of the major devices for indicating plurality in human languages.[21] This plural displays interesting properties: it is compatible with 'numeral+classifier' constructions (30a) and, most importantly, it seems to be compatible with the plural marker 們 *-men*[22] (30b):

(30) a. 200 多個子子孫孫前來祝壽
        èrbǎi duō ge zǐ~zǐ-sūn~sūn   qiánlái zhù-shòu
        200   more CLF son~son-grandson~grandson come   congratulate-longevity

        'More than 200 children and grandchildren came to congratulate [the old woman] on her birthday.'[23]

     b. [...] 讓我們的子子孫孫們還能依靠這個地球生活。
        ràng wǒ-men de zǐ~zǐ-sūn~sūn-men   hái néng yīkào zhè ge
        let   1SG-PL DET son~son-grandson~grandson-PL still can   rely   this CLF
        dìqiú shēnghuó
        earth live

        '[...] to let the future generations still be able to rely on this earth to live.'[24]

From a typological perspective, it is interesting to observe that in languages where reduplication and classifiers are found extensively, plural marking is not well developed and is sensitive to the semantic feature [+human] (Xu 2012a: 12), just like in Mandarin (see Corbett 2000 for a more comprehensive overview of number marking across languages). Xu (2012a) further remarks that the more plural marking is developed, the less this semantic feature ([+human]) is required; also, the more a language possesses developed plural markers, the less it needs reduplication and classifiers.

At the distributional level, the possible co-occurrence of AABB reduplication and of the plural marker 們 *–men* suggests that these two forms of pluralization cannot be equated, and, in a syntactically oriented approach to word formation and inflection, it indicates that these two plurals occupy different syntactic positions in the (extended) nominal projection. In particular, following Wiltschko's (2008) analysis of plural markers in Halkomelem Salish, we will argue that the reduplicative process is a derivational process that operates at the root level, even before root categorization is determined. This analysis allows us to explain the otherwise unexpected occurrence of 們 *–men* plu-

---

[21]Xu (2012b: 48) highlights some general tendencies in the languages of the world: 1) languages with obligatory plural marking tend not to have classifiers (see Greenberg 1972, Sanches & Slobin 1973; but see e.g. Bisang 2012); 2) languages without obligatory plural marking tend to use reduplication to express plurality. In general, languages which do not have plural marking seem to appeal to both reduplication and classifiers.

[22]The plural marker 們 *-men* can be added only to human nouns; it is entirely optional and is generally used "only when there is some reason to emphasize the plurality of the noun" (Li & Thompson 1981: 40). It is obligatorily used only with personal pronouns. Moreover, if the noun is preceded by a 'numeral+CLF', the marker 們 *–men* cannot be used: * 三個老師們 *sān ge lǎoshī-men* 'three CLF teacher-PL, three teachers' (cf. 30a). This can be taken as an indication of the fact that 們 *–men* is a marker of pluralization connected to the determiner/classifier domain, rather than being involved at the NP level.

[23]http://news.xinhuanet.com/society/2007-10/06/content_6833517.htm [2016-11-24].

[24]http://www.china-coop.org/index.php?ac=article&at=read&did=854 [2016-11-24].

ral marking on AABB (animate) nouns, which could be analysed as a modifier in the DP domain. We will go back to this issue in section 4.

3.4 Further remarks on the AABB pattern

To sum up, the data above show that increasing AABB reduplication is sensitive to the morphological makeup of its input, and insensitive to the categorial feature of the base (Adj, V, N) or, semantically, to its ontological/sortal type (whether the base denotes a quality, an event, or an entity/individual). As for the morphological restrictions on the base units, it is worthwhile noting that the requirement of a compound base of a specific type is also category-neutral, since it is found with AABB adjectives, verbs and nouns. In particular, the kind of root combinations we find seem to have much in common with 'co-compounds', in particular, with the following categories singled out by Wälchli (2005: 138): 'additive co-compounds', as e.g. Georgian *xel-p'exi* 'hand-foot'; 'generalizing co-compounds', as e.g. Mordvin *t'ese-toso* 'here-there, everywhere'; collective co-compounds, as e.g. Chuvash *sĕt-śu* 'milk-butter, dairy products'; synonymic co-compounds, as e.g. Uzbek *qadr-qimmat* 'value-dignity, dignity'.

According to Wälchli, additive co-compounds denote pairs consisting of the parts A and B; in a broader sense, they denote sets exhaustively listed by A and B. Generalizing co-compounds denote general notions (as e.g. 'all', 'always'); their parts express the extreme opposite poles of which the whole consists. As for collective co-compounds, they are not always easy to define since they obey to different criteria, which do not always agree: the parts do not exhaustively list the whole; the whole comprises all meanings having the properties shared by A and B; collective co-compounds are co-compounds which denote collectives.[25] Finally, in synonymic co-compounds, the constituents (A and B) and the whole compound have (almost) the same meaning. Wälchli observes that synonymic co-compounds "express homogeneous collection complexes in which (ideally) every element contained in them can be referred to by both parts of the co-compound" (p. 140). This, according to Wälchli, explains the affinity between synonymic co-compounds and plurality, though there is no language in which synonymic compounds work as fully grammaticalized plurals. Synonymic co-compounds may have affinities either to collective, to additive or to generalizing co-compounds. In any case, each type of co-compound described above may be considered as complexes where the referents are joint together to indicate a 'set'.

Interestingly enough, the AABB pattern can apply to AB bases that are not attested as coordinated bases (see sections 3.2, 3.3), and crucially it can be 'category-changing' (see Paul 2010: 145-146; cf. also ex. (9)):

(31) 婆婆媽媽   →   [AABB] = Adj
  pó~po-mā~mā
  old.lady~old.lady-mother~mother
  'kindhearted/sentimental/effeminate'

---

[25]The example from Chuvash reported above meets all the three criteria, but it is not always the case. It is difficult to distinguish between additive and collective co-compounds if the first two criteria do not apply at the same time.

In (31), the AB base is not an existing word, but AABB reduplication applies to two free/non-conjoined lexical roots. Reduplication of two elements independently compatible with a nominal meaning[26] results in an *adjectival* AABB lexeme.

Furthermore, the AABB pattern extends to others categories too, like numerals, place words, coordinated classifiers, onomatopoeias, etc. (see Hu 1994):

(32)   a. 千千萬萬
        qiān~qiān-wàn~wàn
        thousand~thousand-ten.thousand~ten.thousand

        'thousands and thousands'

   b. 前前後後
        qián~qián-hòu~hòu
        front~front-back~back

        'whole story/ins and outs'

   c. 嘻嘻哈哈
        xī~xī-hā~hā
        giggling.onomatopoeia~giggling.onomatopoeia-
        laughter.onomatopoeia~laughter.onomatopeia

        'laughing and joking'

All these facts seem to support the hypothesis that the AABB reduplication pattern applies even before the conjoined bases get their categories (and indeed the constituents can be bound roots too).[27] This is consistent with an analysis according to which word formation can apply to roots, or in this specific case, to combination/coordination of category-less roots, which would explain why, different from ABAB diminishing reduplication, it is a phenomenon found across almost all word classes.[28] We will go back to this in section 4, where we will put forth an analysis for this reduplication pattern.

3.5 On the base units of AABB reduplications

As we have seen in 2.2, diminishing reduplication does not form syntactic atoms and can be analyzed as a syntactic operation whose application is conditioned by structural re-

---

[26]It is worth noticing that when the base is formed by a bound root constituent, like 婆 *pó* 'old.lady' in (31), we cannot determine its lexical category since bound roots do not occupy syntactic slots (see section 1.2); rather, it can be said that these roots are 'noun-like' *semantically,* i.e. they denote entities/individuals (see section 3.5).

[27]A reviewer observed that it is difficult to make such a claim if the cases mentioned in this section are well-established lexicalized formations. Actually, these cases seem to be quite marginal, and for category changing items it is quite expected, since intuitively we expect that reduplication of two roots compatible with the nominal meaning leads to a nominal output. However, these examples further highlight the cross-categoriality of the pattern and further support the hypothesis of the acategoriality of the base roots. In any case, it is undoubtable that bound roots can enter this pattern of reduplication (see e.g. the reduplicated word in the examples (30) above, where both roots are bound), which as mentioned above (footnote 26; see also section 3.5) do not have a lexical category, and this points toward the acategorical nature of the conjoined roots.

[28]Reduplication of non-existent AB bases is not possible with diminishing verbal reduplication; in ABAB verbal reduplication, the AB base must be an existing disyllabic verb.

strictions in the *v*P domain (see Arcodia et al. 2014, Basciano & Melloni 2017). In contrast, we have shown that increasing reduplication is subject to 'morphological' restrictions. Keeping in line with previous research on reduplication and plural marking, we argue that AABB increasing reduplication is the result of the modification of roots (see section 4), understood here, as in most exoskeletal approaches (see Borer 2003), like elements crucially lacking category features. Moreover, as we will show in details in the next section, AABB reduplications are syntactic atoms which cannot allow for the insertion of other material between the iterated units (see e.g. Lapointe 1980).

Different pieces of evidence speak in favour of the hypothesis that AABB reduplication applies to elements smaller than a word, i.e. a root/stem, and possibly lack *per se* a definite category specification. In what follows, we will concentrate on the differences between AABB/increasing reduplication and other reduplicative processes to illustrate our point.

First of all, let us consider the verbal domain, where we find both diminishing reduplication and increasing reduplication. A first crucial difference between the two patterns, namely ABAB and AABB verbs, concerns the distribution of aspectual markers. With AABB reduplicated verbs, if an aspectual marker is present, it follows the whole reduplicated verb (33a), as in the case of resultatives and other kinds of compound verbs (cf. fn. 4). In diminishing reduplication, as we have seen, the aspectual marker 了 *le* is unexpectedly placed between the base and the reduplicant (33b):

(33) a. 連老郭都進進出出了好幾次。
      lián lǎo-Guō dōu jìn~jìn-chū~chū-le   hǎojǐ cì
      even old-Guo all   enter~enter-exit~exit-PFV many time

      'Even old Guo entered and exited from there many times.'[29]

   b. 她試了試那件衣服。
      tā   shì-le   shì nà jiàn yīfu
      3SG.F try-PFV try that CLF dress

      'She tried on that dress.'

A second piece of evidence comes from 'rhotacization' or *erhua* (兒化 *érhuà*), a morpho-phonological phenomenon that is very common in the speech varieties of Northern China, consisting in the addition of a retroflex approximant (兒 *-r*) at the end of a word. More precisely, phonologically, this suffix incorporates into the final syllable of a host stem replacing an existing coda, as e.g. 公園 *gōngyuán* → 公園兒 *gōngyuár* 'park', 鳥 *niǎo* → 鳥兒 *niǎor* 'bird'. The suffix 兒 *-r* can appear in reduplicated adjectives, and in the AABB pattern it occurs after the whole reduplicated adjective:

(34) 高高興興兒
      gāo~gāo-xìng~xìng-r
      'really happy'

---

[29]http://www.cctv.com/program/zoujinkexue/topic/science/C15580/20060413/100489.shtml [2016-11-24].

Lee-Kim (2016) observes that, even if to a lesser extent, this suffix can be also found in the reduplication of modifier-head adjectives (see 3.1). However, in this case the suffix attaches after each AB, i.e. AB-*r* AB-*r*:

(35) 雪白兒雪白兒
      xuě-bái-r~xuě-bái-r
      '(very) snow-white'

According to Lee-Kim (2016), this difference between the AABB pattern and the ABAB pattern, as far as the suffix 兒 –*r* is concerned, suggests that these two types of reduplication have a distinct internal structure. Assuming that 兒 –*r* adjoins to a phrasal node that introduces categorial information (*n, v, a* in DM), since it consistently occurs at the end of a full-fledged category, Lee-Kim argues that the contrast between (34) and (35) indicates that each AB forms an adjective phrase in the adjectival ABAB pattern of reduplication, while AABB as a whole forms a single adjectival phrase. She further argues that modifier-head compounds would undergo *erhua* before reduplication ([AB-*r*]-RED), while coordinate compounds reduplicate before 兒 –*r* adjoins ([AB-RED]-*r*). Since in the ABAB pattern 兒 –*r* adjoins *before* reduplication, the double occurrence of this suffix (AB-*r* AB-*r*) elegantly follows: reduplication applies to the whole suffixed compound AB-*r*, copying it as a whole. According to Lee-Kim, this also suggests that reduplication of modifier-head compounds is phrasal, while reduplication of coordinate compounds targets units smaller than a phrase. A corollary of this analysis might be that reduplication applies both to units below and above X°, but under this view it would be difficult to explain that there are no constraints on the gradability of the base, in the case of ABAB adjectival reduplication.

An alternative and more feasible hypothesis is that the ABAB pattern instantiates another kind of phenomenon, which is well attested across languages (even those ones that lack productive reduplication), *viz. contrastive focus* reduplication/repetition. Different from 'morphological' reduplication, contrastive repetition phenomena involve the copying of full fledge words and sometimes phrases, as in the following examples from Ghomeshi et al. (2004: 308), and typically have no phonological/tone reanalysis or other types of morpho-phonological readjustment phenomena that characterize reduplication in a cross-linguistic perspective:

(36)   a. I'll make the tuna salad, and you make the SALAD–salad.
        b. My car isn't MINE–mine; it's my parents'.
        c. Oh, we're not LIVING-TOGETHER–living-together.

The semantic effect of this construction is, according to Ghomeshi et al., "to focus the denotation of the reduplicated element on a more sharply delimited, more specialized, range" (p. 308). For example, in (36a) *SALAD-salad* denotes green salads as opposed to salads in general.

Although the interpretive difference between increasing reduplication and contrastive repetition is difficult to get from our Mandarin-speaking informants, we suggest that

reduplicated adjectives such as 雪白雪白 *xuě-bái~xuě-bái* 'snow-white~snow-white' might have a similar semantic effect, which is to express a prototypical, standard property denotation in the adjectival domain. As such, ABAB would be a different phenomenon applying at the phrasal level and crucially lacking the morphological constraints found with increasing reduplication. In contrast, the AABB pattern operates below the X° level and affects the gradable property of the base, i.e. it turns a gradable base into a no longer gradable one (see section 3.1).

A further element which seems to support the status of the AABB reduplicated forms as syntactically atomic units[30] is that they are often formed by at least one bound root (either A or B, or both of them) which cannot stand as a syntactic word by itself (see section 3.4, ex. (31) and fn. 26 and 27). For instance, in the example (37) the AB base is formed by two bound roots (cf. the free forms 兒子 *érzi* 'son' and 孫子 *sūnzi* 'grandson'):

(37) 子孫   →   子子孫孫
      zǐ-sūn   *zǐ~zǐ-sūn~sūn*
      son-grandson   son~son-grandson~grandson
      'children and   'heirs'/'generation after generation
       grandchildren'/   of descendants'
      'descendants'

This further corroborates the hypothesis that this process applies to roots, thus to acategorial elements; bound roots, indeed, have 'nouny', 'verby', 'adjective-like', etc. features, but, since they are not able to occupy a syntactic slot by themselves, they do not have a syntactic category proper.

4 Analysis

Given the properties illustrated thus far, in this section we will propose that AABB reduplication is a phenomenon applying at the root level, as we briefly mentioned in section 3.5. In particular, in the previous sections we have shown that the AABB pattern applies across categories and even to non-attested AB units, can be 'category changing' (e.g. a coordination of two noun-like roots may result in an adjective), can be formed by bound roots, and displays syntactic atomicity/lexical integrity.

We thus propose, along the line of Wiltschko (2008) and Zhang (2015), that AABB reduplication constitutes a modification/adjunction process which targets category-less roots.

4.1 Reduplication of (compound) roots

Over the last two decades, frameworks of word formation, especially Distributed Morphology or Borer's exoskeletal framework (2003), have taken very seriously the hypoth-

---

[30]Whether they are category-less roots/stems or standard lexemes endowed with category features will be discussed throughout section 4.

esis that roots, as the invariant core of full-fledged words (stripped away of all morphological formatives) are category-less elements, and that they must be combined in the syntax with category assigning heads (see among others Marantz 2001, Embick & Noyer 2007, Embick & Marantz 2008). Under this view, lexemes/words never are atomic entities, but are the spell-out forms of roots selected by a functional head, i.e. *a, n, v*, determining the corresponding phrasal domain, so that: N = [*n* + √], V = [*v* + √], A = [*a* + √].

Adopting this approach to word formation and its compositional analysis of lexemes, a possibility allowed by the system is that morphological phenomena traditionally described as 'derivational' do not actually target lexemes proper but category-less items, i.e. category-less roots. Increasing reduplication in Mandarin would then fall within the realm of those phenomena that apply at a very 'low' level in the morphosyntactic derivation, namely before categorization takes place. Leaving aside for the moment the complicating factor that the base of increasing reduplication is not a single root but a compound form made up of two roots (see section 4.4 for further discussion on this), under this analysis, it naturally follows that the whole reduplicated AABB form can be assigned to different lexical categories, in accordance with the ontological (/sortal) specification of the root, i.e. whether it denotes objects, events, or (gradable) qualities/attributes.

(38)   a.
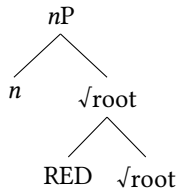
b.
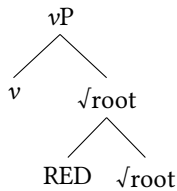
c.
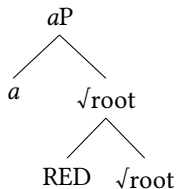
In (38) we limited our representation to nouns, verbs and adjectives, but the analysis can be in principle extended to other categories too, like adverbs. The assumption that roots are atomic, non-decomposable elements virtually independent of the traditional lexical categories (i.e. roots are not associated with categorial information, as e.g.

nouns, verbs, adjectives; see Marantz 1997) allows for a unified analysis of AABB reduplication across categories. Under this approach, reduplication involves acategorial items, and categorization is determined afterwards, in accordance with the type of category-determining heads, i.e. *n*, *v*, *a*, and under the assumption that "whatever category can select for roots can also select for pluralized roots, because pluralized roots are still roots" (see Wiltschko 2008: 60).

While we argue, along the line of Wiltschko (2008) and Zhang (2015), that a single structural analysis is capable to explain for all the category patterns of increasing reduplication, the interpretive outcomes of reduplication are still in need of a satisfactory analysis in the literature.

As can be observed in other languages too, reduplication of nouns and verbs results is a (lexical) means of pluralization. The existence of lexical plurals, in particular, in the nominal domain is well attested across languages, with Italian, for instance, having a class of (feminine) nouns that are lexically specified as being plural (e.g. *braccia* 'arms', see Acquaviva 2008). As for the Chinese cases under consideration, according to Zhang (2015), AABB reduplication expresses overall a 'greater plural' meaning, which can apply both to individual-denoting and to action-denoting elements. In particular, this plural marker, according to Zhang, is integrated in the word-formation domain, where instead of categorial features, semantic features (see Cinque 1990, Lieber 2004, Lieber 2006) and probably phonological features, take part in the selection.

Zhang's analysis relies much on Wiltschko's (2008) analysis of pluralization in Halkomelem Salish. Wiltschko proposes, based on different distributional properties, that in a language like English, with obligatory plural marking, and in a language like Halkomelem, with optional plural marking, plural markers differ in their 'way' and place of merging. While in English, as it is generally assumed, the plural marker spells out the plural value of a functional head selective for a phrasal node such as little *n*, in Halkomelem plural marking functions as a modifier of the category-less root:

(39) a. English

b.   Halkomelem

```
                    D
                   / \
                  /   \
                 D     n
                      / \
                     /   \
                    n    √root
                        /    \
                       /      \
                 PLURALIZER   √root
```
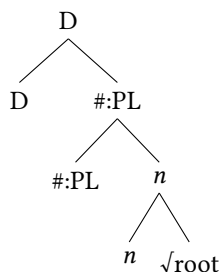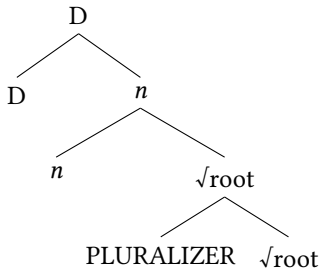
According to Wiltschko (2008: 688), modifying plural markers (39b) have the syntax of adjuncts, rather than of selecting heads, because of a set of properties setting them aside from functional plurals: they are not obligatory; they do not trigger agreement; their absence is not associated with a specific meaning, but instead is truly unmarked; they cannot be selected for; they do not allow for form-meaning mismatches.

We argue that the root-adjoined analysis in (39b) can be the correct analysis for the Mandarin AABB reduplication under examination, where the 'pluralizer' is expressed by means of the reduplicative pattern itself, i.e. by means of independent phonological copying of both base units.[31] This explains for several peculiar features of AABB reduplication, such as its non-obligatoriness and cross-categoriality, as well as its compatibility with the plural marker 們 –*men*, possibly used to emphasize plurality (see fn. 22), and with nominal classifiers. In particular, as we have noticed in section 3.3 (30b), reduplication and pluralization are not incompatible:

(40)   子子孫孫們 (extracted from ex. (30b))
       zǐ~zǐ-sūn~sūn-men
       son~son-grandson~grandson-PL

       'heirs/generation after generation of descendants'

Furthermore, the plural meaning of increasing reduplication is not merely 'plural': since it applies to a coordination of entities/individuals which are *per se* inherently plural (AB means the sum of the entities/individuals denoted by A and those denoted by B, see section 3.4), its meaning is that of 'excessive/greater plural'.

Another striking feature shared by Halkomelem Salish and Mandarin lies in the fact that their 'lexical' plural marking is not restricted to nouns, different from inflectional plural marking which is typically bound to nominal lexemes (not counting agreement plural marking, which can occur wherever it is required). This leads us to discuss the other lexical categories of the outputs of these reduplicative processes.

As for the verbal domain, pluractional meaning of reduplicated verbs is certainly not exceptional in a cross-linguistic perspective. A great deal of reduplicative processes

---

[31]The intriguing issue of the peculiar phonological exponence of disyllabic increasing reduplication is left for future investigation, but we refer to Feng (2003) for an interesting analysis within Optimality Theory framework. See section 4.4. for further remarks on this.

across languages show a pattern close to Mandarin, where (increasing) reduplication in the verbal domain implies repetition/iteration of the event expressed by the base, hence operating over the verb aspectual structure. This means that increasing reduplication has an inherent quantificational meaning, resulting in a plurality of individuals or in a pluractionality of events, in compliance with the (vague) root meaning, ultimately determined by the type of selecting head, *n* vs. *v*, taking the reduplication as its complement (see (38)). Another property in common with nouns and, to the best of our knowledge, specific of Mandarin Chinese, is the need for a base composed of coordinated roots (especially in the case of verbs), standing in a symmetrical relation. We will come back to this intriguing issue in section 4.3.

4.2 Zooming in on adjectives

Whereas the plural analysis seems to nicely fit the nominal and verbal domains of AABB reduplication, it remains to be understood what the interpretive analysis of adjective reduplication is. Interestingly, Wiltschko (2008) observes that in Halkomelem Salish the pluralizer (be it an affix, ablaut or a reduplicated form) occurs productively not only with nouns (41a, 41b), but with verbs (41c) and adjectives (41d) too (Wiltschko 2008: 641, 679-680), conveying a meaning close to the one we find in Mandarin AABB reduplication:[32]

(41) a. *méle*   *mámele*
         child   child.PL
         'child' 'children'

      b. *q'ámi*   *q'álemi*
         girl   girl.PL
         'girl'   'girls'

      c. *qw'óqw-et*   *qw'óleqw-et*
         whip-TRANS   whip.PL-TRANS
         'whip something/someone' 'whip something/someone several times'

      d. *kw'ós*   *kw'ó-kw'es*
         hot   hot.PL
         'hot'   'real hot/very hot'

Wiltschko (2008) argues that, no matter whether it occurs in the context of nouns, verbs or adjectives, the plural marker is exactly the same. She further observes that, if the plural marker is exactly the same, we expect it having exactly the same meaning in each of these contexts. However, to determine what a root pluralizer denotes, we need to know what a root denotes, i.e. what its sortal type is. Wiltschko thus speculates that roots do not have a specific denotation (vs. nouns, which denote individualities, verbs, which denote eventualities, or adjectives, which denote attributes/qualities); they are able to

---

[32]The reader should note that the unmarked form, here glossed as a singular form, is in fact compatible with both singular and plural interpretation; as we have mentioned, the plural marker is not obligatory in Halkomelem.

name "Events, Things, States and Qualities (see Harley 2005), and the pluralizer appears to simply assert that there are a lot of Events, Things, States, Qualities, depending on the nature of the √root" (p. 686).

While this intuitive explanation in principle could work for nouns and verbs, it is nonetheless far less accurate for depicting the increased semantics of reduplicated adjectives. Looking at the semantic effects that reduplication has on Mandarin adjectives, it does not seem the case that it denotes 'lots of Qualities'. Rather, it seems that AABB adjectives express 'increased intensity', thus affecting the gradable property of the base, and this seems to be true also for many other languages that exhibit reduplication with increasing semantics (with Halkomelem *pluralized* adjectives not counting as an exception in this domain, see (41d)).[33] Since reduplication affects gradability, providing a greater/increased degree value expressed by the base root, we might ask what the interpretive relation is between increasing reduplication in the adjectival domain, on the one hand, and increasing reduplication in the verbal and nominal domain on the other, where reduplication is a means of quantification over entities/individuals and events.

4.3 Wellwood's (2014, 2015) analysis of measurement functions across categories

The analysis of adjectives, especially the fact that only gradable adjectives can be reduplicated, sheds light on the core issue of gradability/scalarity in increasing reduplication. However, as we mentioned in the previous section, the relation between increasing reduplication in the adjectival domain and increasing reduplication in the verbal and nominal domain still remains to be explained. In this section, based on the existing literature, we show that concepts of gradability and measurement, rather than being limited to the adjectival domain, may be applied uniformly across categories. This will help to support our hypothesis on the function of Mandarin increasing reduplication, namely that it expresses a unique function, i.e. 'increased measure', as will be discussed in the next section.

While according to some authors gradability is a distinctive property of adjectives (see e.g. Jackendoff 1977), a great deal of research over the last decades found evidence of gradable properties across lexical categories (see e.g. Bolinger 1972, Bresnan 1973, Doetjes 1997, Neeleman et al. 2004, Caudal & Nicolas 2005, Bochnak 2010). As observed by Nicolas (2010), gradable expressions are found among: plural count nouns (*more dogs*), but not singular count nouns (**more dog, *less cup*); mass nouns, concrete (*more water, less wine*) or abstract (*more sadness, less playfulness*); adjectives (*smaller, less sad*); verbs (*to work more/less*).

Wellwood (2015) puts forward a unified account of comparison across categories, challenging those theories that consider gradable adjectives as elements specifying measure functions (see above) vs. nouns and verbs, which allegedly do not express such measure functions. According to this scholar, "which dimensions are possible across domains is a

---

[33] According to Xu (2012a), reduplication is iconically motivated, and 'positive degree' constitutes its core meaning.

consequence of what is measured, rather than which expressions measure" (p. 69). Wellwood (2015: 69) also observes that a noun like *coffee* introduces individuals that can be measured, while a verb like *run* introduces events and an adjective like *tall* introduces states; in any case, they all can be measured along certain types of dimensions, specifically those which respect 'part-whole' relation (e.g. volume and weight for *soup*, but not temperature; time and distance for *run*, but not speed[34]). She posits a variable in nominal and verbal domains "that ranges over measure functions, restricted to just those that are homomorphic to the measured domain" (p. 68). Wellwood (2014, 2015) argues that comparative sentences in the adjectival, nominal and verbal domain all contain instances of a single (phonologically overt or covert) morpheme that compositionally introduces degrees; "this morpheme, sometimes pronounced *much*, contributes a structure-preserving map from entities, events, or states, to their measures along some dimension." (Wellwood 2015: 67).

This approach characterizes the notion of "measurement" uniformly in terms of structure-preservation across comparative constructions and unifies the contrasts existing (within each category) between gradable and non-gradable adjectives, between mass and count nouns, and between atelic and telic verb phrases.[35] Wellwood observes that mass nouns tend to show cumulative reference: "if *coffee* applies to two portions of matter, then it also applies to the mereological sum of those portions" (p. 71). In contrast, count nouns, when interpreted singularly, tend to show non-cumulative reference: "if a *cup* applies to a given object, it fails to apply to any of its (relevant) proper parts" (p. 71). Therefore, the semantics of mass nouns is modelled in terms of a domain structured by the part-of relation, while that of a noun like *cup* lacks such structure. Similarly, atelic predicates (like mass nouns) tend to show cumulative reference, while telic predicates tend to show quantized, non-cumulative reference. If *run in the park* applies to two stretches of activity, it also applies to their sum; thus atelic events have domains structured by the part-of relation on events. In contrast, if *run to the park* applies to an event, it fails to apply to any of its relevant subparts; thus telic events lack the part-of relation (Wellwood 2015: 73).

As for adjectives, Wellwood proposes that non-gradable adjectives, which express quantities that either exist or not (a table is either square or not, it cannot be more or less square) are formally parallel to (singular) count nouns and telic predicates, while gradable adjectives, which express quantities that there may be more or less of (a thing can be more or less hot), are parallel to mass nouns and atelic predicates. They both express predicates of states, the difference being that gradable adjectives, unlike non-gradable ones, predicate of ordered states: they associate directly with sets of ordered degrees, or scales. Besides, Wellwood assumes that the measure functions introduced with gradable adjectives are not only homomorphic to the ordering relations on the measured domain, but to non-trivial part-whole relations.

---

[34]For example, she observes that larger portions of soup have greater measures by volume or weight than smaller portions, but generally this is not the case with measures by temperature.

[35]Gradability presupposes the existence of a scale, and can be seen as related to ±boundedness (see Paradis 2001, Alexiadou 2010).

Therefore, instead of adopting a notion of 'measurement' based on a variety of measure functions acting on the same objects in unpredictable ways, Wellwood proposes that language encodes measurement of different sorts of things in limited ways. Accordingly, she elaborates a uniform account of measurement as a monotonic mapping from ordered sets of entities, events, or states to degrees.

4.4 Reduplication as increased measure

Let us now try to combine the structural analysis of increasing reduplication proposed in section 4.2 with the cross-categorial (strictly compositional) analysis of measurement functions proposed by Wellwood (2014, 2015). Keeping with Wellwood's proposal that there are no differences in the type of measurement functions among the lexical categories at a higher level of syntactic/semantic composition, we speculate that reduplication conveys a similarly stable/unique function but it targets elements lacking any specification in terms of formal features.[36] In particular, we wish to argue that reduplication expresses a unique function, i.e. 'increased measure', that constantly applies to roots, only differing in their ontological denotation. Therefore, increasing reduplication is a very low-level ('morphological') adjunction operation which conveys the function 'increased measure' to the roots it applies to: the semantic effects obtained (pluralization, pluractionality, intensification of the base gradable property) ultimately depend on the different sort of things reduplication modifies, and arguably emerge constructionally, that is, after root categorization applies. It should be noticed that, semantically, similar results might be obtained at higher level of syntactic composition via different means, depending on the categorial domain of application, i.e. through fully-fledged degree phrases in the adjectival domain (see En. 'very Adj', e.g. *very good*; Ch. '很 *hěn* Adj', e.g. 很高興 *hěn gāoxìng* 'very happy'), and through the use of plural affixes and aspectual markers in the nominal and verbal domain respectively.

This analysis, however, does not account for some relevant asymmetries across lexical categories previously noted in the literature (see Zhang 2015). As it has been argued in section 3, the main difference at the structural level between adjectives, on the one hand, and nouns and verbs, on the other, concerns the obligatoriness of disyllabic bases for the latter. That is, whereas increasing reduplication applies to quality-denoting roots that may be either mono- or disyllabic, resulting in AA and AABB patterns interpretively equivalent, with entity and event denoting roots it targets disyllabic units, resulting exclusively in the AABB pattern.[37]

As we have seen in 3.3, the AABB reduplication pattern requires a coordinate base, i.e. two elements related in a symmetrical fashion, either in a logical coordination, or synonyms or antonyms; thus, instead of having a single root we have a combination

---

[36]It is worth reminding that roots have a strongly underspecified semantics which allows them to be compatible with the semantics of adjectives (as properties of attributes), verbs (as properties of events), nouns (as properties of individuals).

[37]The generalization holds under the assumption that AA monosyllabic reduplication in the nominal domain should be rather understood as reduplication of classifiers (see section 3.3). We do not have an analysis of this type of reduplication yet, and we leave the issue for future research.

of roots. These roots are joined together to form a set, whereby the two constituents equally contribute to the semantics of the whole complex stem, i.e. they are in a symmetrical relation. Structurally, it is worth emphasizing that these operations all apply at the root level, resulting in a recursive application of 'morphological' phenomena, with (symmetrical) compounding and reduplication rigidly ordered in the derivation, yet both applying before categorization (see Zhang 2015):

(42)
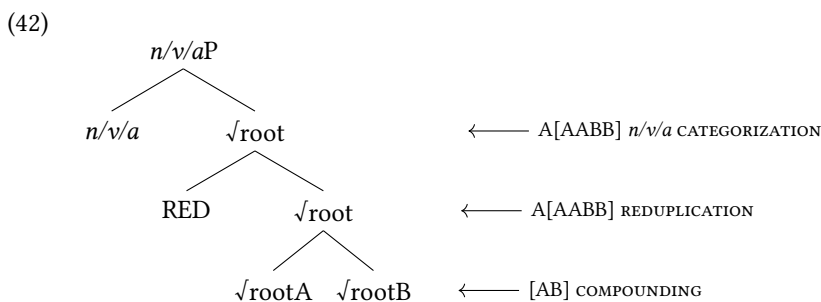
This analysis seems to produce the surface pattern ABAB, since reduplication applies to a compound base AB. However, prosodic patterns within AABB structures actually seem to support the structural analysis in (42). In particular, Feng (2003) examines tone *sandhi* rules within disyllabic reduplication and, for AABB, he argues that these rules apply first between the second A and first B and then between the first B and second B. On this basis, Feng argues that AB is the actual morphological unit, whereas AA and BB are not, resulting in the structural analysis [A[AB]B] (Feng 2003: 7-8). The issue deserves further investigation especially aimed at explaining the reason for the mismatch between underlying structure, supra-segmental patterns and surface order of morphemes, for which at the moment we cannot offer an explanation. Suffice it to say that the prosodic pattern of AABB provides evidence in favour of the analysis in (42).

At the interpretive level, we put forward that the combination of two roots which act as the base for the AABB reduplication process forms itself a sort of 'plural/collective' expression and reduplication provides an *increased measure* for this kind of expressions. It has been noted that AABB nouns express greater plural (possibly differing in the semantics from AA reduplication of nouns/classifiers, most typically expressing a distributive meaning), and a similar effect is obtained with AABB verbs (ex. in (43a) and (43b) are adapted from examples (22, 24) in Zhang 2015):

(43)   a.   枝枝葉葉
            zhī~zhī-yè~yè
            twig~twig-leaf~leaf
            'twigs and leaves'
       b.   縫縫補補
            féng~féng-bǔ~bǔ
            sew~sew-repair~repair
            'sew and repair repeatedly'

A possible explanation for this structural requirement might lie in the different onto-logical type of roots: in particular, individual and event denoting roots, different from quality denoting roots, seem to require an inherently plural interpretation in order to be measured. As a matter of fact, typically comparative expressions with *more* in English require either mass nouns or plural nouns, but exclude singular nouns (*more dogs* vs. *more dog). Similar effects obtain in the domain of verbs with the contrasts between telic and atelic verbs discussed by Wellwood (2015).

Although at this point the present analysis becomes very speculative, we put forward here that a principled reason for the necessary disyllabicity of nominal and verbal bases might have the same source of the asymmetry observed in the domain of comparative expressions. Specifically, if the semantics of roots is very vague and compatible with any interpretation which eventually emerges at higher levels of syntactic composition, a way to introduce gradability at the level of roots is to merge them directly, so to create a collection of individuals, like e.g. 男女 *nán-nǚ* 'man and woman' (which is reduplicated as 男男女女 *nán~nán-nǚ~nǚ* 'men and women'), or of events, e.g. 起伏 *qǐ-fú* 'rise and fall' (which is reduplicated as 起起伏伏 *qǐ~qǐ-fú~fú* 'rise and fall repeatedly'). In this view, the first merger provides reduplication with the 'gradable base' over which it can apply its increased measure function. On the contrary, roots that are selected by an adjectival head (i.e. *a*) would inherently express a gradable property and, accordingly, reduplication would not pose specific disyllabic requirements on these base units. Furthermore, if this is the case, we expect no difference in meaning between the reduplication of AA and AABB adjectival forms, as confirmed by the data (see examples (6a) and (6b) in section 2.1, repeated below for the reader's convenience):

(44) a. 小 (A)   →   小小 (AA)
          xiǎo   *xiǎo~xiāo*
          small   small~small
          'small'   'very/really small'

      b. 高興 (AB)   →   高高興興 (AABB)
          gāoxìng   *gāo~gāo-xìng~xìng*
          'happy'   'very/really happy'

5 Conclusion

Reduplication is a challenging phenomenon in many respects: it is hardly amenable to a uniform characterization in a cross-linguistic perspective, given the extreme variety of forms and functions it is associated with; further, it can surface with different forms and meanings within a single language too, as we have shown with the reduplicative processes of Mandarin under consideration; it can manifest semantic functions closely related to the inflectional/functional domain, but it approaches more closely the domain of derivation/word formation; finally, it can take as its base units elements of different size, ranging from lexeme/word-like units in one domain (diminishing reduplication,

which implies *verbal* reduplication in Mandarin) to category-less units in the other (increasing reduplication).

The case of diminishing reduplication seems to involve units as 'big' as lexemes, i.e. stems endowed with category features and with specific (aspectual) semantics, as we have shown in section 2.1. The case of increasing reduplication, however, points to the existence of word formation phenomena that applies below the lexeme level. In particular, increasing reduplication seems to suggest that it is a phenomenon that can apply at a very 'low level', namely, that it can merge with roots/stems lacking category specification. Further, it is *per se* unable to express a definite category, given its presence across all major lexical categories at both input and output levels. Therefore, the present case study sheds some light on the existence of word formation that does not take lexemic inputs and does not give lexemic outputs either.

On the one hand, this study brings further evidence in favor of a neo-constructionist/ DM-like view of the lexemes or word units as syntactically complex elements, and ultimately for the very existence of category-less roots. On the other hand, the curious asymmetries observed in the domain of increasing AA and AABB reduplication, whereby adjectives seem to part company from verbs and nouns, call into question the semantic (ontological?) character of roots and their alleged requirements for insertion in the syntactic structure responsible for category assignment and, overall, for their morphosyntactic properties and distribution. This is a very complex issue on which we hope to have contributed some further empirical and theoretical basis but that, it goes without saying, needs further research and ampler empirical coverage to be satisfactorily addressed.

To conclude, our research has explored the structural and interpretive effects of reduplication, so productive in Mandarin (see Basciano & Melloni 2017) and broadly attested across Sinitic (see Arcodia et al. 2015) yet still lacking a satisfying analysis, despite of a growing interest in the last years. So doing, we hope to have paved the way for a better understanding of Mandarin reduplication specifically, and more in general for an approach to word formation which seeks to reinterpret morphology-specific properties and restrictions within a more integrated model of grammar, where syntax is also responsible for word formation.

Acknowledgments

Author contributions   The paper is the result of close collaboration between the two authors, who are listed in random order. For academic purposes only, Chiara Melloni

takes responsibility for Sections 1, 3, 3.1, 3.2, 3.5, 4.3, 4.4, 5, and Bianca Basciano takes responsibility for Sections 2, 3.3, 3.4, 4, 4.1, 4.2.

References

Acquaviva, Paolo. 2008. *Lexical plurals: A morphosemantic approach.* Oxford: Oxford University Press.

Alexiadou, Artemis. 2010. Reduplication and doubling contrasted: Implications for the structure of the DP and the AP. *Linguística – Revista de Estudos Linguísticos da Universidade do Porto* 5. 9–25.

Arcodia, Giorgio F. 2014. The Chinese adjective as a word class. In Raffaele Simone & Francesca Masini (eds.), *Word classes: Nature, typology and representations*, vol. 332, 95. Amsterdam: John Benjamins.

Arcodia, Giorgio F. & Bianca Basciano. 2017. Morphology, modern. In Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang & James Myers (eds.), *Encyclopedia of Chinese language and linguistics*, vol. 3, 104–118. Leiden: Brill.

Arcodia, Giorgio F., Bianca Basciano & Chiara Melloni. 2014. Verbal reduplication in Sinitic. In Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé & Fabio Montermini (eds.), *Proceedings of the 8th Décembrettes*, vol. 22 (Carnets de grammaire), 15–45. Bordeaux: CLLE-ERSS.

Arcodia, Giorgio F., Bianca Basciano & Chiara Melloni. 2015. Areal perspectives on total reduplication of verbs in Sinitic. *Studies in Language* 39(4). 836–872.

Basciano, Bianca. 2017. Word classes, Modern. In Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang & James Myers (eds.), *Encyclopedia of Chinese language and linguistics*, vol. 4, 554–566. Leiden: Brill.

Basciano, Bianca & Chiara Melloni. 2017. Event delimitation in Mandarin: The case of diminishing reduplication. *Italian Journal of Linguistics / Rivista di linguistica* 29(1). 147–170.

Bisang, Walter. 2008. Precategoriality and argument structure in Late Archaic Chinese. In Jaako Leino (ed.), *Constructional reorganization*, 55–88. Amsterdam: John Benjamins.

Bisang, Walter. 2012. Numeral classifiers with plural marking. A challenge to Greenberg. In Dan Xu (ed.), *Plurality and classifiers across languages in China*, 23–42. Berlin: Walter de Gruyter.

Bochnak, M Ryan. 2010. Quantity and gradability across categories. In *Proceedings of semantics and linguistic theory 20*, 251–268.

Bolinger, Dwight. 1972. *Degree words.* The Hague: Mouton.

Borer, Hagit. 1994. The projection of arguments. *University of Massachusetts occasional papers in linguistics* 17. 19–47.

Borer, Hagit. 2003. Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In John Moore & Maria Polinsky (eds.), *The nature of explanation in linguistic theory*, 31–66. Stanford: CSLI Publications.

Borer, Hagit. 2005. *Structuring sense. In the name only*. Vol. 1. Oxford: Oxford University Press.

Bresnan, Joan. 1973. Syntax of the comparative clause construction in English. *Linguistic inquiry* 4(3). 275–343.

Cai, Chaohui. 2007. Tanxi hanyu zhong mingci de chongdie 探析汉语中名词的重叠 [an analysis of noun reduplication in Mandarin]. *Yangzhou daxue xuebao* 扬州大学学报 11. 51–54.

Caudal, Patrick & David Nicolas. 2005. Types of degrees and types of event structures. In Claudia Maienborn & Angelika Wöllstein (eds.), *Event arguments: Foundations and applications*, 277–300. Tübingen: Niemeyer.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*.

Cinque, Guglielmo. 1990. Ergative adjectives and the lexicalist hypothesis. *Natural Language & Linguistic Theory* 8(1). 1–39.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective.* Oxford: Oxford University Press.

Ding, Yongshou. 2010. *A complete guide to teaching Chinese as a second language-* 对外汉语教学参考. Beijing: Beijing Language & Culture University Press.

Doetjes, Jenny. 1997. *Quantifiers and selection*. Leiden: Rijksuniversiteit Leiden Doctoral dissertation.

Embick, David & Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39. 1–53.

Embick, David & Rolf Noyer. 2007. Distributed morphology and the syntax-morphology interface. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 289–324. Oxford: Oxford University Press.

Feng, Guanjun Bella. 2003. Lexical category specific constraints: Mandarin verb versus adjective reduplication. *USC Working Papers in Linguistics* 1. 1–12.

Forza, Francesca. 2011. *Regarding reduplication and repetition: A separate treatment in a unified approach*. Verona: Università degli Studi di Verona Doctoral dissertation.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Ghomeshi, Jila, Ray Jackendoff, Nicole Rosen & Kevin Russell. 2004. Contrastive focus reduplication in English (The salad-salad paper). *Natural Language & Linguistic Theory* 22(2). 307–357.

Greenberg, Joseph H. 1972. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Working Papers on Language Universals* 9.

Guo, Jimao. 1999. Zai tan liangci chongdie xingshi de yufa yiyi 再谈量词重叠形式的语法意义 [revisit the grammatical meaning of classifier reduplication]. *Hanyu Xuexi* 汉语学习 8(4). 6–9.

Harley, Heidi. 2005. How do verbs get their names? Denominal verbs, Manner incorporation and the ontology of verb roots in English. In Tova Rapoport & Nomi Erteschik-Shir (eds.), *The syntax of aspect*, 42–64. Oxford: Oxford University Press.

Hu, Xiaobin. 2006. Dongci chongdie aabb-shi de yufahua 动词重叠 aabb 式的语法化 [grammaticalization of the aabb pattern of verbal reduplication]. *Hanyu xuexi* 汉语学习 8(4). 18–25.

Hu, Xizhi. 1994. Shilun xiandai hanyu mingci chongdie 试论现代汉语名词重叠 [on the reduplication of nouns in modern Chinese]. *Fuyang shifan xueyuan xuebao* 阜阳师范学院学报 1. 101–115.

Jackendoff, Ray S. 1977. X̄ *syntax: A study of phrase structure*. Cambridge: MIT Press.

Kwong, Oi Yee & Benjamin K Tsou. 2003. Categorial fluidity in Chinese and its implications for part-of-speech tagging. In *Proceedings of the tenth conference on European chapter of the association for computational linguistics-volume 2*, 115–118.

Lapointe, Steven Guy. 1980. *A theory of grammatical agreement*. Amherst: University of Massachusetts dissertation.

Lee-Kim, Sang-Im. 2016. Syntax-based phonological asymmetries: The case of adjective reduplication in Mandarin Chinese. *Lingua* 179. 1–23.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese. A functional reference grammar*. Berkeley: University of California Press.

Li, Sijun. 2009. Hanyu mingci chongdie yanjiu zongshu 汉语名词重叠研究综述 [an overview of the research on nominal reduplication in Mandarin]. *Chongqing wenli xueyuan xuebao* 重庆文理学院学报 28(6). 175–188.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle. 2006. The category of roots and the roots of categories: What we learn from selection in derivation. *Morphology* 16(2). 247–272.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *Penn Working Papers in Linguistics* 4. 201–225.

Marantz, Alec. 2001. Words and things. Ms. MIT.

Mattes, Veronika. 2014. *Types of reduplication: A case study of Bikol*. Vol. 16. Berlin: Walter de Gruyter.

McClure, William Tsuyoshi. 1995. *Syntactic projections of the semantics of aspect*. Ithaca: Cornell University Doctoral dissertation.

Neeleman, Ad, Hans Van de Koot & Jenny Doetjes. 2004. Degree expressions. *Linguistic review* 21(1). 1–66.

Nicolas, David. 2010. Towards a semantics for mass expressions derived from gradable expressions. *Recherches linguistiques de Vincennes* (39). 163–198.

Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Paradis, Carita. 2001. Adjectives and boundedness. *Cognitive Linguistics* 12(1). 247–271.

Paris, Marie-Claude. 1979. *Nominalization in Mandarin Chinese: The morpheme 'de' and the 'shi'…'de' constructions*. Paris: Université Paris VII Thèse de doctorat.

Paris, Marie-Claude. 2007. Un aperçu de la réduplication nominale et verbale en mandarin. *Faits de langue* (29). 65–76.

Paul, Waltraud. 2010. Adjectives in Mandarin Chinese: The rehabilitation of a much ostracized category. In Patricia Cabredo Hofherr & Ora Matushansky (eds.), *Adjectives: Formal analyses in syntax and semantics*, 115–152. Amsterdam: John Benjamins.

Ramchand, Gillian Catriona. 2008. *Verb meaning and the lexicon.* Cambridge: Cambridge University Press.

Sagart, Laurent. 1999. *The roots of old Chinese.* Amsterdam: John Benjamins.

Sanches, Mary & Linda Slobin. 1973. Numeral classifiers and plural marking: An implicational universal. *Working Papers on Language Universals* 11. 1–22.

Shi, Y. 2002. *The establishment of modern Chinese grammar. The formation of the resultative construction and its effects.* Amsterdam: John Benjamins.

Tang, Ting-Chi. 1979. *Guoyu yufa yanjiu lunji* 國語語法研究論集 *[studies in Chinese syntax].* Taipei: Taiwan Xuesheng Shuju.

Tang, Ting-Chi. 1988. Guoyu xingrongci de chongdie guilü 國語形容詞的重疊規律 [adjectives reduplication rules in Mandarin Chinese]. In Ting-Chi Tang (ed.), *Hanyu cifa jufa lunji* 漢語詞法句法論集 *[studies on Chinese morphology and syntax]*, 29–57. Taipei: Student Book.

Travis, Lisa. 2000. Event structure in syntax. In Carol Tenny & James Pustejovsky (eds.), *Events as grammatical objects*, 145–185. Stanford: CSLI Publications.

Travis, Lisa. 2010. *Inner aspect: The articulation of VP*. Dordrecht: Springer.

Tsao, Feng-fu. 2001. Semantics and syntax of verbal and adjectival reduplication in Mandarin and Taiwanese Southern Min. In Hilary Chappell (ed.), *Sinitic grammar: Synchronic and diachronic perspectives*, 285–308. Oxford: Oxford University Press.

Wälchli, Bernhard. 2005. *Co-compounds and natural coordination.* Oxford: Oxford University Press.

Wellwood, Alexis. 2014. *Measuring predicates.* University of Maryland, College Park Doctoral dissertation.

Wellwood, Alexis. 2015. On the semantics of comparison across categories. *Linguistics and Philosophy* 38(1). 67.

Wiltschko, Martina. 2008. The syntax of non-inflectional plural marking. *Natural Language & Linguistic Theory* 26(3). 639–694.

Wu, Yin & Jingmin Shao. 2001. Shilun mingci chongdieshi yufa yiyi ji qita 试论名词重叠式语法意义及其他 [on the grammatical meaning of nominal reduplication and other issues]. *Yuwen yanjiu* 语文研究 78. 12–16.

Xiao, Richard & Tony McEnery. 2004. *Aspect in Mandarin Chinese: A corpus-based study.* Amsterdam: John Benjamins.

Xu, Dan. 2012a. Introduction. In Dan Xu (ed.), *Plurality and classifiers across languages in china*, 1–22. Berlin: Mouton de Gruyter.

Xu, Dan. 2012b. Reduplication in languages: A case study of languages of China. In Dan Xu (ed.), *Plurality and classifiers across languages in China*, 43–66. Berlin: Mouton de Gruyter.

Zádrapa, Lukáš. 2017. Word classes, premodern. In Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang & James Myers (eds.), *Encyclopedia of Chinese language and linguistics*, vol. 4, 566–576. Leiden: Brill.

Zhang, Niina Ning. 2014. Expressing number productively in Mandarin Chinese. *Linguistics* 52(1). 1–34.

Zhang, Niina Ning. 2015. The morphological expression of plurality and pluractionality in Mandarin. *Lingua* 165. 1–27.

Zhang, Xiaoqian. 2016. *Expressions duratives en chinois mandarin : Une étude sur l'aspect*. Université Paris Diderot (Sorbonne Paris Cité) Thèse de doctorat.

Zhu, Jingsong. 2003. Xingrongci chongdieshi de yufa yiyi 形容词重叠式的语法意义 [the grammatical meaning of reduplicated adjectives]. *Yuwen yanjiu* 语文研究 3. 9–17.

Chapter 15

La parasynthèse à travers les modèles : Des RCL au ParaDis

Nabil Hathout

CLLE. Université de Toulouse & CNRS

Fiammetta Namer

Université de Lorraine & ATILF CNRS

Cet article est consacré à l'analyse des formes dites parasynthétiques, à la façon dont cette analyse a évolué avec les modèles théoriques qui l'ont appréhendée, et à la manière dont, en retour, elle a contribué à leur changement. L'évolution de l'analyse de la dérivation parasynthétique peut en effet être perçue comme un indicateur des transformations et des progrès des théories morphologiques et des modèles dérivationnels. Nous montrons notamment comment les propositions successives pour l'analyse de ce phénomène ont conduit à un assouplissement progressif des cadres théoriques, à partir des modèles morphémiques où formes et sens sont totalement associés au sein des morphèmes, en passant par les lexèmes et les Règles de Construction de Lexèmes (RCL) qui procèdent à une première séparation entre les trois dimensions du lexème (forme, catégorie et sens), pour arriver aux modèles paradigmatiques de la morphologie dérivationnelle où la relation binaire entre base et dérivé est généralisée à des réseaux de lexèmes connectés à des réseaux de propriétés. Cette progression nous conduit, enfin, à notre objectif final : la présentation du modèle d'analyse constructionnel ParaDis, dont la genèse résulte de l'aboutissement des transformations théoriques successives ou parallèles qui ont façonné les différents courants en morphologie dérivationnelle. Les principes d'analyse de ParaDis combinent les principes formels qui sous-tendent les RCL et la structure tridimensionnelle des lexèmes à une approche en réseau de la construction lexicale. A travers l'exemple de la préfixation en *anti-*, nous montrons comment cette association originale fait de ParaDis un cadre qui dispose des propriétés et des clés nécessaires pour analyser de manière simple et intuitive les constructions parasynthétiques.

1 Introduction

La morphologie dérivationnelle, bien plus que la morphologie flexionnelle, comporte une quantité importante de constructions difficiles à décrire du fait du nombre et de

la diversité des variations observées. Il existe en morphologie flexionnelle des modèles, comme celui de (Stump 2016), capables de décrire la totalité du système pour la plupart des langues européennes. Une grande partie de l'effort de recherche dans ce domaine porte sur l'optimisation des systèmes du point de vue de leur complexité computationnelle, des notions dont ils font usage ou de leur plausibilité psychologique. La situation est très différente en morphologie dérivationnelle où un grand nombre de phénomènes n'ont toujours pas reçu une analyse complète satisfaisante. C'est le cas des très nombreuses formations non canoniques, au sens de Corbett (2010), dont les constructions parasynthétiques constituent un exemple bien connu. Ces constructions sont un objet d'étude à la fois récurrent et ancien en morphologie. En particulier, ce phénomène, qui a intéressé les chercheurs français depuis Darmesteter (1877, 1894), a largement été traité dans le cadre des modèles génératifs des années 1970 (Dell 1970, 1979), puis par les spécialistes majeurs de la morphologie en France, notamment Corbin (1980, 1987) et Fradin (1997a, 1997b, 2003).

Nous nous intéressons dans cet article à l'analyse de ces formes (section 2), à la façon dont elle a évolué avec les modèles théoriques qui l'ont appréhendée, et à la manière dont, en retour, elle a contribué à leur changement. L'évolution de l'analyse de la dérivation parasynthétique peut en effet être perçue comme un indicateur des transformations et des progrès des théories morphologiques et des modèles dérivationnels. Nous montrons notamment comment les propositions successives pour l'analyse de ce phénomène ont conduit à un assouplissement progressif des cadres théoriques, à partir des modèles morphémiques (section 3) où formes et sens sont totalement associés au sein des morphèmes, en passant par les lexèmes et les Règles de Construction de Lexèmes (RCL ; section 4) qui procèdent à une première séparation entre les trois dimensions du lexème (forme, catégorie et sens), pour arriver aux modèles paradigmatiques de la morphologie dérivationnelle (section 5) où la relation binaire entre base et dérivé est généralisée à des réseaux de lexèmes connectés à des réseaux de propriétés.

Cette progression nous conduit, en section 6, à notre objectif final : la présentation du modèle d'analyse constructionnel ParaDis, dont la genèse résulte de l'aboutissement des transformations théoriques successives ou parallèles qui ont façonné les différents courants en morphologie dérivationnelle. ParaDis hérite en particulier de deux approches dont la contribution a été décisive dans l'évolution de la prise en compte des dérivés parasynthétiques en particulier, et, plus généralement, des constructions dérogeant aux principes de canonicité dérivationnelle. Il s'agit d'une part des travaux présentés dans Fradin (2003), et, d'autre part, des analyses développées à Toulouse en réaction à ces propositions théoriques.

ParaDis est développé comme une articulation de l'approche toulousaine avec la dissociation des niveaux formel, catégoriel et sémantique que permet la formalisation du lexème et des RCL développée par Fradin (2003). Le socle du modèle ParaDis est élargi aux patrons cumulatifs de Bochner (1993) et fait des relations morphologiques dérivationnelles l'une de ses unités fondamentales. Ses principes d'analyse combinent ainsi les solutions de l'approche toulousaine, les principes formels qui sous-tendent les RCL et la structure tridimensionnelle des lexèmes à une approche en réseau de la construction

lexicale. A travers l'exemple de la préfixation en *anti-*, nous montrons comment cette association originale fait de ParaDis un cadre qui dispose des propriétés et des clés nécessaires pour analyser de manière simple et intuitive les constructions parasynthétiques.

2 Constructions dites parasynthétiques

On parle de dérivation « parasynthétique », terme introduit par Darmesteter (1875, 1877), pour décrire les structures dérivées (i) qui sont une instance du patron *pref-X-suf* et qui (ii) présentent un décalage entre leur sens et leur forme. En français, les dérivés parasynthétiques sont essentiellement adjectivaux (ex. $grève_N \rightarrow antigréviste_A$), ou verbaux (ex. $sensible_A \rightarrow désensibiliser_V$, $rat_N \rightarrow dératiser_V$), même si des études font également état de parasynthétiques nominaux (ex. $col_N \rightarrow encolure_N$)[1].

En dehors du français, les dérivations parasynthétiques s'observent très largement dans les langues romanes (Reinheimer-Ripeanu 1974, Serrano Dolader 2015) : en portugais (1a), (Basílio 1991) ; en italien (1b), (Guevara 2007, Iacobini 2004, Melloni & Bisetto 2010, Scalise 1994) ; en espagnol (1c), (Serrano Dolader 1995, Schroten 1997) ; mais aussi en grec (1d), (Efthymiou 2014) et dans les langues slaves comme le slovaque (1e) ou germaniques comme l'allemand (1f), où certains types de composés dits « synthétiques » ou « exocentriques » présentent une configuration analogue (voir entre autres Neef (2015), Gaeta (2010), Crocco-Galéas (2003), Chovanová (2010) et pour un panorama complet, Lieber & Štekauer (2009)).

(1)   a.   *a-X-cer*
           apodrecer 'pourrir' où *X = podre* 'pourri'

      b.   *extra-X-ale*
           extramatrimoniale 'hors mariage' où *X = matrimonio* 'mariage'

      c.   *sub-X-ino*
           submarino 'sous-marin' où *X = mar* 'mer'

      d.   *apo-X-izo*
           apokefalizo 'décapiter' où *X = kefale* 'tête'

      e.   *Y-X-ý*
           bosonohý 'aux pieds nus' où *Y = bosý* 'nu' et *X = noha* 'pied'

      f.   *Y-X-ig*
           blauaügig 'aux yeux bleus' où *Y = blau* 'bleu' et *X = Auge* 'oeil'

Une propriété commune à ces constructions est la variabilité des valeurs que peut prendre la séquence *X-suf* à *pref-* constant. Les exemples sous (2) illustrent en français

---

[1]Un autre type de construction a longtemps été considéré comme faisant partie de cette classe de dérivés. Il s'agit des verbes formés par préfixation, comme *dépoussiérer* en français ou son équivalent *spolverare* en italien. Pour ses défenseurs, cette analyse repose sur deux justifications : (i) la préfixation serait dépourvue de pouvoir catégorisateur, et (ii) la marque flexionnelle suffixale qui apparaît systématiquement sur les verbes dans les langues romanes possède un pouvoir dérivationnel dont l'explication fait intervenir des facteurs diachroniques (Crocco-Galéas & Iacobini 1993, Iacobini 2010, Acedo-Matellán & Mateu 2009).

la construction d'adjectifs adversatifs qui comportent tous le préfixe *anti-*. La séquence suffixale (*-al*, *-ique*, *-aire*, *-eux*, *-ique*) varie, sans que cette variation n'ait un impact sur l'interprétation de l'adjectif. Pour un *X* donné, on remarque que cette séquence est isomorphe à l'exposant de la règle formant l'adjectif dénominal *X-suf* (ex. *gouvernemental* 'en relation avec le gouvernement').

(2) *anti-X-al* (*antigouvernemental* où *X = gouvernement*),
  anti-X-ique (*antialcoolique* où *X = alcool*),
  anti-X-aire (*antiparlementaire* où *X = parlement*),
  anti-X-eux (*anticancéreux* où *X = cancer*).

Relativement aux critères de canonicité énoncés par Corbett (2010), les dérivés parasynthétiques présentent un éloignement clair vis-à-vis de la situation idéale, représentée par l'observation concomitante de deux propriétés sur un dérivé : transparence formelle et compositionnalité du sens. Dans les formations parasynthétiques en effet, les deux découpages formels possibles, *pref- +X-suf* et *pref-X+ -suf*, sont incompatibles avec la décomposition sémantique : il y a dans le dérivé une marque formelle (i.e. *-suf*) non corrélée à un élément servant la construction du sens. En d'autres termes, on a affaire ici à un cas de ce que Hathout & Namer (2014b) nomment « surmarquage formel », exprimé par une séquence phonologique suffixale dont la forme est variable.

Pour résoudre cette divergence, les modèles morphologiques développent trois types de stratégies complémentaires :

1. ils privilégient l'interprétation du dérivé et amendent les principes théoriques pour rendre compte de sa forme ;

2. ils favorisent la régularité formelle au détriment de la construction du sens, pour laquelle ils mettent en œuvre des aménagements particuliers ;

3. ils interviennent sur les deux niveaux pour que les représentations se correspondent.

Nous allons voir (section 3) que les modèles morphématiques, que ce soit ceux qui relèvent du cadre Item et Arrangement ou ceux qui adoptent une conception plus fonctionnelle du morphème affixal (Corbin 1987), choisissent la première option. Nous montrons ensuite, dans la section 4, comment Fradin (2003), qui inscrit son modèle dans le courant lexématique de la morphologie, opte pour la deuxième solution. Enfin, nous expliquons en section 6, comment le système ParaDis, conçu comme une synthèse des principes présentés en section 4 et des propositions toulousaines (section 5), s'efforce de suivre la troisième des stratégies listées ci-dessus.

3 Parasynthèse et morphologie morphématique

Les principes à l'œuvre dans le courant morphématique traditionnel de la morphologie dérivationnelle conçoivent la construction d'un mot comme le fruit de concaténations

binaires successives conformément à la *Binary Branching Hypothesis* héritée du structuralisme et adoptée en morphologie par Aronoff (1976), Booij (1977) entre autres[2]. L'un des deux constituants réunis dans une règle est un morphème affixal, unité minimale de forme et de sens, qui contraint les propriétés phonologiques, catégorielles et sémantiques de l'autre constituant, i.e. la base à laquelle il se combine. Ces contraintes affectant simultanément ses dimensions sémantique et formelle, deux difficultés apparaissent pour l'analyse des parasynthétiques. Les deux illustrent le décalage entre forme et sens mais correspondent chacune à l'une des réalités que recouvre la notion de parasynthèse :

1. La base du dérivé parasynthétique est non attestée. Pour le verbe *dératiser* par exemple, aucune des deux bases que l'on obtiendrait en supprimant le préfixe ou le suffixe n'est attestée. Ce verbe n'est ni préfixé sur *ratiser* ni suffixé sur *dérat*.

2. Le sens du dérivé est non compositionnel, comme l'illustrent les exemples *antigréviste*$_A$ et *désensibiliser*$_V$. À première vue, le verbe *désensibiliser* pourrait avoir pour base *sensibiliser*$_V$, mais le sens de *désensibiliser* n'est pas une fonction de celui de *sensibiliser* ; en effet, « désensibiliser une dent » ne signifie pas 'annuler le résultat obtenu à l'issue de l'acte de sensibiliser la dent', mais 'priver une dent de son caractère sensible'. De la même manière, le sens de *antigréviste* 'qui s'oppose à la grève' ne fait intervenir ni celui de *gréviste* (ce qui supposerait une préfixation en *anti-* appliquée à une base suffixée), ni celui d'*antigrève* (ce qui correspondrait à la suffixation par *-iste* d'un mot préfixé en *anti-*).

Citant Serrano Dolader (1995 : 23-74), Iacobini (2004 : 167) résume en trois schémas d'analyse les solutions que les tenants du cadre morphématique adoptent pour les dérivés parasynthétiques. Outre la solution consistant à préserver la binarité des règles de dérivation, et à activer successivement les procédés de suffixation puis de préfixation ou vice-versa, la démarche parasynthétique préconise soit la concaténation simultanée du préfixe et du suffixe au morphème de base, soit l'attribution du statut de circonfixe à la séquence formée par le préfixe et le suffixe ; la troisième approche, défendue dans Corbin (1980, 1987), passe par l'attribution d'un pouvoir catégorisateur au processus de préfixation, ce que récusent par exemple Scalise (1984) ou Alcoba-Rueda (1987). Nous illustrons ci-dessous ces types d'analyse à travers les exemples des verbes *dératiser* (« dératiser *Y* » signifie 'enlever les rats de *Y*') et de *désensibiliser* (« désensibiliser *Y*» signifie 'priver *Y* de son caractère sensible').

3.1 Application séquencielle de *dé-* et *-iser*

Pour préserver la nature à la fois homocatégorielle et binaire des règles de combinaison préfixe · base, l'analyse des dérivés parasynthétiques que défend par exemple Alcoba-Rueda (1987) consiste à voir dans *dératiser* et *désensibiliser* le résultat de la concaténation du préfixe *dé-* appliqué, respectivement, au nom *rat* et à l'adjectif *sensible*, suivi de

---

[2]Voir aussi Guevara (2007) pour une justification théorique et Heyna (2014) pour un panorama complet des traitements proposés en français dans ce cadre théorique, ainsi que pour une proposition d'analyse des dérivés parasynthétiques adjectivaux en *anti-* et verbaux en *dé-*.

celle du suffixe *-iser*, sélectionnant la base non attestée nominale °*dérat-* ou adjectivale °*désensibl-* obtenue à l'issue de la première étape. L'analyse de *dératiser*, traduite dans une notation parenthésée permettant de coder son histoire dérivationnelle, est donnée en (3) et celle de *désensibiliser* en (4). Chez Scalise (1984), le raisonnement est le même, à l'ordre de l'application des règles de combinaison affixe · base près. Pour cet auteur, la dernière étape de la construction de *désensibiliser* (resp. *dératiser*) est la concaténation du préfixe *dé-* à une base suffixée par *-iser* (*sensibiliser*) éventuellement non attestée (*ratiser*). Ces dérivations correspondent, respectivement, aux représentations données en (5) et (6).

(3) $[[\text{ } dé\text{-} \text{ } [\text{rat}]_N]_N \text{ } \text{-}iser \text{ }]_V$

(4) $[[\text{ } dé\text{-} \text{ } [\text{sensible}]_A]_A \text{ } \text{-}iser \text{ }]_V$

(5) $[\text{ } dé\text{-} \text{ } [[\text{rat}]_N \text{ } \text{-}iser \text{ }]_V]_V$

(6) $[\text{ } dé\text{-} \text{ } [[\text{sensible}]_A \text{ } \text{-}iser \text{ }]_V]_V$

3.2 Application simultanée de *dé-* et *-iser*

Une autre des solutions proposées pour expliquer la construction de *dératiser* et *désensibiliser* est l'adjonction simultanée de *dé-* et de *-iser* sur le nom *rat* ou l'adjectif *sensible*. Les représentations schématiques de l'analyse de ces deux verbes sont (7) et (8), respectivement.

(7) $[\text{ } dé\text{-} \text{ } [\text{rat}]_N \text{ } \text{-}iser \text{ }]_V$

(8) $[\text{ } dé\text{-} \text{ } [\text{sensible}]_A \text{ } \text{-}iser \text{ }]_V$

Il s'agit de ce que Booij (2002) appelle « synaffixation » et qui revient à admettre l'existence de règles ternaires : deux opérateurs s'appliquent simultanément à la même base. L'inconvénient de cette solution est que chacun de deux affixes selectionne en temps normal un constituant nominal ou adjectival : *dé-* sélectionne *herbe* pour former *désherber* et *-iser*, *cristal* pour *cristalliser* ; *dé-* se combine avec *saoul* dans *dessaouler* et *-iser* est concaténé à *fertile* dans *fertiliser*. L'application simultanée des deux morphèmes à une même base est corrélée à une contribution combinée de leur contenu sémantique : la privation pour *dé-* et la cause pour *-iser*. Mais cette analyse contredit le principe d'unicité sémantique des morphèmes —le sens de *dé-* dans *désherber* cumule par exemple les opérateurs de privation et de cause— et celui de combinaison des morphèmes, selon lequel une règle de réécriture est binaire et n'associe qu'une tête affixale au constituant régi par cet affixe.

3.3 Circonfixation

Des auteurs comme Bosque (1983) proposent d'analyser les séquences affixales comme des « morphèmes discontinus » ou « circonfixes » ; cette approche revient à considérer *dé-…-iser* comme un affixe unique dont la combinaison avec *rat* ou *sensible* respecte le principe de binarité des règles de réécriture du modèle. On obtient alors une construction

en une étape : (9) est la représentation de l'analyse de *dératiser*, et (10), celle de *désensibiliser*. Cependant, l'utilisation d'un circonfixe pose plusieurs problèmes : (i) elle n'explique pas comment est choisie la valeur de la séquence suffixale (par exemple, pourquoi at-on *-iser* dans *dératiser*, mais *-ifier* dans *dégazéifier* ?) ; (ii) elle contrevient au principe d'unicité du morphème : *dé-...-iser*, *dé-...-ifier* et *dé-* sont des morphèmes synonymes, mais la variation allomorphique qui les distingue n'est pas imputable à des contraintes morphophonologiques.

(9) [*dé-...-iser* [rat]$_N$]$_V$

(10) [*dé-...-iser* [sensible]$_A$]$_V$

3.4 Préfixation et intégrateur paradigmatique

L'analyse des parasynthétiques comme dérivés préfixés dans lesquels le suffixe est un intégrateur paradigmatique s'inscrit elle aussi dans la tradition binaire des règles de dérivation, mais récuse l'absence de pouvoir catégorisateur du préfixe. Elle a été proposée par Danielle Corbin, dont les travaux, au cours des 30 dernières années du 20ᵉ siècle, ont impulsé à la morphologie dérivationnelle des évolutions théoriques fondamentales qui dépassent l'étude du lexique du français auxquelles l'auteur se consacre. Danielle Corbin développe en effet avec sa thèse (Corbin 1987) un système génératif de représentation du lexique construit qui s'éloigne du principe de concaténation de morphèmes. Ce système comporte un composant dérivationnel qui utilise des Règles associatives de Construction de Mots (RCM) ; une RCM est un processus morphologique qui s'applique à une base. Les principes de fonctionnement des RCMs offrent de nouvelles perspectives pour l'analyse des dérivés parasynthétiques. Corbin (1980) déjà, repris ensuite dans Corbin (1987), défend l'idée que le préfixe a la faculté de produire des dérivés ayant une catégorie grammaticale différente de la base.

Dans cette analyse, la construction morphologique de *dératiser*, donnée en (11), comme celle de *désensibiliser*, en (12), est réalisée en deux étapes : une préfixation en *dé-* sur base nominale ou adjectivale, suivie d'une modification formelle affectant respectivement les séquences verbales obtenues *dérat-* et *désensibl-*, qui consiste en l'ajout, à la sortie du composant dérivationnel, du segment dépourvu de sens *-is(er)*.

(11) [*dé-* [rat]$_N$ *+is(er)*]$_V$

(12) [*dé-* [sensible]$_A$ *+is(er)*]$_V$

La séquence suffixale, identifiée dans la notation ci-dessus par le signe « + », est nommée « intégrateur paradigmatique » car son rôle est d'insérer le mot auquel elle s'applique dans un paradigme, ici, la classe des verbes de changement d'état. Le Principe de Copie, auquel obéit l'emploi de cet intégrateur permet de donner au segment ajouté une fonction purement analogique : *-is(er)* est le suffixe verbal le plus disponible et correspond, le cas échéant, à la valeur du suffixe verbal utilisé dans la famille du préfixé (ex. *sensibiliser*). Le recours à ce principe ne suffit cependant pas à expliquer l'absence de copie pour certains verbes préfixés en *dé-* comme *désherber*.

3.5 Bilan

Ces propositions d'analyse sont toutes motivées par la volonté de rendre compte du lien sémantique direct existant entre *rat* ou *sensible*, et le verbe préfixé apparenté. Chacune à sa manière, elles cherchent à représenter la séquence *iser* d'une manière permettant de court-circuiter le décalage sémantique : soit *+is(er)* est vidé de son sens et n'est plus qu'un marqueur catégoriel, soit les deux affixes se partagent les propriétés sémantico-catégorielles, soit encore ils fusionnent pour ne constituer qu'un seul morphème.

4 Parasynthèse et RCL

Les dérivés parasynthétiques font partie des structures dérivées qui tirent le bénéfice le plus substantiel de la démarche lexèmatique en morphologie, et plus particulièrement des innovations du modèle des Règles de Construction de Lexèmes (RCL) tel qu'il est développé, motivé, détaillé, formalisé et largement illustré dans Fradin (2003). Dans cet ouvrage, la seule unité manipulée est le *lexème*, objet pour lequel l'auteur développe sa propre définition à la suite, entre autres, des travaux de Anderson (1992), Aronoff (1976), Beard (1995), Carstairs-McCarthy (1992), Matthews (1974), Scalise (1984) qui chacun proposent une alternative à l'approche Item et Arrangement (Hockett 1954) dont relèvent les modèles à base morphème. La conception du lexème est défendue à travers une série d'exemples que la section 4.1 ci-dessous résume brièvement[3].

4.1 Lexème et RCL : principes fondamentaux

Fradin (2003) bâtit son modèle dans le cadre lexèmatique de la morphologie constructionnelle. Son originalité se manifeste à travers les propriétés suivantes :

— Un lexème est une unité tridimensionnelle, dont les rubriques, indépendantes entre elles, consignent ses propriétés formelles, syntaxico-catégorielles (ou « syntactiques » dans la terminologie de Fradin (2003), inspirée de celle de Mel'čuk (1993)), et sémantiques.

— Le lexique obéit à une organisation hiérarchique, de sorte que, pour tous les lexèmes, chaque dimension hérite, à travers un partage de type, des propriétés de l'élément qui la domine dans la hiérarchie. Cette organisation est inspirée des travaux de Koenig (1999) et Davis & Koenig (2000).

— Contrairement à ce qu'implique la hiérarchisation lexicale chez Koenig (1999), le lexème est une entité monosémique dont le contenu sémantique est entièrement spécifié. L'argument, repris dans Fradin & Kerleroux (2003a,b), est que les procédés constructionnels s'appliquent à des entrées lexicales sémantiquement non ambiguës mais qu'ils forment des lexèmes construits dont le contenu sémantique peut être sous-spécifié, la spécification de ce contenu dépendant d'autres facteurs.

---

[3]Dans la suite du chapitre, nous représentons le lexème en petites capitales, conformément à la notation proposée par Matthews (1974).

— Une RCL met en relation un lexème de base (ou deux lexèmes de base, dans le cas de la composition), et un lexème construit. Cette relation se traduit par l'application simultanée, sur chacune des trois composantes constitutives des lexèmes concernés, d'un faisceau de fonctions indépendantes les unes des autres.

— Une attention particulière est portée à l'expression formelle des RCL, et, en particulier, au traitement sémantique. Celui-ci fait appel à un formalisme combinant logique et *λ*-calcul qui permet de représenter le contenu sémantique des patrons de lexèmes en entrée et sortie des RCL, et des lexèmes qui instancient ces patrons. (voir *infra* les figures 1 à 3).

La principale rupture à laquelle le modèle conduit vis-à-vis des systèmes théoriques qui l'ont précédé, notamment Corbin (1987), réside en la description en trois niveaux de la relation établie par une RCL entre les lexèmes qu'elle connecte. Tout en s'affranchissant de l'assemblage de morphèmes imposé par les théories relevant du cadre Item et Arrangement, ce principe rend également obsolètes les *règles associatives*, le Principe de Copie et le Principe d'Unicité sémantique des procédés morphologiques de Corbin (1987) et supprime ainsi la nécessité de recourir aux « mots possibles » comme des étapes indispensables dans l'analyse constructionnelle de certains dérivés. Cette propriété fondamentale des RCL ouvre des perspectives nouvelles dans l'analyse des constructions parasynthétiques, comme nous le détaillons en section 4.2 à travers les exemples des verbes DÉRATISER et DÉSENSIBILISER.

4.2 Analyse des verbes en *dé-X-iser*

L'analyse des verbes vérifiant le patron *dé-X-iser* dans Fradin (2003), que l'on retrouve aussi dans Fradin (1997a) à propos des préfixés adjectivaux en *anti-*, démontre la nécessité de déconnecter les opérations formelles et sémantiques des RCL : l'interprétation de *dé-X-iser* fait intervenir le sens de *X* (qu'il soit adjectival, comme dans DRAMATIQUE → DÉDRAMATISER ou SENSIBLE → DÉSENSIBILISER, ou nominal, comme dans RAT → DÉRATISER), alors que sa forme est motivée par celle de *X-iser*. La solution de Fradin (2003 : 297) consiste à faire du verbe *X-iser* l'entrée de la règle de préfixation en *dé-*. Il s'agit donc d'une relation de préfixation entre deux verbes, la base étant formellement suffixée en *-iser*. Une seule et même RCL s'applique quelle que soit la catégorie de *X* (nom ou adjectif) et quel que soit le sens de *dé-X-iser* : annulation de la propriété *X* (SENSIBLE → DÉSENSIBILISER), ou dissociation de la partie et du tout auquel cette partie est attachée initialement (RAT → DÉRATISER ou NICOTINE → DÉNICOTINISER ; dans ces deux cas, la partie dissociée est exprimée par *X*, *X* = RAT et *X* = NICOTINE respectivement). Les mécanismes de cette RCL sont détaillés *infra* pour l'analyse de DÉSENSIBILISER et DÉRATISER. La même RCL s'applique aux dérivés *dé-X-iser* pour lesquels le nom *X* dénote le tout qui sera privé de l'une de ses parties à l'issue du déroulement du procès décrit par *dé-X-iser*, comme DÉBUDGÉTISER 'faire sortir du budget'. Dans tous les cas, l'objectif principal de l'analyse est de légitimer la présence du suffixe *-iser*. Les différences entre forme et sens dans le fonctionnement de la RCL sont dues au contenu sémantique de *X-iser*. Elles influent naturellement sur le contenu sémantique du dérivé *dé-X-iser* produit par la RCL.

Quand la RCL est appliquée à ѕᴇɴѕɪʙɪʟɪѕᴇʀ (figure 1), le patron sémantique de *X-iser* décrit un prédicat d'accomplissement (représenté par la primitive factitive « CAUSE ») conduisant à un changement d'état (représenté par la primitive « become »). L'argument patient du verbe *X-iser* qui subit le changement d'état est représenté par la variable y et l'agent (i.e. le causateur) par la variable x. Le contenu sémantique de *X-iser* fait intervenir un prédicat représenté par la variable P' et appliqué à y : il s'agit de la propriété caractérisant le référent de y, c'est-à-dire le contenu sémantique de X. On voit par là que le sens de *X-iser* fait clairement apparaître celui de sa base X ce qui le rend disponible pour la construction du sens de *dé-X-iser*. La notation formelle adoptée par Fradin (2003) dans la rubrique sémantique des lexèmes met en évidence la combinaison des différents maillons (opérateurs, prédicats et primitives sémantiques) qui construisent le sens d'un lexème, en particulier quand il est morphologiquement construit.

	X-iser	
form	F	$\rightarrow$
cat	V, StrArg = <SN,SN>	$\rightarrow$
sem	$\lambda y \lambda x \lambda e (\text{CAUSE}(e, x, (\text{become } P'(y))))$	$\rightarrow$

	dé-X-iser
	de$F$
	V, StrArg = <SN,SN>
	$\lambda y \lambda x \lambda e (\text{CAUSE}(e, x, (\text{become NOT}(P'(y)))))$

Figure 1 : RCL : *X-iser*$_V$ $\rightarrow$ *dé-X-iser*$_V$ où X est un adjectif et F est la forme de *X-iser*

La sortie de la RCL décrit un prédicat également télique exprimant un changement d'état : la structure logique qui décrit le sens de *dé-X-iser* met en jeu les mêmes primitives « CAUSE » et « become » que sa base. Mais on observe que, pour représenter la privation de la propriété P' qui correspond au sens de l'adjectif X et qui qualifie y avant le début du procès, la RCL extrait le prédicat $P'(y)$ intervenant dans la rubrique sémantique de *X-iser* et lui applique l'opérateur de négation "NOT". En d'autres termes, la RCL ne construit pas le sens de *dé-X-iser* à partir de celui de *X-iser*, mais bien directement à partir de celui de X. De cette manière, elle signale que la propriété annulée n'est pas nécessairement le résultat d'un procédé antérieur (par exemple, « désensibiliser une dent » consiste à ôter à la dent la sensibilité à la douleur, qui est une propriété physiologique inhérente des parties du corps). L'emploi d'une représentation formelle du sens permet ainsi à Fradin (2003) de connecter directement la structure sémantique de *pref-X-suf* au prédicat exprimant le sens de X (i.e. $P'(y)$ quand X désigne une propriété adjectivale), auquel la RCL accède à travers la combinaison de primitives qui définissent *X-iser*. Ce que par ailleurs sous-entend cette représentation, c'est l'existence d'un premier chaînon permettant d'expliquer le fonctionnement de la RCL. En d'autres termes la RCL ne connecte pas deux lexèmes, mais trois, ce que l'on pourrait représenter par le patron de la figure 2, le sens de X motivant à la fois celui de *X-iser* et celui de *dé-X-iser*.

$$
\begin{array}{c}
\begin{array}{ccc}
& X & X\text{-}iser \\
\text{form} & \boxed{F} \rightarrow & \boxed{\textit{Fiz}} \rightarrow \\
\text{cat} & \boxed{\text{A}} \rightarrow & \boxed{\text{V, StrArg} = \text{<SN,SN>}} \rightarrow \\
\text{sem} & \boxed{\lambda y P'(y)} \rightarrow & \boxed{\lambda y \lambda x \lambda e (\text{CAUSE}(e, x, (\text{become } P'(y))))} \rightarrow
\end{array}
\end{array}
$$

$$
\begin{array}{l}
d\acute{e}\text{-}X\text{-}iser \\
\boxed{\begin{array}{l}
\text{de}\textit{Fiz} \\
\text{V, StrArg} = \text{<SN,SN>} \\
\lambda y \lambda x \lambda e (\text{CAUSE}(e, x, (\text{become NOT}(P'(y)))))
\end{array}}
\end{array}
$$
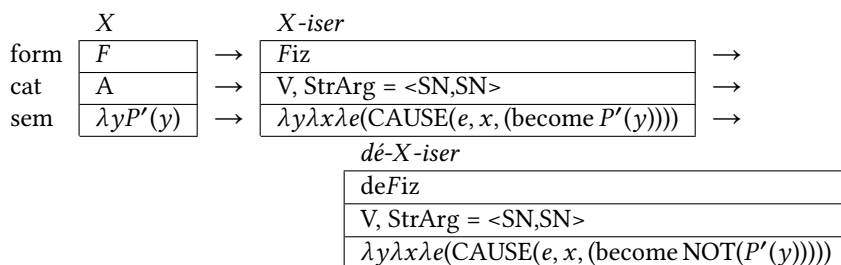
Figure 2 : Combinaison de deux RCL : $X_{\text{A}} \rightarrow X\text{-}iser_{\text{V}} \rightarrow d\acute{e}\text{-}X\text{-}iser_{\text{V}}$ où F est la forme de X

Pour l'analyse de DÉRATISER, le même schéma de règle est appliqué : il connecte de façon implicite X et X-*iser*, et de façon explicite X-*iser* et *dé-X-iser*. Comme pour DÉ-SENSIBILISER, le verbe préfixé est traité comme un simple cas de suffixation suivi d'une préfixation (Fradin 2003 : 298), à ceci près que le maillon intermédiaire X-*iser*, ici RATISER est pragmatiquement peu plausible. À l'image de la figure 2, nous reprenons en figure 3 le schéma de (Fradin 2003 : 297), en le modifiant pour faire apparaître le rôle de X. La figure montre que le contenu sémantique de X-*iser* décrit la localisation du référent de X ($\lambda y \, \text{rat}'(y)$) sur ou dans ce que dénote le patient z du verbe ; le mécanisme qui conduit à la construction du contenu sémantique de *dé-X-iser* procède comme pour DÉSENSIBILI-SER : le sens de DÉRATISER, qui décrit l'état final du référent du patient z, débarassé de ce que dénote X, n'est pas élaboré à partir du contenu sémantique de X-*iser*, mais exploite directement le prédicat $\lambda y \, \text{rat}'(y)$ qu'il extrait de la rubrique sémantique de ce verbe. L'analyse convoque ici un raisonnement légèrement différent puisque l'attestation de X-*iser* est optionnelle. Cette étape, motivée sémantiquement, est également justifiée par l'uniformisation du traitement des *pref-X-suf*. Nous verrons d'ailleurs, dans la section 6, que l'analyse proposée dans le cadre de ParaDis intègre explicitement X dans l'analyse de DÉRATISER et DÉSENSIBILISER, en y incluant les relations dérivationnelles binaires dans des modules qui donnent accès à une partie de la famille dérivationnelle des lexèmes construits.

$$
\begin{array}{c}
\begin{array}{ccc}
& X & X\text{-}iser \\
\text{form} & \boxed{F} \rightarrow & \boxed{\textit{Fiz}} \rightarrow \\
\text{cat} & \boxed{\text{N}} \rightarrow & \boxed{\text{V, StrArg} = \text{<SN,SN>}} \rightarrow \\
\text{sem} & \boxed{\lambda y P'(y)} \rightarrow & \boxed{\lambda x \lambda z \lambda e \exists y (\text{CAUSE}(e, x, (\text{LOC}(y, \text{in}(z)) \wedge P'(y))))} \rightarrow
\end{array}
\end{array}
$$

$$
\begin{array}{l}
d\acute{e}\text{-}X\text{-}iser \\
\boxed{\begin{array}{l}
\text{de}\textit{Fiz} \\
\text{V, StrArg} = \text{<SN,SN>} \\
\lambda x \lambda z \lambda e \exists y (\text{CAUSE}(e, x, (\text{NOT}(\text{LOC}(y, \text{in}(z)) \wedge P'(y)))))
\end{array}}
\end{array}
$$
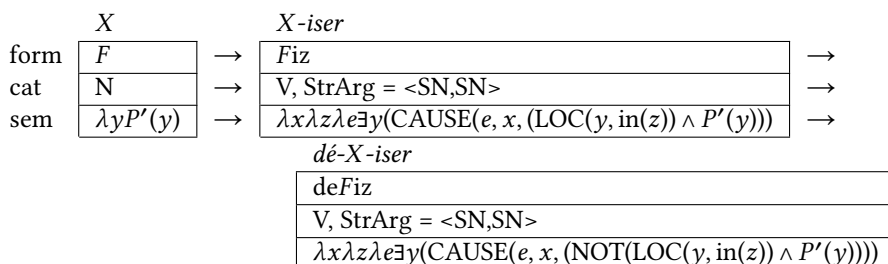
Figure 3 : Combinaison de deux RCL : $X_{\text{N}} \rightarrow X\text{-}iser_{\text{V}} \rightarrow d\acute{e}\text{-}X\text{-}iser_{\text{V}}$ où F est la forme de X

Puisqu'une RCL peut sélectionner de façon autonome les caractéristiques formelles d'un verbe et le contenu sémantique de l'adjectif base de celui-ci, cette solution constitue de fait un premier pas vers une conception en réseau de la morphologie dérivationnelle : la construction du préfixé est tributaire de la forme d'un membre de sa famille dérivationnelle, et du sens d'un autre. Relativement aux analyses décrites dans la section 3, celle de Fradin (2003) repose sur un rapport formel uniforme, et transforme le « casse-tête » parasynthétique en une simple relation entre un lexème verbal de base (éventuellement non attesté) et un lexème verbal construit par préfixation. Elle règle par ailleurs le problème du décalage entre forme et sens grâce à l'utilisation de représentations formelles qui permettent de construire un sens approprié pour le dérivé à partir des éléments de sens pertinents présents dans la représentation sémantique de la base.

Néanmoins l'analyse proposée n'est pas totalement satisfaisante : le traitement de DÉRATISER ou d'un néologisme verbal comme EMPUISSANTISER dans « Or il y a deux moyens d'*empuissantiser* les idées. » (citation de l'économiste Frédéric Lordon[4] entendue sur France Culture en 2015), impose de recourir à l'artifice qui consiste à reconstruire un verbe non attesté[5]. Elle ne permet pas non plus de connaître *a priori* les formes que peut prendre la séquence suffixale, car la démarche est descriptive et à visée d'analyse : étant donné une forme vérifiant le patron *pref-X-suf*, la RCL permet d'en expliquer le sens et la forme. En revanche, le dispositif n'est pas conçu pour rendre compte de la variation dans le nombre et la diversité des inputs possibles de la RCL, ni pour prédire le fait que plusieurs *pref-X-suf* synonymes peuvent être construits à partir du même *X*. En d'autres termes, les RCL ne permettent pas, par exemple, de décrire l'ensemble des mécanismes à l'origine de la régularité qui explique que 'contre le cancer' est une paraphrase du sens de ANTICANCER, ANTICANCÉREUX, ANTICANCÉRIGÈNE ou ANTIONCOLOGIQUE ni ceux qui font que ANTIVIBRATION, ANTIVIBRATOIRE, ANTIVRIBEUR, ANTIVIBRATEUR, ANTIVIBRATIF, ANTIVIBRANT, ANTIVIBRATILE sont autant de dérivés concurrents signifiant 'contre les vibrations'. Le principe fondamental des RCL qui consiste en une action indépendante et simultanée de leurs opérations constitutives est donc nécessaire, mais ne suffit pas à expliquer complètement les constructions dites parasynthétiques.

4.3 Bilan

Les principes théoriques défendus dans Fradin (2003) comportent des propositions centrales pour le modèle ParaDis, objet de la section 6. Certaines sont formulées de façon explicite : le lexème supplante le morphème comme unité de traitement dans la construction du lexique ; il s'agit d'une unité tridimensionnelle sémantiquement spécifiée, disposant d'un ensemble organisé de radicaux libres et supplétifs dont Fradin (2003 : 138-140) propose une première structuration relative à leur statut « libre » ou « savant » ; les RCL qui relient ces lexèmes font intervenir des fonctions agissant de façon indépendante sur chacune des trois dimensions connectées. On verra que dans ParaDis cet aspect modulaire de la construction lexicale est étendu aux relations entre les éléments du lexique.

---

[4]Lordon, Frédéric (2016). *Les affects de la politique*, Seuil, Paris.
[5]La requête Google ″ratiser″ ne ramène aucune page utile (08/10/2016).

Mais on montrera que l'élaboration de ParaDis profite également d'avancées du modèle de Fradin (2003) que l'auteur ne met pas en avant. D'une part, son analyse des parasynthétiques met en jeu une morphologie en réseau qui ne dit pas son nom : le fait que la RCL qui construit le dérivé *dé-X-iser* puisse utiliser directement la sémantique de l'adjectif base *X* du verbe *X-iser* suppose que la relation dérivationnelle entre *X-iser* et sa propre base *X* soit accessible, via la structure interne du verbe suffixé. L'analyse des parasynthétiques, à travers l'exemple des verbes préfixés en *dé-*, montre que les procédés de construction ont accès, au delà du couple base/dérivé, aux autres membres de leur famille dérivationnelle. D'autre part, même si Fradin (2003) ne l'indique pas explicitement, la manière dont la RCL organise la mise en relation entre deux membres d'une famille constructionnelle fait sauter les verrous de la nécessaire orientation lexème(s) base(s) → lexème construit des procédés dérivationnels : dans la mesure où le mécanisme d'application d'une RCL n'impose aucune contrainte sur la complexité relative que doivent respecter la (ou les) base(s) et le construit connectés par la RCL, chacun de ces lexèmes peut être plus (ou aussi) complexe que l'autre, formellement, mais aussi sémantiquement. Finalement, les principes d'indépendance des fonctions qui constituent les RCL ouvrent la voie à des analyses mettant en jeu des relations constructionnelles a-directionnelles et bi-directionnelles.

De nombreux morphologues francophones ont adhéré aux idées défendues dans Fradin (2003), et les ont fait évoluer. C'est ainsi que les années qui ont suivi la parution de cet ouvrage ont vu se développer de nombreuses analyses fondées sur le modèle de RCL, dont certaines étendent ses principes théoriques : en particulier, différents travaux se sont intéressés à la structure formelle du lexème (Bonami & Boyé 2007), à l'incorporation des radicaux supplétifs ou savants (Amiot & Dal 2005, Bonami et al. 2009), ou à leur extension à des thèmes dérivationnels supplémentaires (Tribout 2012). Dans le même temps, les RCL et la notion de lexème suscitent des réactions et des critiques qui conduisent à l'élaboration de travaux s'appuyant sur les principes qui ont émergé de ces confrontations. C'est ce que présente la section 5.

5 Vers une morphologie dérivationnelle en réseau

Le modèle des RCL constitue, comme nous venons de voir, un progrès déterminant dont les bénéfices pour l'analyse de la parasynthèse sont importants. Comme nous l'avons évoqué *supra* (section 4.1), le cadre théorique développé par Fradin (2003) a constitué une référence forte pour la plupart des recherches en morphologie dérivationnelle qui ont été menées en France et ailleurs dans les années 2000. C'est notamment le cas des travaux réalisés à Toulouse au sein de l'axe DUMAL ("Des Unités Morphologiques Au Lexique") et plus généralement par les morphologues de l'ERSS. Le livre DUMAL (Roché et al. 2011) en propose une synthèse. Nous présentons dans cette section les principes qui ont guidé ces travaux et les avancées qu'ils ont rendues possibles, en particulier dans l'analyse des parasynthétiques.

5.1 Variabilité des dérivés morphologiques

Le cadre théorique de Fradin (2003) se caractérise par sa nature formelle qui le place dans la lignée des recherches menées au sein du laboratoire LLF. Les analyses développées dans ce cadre portent essentiellement sur les aspects sémantiques de la dérivation morphologique. Fradin (2003) propose un système formel à la fois original par son utilisation du λ-calcul pour la description du sens lexical, et relativement classique par le mécanisme d'héritage multiple de lexèmes sous-spécifiés guidé par une structure hiérarchique du lexique, en l'occurrence un treillis. La formalisation de la construction du sens dérivationnel est par certains aspects supérieure aux descriptions faites au moyen de paraphrases ou de gloses. On peut en effet considérer que la nature informelle de ces dernières rend indécidables (ou irréfutables) les démonstrations qui les utilisent parce qu'elles empêchent toute preuve des propriétés et généralisations avancées. Cependant, la linguistique n'est pas un système purement formel mais manipule un matériau naturel, ce qui fait que l'instanciation des variables et des prédicats lors de l'interprétation des représentations formelles constitue un passage dans l'informel qui entache et affaiblit les démonstrations. Par ailleurs, la nature formelle et explicite de ces descriptions présente quelques limites qui expliquent probablement que ces aspects du modèle de Fradin (2003) n'ont pas reçu le même niveau d'adhésion que la description des dérivations au moyen de RCL. La description du sens dans le formalisme du λ-calcul comporte en effet différentes faiblesses :

— elle est peu familière à la plupart des morphologues ;

— les représentations sont difficiles à construire, à combiner, à exploiter manuellement ou par programme : peu de morphologues sont capables de les « faire fonctionner » ;

— elle est trop rigide pour rendre compte de la variété des propriétés sémantiques impliquées dans la construction morphologique et de la plasticité du sens des dérivés.

Fradin (2003) est ainsi amené à multiplier les instructions sémantiques. Ces instructions sont disjointes et le cadre formel ne prévoit aucun mécanisme simple permettant d'exprimer leur similarité comme dans le cas de la suffixation en *-ette* où les dérivés féminins (ex. FLIC → FLIQUETTE dont le sens est construit au moyen de l'instruction sémantique (13)) et les noms de lieu déverbaux (ex. COUCHE → COUCHETTE dont le sens est construit au moyen de (14)) ne partagent aucune propriété (Fradin et al. 2003).

(13) $\lambda y \lambda P', P'(y) \wedge \text{femelle}'(y)$

(14) $\lambda y \lambda x \lambda P' \exists e, P'(e, x, ...) \wedge \text{dans}'(e, y) \wedge \text{temporaire}'(e)$

Les morphologues toulousains ont proposé différents aménagements du cadre de Fradin (2003) pour répondre à ces limitations, et notamment disposer d'une souplesse adaptée à la variabilité sémantique, formelle et catégorielle de la construction morphologique. Ils ont repris le principe fondamental de dissociation entre les représentations formelles,

catégorielles et sémantiques proposé dans Fradin (2003 : 9), indispensable pour rendre compte des décalages entre forme et sens puisqu'il permet de faire coopérer plusieurs lexèmes pour former un dérivé en utilisant la forme de l'un et le sens d'un autre. La notion de "base" se retrouve ainsi redéfinie et correspond au lexème qui motive sémantiquement le dérivé. C'est le cas par exemple dans la construction de l'adjectif PIANISTIQUE, dont la forme est construite relativement à PIANISTE, mais dont l'interprétation, dans un énoncé comme « un concerto/une sonate/une sonorité pianistique » renvoie directement au contenu sémantique du nom PIANO. Ce découplage a été mis en œuvre de deux manières différentes. Roché (2009) a proposé de considérer que la construction d'un lexème dérivé se compose d'un ensemble d'opérations phonologiques, syntaxiques et sémantiques (élémentaires) indépendantes. Il n'y a aucune contrainte a priori sur cet ensemble sinon qu'il ne doit pas être vide. Il propose par exemple de considérer la dérivation RAT → DÉRATISER[6] comme étant composée de quatre opérations élémentaires :

1. une opération catégorielle N → V ;

2. une opération formelle de suffixation en *-iser* qui signale l'opération catégorielle ;

3. une opération sémantique 'rat' → 'éliminer toutes les instances de rat qui sont dans un lieu' ;

4. une opération formelle de préfixation en *dé-* qui signale l'opération sémantique.

Les dérivés parasynthétiques comme dans PARLEMENT → ANTIPARLEMENTAIRE peuvent être analysés strictement de la même manière : une opération catégorielle (N → A), une opération sémantique ('P' → 'qui est contre P') et deux opérations formelles (une préfixation en *anti-* et une suffixation en *-aire*), la préfixation signalant l'opération sémantique et la suffixation l'opération catégorielle. Cette proposition n'a pas été élaborée davantage et Roché (2009) ne dit rien des contraintes qui portent sur ces outputs ni sur les associations entre ces contraintes.

Hathout (2011) propose une autre mise en œuvre, plus élaborée, fondée sur un modèle à quatre niveaux de représentation et sur plusieurs jeux de contraintes. Certaines portent sur les représentations de l'un des quatre niveaux tandis que d'autres sont destinées à contrôler la correspondance entre les représentations des niveaux phonologique, syntaxique et sémantique avec celles du niveau lexical. Dans ce modèle, une grande partie des contraintes sont exprimées en termes d'analogie et la construction morphologique est vue comme le calcul d'une solution optimale relativement à l'ensemble des contraintes qui portent sur les lexèmes qui participent à cette opération constructionnelle. Le découplage des quatre niveaux fournit au système les degrés de liberté nécessaires pour rendre compte de l'association d'une même forme construite à plusieurs sens comme *antipaternel*, qui peut signifier 'contre les pères' ou 'relatif aux antipères', et la multiplicité des formes qui peuvent exprimer un même sens comme *antivibration*, *antivibratoire*, *antivibrant*, *antivibreur*, etc. qui toutes peuvent être associées au même sens

---

[6]Rappelons que dans l'analyse de DÉRATISER de Fradin (2003), l'intervention du nom RAT est seulement sous-entendue, même si la représentation que nous en donnons dans la figure 3 la fait apparaître de manière explicite.

construit 'contre les vibrations'. Cette proposition est en grande partie reprise dans ParaDis.

On le voit, l'assouplissement du cadre défini par Fradin (2003) passe par le remplacement des représentations formelles par des ensembles de contraintes. Inspirées de la Théorie de l'Optimalité (Prince & Smolensky 1993 ; McCarthy & Prince 1993), elles sont contradictoires et violables. Initialement définies sur les caractéristiques morphophonologiques, comme les contraintes dissimilatives (Plénat 2011), et prosodiques (Plénat 2009b), elles ont ensuite été étendues à des propriétés plus structurelles, portant sur les familles et les séries dérivationnelles (Hathout 2011 ; Roché & Plénat 2014). Le modèle ParaDis que nous détaillons dans la section 6 reprend à la fois ce principe de contrôle des constructions morphologiques par des contraintes et la représentation formelle du sens. Cette formalisation sémantique semblable à celle qui est utilisée dans la base de données morphologique Démonette (Hathout & Namer 2014a ; Hathout & Namer 2016) diffère de celle de Fradin (2003). Elle comporte d'une part un typage sémantique des variables qui représentent le sens des lexèmes en jeu dans une construction donnée, et d'autre part une représentation formelle des relations de sens qui existent entre ces lexèmes.

Sur le plan méthodologique, les morphologues toulousains ont défendu et utilisé une approche extensive (Plénat et al. 2002 ; Hathout et al. 2003, 2008, 2009) qui, complètement en accord à leur intérêt pour la variation et la variabilité, consiste à collecter le plus grand nombre possible d'attestations et d'exemples des phénomènes étudiés, notamment sur la Toile, et à proposer des analyses rendant compte de l'ensemble des données collectées. La démarche extensive a notamment été utilisée par Hathout et al. (2009), Hathout (2011) pour l'analyse de la préfixation en *anti-* (voir section 5.3). Elle a permis de mettre au jour des dérivés inattendus comme ANTIDÉSHERBANT, dérivé sur HERBE et synonyme de DÉSHERBANT. Ce lexème est formé par trois opérations formelles : une suffixation en *-ant* qui signale l'opération catégorielle N → A et deux opérations formelles (une préfixation en *dé-* et une préfixation en *anti-*) qui signalent la même opération sémantique.

5.2 Inscription dans le lexique

L'approche DUMALienne de la morphologie dérivationnelle (Roché et al. 2011) a aussi mis fortement en avant l'inscription de la morphologie dérivationnelle dans le lexique. Cette relation essentielle est notamment l'objet d'un important article de Michel Roché (2009). Outre les faits, aujourd'hui consensuels, que (i) l'une des fonctions de la morphologie dérivationnelle est de construire des mots capables d'entrer dans le lexique (mais ce n'est pas toujours le cas comme l'ont montré Dal & Namer (2016)) et (ii) la morphologie utilise le lexique comme une ressource dans laquelle elle trouve les bases et plus largement les lexèmes dont elle a besoin, la construction morphologique est soumise à la pression du lexique existant. Cet état de fait permet d'expliquer des décalages exceptionnels dus à la présence dans le lexique de mots à consonance proche comme dans la suffixation en *-esque* BAMBOU → BAMBOULESQUE, où l'épenthèse en /l/, très rare, est légitimée par la présence dans le lexique du lexème BAMBOULA et de son adjectif relationnel BAMBOULESQUE (Plénat 2009a).

La prise en compte du lexique existant n'est pas en elle-même une innovation. Elle fonde notamment le Principe de Copie introduit par Dell (1970) et repris par Corbin (1980, 1987) qui, comme nous l'avons rappelé *supra* (section 3.4), est utilisé pour expliquer la sélection du suffixe dans les dérivés parasynthétiques. Ce principe a été étendu par Roché (2007) en un Principe d'Économie qui stipule que la « langue [tend à réutiliser] une forme déjà existante dans le paradigme dérivationnel , en violation de l'instruction propre à l'affixe [plutôt que de construire une forme nouvelle] ».

Ces deux principes sont destinés à préserver et renforcer les régularités qui existent dans le lexique —ou la simplicité du lexique dans les termes de Dell (1970), régularités qui en déterminent l'organisation morphologique. L'approche toulousaine se distingue nettement de celle de Fradin (2003) sur ce plan. Comme nous l'avons indiqué en section 4.1, ce dernier adhère en effet à une conception hiérarchisée du lexique où les diverses catégories sont reliées par des relations d'héritage multiple (voir aussi Koenig (1999), Davis & Koenig (2000)). À l'inverse, les structures envisagées au sein de l'axe DUMAL sont d'une nature plus paradigmatique, et s'inscrivent dans un cadre « orienté output » alors que Fradin (2003) est l'héritier des traditions génératives « orientées input », même si son modèle a joué un rôle de tremplin qui permet de s'en détacher. Ainsi, dans l'approche développée à l'ERSS, la pression du lexique existant s'exerce dans des directions définies par deux types de structures : les familles dérivationnelles et les séries dérivationnelles. Si la notion de famille dérivationnelle, traditionnellement appelée « famille morphologique », est bien connue, elle ne joue aucun rôle dans les modèles théoriques antérieurs de la morphologie dérivationnelle. Sa formalisation est initiée dans Hathout (2011) qui en fait le fondement du modèle qu'il propose. Une famille dérivationnelle regroupe un ensemble de mots connectés par des relations de construction morphologique (ex. la famille dérivationnelle de Laver contient les mots qui lui sont reliés directement ou indirectement : laveur, laveuse, lavoir, lavage, laverie, lavette, délaver, etc.) ; une série réunit un ensemble de mots du lexique formés par un même procédé dérivationnel (ex. par la suffixation en *-able*). Ces structures sont essentielles pour l'analyse des dérivés parasynthétiques car elles donnent accès aux différents lexèmes impliqués dans leur construction. Ces lexèmes guident l'opération constructionnelle et lui fournissent les éléments de forme et de sens dont elle a besoin. Familles et séries sont à la base de l'organisation paradigmatique du lexique dérivationnel. À un niveau relationnel, l'inscription de la morphologie dérivationnelle dans le lexique permet de rendre compte du fait que les relations constructionnelles forment des analogies qui connectent des séries dérivationnelles. Ces connexions s'agrègent dans des graphes qui définissent des paradigmes dérivationnels comme nous le détaillons dans la section 6.

5.3 Améliorations dans l'analyse des parasynthétiques

Hathout (2011) ébauche un modèle de la morphologie dérivationnelle qu'il utilise pour décrire la dérivation en *anti-*, et notamment les correspondances multiples entre formes et sens. Il propose notamment que, lors de la construction d'un dérivé, le radical soit choisi dans un ensemble étendu qui contient les thèmes de la base, mais également ceux de

tous les autres membres de sa famille dérivationnelle. Les propriétés du dérivé et de ses relations constructionnelles permettent de sélectionner un thème optimal qui convient à la plus forte des coalitions de contraintes capable de se constituer. Celles-ci portent sur les caractéristiques morphophonologiques de la forme des lexèmes (phonation, dissimilation, taille), sur leur intégration dans le lexique existant (maximalisation de la ressemblance avec les formes présentes, inclusion dans la famille et la série dérivationnelles), sur la transparence sémantique et catégorielle, etc. Par ailleurs, ce modèle prédit que cette sélection dépend de l'importance accordée par le locuteur à chacune des contraintes et que cette pondération varie en fonction du contexte dans lequel le dérivé est utilisé. C'est ainsi que l'on peut observer au moins neuf dérivés en *anti-* formés sur VIBRATION (ANTIVIBRATION ; ANTIVIBRANT ; ANTIVIBRATOIRE ; ANTIVIBRATIF ; ANTIVIBRATILE ; ANTIVIBRATEUR ; ANTIVIBREUR ; ANTIVIBRABLE ; ANTIVIBRE). Il prédit également que, si les locuteurs peuvent ponctuellement favoriser l'une ou l'autre de ces contraintes, la structure paradigmatique du lexique existant partagée par l'ensemble de la communauté exerce une pression forte qui permet de prédire lequel des lexèmes en concurrence sera le plus fréquemment choisi. Par exemple, dans la compétition entre les lexèmes ANTICANCER, ANTICANCEREUX et ANTICANCÉRIGÈNE, qui tous signifient 'contre le cancer', ANTICANCER est préféré aux deux autres car il satisfait presque toutes les contraintes identifiées par Hathout (2011) :

— sa forme a une structure CV satisfaisante ;

— elle ne comporte pas de problèmes de dissimilation ;

— la taille de son radical est optimale (deux syllabes) ;

— sa forme permet une identification optimale de la base (sémantique), de la famille et de la série dérivationnelle du dérivé ;

• le lexème est sémantiquement transparent ; etc.

Seule la contrainte de transparence catégorielle est enfreinte, car ANTICANCER, qui est un adjectif, a une forme de nom que lui confère sa finale *cancer*, le lexique du français ne contenant que très peu de formes adjectivales finissant en /sɛʁ/. À l'inverse, les deux autres concurrents satisfont cette contrainte puisque leurs radicaux sont des formes adjectivales et que leurs finales (/ø/et /ʒɛn/) sont fréquentes parmi les adjectifs construits. Des deux, ANTICANCÉREUX est clairement préféré par les locuteurs à ANTICANCÉRIGÈNE, parce qu'il satisfait davantage une des contraintes fortes du système, à savoir la transparence sémantique : en effet, la similarité interprétative de CANCER est plus forte avec CANCÉREUX qu'avec CANCÉRIGÈNE.

5.4 Bilan

L'approche de la morphologie dérivationnelle développée à l'ERSS se distingue ainsi nettement de celle que propose Fradin (2003) : elle est orientée output, met en place une architecture paradigmatique où famille et série dérivationnelles complètent la notion

de lexème, s'inscrit dans le lexique, prend en compte la pression du lexique existant et définit un ensemble étendu de contraintes qui contrôlent la construction morphologique tout en donnant au modèle suffisamment de souplesse pour rendre compte de la plasticité du sens et des variations formelles. Ces propriétés en font un cadre particulièrement bien adapté à l'analyse des dérivés parasynthétiques. Une dernière différence avec Fradin (2003) concerne l'attitude de l'axe DUMAL vis-à-vis de la formalisation du sens. Sur ce plan, Fradin (2003) correspond davantage au canon de la recherche en linguistique. En revanche, Fradin (2003) et Roché et al. (2011) se rejoignent sur l'organisation tripartite du lexème, des RCL et des relations dérivationnelles. Les principales propositions de ces deux conceptions de la morphologie dérivationnelle sont intégrées au modèle ParaDis qui les articule dans une organisation radicalement paradigmatique.

6 La dérivation modulaire dans le modèle ParaDis

Comme nous venons de le faire pour les théories qui l'ont précédé en morphologie constructionnelle, nous présentons le modèle ParaDis (*Paradigms and Discrepancies*) en montrant comment il permet d'analyser, d'expliquer et de prédire la construction des lexèmes et notamment de ceux qui, à l'image des dérivés parasynthétiques, dérogent aux principes canoniques de transparence formelle et de compositionnalité sémantique. ParaDis est une synthèse entre un ensemble de propositions qui incluent les triangles proposés dans Lignon et al. (2014), les patrons cumulatifs de Bochner (1993) ainsi que les deux courants de la morphologie développés en France qui viennent d'être présentés : l'approche défendue dans Fradin (2003) fondée sur la notion de lexème et l'indépendance des opérations qui affectent chacune des trois dimensions constitutives des RCL, et l'approche développée à Toulouse au sein de l'ERSS (Roché et al. 2011) qui se fonde sur l'observation de données authentiques et qui prône une conception en réseau de la morphologie dont les mécanismes reposent sur la compétition des outputs arbitrée par un jeu de contraintes étendu, plutôt que sur l'application de règles. ParaDis intègre ces différentes propositions dans un cadre paradigmatique de la morphologie dérivationnelle et s'inscrit dans la lignée de Roché (2009, 2010, 2011b), Plénat & Roché (2012) ou Hathout (2008) dont les analyses intègrent les notions de série et famille dérivationnelles. La section 6.1 propose un bref rappel de ces notions et plus généralement de celle de paradigme. Nous présentons ensuite ParaDis en section 6.2 et illustrons son fonctionnement sur des exemples de dérivation parasynthétique.

6.1 Paradigmes dérivationnels

La notion de paradigme est fortement associée à la morphologie flexionnelle où elle a été définie clairement par des auteurs comme Wunderlich & Fabri (1995 : 266) :

> *A paradigm is an n-dimensional space whose dimensions are the attributes (or features) used for the classification of word forms. In order to be a dimension, an attribute must have at least two values. The cells of this space can be occupied by word forms of appropriate categories.*

ou comme Carstairs-McCarthy (1994 : 739) qui propose de distinguer les traits qui définissent les cellules des paradigmes (et qu'il nomme « paradigmes abstraits ») des formes qu'elles contiennent (et qu'il nomme « paradigmes concrets ») :

> *Let us call the abstract notion 'paradigm$_1$' and the more concrete one 'paradigm$_2$', and define them as follows :*
>
> (1) PARADIGM$_1$: *the set of combinations of morphosyntatic properties or features (or the set of 'cells') realized by inflected forms of words (or lexemes) in a given word-class (or major category or lexeme- class) in a given language.*
>
> (2) PARADIGM$_2$: *the set of inflectional realizations expressing a paradigm1 for a given word (or lexeme) in a given language.*

L'approche paradigmatique est devenue tout à fait standard voire dominante en morphologie flexionnelle (Stump 2001, 2006a,b, Ackerman et al. 2009, Baerman et al. 2010, Bonami & Stump 2016, Stump & Finkel 2013). Ce développement a été rendu possible par l'acceptation des modèles morphologiques basés sur les mots comme ceux de Blevins (2003-12, 2006). Dans ces modèles, les formes fléchies sont vues comme des réalisations d'un lexème et non plus comme des formes générées par des ensembles de règles opérant sur une forme de base. Cela permet de recentrer les études sur les relations qui existent entre les lexèmes et leurs formes fléchies et de regrouper dans des paradigmes les lexèmes qui partagent les mêmes relations avec leurs formes.

La situation est en revanche nettement différente en morphologie dérivationnelle où il n'existe pas de consensus sur le concept de paradigme. Certains comme Stump (1991) proposent de transposer à la dérivation les définitions établies pour la flexion, mais cette traduction ne va pas de soi et la question de l'élaboration d'une définition mieux adaptée à la dérivation demeure. Il existe en effet des différences notables entre flexion et dérivation. En particulier, comme le rappelle Stump (2001), la correspondance entre forme et sens n'intervient pas en flexion alors qu'elle est centrale en dérivation ; de plus, la régularité et la cohérence paradigmatique est intrinsèquement plus grande en flexion qu'en dérivation (Pounder 2000, Štekauer 2014).

Ceci dit, la notion de paradigme connaît depuis quelques années un intérêt grandissant en morphologie dérivationnelle (Štekauer 2014, Boyé & Schalchli 2016). Les morphologues qui travaillent dans cette approche s'intéressent notamment à la dimension paradigmatique de la dérivation, à la définition de modèles morphologiques paradigmatiques et au rapprochement de l'organisation de la morphologie flexionnelle et de la morphologie dérivationnelle (Van Marle 1985, Stump 1991, Bochner 1993, Booij 1996, Pounder 2000, Hathout et al. 2009, Roché 2009, Hathout 2011, Roché 2011b, Roché & Plénat 2014, Strnadová 2014a,b). Ainsi, certains de ces auteurs comme Van Marle (1985), Stump (1991), Pounder (2000) conçoivent les paradigmes dérivationnels comme de simples extensions des paradigmes flexionnels. Les paradigmes dérivationnels se distinguent cependant des paradigmes flexionnels par exemple parce qu'ils peuvent rendre compte des régularités sémantiques de dérivés construits par des affixations concurrentes comme les noms d'agents en français en *-eur* (VOLEUR), *-ant* (REPRÉSENTANT) ou *-iste* (JOURNALISTE) dont

les propriétés sémantiques sont similaires et qui entretiennent des relations analogues avec les membres de leurs familles dérivationnelles respectives.

Par ailleurs, les paradigmes dérivationnels ont été, dans le sillage des modèles *word-based*, une réponse à la conception générative de la construction morphologique et à ses règles binaires et orientées. Les modèles paradigmatiques mettent en jeu des relations dérivationnelles qui peuvent être orientées dans les deux sens (base → dérivé ou dérivé → base) ou ne pas être orientées du tout (Jackendoff 1975). D'autres part, ces relations ne sont pas limitées aux couples base-dérivé. Les paradigmes dérivationnels sont ainsi particulièrement adaptés à la description des relations transversales (*cross-formations*) qui caractérisent par exemple les couples de dérivés en *-isme* et en *-iste*, ou les affixations multiples, par exemple en *-isation* ou en *-ologique* (Lasserre & Montermini 2014), etc.

Les paradigmes dérivationnels sont des réseaux de mots interconnectés qui reproduisent les motifs (i.e. les régularités) formés par les nombreuses relations, de toute nature, que chacun des membres du paradigme entretient avec les autres. Ces réseaux s'agrègent au sein des familles dérivationnelles, se superposent pour former des séries dérivationnelles connectées au sein d'analogies (Skousen 1989, 1992, Krott et al. 2001, Dal 2003, Blevins & Blevins 2009, Arndt-Lappe 2015). Pour certains auteurs comme Stump (1991) ou Spencer (2013), les paradigmes dérivationnels décrivent des relations formelles entre deux classes sémantiques tandis que Štekauer (2014) propose qu'ils s'organisent autour de catégories cognitives. La plupart des auteurs considèrent que les paradigmes se composent de relations qui impliquent plus de deux éléments (Van Marle 1985, Booij 2010) même si pour certains, comme Spencer (2013), ils ne contiennent que des relations binaires.

6.2 Les principes de **ParaDis**

Le modèle ParaDis n'est pas une formalisation directe des paradigmes dérivationnels, mais plutôt un système qui met en jeu un ensemble de dispositifs permettant d'envisager les procédés constructionnels sous l'angle de leur dimension paradigmatique. L'architecture de ParaDis articule ainsi deux principes : la séparation des niveaux de description des lexèmes, et la conception modulaire de la construction morphologique. Le premier s'inscrit dans la droite ligne de l'analyse de Fradin (2003) : le lexème est une entité tri-dimensionnelle ; les trois dimensions fonctionnent de façon simultanée et indépendante dans chaque règle de construction. Le second correspond à un changement d'échelle : l'unité de traitement est étendue à un (sous-)ensemble des membres de la famille dérivationnelle du couple base-dérivé, ce qui donne au système la capacité d'analyser des constructions pour lesquelles la forme et le sens ne sont pas coordonnés, et notamment les formations parasynthétiques (section 4), mais aussi la concurrence affixale, comme dans le cas de la formation des noms de plantation dont la base dénote une plante (ex. CERISE → CERISAIE vs CERISE → CERISERAIE) ou les formations rivales d'adjectifs dénominaux en anglais en *-ic* et *-ical* (ex. HISTORY → HISTORIC vs HISTORY → HISTORICAL étudiés par Lindsay & Aronoff (2013)), ou encore de schémas dérivationnels polysémiques,

comme celui auquel appartiennent les adjectifs en *-istique*, comme FOOTBALLISTIQUE qui signifie 'relatif au football' ou 'relatif aux footballeurs' (voir Strnadová (2014b) pour une analyse des adjectifs dénominaux dont l'interprétation est ambiguë).

6.3 Quatre composants

La différence essentielle entre ParaDis et les modèles morphologiques *lexeme-based* est l'unité descriptive du mécanisme constructionnel. C'est le couple formé par un dérivé et sa base dans le courant lexématique de la morphologie, alors que dans ParaDis il s'agit du *module*, un dispositif qui opère au niveau du réseau de lexèmes. La notion de module s'inspire des *Patrons Cumulatifs* introduits dans Bochner (1993), qui propose de fusionner en patrons *n*-aires des schémas de lexèmes régulièrement connectés entre eux. Ces patrons résultent du recouvrement —autrement dit, du cumul— de relations élémentaires entre schémas de lexèmes partagés. Ces relations sont comparables à des RCL non orientées, en ce qu'elles inter-définissent collectivement les propriétés des lexèmes qu'elles mettent en relation. Ainsi, le patron cumulatif (15) qui exprime la relation ternaire qui connecte de façon régulière les noms d'idéologie en *-isme*, les noms d'adeptes en *-iste* et l'objet valorisé, est-il le produit de la superposition des structures binaires (16), (17) et (18), chacune exprimant un fragment de module (ces exemples sont empruntés à Strnadová (2014a)). En d'autres termes, comme pour Bochner (1993), un module est une structure de graphe connexe dont les sommets décrivent des ensembles de lexèmes dont les éléments entretiennent des relations d'interprédictibilité. L'un des corollaires de cette définition est que tout sous-module est un module.

$$(15) \quad \left\{ \begin{bmatrix} /X/ \\ N \\ \text{'Z'} \end{bmatrix}, \begin{bmatrix} /Xisme/ \\ N \\ \text{'mouvement favorisant Z'} \end{bmatrix}, \begin{bmatrix} /Xiste/ \\ A \\ \text{'qui relève de Z, du mouvement favorisant Z'} \end{bmatrix} \right\}$$

$$(16) \quad \left\{ \begin{bmatrix} /X/ \\ N \\ \text{'Z'} \end{bmatrix}, \begin{bmatrix} /Xisme/ \\ N \\ \text{'mouvement favorisant Z'} \end{bmatrix} \right\}$$

$$(17) \quad \left\{ \begin{bmatrix} /X/ \\ N \\ \text{'Z'} \end{bmatrix}, \begin{bmatrix} /Xiste/ \\ A \\ \text{'qui relève de Z'} \end{bmatrix} \right\}$$

$$(18) \quad \left\{ \begin{bmatrix} /Xisme/ \\ N \\ \text{'Y'} \end{bmatrix}, \begin{bmatrix} /Xiste/ \\ A \\ \text{'qui relève de Y'} \end{bmatrix} \right\}$$

L'une des différences entre le formalisme de Bochner (1993) et ParaDis est que dans ce dernier le fonctionnement modulaire se distribue suivant quatre niveaux de description lexicale, de sorte qu'un module se définit comme le produit de quatre composants interconnectés ayant chacun la structure d'un graphe connexe :

CS : Le *composant sémantique* est un réseau de classes sémantico-conceptuelles qui décrit la manière dont celles-ci interagissent.

CF : Le *composant formel* est un graphe de corrélations entre des schémas phonologiques ou graphémiques.

CC : Le *composant catégoriel* connecte les parties du discours impliquées.

CL : Le *composant lexical* réunit les lexèmes d'une même famille dérivationnelle qui vérifient l'ensemble des relations exprimées dans les trois autres composants.

Autrement dit, un module est l'expression des relations morphologiques qui existent entre certains lexèmes d'une même famille dérivationnelle, examinées indépendamment et simultanément à chacun des quatre niveaux de représentation lexicale. Le réseau qui réalise le composant lexical peut être considéré comme concret. Chacun des lexèmes qui le composent instancie une description abstraite dans chacun des trois autres niveaux. Pour le dire autrement, le niveau lexical est celui des familles dérivationnelles, c'est-à-dire des réalisations concrètes, alors que les trois autres niveaux décrivent les séries dérivationnelles, sous forme abstraite.

Un module structure les relations entre les membres d'une sous-famille dérivationnelle en quatre plans descriptifs. La notion de composant lexical, et plus globalement celle de module, permet d'affiner la définition des familles dérivationnelles. Alors que traditionnellement, une famille est définie comme l'ensemble des lexèmes partageant un même ascendant, nous considérons dans ParaDis que deux lexèmes appartiennent en effet à la même famille s'ils sont reliés par un chemin à travers un ou plusieurs composants lexicaux connexes. Une famille dérivationnelle devient ainsi une collection connexe de composants lexicaux. Prenons l'exemple du nom d'activité VIDAGE. Il entretient une relation régulière avec le prédicat verbal VIDER dont il constitue la nominalisation de procès, et avec le nom VIDEUR, qui s'interprète comme l'agent de cette activité, et dont la base est le même verbe VIDER. Les trois lexèmes entretiennent la même relation paradigmatique que par exemple (19) ou (20).

(19)   BRACONNER, BRACONNEUR, BRACONNAGE,

(20)   COLLECTER, COLLECTEUR, COLLECTAGE

Dans la terminologie de ParaDis, (VIDER, VIDEUR, VIDAGE) constitue le composant lexical du module représenté dans la figure 4. Le paradigme qu'il décrit inclut également les triplets (19) et (20). Ce module est régulier : il implique des catégories sémantiques (conceptuelles) logiquement connectées —un prédicat (PRED) se nominalise en une activité (ACT) et requiert un AGENT— et des schémas dérivationnels formellement interprédictibles : le thème *Y* du verbe utilisé en flexion pour construire les formes de l'imparfait l'est aussi pour construire les noms en *-age* et en *-eur*. Chaque sommet dans un des composants est connecté à un sommet au moins dans chacun des trois autres. La figure 4 rend compte de la régularité paradigmatique qui caractérise le triplet (VIDER, VIDEUR, VIDAGE), qui se manifeste dans la géométrie isomorphe (ici, triangulaire) des structures qui réalisent les composants formel (CF), sémantique (CS) et lexical (CL).

Pour alléger les graphiques des figures 4 à 8, le composant catégoriel n'est pas représenté explicitement. Nous avons indiqué sous forme d'indices dans le composant lexical
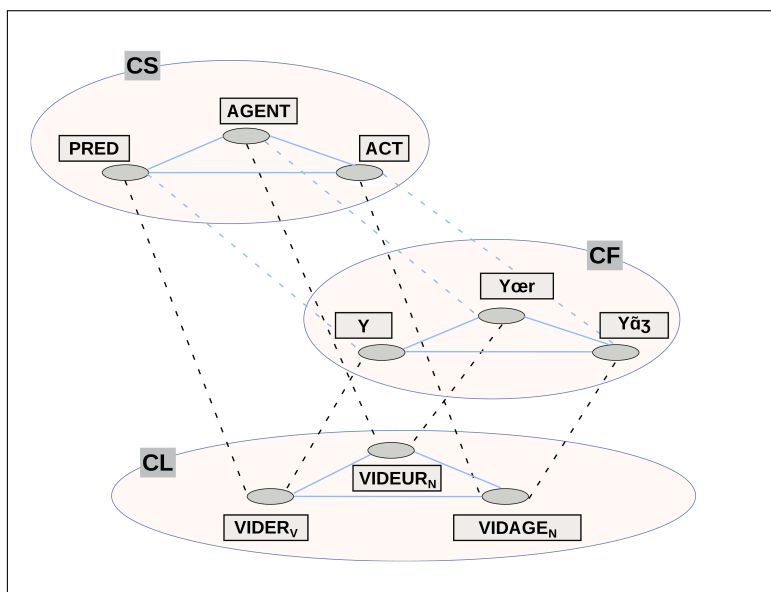
Figure 4 : Module correspondant à l'analyse de (VIDER, VIDEUR, VIDAGE). Le niveau catégoriel est omis.

les catégories grammaticales auxquelles appartiennent les lexèmes connectés. Les lignes continues représentent les connexions entre les éléments au sein d'un composant, et les lignes en pointillé relient les composants entre eux. La régularité de la construction de (VIDER, VIDEUR, VIDAGE) se traduit par une connexion doublement motivée dans le CL entre les éléments du triplet. Chacune de trois relations concrètes dans le CL est en effet l'instance de la relation abstraite correspondante dans les deux autres composants.

6.4 L'analyse « **ParaDisiaque** » des adjectifs en *anti-X-suf*

Nous avons montré que la dérivation parasynthétique était un modèle de préfixation répandu dans les langues, fréquemment observable pour une grande variété de suffixes, et, comme a pu le montrer Hathout (2011), extrêmement productif. Pour un dérivé *pref-X-suf*, la marque suffixale *suf* coïncide avec l'exposant de l'un des dérivés suffixés de *X*, i.e. *X-suf* quand celui-ci est attesté, témoignant ainsi du fait que, si *pref-X-suf* se définit par rapport à *X*, sa forme emprunte le segment *suf* au lexème *X-suf* dérivationnellement apparenté à *X*. La modélisation du schéma de construction de ces formes doit donc inclure un dispositif d'accès aux membres de la famille de *X*. En nous servant de l'analyse de l'adjectif ANTIMILITARISTE, voyons comment ce mécanisme est réalisé dans ParaDis.

La représentation d'ANTIMILITARISTE, dans la figure 5, se distribue suivant quatre dimensions : c'est un adjectif ; il instancie la classe conceptuelle d'opposition comme l'indique l'étiquette CONTRE dans le CS ; il vérifie le patron formel ãtiXist dans le CF. Le

module d'ᴀɴᴛɪᴍɪʟɪᴛᴀʀɪꜱᴛᴇ inclut dans son CL le nom ᴍɪʟɪᴛᴀɪʀᴇ avec lequel ᴀɴᴛɪᴍɪʟɪᴛᴀ-
ʀɪꜱᴛᴇ entretient une relation sémantiquement motivée : « une chanson antimilitariste »
est 'une chanson contre les militaires', et plus généralement 'une chanson contre l'ar-
mée'. La connexion entre les deux lexèmes est donc héritée du composant sémantique
où le concept CONTRE requiert nécessairement l'existence d'une entité (ENTITÉ) qui
est l'objet de cette opposition. Cette relation est régulière : toute entité (concrète ou abs-
traite) peut déclencher une réaction d'opposition, et à toute attitude hostile correspond
nécessairement l'objet rejeté.

En revanche, il n'existe pas de justification formelle à cette relation : ᴍɪʟɪᴛᴀɪʀᴇ dont
la forme est une instance du patron *Yɛʁ* (en considérant *militaire* comme formé sur le
thème supplétif º/milit/ de ᴀʀᴍᴇ́ᴇ), ne permet pas la prédiction de ãti*X*ist, et réciproque-
ment. Il apparaît ainsi un décalage entre la régularité sémantique et l'absence de lien
formel entre ᴀɴᴛɪᴍɪʟɪᴛᴀʀɪꜱᴛᴇ et ᴍɪʟɪᴛᴀɪʀᴇ, ce qu'illustre la figure 5 : la ligne continue
qui connecte CONTRE et ENTITÉ dans le CS n'a pas de correspondant dans le CF. La mo-
tivation sémantique justifie donc seule la relation qui unit, dans le CL, ᴀɴᴛɪᴍɪʟɪᴛᴀʀɪꜱᴛᴇ
et ᴍɪʟɪᴛᴀɪʀᴇ.
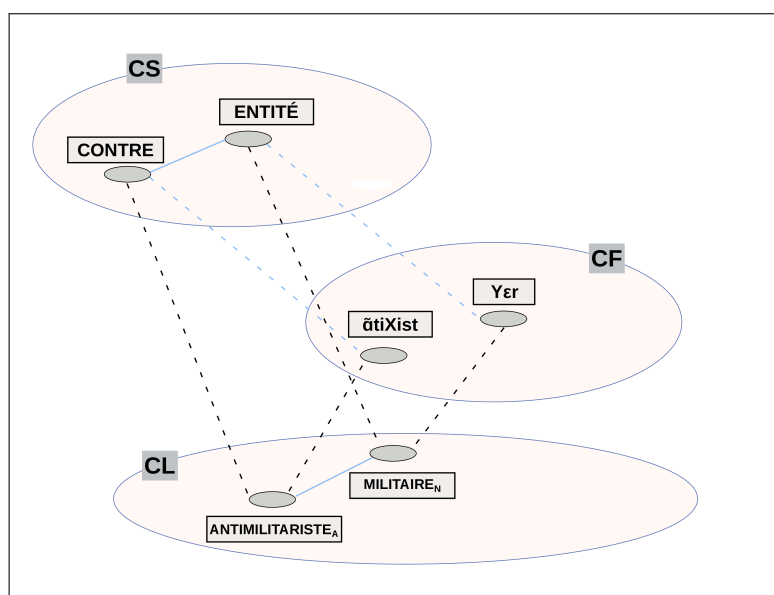
Figure 5 : Élément de l'analyse de ᴀɴᴛɪᴍɪʟɪᴛᴀʀɪꜱᴛᴇ : la motivation sémantique
ᴀɴᴛɪᴍɪʟɪᴛᴀʀɪꜱᴛᴇ ⟵ ᴍɪʟɪᴛᴀɪʀᴇ

Puisque la forme de ᴀɴᴛɪᴍɪʟɪᴛᴀʀɪꜱᴛᴇ ne coïncide pas avec sa construction sémantique,
c'est dans le voisinage dérivationnel de l'adjectif que l'on va chercher la motivation de
sa structure morphologique. Le nom (et adjectif) ᴍɪʟɪᴛᴀʀɪꜱᴛᴇ répond à cette exigence.
En effet, formellement, ᴍɪʟɪᴛᴀʀɪꜱᴛᴇ est une instance du patron *X*ist, et entretient une re-
lation d'interprédictibilité avec ᴀɴᴛɪᴍɪʟɪᴛᴀʀɪꜱᴛᴇ, *anti-* apparaissant fréquemment dans

des structures comportant une finale en /ist/[7]. C'est ce qui est représenté dans la figure 6. En revanche, la relation entre ANTIMILITARISTE et MILITARISTE ne répond à aucune motivation sémantique comme l'indique l'absence de relation d'interprédictibilité entre les catégories sémantiques CONTRE et PARTISAN dans la figure 6 : en l'occurrence, l'émergence d'un comportement adversatif (CONTRE) ne requiert pas l'existence d'un PARTISAN.
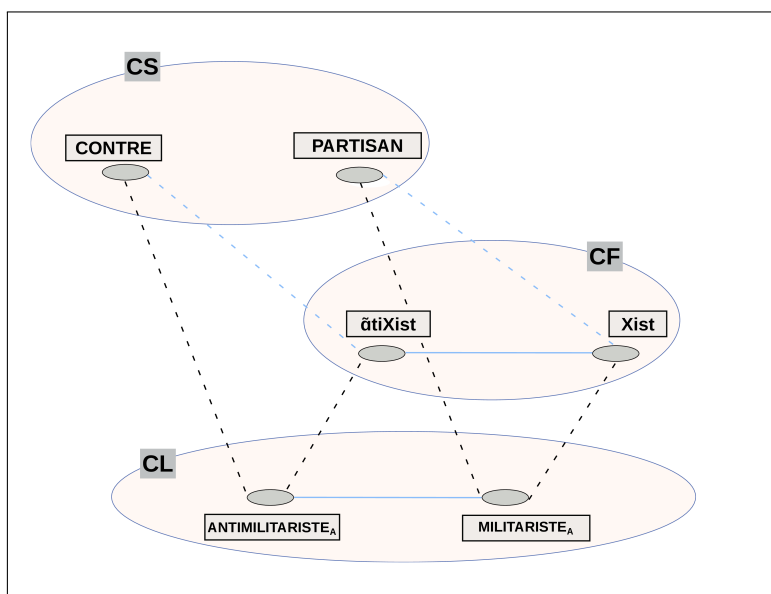
Figure 6 : Élément de l'analyse de ANTIMILITARISTE : la motivation formelle
ANTIMILITARISTE ⟵ MILITARISTE

PARTISAN, la catégorie sémantique de MILITARISTE, est en contrepartie indissociable de celle de l'objet valorisé, qui peut être une idéologie (le *pointillisme*, pour le *pointilliste*), un individu (*Sarkozy*, pour le *Sarkoziste*), une fonction (le *pape*, pour un *papiste*), une activité (*bouger*, pour le *bougiste*), un objet concret (la *viande*, pour le *viandiste*), etc. C'est en d'autres termes une entité conceptuelle non contrainte, que nous représentons par la classe ENTITÉ (voir Roché (2007, 2011a) pour une analyse détaillée des suffixations en *-isme* et *-iste* en français). La relation est également prédictible dans le CF : la suffixation en *-iste* présente une affinité notable avec les structures comportant une finale en /ɛʁ/[8]. L'assemblage des quatre composants, illustré par la figure 7, montre que MILITARISTE forme avec MILITAIRE un module sémantiquement et formellement régulier : la géométrie dans les quatre composants est isomorphe.

---

[7]Dans TLFindex par exemple, 11% des adjectifs de la forme ɑ̃ti*X* finissent en *-iste* (i.e. sont des instances de ɑ̃ti*X*ist).

[8]Les noms et adjectifs en *X*aʁist forment 4% des entrées en *Y*ist dans TLFindex.
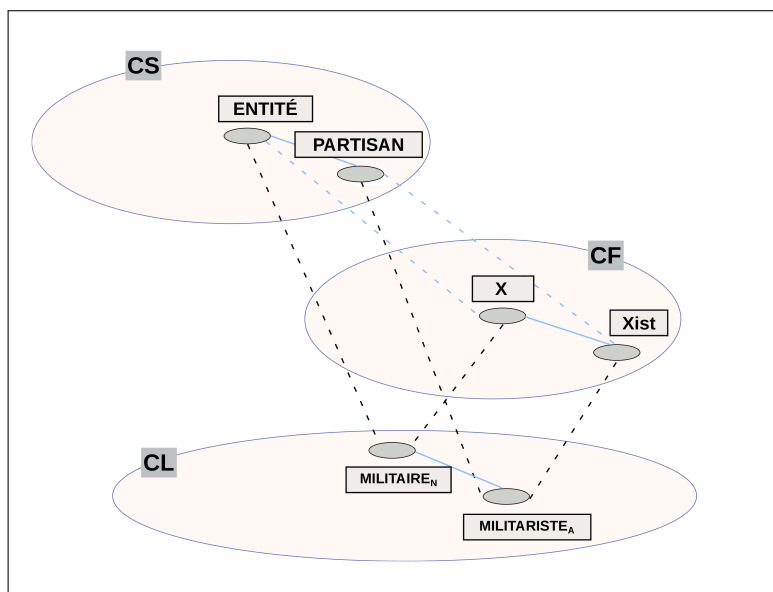
Figure 7 : Module régulier (MILITARISTE, MILITAIRE)

En rassemblant les éléments d'analyse que nous venons de présenter, on voit que la forme et le sens de ANTIMILITARISTE résultent d'une combinaison de facteurs qui interviennent de façon inégale :

1. ANTIMILITARISTE et MILITAIRE sont sémantiquement motivés l'un par l'autre (figure 5) ;

2. ANTIMILITARISTE et MILITARISTE sont formellement motivés l'un par l'autre (figure 6) ;

3. MILITARISTE et MILITAIRE sont sémantiquement et formellement connectés (figure 7).

Cette convergence de propriétés fait intervenir l'unification, au niveau du CF, du *X* de la figure 7 avec le *Y*ɛʁ de la figure 5, ce qui conduit à la spécification (21b) de la relation formelle (21a) de la suffixation en /ist/. La variation /ɛʁ/-/ɑʁ/ en (21b) est due à la proximité de la voyelle /ɛ/ avec le /i/ dans /ist/ :

(21)   a. X — Xiste
            ‖   ‖
       b. Yaire   — Yariste

Le résultat, présenté dans la figure 8, est un module dont les trois composants sont entièrement interconnectés, avec un composant lexical formant un graphe complet, et les composants sémantique et formel constituant chacun un graphe connexe acyclique

dont les sommets reliés sont différents. Comme on peut le voir, la figure 8 est une simple superposition des sous-modules des figures 5, 6 et 7. La non-coïncidence entre les trois composants abstraits dans la figure 8 se manifeste dans la géométrie des composants du module complet de ANTIMILITARISTE. Elle contraste avec la géométrie régulière du module de (VIDER, VIDEUR, VIDAGE) illustré par la figure 4 dont la canonicité paradigmatique se traduit par la coprésence de trois triangles isomorphes.
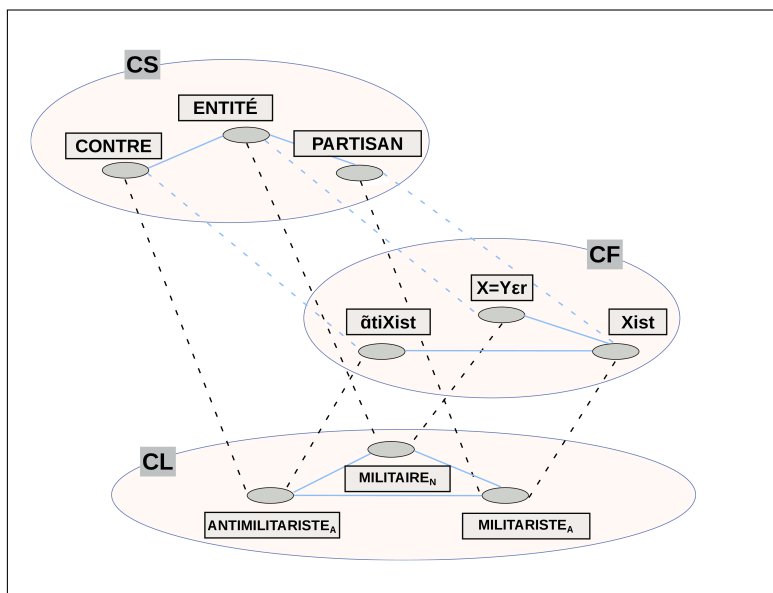
Figure 8 : Module décrivant à l'analyse de ANTIMILITARISTE

6.5 Pour récapituler

Le modèle ParaDis résulte d'un triple héritage : il s'inspire des *Patrons Cumulatifs* de Bochner qui essentiellement décrivent les composantes formelles et catégorielles de la dérivation morphologique. ParaDis les étend à la dimension sémantique des paradigmes et tire parti du fonctionnement indépendant et simultané des composants formel, catégoriel et sémantique des RCL et de la nature tri-dimensionnelle des lexèmes sur lesquels elles s'appliquent. Enfin, ParaDis adopte, dans le but de la formaliser, l'organisation en réseau de la morphologie constructionnelle initiée par l'axe DUMAL qu'il complète en les articulant avec les structures paradigmatiques de famille et de série dérivationnelles.

De cette manière, la distribution et le traitement des informations morphologiques de ParaDis servant à réaliser l'analyse des constructions morphologiques, et notamment des dérivés parasynthétiques, s'effectue sur trois plans :

— suivant les trois dimensions classiques du lexème ;

— auprès des membres de la famille dérivationnelle du dérivé que l'on souhaite décrire ;

— à travers les relations entre les séries dérivationnelles dans lesquelles s'insèrent les lexèmes de cette famille.

Avec cette organisation multi-niveaux ParaDis peut appréhender la construction morphologique aussi bien sous forme de relations binaires, que du point de vue de modules plus complexes instanciant les réseaux de motivation paradigmatiques des dérivés morphologiques ; l'organisation proposée permet de traiter de manière uniforme tous les types de dérivés, quel que soit leur éloignement vis-à-vis de la situation idéale de transparence formelle et sémantique. Relativement aux modèles qui l'on précédé, ParaDis peut donc traiter les apparentes anomalies constructionnelles que manifestent les dérivés parasynthétiques, sans recourir à des artéfacts analytiques : les mécanismes qui servent à les analyser sont strictement identiques à ceux qui permettent d'analyser les dérivations canoniques. Les relations formelles et sémantiques asynchrones qui induisent leur écart relativement à la situation canonique sont envisagées de manière disjointe, se traduisant, dans le cas de la parasynthèse, par une autonomisation de la motivation du préfixe et de la séquence suffixale. La disponibilité de la famille du dérivé parasynthétique, distribuée dans les différents composants, et sa structure en réseaux permettent de calculer la forme appropriée de la séquence finale.

Références

Acedo-Matellán, Victor & Jaume Mateu. 2009. L'expressió dels esdeveniments de canvi : del llatí al català. In Joan Rafel (éd.), *Diachronic linguistics*, 419–442. Girona : Documenta Universitaria.

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes : Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (éds.), *Analogy in grammar : Form and acquisition*, 54–82. Oxford : Oxford University Press.

Alcoba-Rueda, Santiago. 1987. Los parasintéticos : constituyentes y estructura léxica. *Revista española de lingüística* 17. 245–267.

Amiot, Dany & Georgette Dal. 2005. Integrating combining forms into a lexeme-based morphology. In Geert Booij, Bernard Fradin, Angela Ralli & Sergio Scalise (éds.), *Proceedings of the Fifth Mediterranean Meeting on Morphology (MMM5)*, 323–336. Fréjus, France.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge : Cambridge University Press.

Arndt-Lappe, Sabine. 2015. Word-formation and analogy. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, 822–841. Berlin : de Gruyter Mouton.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge : MIT Press.

Baerman, Matthew, Greville G. Corbett & Dunstan Brown. 2010. *Defective paradigms : Missing forms and what they tell us*. Oxford : Oxford University Press.

Basílio, Margarida. 1991. *Teoria lexical*. São Paulo : Ática.

Beard, Robert. 1995. *Lexeme morpheme-based morphology*. Albany : State University of New York Press.

Blevins, James P. 2003-12. Stems and paradigms. *Language* 79(4). 737–767.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.

Blevins, James P. & Juliette Blevins. 2009. *Analogy in grammar : Form and acquisition.* Oxford : Oxford University Press.

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin : Mouton de Gruyter.

Bonami, Olivier & Gilles Boyé. 2007. Remarques sur les bases de la conjugaison. In Elisabeth Delais-Roussarie & Laurence Labrune (éds.), *Des sons et des sens*, 77–90. Paris : Hermès.

Bonami, Olivier, Gilles Boyé & Françoise Kerleroux. 2009. L'allomorphie radicale et la relation dérivation-construction. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 103–125. Saint-Denis : Presses Universitaires de Vincennes.

Bonami, Olivier & Gregory T. Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley & Gregory T. Stump (éds.), *The Cambridge Handbook of Morphology*, 449–481. Cambridge : Cambridge University Press.

Booij, Geert. 1977. *Dutch morphology*. Dordrecht : Foris.

Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1995*, 1–16. Dordrecht : Kluwer Academic Publishers.

Booij, Geert. 2002. *The morphology of Dutch*. Oxford : Oxford University Press.

Booij, Geert. 2010. *Construction morphology*. Oxford : Oxford University Press.

Bosque, Ignacio. 1983. La morfología. In Francisco Abad & Antonio García Berrio (éds.), *Introducción a la lingüística*, 115–153. Madrid : Alhambra.

Boyé, Gilles & Gauvain Schalchli. 2016. The status of paradigms. In Andrew Hippisley & Gregory T. Stump (éds.), *The Cambridge handbook of morphology*, 206–234. Cambridge : Cambridge University Press.

Carstairs-McCarthy, Andrew. 1992. *Current morphology*. London : Routledge.

Carstairs-McCarthy, Andrew. 1994. Inflection classes, gender, and the principle of contrast. *Language* 70(4). 737–788.

Chovanová, Iveta. 2010. On the semantics of Slovak AN compound adjectives. *Italian Journal of Linguistics / Rivista di linguistica* 22(2). 229–264.

Corbett, Greville G. 2010. Canonical derivational morphology. *Word Structure* 3(2). 141–155.

Corbin, Danielle. 1980. Contradictions et inadéquation de l'analyse parasynthétique en morphologie dérivationnelle. In Anne-Marie Dessaux-Berthonneau (éd.), *Linguistiques et traditions grammaticales*, 181–224. Villeneuve-d'Asq : Presses universitaires de Lille.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Max Niemeyer Verlag.

Crocco-Galéas, Grazia. 2003. Compound adjectives in English : The type of lion-hearted and good-natured. *Studi italiani di linguistica teorica ed applicata* 32(1). 31–43.

Crocco-Galéas, Grazia & Claudio Iacobini. 1993. Parasintesi e doppio stadio derivativo nella formazione verbale del latino. *Archivio Glottologico Italiano* 78. 167–199.

Dal, Georgette. 2003. Analogie et lexique construit : quelles preuves ? *Cahiers de Grammaire* 28. 9–30.

Dal, Georgette & Fiammetta Namer. 2016. À propos des occasionalismes. In Franck Neveu, Gabriel Bergounioux, Marie-Hélène Côté, Jean-Marc Fournier, Linda Hriba & Sophie Prévost (éds.), *Actes du 5e Congrès Mondial de Linguistique Française. Tours, 4-8 juillet 2016.* Paris : Institut de Linguistique Française.

Darmesteter, Arsène. 1875. *Traité de la formation des mots composés dans la langue française.* Paris : Honoré Champion.

Darmesteter, Arsène. 1877. *De la création actuelle de mots nouveaux dans la langue française et des lois qui la régissent.* Paris : F. Vieweg.

Darmesteter, Arsène. 1894. *Traité de la formation des mots composés dans la langue française.* 2ᵉ éd. Paris : Champion.

Davis, Anthony R. & Jean-Pierre Koenig. 2000. Linking as constraints on word classes in a hierarchical lexicon. *Language* 76(1). 56–91.

Dell, François. 1970. *Les règles phonologiques tardives et la morphologie dérivationnelle du français.* Cambridge : MIT thèse de doct.

Dell, François. 1979. La morphologie dérivationnelle du français et l'organisation de la composante lexicale en grammaire générative. *Revue romane* XIV(2). 185–216.

Efthymiou, Angeliki. 2014. Modern Greek parasynthetic verbs : A hierarchical relationship between prefixes and suffixes ? In Stela Manova (éd.), *Affix ordering across languages and frameworks*, 82–110. Oxford : Oxford University Press.

Fradin, Bernard. 1997a. Esquisse d'une sémantique de la préfixation en *anti-*. *Recherches linguistiques de Vincennes* 26. 87–112.

Fradin, Bernard. 1997b. Une préfixation complexe : le cas de *anti-*. *Neuphilologische Mitteilungen* XCVIII(4). 333–349.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie.* Paris : Presses Universitaires de France.

Fradin, Bernard, Nabil Hathout & Fanny Meunier. 2003. La suffixation en *-ET* et la question de la productivité. *Langue Française* 140. 56–78.

Fradin, Bernard & Françoise Kerleroux. 2003a. Quelle base pour les procédés de la morphologie constructionnelle ? In *Les unités morphologiques. Actes du 3ᵉ forum de morphologie*, t. 3 (Silexicales), 76–83. Villeneuve-d'Asq : Université de Lille 3.

Fradin, Bernard & Françoise Kerleroux. 2003b. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (éds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona : IULA – Universitat Pompeu Fabra.

Gaeta, Livio. 2010. Synthetic compounds. With special reference to German. In Sergio Scalise & Irene Vogel (éds.), *Cross-disciplinary issues in compounding*, 219–235. Amsterdam : Benjamins.

Guevara, Emiliano. 2007. Binary branching and linguistic theory : Morphological arguments. In M. Cecilia Picchi & Alan Pona (éds.), *Proceedings of the "XXXII incontro di grammatica generativa''*, 93–106. Alessandria : Edizioni dell'Orso.

Hathout, Nabil. 2008. Acquisition of the morphological structure of the lexicon based on lexical similarity and formal analogy. In *Proceedings of the COLING workshop Textgraphs-3*, 1–8. Manchester.

Hathout, Nabil. 2011. Une approche topologique de la construction des mots : propositions théoriques et application à la préfixation en *anti-*. In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.), *Des unités morphologiques au lexique*, 251–318. Paris : Hermès Science-Lavoisier.

Hathout, Nabil, Fabio Montermini & Ludovic Tanguy. 2008. Extensive data for morphology : Using the World Wide Web. *Journal of French Language Studies* 18. 67–85.

Hathout, Nabil & Fiammetta Namer. 2014a. Démonette, a French derivational morphosemantic network. *Linguistic Issues in Language Technology* 11(5). 125–168.

Hathout, Nabil & Fiammetta Namer. 2014b. Discrepancy between form and meaning in word formation : The case of over- and under-marking in French. In Franz Rainer, Wolfgang U. Dressler, Francesco Gardani & Hans Christian Luschützky (éds.), *Morphology and meaning*, 177–190. Amsterdam : John Benjamins.

Hathout, Nabil & Fiammetta Namer. 2016. Giving lexical resources a second life : Démonette, a multi-sourced morpho-semantic network for French. In *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*. Portorož, Slovenia.

Hathout, Nabil, Fiammetta Namer, Marc Plénat & Ludovic Tanguy. 2009. La collecte et l'utilisation des données en morphologie. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 266–287. Saint-Denis : Presses Universitaires de Vincennes.

Hathout, Nabil, Marc Plénat & Ludovic Tanguy. 2003. Enquête sur les dérivés en *-able*. *Cahiers de grammaire* 28. 49–90.

Heyna, Franziska. 2014. *Etudes morpho-syntaxiques des parasynthétiques. Les dérivés en dé- et en* anti-. Louvain : De Boeck – Duculot.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–234.

Iacobini, Claudio. 2004. Parasintesi. In Maria Grossmann & Franz Rainer (éds.), *La formazione delle parole in italiano*, 165–188. Tübingen : Niemeyer.

Iacobini, Claudio. 2010. Les verbes parasynthétiques : de l'expression de l'espace à l'expression de l'action. *De Lingua Latina* 3. 1–16.

Jackendoff, Ray S. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3). 639–671.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford : CSLI Publications.

Krott, Andrea, R. Harald Baayen & Robert Schreuder. 2001. Analogy in morphology : Modeling the choice of linking morphemes in Dutch. *Linguistics* 39. 51–94.

Lasserre, Marine & Fabio Montermini. 2014. Pour une typologie des lexèmes construits : entre composition, composition néoclassique et affixation. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberger, Judith Meinschaefer & Sophie Prévost

(éds.), *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1797–1812. Paris : Institut de Linguistique Française.

Lieber, Rochelle & Pavol Štekauer. 2009. *The Oxford handbook of compounding*. Oxford : Oxford University Press.

Lignon, Stéphanie, Fiammetta Namer & Florence Villoing. 2014. De l'agglutination à la triangulation ou comment expliquer certaines séries morphologiques. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1813–1835. Paris : Institut de Linguistique Française.

Lindsay, Mark & Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In Fabio Montermini, Gilles Boyé & Jesse Tseng (éds.), *Morphology in Toulouse*, 133–153. Munich : Lincom Europa.

Matthews, P. H. 1974. *Morphology*. Cambridge : Cambridge University Press.

McCarthy, John & Alan Prince. 1993. *Prosodic morphology : Constraint interaction and satisfaction*. Technical report 3. New Brunswick : Rutgers University Center for Cognitive Science.

Mel'čuk, Igor. 1993. *Cours de morphologie générale. Première partie : Le mot*. T. 1. Montréal /Paris : Les Presses de l'Université de Montréal/CNRS Éditions.

Melloni, Chiara & Antonietta Bisetto. 2010. Parasynthetic compounds : Data and theory. In Sergio Scalise & Irene Vogel (éds.), *Cross-disciplinary issues in compounding*, t. 311, 199–218. John Benjamin Publishers company.

Neef, Martin. 2015. Synthetic compounds in German. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, t. 1, chap. 35, 582–593. Berlin : de Gruyter Mouton.

Plénat, Marc. 2009a. Le conditionnement de l'allomorphie radicale en français. In *La morphologie lexicale : un domaine autonome de la grammaire ?* (Mémoires de la Société de Linguistique, Nouvelle série), 119–140. Leuven : Peeters.

Plénat, Marc. 2009b. Les contraintes de taille. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 47–63. Saint-Denis : Presses Universitaires de Vincennes.

Plénat, Marc. 2011. Enquête sur divers effets des contraintes dissimilatives en français. In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.), *Des unités morphologiques au lexique*, 145–190. Paris : Hermès Science-Lavoisier.

Plénat, Marc & Michel Roché. 2012. Éléments de morphologie quotidienne - contribution à l'étude des conditionnements de l'épenthèse présuffixale. In Franck Neveu, Valéria Muni-Toke, Peter Blumenthal, Thomas Kingler, Pierluigi Ligas, Sophie Prévost & Sandra Teston-Bonnard (éds.), *Actes du 3e congrès mondial de linguistique française*, 1369–1386. Institut de Linguistique Française.

Plénat, Marc, Ludovic Tanguy, Stéphanie Lignon & Nicole Serna. 2002. La conjecture de Pichon. *Corpus* 1. 105–150.

Pounder, Amanda. 2000. *Process and paradigms in word-formation morphology*. Berlin : Walter de Gruyter.

Prince, Alan & Paul Smolensky. 1993. *Optimality theory : Constraint interaction in generative grammar*. Technical report 2. New Brunswick : Rutgers University Center for Cognitive Science.

Reinheimer-Ripeanu, Sanda. 1974. *Les dérivés parasynthétiques dans les langues romanes.* La Haye : Mouton.

Roché, Michel. 2007. Logique lexicale et morphologie : La dérivation en *-isme.* In Fabio Montermini, Gilles Boyé & Nabil Hathout (éds.), *Selected proceedings of the 5th Décembrettes : Morphology in Toulouse*, 45–58. Somerville : Cascadilla Press.

Roché, Michel. 2009. Pour une morphologie *lexicale.* In *La morphologie lexicale est-elle possible ?* (Mémoires de la Société de Linguistique, Nouvelle Série), 65–87. Leuven : Peeters.

Roché, Michel. 2010. Base, thème, radical. *Revue linguistique de Vincennes* 39. 95–134.

Roché, Michel. 2011a. Quel traitement unifié pour les dérivations en *-isme* et en *-iste* ? In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.), *Des unités morphologiques au lexique*, 69–143. Paris : Hermès Science-Lavoisier.

Roché, Michel. 2011b. Quelle morphologie ? In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.), *Des unités morphologiques au lexique*, 15–39. Paris : Hermès Science-Lavoisier.

Roché, Michel, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat. 2011. *Des unités morphologiques au lexique*. Paris : Hermès Science-Lavoisier.

Roché, Michel & Marc Plénat. 2014. Le jeu des contraintes dans la sélection du thème présuffixal. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberger, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1863–1878. Paris : Institut de Linguistique Française.

Scalise, Sergio. 1984. *Generative morphology*. Dordrecht : Foris.

Scalise, Sergio. 1994. *Morfologia*. Bologna : Il Mulino.

Schroten, Jan. 1997. On denominal parasynthetic verbs in Spanish. In Jane A. Coerts & Helen de Hoop (éds.), *Linguistics in the Nederlands*, 195–206. Amsterdam : Benjamins.

Serrano Dolader, David. 1995. *Las formaciones parasintéticas en español*. Madrid : Arco Libros.

Serrano Dolader, David. 2015. Parasynthesis in Romance. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (éds.), *Word-formation. An international handbook of the languages of Europe*, t. 1, chap. 29, 524–536. Berlin : de Gruyter Mouton.

Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht : Kluwer.

Skousen, Royal. 1992. *Analogy and structure*. Dordrecht : Kluwer Academic Publishers.

Spencer, Andrew. 2013. *Lexical relatedness : A paradigm-based model*. Oxford : Oxford University Press.

Štekauer, Pavol. 2014. Derivational paradigms. In Rochelle Lieber & Pavol Štekauer (éds.), *The Oxford handbook of derivational morphology*, 354–369. Oxford : Oxford University Press.

Strnadová, Jana. 2014a. *Les réseaux adjectivaux : sur la grammaire des adjectifs dénominaux en français*. Université Paris Diderot & Univerzita Karlova v Praze Thèse de doctorat.

Strnadová, Jana. 2014b. Multiple derivation in French denominal adjectives. In Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé & Fabio Montermini (éds.), *Proceedings of the 8th Décembrettes*, 327–346. Bordeaux.

Stump, Gregory T. 1991. A paradigm-based theory of morphosemantic mismatches. *Language*. 675–725.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge : Cambridge University Press.

Stump, Gregory T. 2006a. Heteroclisis and paradigm linkage. *Language* 82. 279–322.

Stump, Gregory T. 2006b. Paradigm function morphology. In Keith Brown (éd.), *Encyclopedia of language and linguistics*, t. 9, 171–173. Oxford : Elsevier.

Stump, Gregory T. 2016. *Inflectional paradigms : Content and form at the syntax-morphology interface*. Cambridge : Cambridge University Press.

Stump, Gregory T. & Raphael A. Finkel. 2013. *Morphological typology : From word to paradigm*. T. 138. Cambridge University Press.

Tribout, Delphine. 2012. Verbal stem space and verb to noun conversion in French. *Word Structure* 5(1). 109–128.

Van Marle, Jaap. 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht : Foris.

Wunderlich, Dieter & Ray Fabri. 1995. Minimalist morphology : An approach to inflection. *Zeitschrift für Sprachwissenschaft* 14(2). 236–294.

Chapter 16

Much ado about morphemes

Hélène Giraudo

CLLE, Université de Toulouse, CNRS, Toulouse, France

Most of the psycholinguists working on morphological processing nowadays admit that morphemes are represented in long-term memory and the predominant hypothesis of lexical access is morpheme-based as it supposes a systematic morphological decomposition mechanism taking place during the very early stages of word recognition. Consequently, morphemes would stand as access units for any item (i.e., word or nonword) that can be split into two morphemes. One major criticism of this prelexical hypothesis is that the mechanism can only be applied to regular and perfectly segmentable words and, more problematic, it reduces the role of morphology to surface/formal effects. Recently, Giraudo & Dal Maso (2016) discussed the issue of morphological processing through the notion of morphological salience – as defined as the relative role of the word and its parts – and its implications for theories and models of morphological processing. The issue of the relative prominence of the whole word and its morphological components has indeed been overshadowed by the fact that psycholinguistic research has progressively focused on purely formal and superficial features of words, drawing researchers' attention away from what morphology really is: systematic mappings between form and meaning. While I do not deny that formal features can play a role in word processing, an account of the general mechanisms of lexical access also needs to consider the perceptual and functional salience of lexical and morphological items. Consequently, if the sensitivity to the morphological structure is recognized, I claim that it corresponds to secondary and derivative units of description/analysis. Focusing on salience from a mere formal point of view, I consider in the present contribution how a decompositional hypothesis could deal with some phonological endings whose graphemic transcriptions are various. To this end, a distributional study of the final sound [o] in French is presented. The richness and the diversity of the distributions of this ending (in terms of type of forms, size and frequency) reveal that paradigmatic relationships are more suitable to guide morphological processing than morphological parsing as suggested by the lexeme-based approach of morphology (see Fradin 2003).

1 Introduction

In the domain of linguistics, morphological analysis is conceived according to two antagonistic approaches. On the one hand, the morpheme-based approach (exemplified by

the theoretical framework of Distributed Morphology, see Halle & Marantz 1993, 1994) integrates morphology with syntax and considers morphemes as basic minimal forms. On the other hand, the lexeme-based approach postulates that words are the first units of analysis (e.g. Corbin 1987, Aronoff 1994, Fradin 1996) . Psycholinguistic research aiming to understand the cognitive processes underlying word processing has broadly explored the effects of morphological processing on the underlying processes of lexical access. Whereas it was widely admitted that morphological information plays a crucial role during word processing, its representation is still controversial. Nowadays most psycholinguists support a decompositional view of morphological processing (see Rastle & Davis 2008 for a review) that can be linked to the morpheme-based hypothesis, while a few of them defend an opposing view according to which words are recognized holistically. This last procedural hypothesis, which is clearly in line with the lexeme-based approach, is tested in the present chapter through a qualitative and quantitative study of words ending in [o]. The distribution of this ending is so diverse that it would cause a huge number of procedural errors of morpheme decomposition. Conversely, the lexeme-based/holistic approach to morphology seems to be much more appropriate to encompass the diversity.

2 Studying morphological processing

In a seminal experimental study carried out by Taft & Forster (1975) on the recognition of nonwords, the idea of morphological decomposition was first introduced. They showed that 1) nonwords (e.g., *juvenate*) corresponding to an English stem induced longer rejection latencies than nonwords that were not stems (e.g., *pertoire*) and 2) prefixed nonwords constructed with an English prefix and stem (e.g., *dejuvenate*) took longer to be classified compared to morphologically simple control items (e.g., *depertoire*). Longer decision latencies were interpreted as reflecting a pre-lexical mechanism of morphological decomposition by which all the words (real or possible) would be accessed via the first activation of their stem. Forty years of experimental research have been focused on testing this decomposition hypothesis by manipulating the characteristics of morphologically complex words and nonwords (i.e., their form in terms of decomposability and interpretability, their lexical frequency and more rarely their lexical environment) in various perceptual tasks (with a large dominance of the lexical decision task which consists in a word/nonword discrimination) and numerous languages (most studies focusing on English, however). Most of the results have been interpreted as supporting the decompositional view (see the reviews of Amenta & Crepaldi 2012, Diependaele et al. 2012) without really questioning the linguistic processes underlying the construction of complex words. An overview of the tested hypotheses and the materials used to explore complex word recognition indeed reveals a lack of consideration of the paradigmatic characteristic of words for understanding the cognitive mechanisms underlying lexical access. Numerous studies mainly focused on the formal properties of the word and extended the morphological sensitivity effects observed with complex nonwords to complex words (e.g. Taft & Forster 1976, Caramazza et al. 1988, Laudanna et al. 1997, Crepaldi et al. 2010) failing to consider semantic aspects of morphological complexity. Many experimental re-

ports examined morphological processing using the masked priming paradigm (Forster & Davis 1984) that is supposed to reflect the automatic and nonconscious processes engaged in the very early stages of word recognition. In this paradigm, two visually related items are presented successively and participants are asked to perform a lexical decision indicating whether the second item is a word or not. However, because the prime word is presented masked and very briefly, the reader is not even aware of its presence before seeing the target item.[1] Hence, the paradigm allows examination of the effects of the unconscious processes of the prime processing on the target recognition (see Kinoshita & Lupker 2004 for a review on masked priming) . Many masked priming studies demonstrated that when two words are morphologically related (e.g., *singer–sing*), the prior presentation of the prime shortens the recognition latency of the target relative to both a baseline condition in which the prime is completely unrelated to the target (e.g., *baker–sing*) and an orthographic condition that uses a prime that is only formally related to the target (e.g., *single–sing*). Accordingly, morphological priming effects do not result from the mere formal overlap shared by prime–target. Other studies showed that semantic priming effects (e.g., *cello–violin*) only arise when the prime duration is sufficiently high (i.e., > 72 ms, see Rastle et al. 2000 for a comparison between morphological, orthographic and semantic priming effects using different Stimulus–Onset Asynchronies). This general result suggests that priming effects result from morphological relationships shared by prime–target pairs and that morphologically related words are connected by some kind of excitatory links. Most of the models of lexical access have tried to account for these morphological effects.

3   Psycholinguistic models of morphological processing

The architecture of psycholinguistic models of word recognition is mostly based on symbolic interactive activation models (e.g. McClelland & Rumelhart 1981). This type of model is organized in hierarchical levels of processing containing symbolic units. Each level corresponds to a linguistic characteristic of words, from letter features to semantics. During word recognition, activation spreads from the lowest to the highest levels. Within-level units are connected by inhibitory links whereas inter-level units are by excitatory links. Consequently, the model functions according to a principle of competition between within-level units that is compensated by both bottom-up and top-down excitations. The independence of the morphological effects relative to mere formal and semantic effects being established, morphological information was usually represented as a separate level of processing. However, its locus relative to the formal level (phonological and orthographic descriptions of the words) and the semantic level is still controversial. Morphological units have been situated variously: before the formal level and stand as access units to the mental lexicon (see Figure 1a depicting the sublexical model, Taft 1994), at the interface of the formal and the semantic level, organizing the word representations in morphological families (see Figure 1b, the supralexical model, Giraudo

---

[1]The *Stimulus Onset Asynchrony* is usually less than 50 milliseconds, it corresponds to a subliminal processing.

& Grainger 2001) or at either places, before and after the formal level (see Figure 1c, the hybrid/dual route model, Diependaele et al. 2009 ; see also Diependaele et al. 2013).
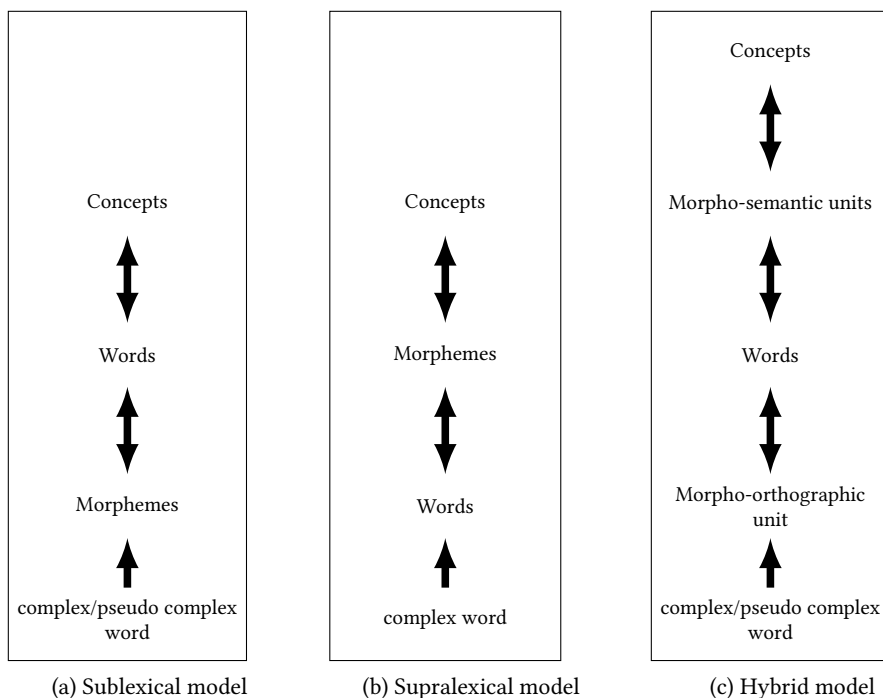
Figure 1: Alternative hierarchical models of morphological processing.

These three options nevertheless assume morpheme representations and by extension, propose a decompositional view of morphology. The sublexical and the hybrid models of morphological processing actually state very clearly that complex words are systematically decomposed into morphemes during lexical access. This decomposition mechanism is reflected by the obligatory activation of morphemes to gain the word representations coded within the mental lexicon. Each time a complex or a pseudo complex word (i.e., a word with a surface morphological structure like for example the word *corner* which comprises a surface stem *corn-* and a surface suffix *-er*) is processed by our cognitive system, it triggers the activation of its constituent morphemes that successively activate the wordforms containing it. Moreover, the hybrid model supposes that "In a priming context, opaque morphological relatives will only be able to prime each other through shared representations at the morpho-orthographic level, whereas transparent items will also be able to do this via shared representations at the morpho-semantic level" (Diependaele et al. 2009: 896). Even if the authors claim that morphological representations *per se* are not simply represented at both levels – the first being orthographically constrained and the second semantically constrained – these two levels actually correspond to surface

morphemes at least as far as the contained units are concerned. In these two frame-works (sublexical and hybrid models), morphological priming effects result from the pre-activation of the morpheme shared by the prime and the target before accessing the word representations. These morphemic units pre-select in a way the wordforms that can potentially match with the target to be recognized. Lexical access takes place via the obligatory activation of surface morphemes.

One major criticism of the prelexical hypothesis is that this mechanism can only be applied to regular and perfectly segmentable words. Even more problematic is the fact that it reduces the role of morphology to surface/formal effects. This is certainly why Diependaele and colleagues proposed a second level of representation for morphology, as numerous experimental studies showed that two morphologically related but ortho-graphically unrelated words (e.g., *bought–buy*) prime each other. However, this solution only considers morphology from its syntagmatic dimension: that is according to the word internal structure. Therefore, nothing is said about the influences of family and series[2] on word representations.

The original version of the supralexical model (Giraudo & Grainger 2001) also inte-grated morphemes even though it did not suppose a decomposition mechanism by which word representations are decomposed properly in order to activate their semantic rep-resentations. On the contrary, the morphological level contained "emerging" base mor-phemes, that is, morpheme representations resulting from the acquisition of complex words that are derived from the same base or the same series. Accordingly the mor-phological node organizes the word level in paradigms (i.e., morphological families and series), morphologically related words being connected together thanks to a supralexi-cal node. Concretely, when the system processes a complex word, it first activates all the word representations that match formally with it while at the same time the com-plex forms activate their common nodes that feed back positively these forms. As all units belonging to the same level compete with each other, the activated formally re-lated words inhibit each other, but those which are also morphologically related receive facilitation from their shared node. Words from the same family are then less inhibited than the other representations at the word level. In masked priming, the morphological facilitation between two morphologically related words observed relatively to two un-related words is explained in terms of a reduced inhibition effect compared to a regular inhibition effect for unrelated items.

4 The benchmark effects: lexicality, frequency, regularity

Among the factors that have been manipulated in order to better understand the nature of morphological relationships and the locus of morphological priming effects within the mental lexicon, one can cite *lexicality*, *frequency* and *regularity*. Starting from the dominant hypothesis according to which words are first decomposed before accessing

---

[2]The term 'series' was, to our knowledge, first introduced by Hathout (2005, 2008) and refers to groups of words sharing the same affix.

the mental lexicon, some authors used the masked priming paradigm to study the influence of lexicality (i.e., comparing the processing of existing words coded in the mental lexicon relative to non-existing but morphologically structured items) in word recognition. A series of masked priming studies examined the effect of complex nonword primes during the early processes of lexical access. For example, Longtin & Meunier (2005) have tested the effects of nonwords constructed using legal and illegal combinations of existing stems and suffixes in French (e.g., legal: *infirmiser–infirme 'disabled+er'– 'disabled'*; illegal: *garagité–garage 'garage+ité'–'garage'*) and found that both types of nonwords produced facilitation relative to orthographic control primes (e.g., *rapiduit– rapide, 'fast+uit'–'fast'*), that did not induce any significant effect on word recognition (see also, McCormick et al. 2009, Morris et al. 2013 for English materials). Giraudo & Voga (2013) replicated these results using French prefixed nonwords (e.g., *infaire–faire, 'un-do'–'do'*) suggesting that these effects apply to all affixed items. Andoni Dunabeitia et al. (2008), focused on affix priming in Spanish and showed that isolated suffixes (e.g., *dad–igualdad, 'ity'–'eguality'*) and suffixes in neutral context (e.g*., #####dad–igualdad*) were also able to induce positive priming effects (see also Crepaldi et al. 2016 using English suffixed related nonword pairs like *sheeter–teacher*). Finally, Crepaldi et al. (2013) examined reversed compounds like *fishgold–goldfish* and observed facilitation within related prime–targets pairs.

Taken together these studies suggest that in the early stages of word recognition – in masked priming conditions in which primes are presented less than 50-60 ms – lexicality does not impact lexical access as far as complex nonwords are considered. Moreover, none of these studies found priming effects using orthographic nonword primes (e.g., *blunana–blunt* tested by McCormick et al. 2009) suggesting a pre-lexical morphological analysis of the primes, blind to lexicality. However, even if these data seem to strengthen the pre-lexical decomposition hypothesis, results obtained using nonword primes created by letter transpositions have to be considered. Following, the discovery in Cambridge University according to which "*it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae... it doesn't matter in what order the letters in a word are, the only important thing is that the first and last letter be at the right place*" (see *http://www.mrc-cbu.cam.ac.uk/personal/matt.davis/Cmabrigde/*), a series of masked priming experiments aimed to explore this effect. Some studies showed that reading comprehension of jumbled words are more or less costly (as demonstrated for example by Rayner et al. 2006), this effect still constitutes a challenge for the decompositionalists. It indeed contradicts the hypothesis according to which lexical access takes place via the obligatory decomposition of complex words into morphemes. Masked priming experiments explored repetition priming effects (i.e., the same stimulus is presented as prime and target, like in *table–table*) and morphological priming effects using jumbled primes and Beyersmann et al. (2012) first found that relative to unrelated primes, both repeated simple primes (e.g., *wran–warn*) and morphological primes (e.g., *wranish–warn*) reduced the latencies of target word recognition (see also Christianson et al. 2005, Duñabeitia et al. 2007 for Spanish and Basque). However, when orthographic primes (e.g., *wranel–warn*) were ma-

nipulated, no facilitation priming was observed highlighting the need for priming effects to keep the morpheme boundary intact. Then, a series of experiments compared of primes with Transposed Letters (TL) at the morpheme boundary (e.g., *speaekr–speak*) vs. outside the morpheme boundary (e.g., *spekaer–speak*). Only one experiment in the literature reported a benefit for TL primes when the transposition fell within the morpheme; no benefit was observed when the transposition fell across the morpheme boundary (Duñabeitia et al. 2007 , using Spanish materials). Subsequent investigations in both English and Spanish failed to replicate these findings (Beyersmann et al. 2012, 2013, Rueckl & Rimzhim 2011, Sánchez-Gutiérrez & Rastle 2013) and obtained equivalent facilitation when the transposed letters appeared within a stem or across a morpheme boundary.

Because TL benefit is not affected by the position of the TL relative to the morpheme boundary, I consider this result as a strong challenge for any decompositionalist model. If morphologically complex stimuli are indeed systematically decomposed into morphemes before gaining the mental lexicon, the main predictions of such models is that when the morphemes boundary is disrupted, no priming effect is expected since the cognitive system cannot parse the item into potential morphemes.

Diependaele et al. (2013) furthermore investigated the TL effect by comparing semantically transparent vs. opaque complex primes. Their first experiment showed that relative to formal primes, both transparent and opaque primes induced positive priming (e.g., *banker–bank = corner–corn > scandal–scan*). However, when morphological primes with TL were used, the transparent ones produced priming while the opaque ones did not (e.g., *baneker–bank > corenr–corn = scandal–scan*). A second experiment manipulated derived nonword primes in order to examine the effect of lexicality on the TL effect. Materials were selected from Longtin & Meunier's 2005 study and the authors found, on the one hand, that relative to unrelated primes, both intact derived word primes and intact derived nonword primes facilitated target recognition equally (e.g., *garagiste–garage = garagité–garage > diversion–garage*). On the other hand, when comparable morphological primes with TL were manipulated, a different pattern of priming emerged: only derived primes induced priming (e.g., *garaigste–garage > garaigté–garage = diverison–garage*). According to the authors, these data are line with the predictions of their hybrid/dual route model of morphological processing (presented above in Figure 3) in which complex items are automatically parsed within two morphological levels: morpho-orthographically and morpho-semantically, reflecting two sources of morphemic activation in word recognition. Morphological complex words (e.g., *banker*) are actually supposed to be processed twice at both morphemic levels, and pseudo-complex words (e.g., *corner*) once at the morpho-orthographic level, letter transposition across the morpheme boundary should interfere more with morpho-orthographic than morpho-semantic processing. Accordingly, transparent words and nonwords with TL are supposed to resist letter transpositions thanks to the morpho-semantic activation while opaque words and nonwords with TL did not because the morphemic activation at the morpho-orthographic level would be skipped.

According to me, the dual route model and the way masked priming effects are interpreted in this study are far from being convincing. "The key prediction of this account

is that fast-acting effects of morphology are not only morpho-orthographic in nature, but also morpho-semantic, and most importantly, that these effects reflect two separate sources of morphemic activation in word recognition" (p. 989).

If genuine complex words benefit from two sources of activation (morpho-ortho-graphic and morpho-semantic) and pseudocomplex words from one only (morpho-or-thographic), words like *banker* should be more efficient primes than *corner*. Nevertheless, their results (experiment 1) and the ones obtained so far in the literature demonstrate on the contrary that prime-target pairs like *banker-bank* and *corner-corn* produce equivalent priming effects (cf. surface morphology effects, see Rastle & Davis 2008 for a review). When TL effects are considered, it has been shown that primes with TL at the mor-pheme boundary (e.g., *banekr–bank*) and within the stem (e.g., *bakner–bank*) both induce equivalent facilitation effects. If the morpho-orthographic level is much more sensitive to letter order than the morpho-semantic level is, then one should have observed greater priming effects when the morpheme boundary of the prime is intact (e.g., *bakner–bank*) because two sources of activation could operate while for jumbled morpheme boundary (e.g., *banekr–bank*) only one source is active. The results obtained so far did not show any difference between these two types of primes, neither in the present paper, nor in the literature. Moreover, Diependaele et al. (2013) found in their experiment 2 that TL let-ter derived primes (e.g., *banekr–bank*) produced faster reaction times than intact primes (e.g., *banker–bank*). This surprising result is also very problematic for a decompositional account since the letter recoding for the TL primes that is necessary to activate mor-phemic representations should have delayed lexical access, therefore reducing priming.

Word processing is also closely linked to input frequency. This factor that has been broadly studied in the psycholinguistic literature on word recognition showing a strong and very robust correlation between lexical frequency and recognition latencies: the higher the frequency, the shorter the reaction time (see Ellis 2002 for a review). Gener-ally, these experimental studies oppose derived or inflected words of comparable surface frequency, but crucially differing in their stem frequency (high vs. low). In this kind of study, when reaction times (RTs) were found to be a function of the stem frequency, this is considered as evidence of the fact that word recognition implies the activation of the stem. For example in Italian, Burani & Caramazza (1987) investigated derived suf-fixed forms (verbal roots combined with highly productive suffixes such as *-mento, -tore, -zione*) by opposing stimuli matched for whole word frequency, but differing in root fre-quency (experiment 1), to stimuli matched for root frequency but differing in whole word frequency (experiment 2). Their results indicated that reaction times were influenced by both root and whole word frequencies (faster RTs were obtained for items containing a high frequency root in experiment 1 and for higher whole word frequency items in experiment 2), the authors suggested that the access procedure crucially operates with both whole word and morpheme access units. Frequency effects have been observed also in French by Colé et al. (1989), who similarly considered derived words matched for sur-face frequency but differing in their cumulative root frequency (e.g., *jardinier 'gardener'*, containing a high frequency root, vs. *policier 'policeman'*, containing a low frequency root). Since a clear cumulative root effect was observed only for suffixed words but not

for prefixed ones, Colé and colleagues suggest that only the former are accessed through decomposition via the root.

More recently, Burani & Thornton (2003) conducted a study on the interplay between the frequency of the root, the frequency of the suffix and the whole word frequency in processing Italian derived words. More precisely, in experiment 3, they considered low frequency suffixed words that differed with respect to the frequency of their morphemic constituents. As expected, the results showed that lexical decisions were faster and more accurate when the derived words included two high-frequency constituents (e.g., *pensatore* 'thinker') and slowest and least accurate when both constituents had low frequency (e.g., *luridume* 'filth'). Interestingly, when the derived words included only one high-frequency constituent (either the root or the suffix), the lexical decision rate was found to be a function of the frequency of the root only, irrespective of suffix frequency. The authors conclude that access through activation of morphemes is beneficial only for derived words with high frequency roots, while lexical decision latencies to suffixed derived words are a function of their surface frequency when they contain a low frequency root.

To sum up, frequency effects have been considered as a diagnostic for determining whether an inflected or derived form is recognized through a decompositional process that segments a word into its morphological constituents or through a direct look-up of a whole word representation stored in lexical memory. Frequency has therefore played a crucial role in the debate which opposed full parsing models, which assume a prelexical treatment of the morphological constituents with a consequent systematic and compulsory segmentation of all complex words (Taft & Forster 1975, Taft 1979), and full listing models, which defend a non-prelexical processing of the morphological structure and a complete representation of all morphologically complex words (see McClelland & Rumelhart 1981).

Despite the importance of the frequency for lexical access (the more frequent a word, the faster its recognition, see Solomon & Postman 1952) and the number of priming studies focused on its impact for word recognition (see Kinoshita 2006 for a review), very few studies manipulated frequencies using masked morphological priming. In a paradigm such as masked priming in which the prime is presented for a very brief duration, frequency is nevertheless a crucial factor since it determines the access speed to lexical representations. Moreover, clear opposite predictions can be derived for the two main approaches of morphological processing. According to the decompositional approach, only the root/stem frequency should interact with morphological priming effects since complex words are supposed to be accessed via the activation of their stem. The holistic hypothesis predicts no stem frequency effect but that surface frequency strongly determines masked morphological priming effects because lexical access takes place on the whole word. Giraudo & Grainger (2000) investigated the interaction of both frequencies with morphological processing through a series of masked priming experiments conducted in French. They manipulated the surface frequencies of derivatives used as primes for the same target (high frequency primes like *amitié–ami* 'friendship'–'friend'; low frequency primes like *amiable–ami* 'friendly'–'friend'). They found an interaction

between priming effects and the prime surface frequency (experiment 1), but no effect for the base frequency. Experiments 1 and 3 demonstrated that the surface frequency of morphological primes affects the size of morphological priming: high surface frequency derived primes showed significant facilitation relative to orthographic control primes (e.g., *amidon–ami* 'starch'–'friend'), whereas low frequency primes did not. The results of experiment 4 revealed, conversely, that cumulative root frequency does not influence the size of morphological priming on free root targets. Suffixed word primes facilitated the processing of free root targets with low and high cumulative frequencies. These data suggest that during the early processes of visual word recognition, words are accessed via their whole form (as reflected by surface frequency effects) and not via decomposition (since the base frequency did not interact with priming).

Another piece of evidence against the decompositional hypothesis comes from the study conducted by Giraudo & Orihuela (2015), which considered the effects of the relative frequencies of complex primes and their base target opposing the configuration with high frequency primes/low frequency targets to the configuration with low frequency primes/high frequency targets in French. Their results revealed that, relative to both the orthographic and unrelated conditions, morphological priming effects emerged only when the surface frequency of the primes is higher than the surface frequency of the targets (see also Voga & Giraudo 2009 for a similar conclusion). Again, these data contradict the prediction of the classical decomposition hypothesis, according to which the reverse effects would be expected.

The interpretation of frequency effects with respect to psycholinguistic models, however, remains very controversial. McCormick, Brysbaert, et al. (2009) defend a completely opposite position, in favour of an obligatory decomposition of all kinds of stimuli (even for the non-morphologically structured ones). They carried out a masked priming experiment manipulating the frequency of the primes, thus comparing high frequency, low frequency and nonword primes. Their hypothesis was that if morphological decomposition was limited to unfamiliar words, as predicted by the horse-race style of dual-route models, then priming should be limited to the last two conditions. On the contrary, if morphological decomposition was routine, an obligatory process applying to all morphologically structured stimuli should occur in all three conditions. The results showed that the priming effect observed with high frequency primes was equivalent to the one observed with low frequency primes and with nonword primes. Such findings seem to confirm the claim that a segmentation process is not restricted to low frequency words or nonwords, as assumed by horse-race models.

Very recently, the masked priming study carried by Giraudo et al. (2016) on Italian materials explored the role stem frequency in morphological processing even more deeply. They focused on the surface frequencies of base targets (comparing high vs. low surface frequency targets, e.g., *trasfire* 'to transfer' vs. *motivare* 'to motivate') primed by either the same base (e.g., *trasfire–trasfire*), a derivation of the base (e.g., *trasferimento–trasfire* 'transfer'–'to transfer'), an orthographic control (e.g., *trasparenza–trasfire* 'transparence' –'to transfer') and an unrelated control (e.g., *sacrificio–trasfire* 'sacrifice'–'to transfer). The data showed that full morphological priming effects were obtained whatever the fre-

quency of the targets (high or low). Accordingly, the frequency of the base contained in the derived primes (e.g., *trasferire* in *trasferimento*) did not interfere with morphological facilitation: primes whose base had a high frequency did not induce stronger facilitation than primes with a low frequency base. As a consequence, contrary to the predictions of a decompositional approach of lexical access to complex words, the prior presentation of a complex prime whose stem had a high surface frequency did not accelerate the access to its lexical representation relative to primes whose stem frequency was low.

Taken together, the frequency effects obtained using masked priming suggest that lexical access depends much more on the lexical frequency of the prime (that determines its activation threshold) than on its the stem frequency. Stem frequency does not seem to interfere with the access to the mental lexicon and morphological priming effects reveal instead that, as soon as a lexical representation is activated within the mental lexicon, such a representation automatically triggers the activation of all its family members. The result of the overall activation of the morphological family is revealed in those LDT experiments in which it has been observed that both the lexical and the base frequencies determine the recognition latencies of suffixed words. Only models that consider the word as the main unit of analysis, be it morphological (e.g., Giraudo & Voga 2014) or not (e.g., Baayen et al. 2011), are able to account for these findings.

Finally regularity is another factor from which opposite predictions can be drawn by the two views of morphological processing. In the psycholinguistic literature, this issue is intimately linked with the ease with which a complex word can be segmented into morphemes. Most of these studies consider morphology under the single angle of the word internal structure and the reported experiments carried out with irregular words aimed to test the predictions of decomposition hypothesis according to which parsability should interact with the magnitude of morphological priming effects. Regularity has been mainly tested with irregular materials like the irregular verbs in English (e.g., *bought–buy*) and with complex words containing various orthographic alterations (e.g., *bigger–big*). Pastizzo & Feldman (2002) carried a series of masked priming experiments on English irregular verbs (viz. allomorphs). They found that allomorphs (e.g., *fell*) whose construction enables decomposition, primed their verbal base (e.g., *fall*) more than orthographically matched (e.g., *fill*) and unrelated control words (e.g., *hope*) did. Contrary to the predictions of the decompositional hypothesis, non-segmentable complex words then induce priming effects that cannot be attributed to the formal overlap between prime–target pairs but depend on the morphological relationships they share. These results have been replicated later by Crepaldi et al. (2010; see also the MEG study carried by Fruchter et al. 2013 leading to the same pattern of data) who were forced to admit the "existence of a second higher-level source of masked morphological priming" and proposed a lemma-level composed of inflected words acting "at an interface between the orthographic lexicon and the semantic system" (p. 949).

McCormick et al. (2008) manipulated another category of derived stimuli that cannot be segmented perfectly into their morphemic components (for example, missing 'e' (e.g., *adorable–adore*), shared 'e' (e.g., *lover–love*), and duplicated consonant (e.g., *dropper–drop*) in order to test the flexibility of the morpho-orthographic segmentation

process described by morpheme-based models. Once again, their results demonstrate the robustness of this segmentation process in the case of various orthographic alterations in semantically related (e.g., *adorable–adore*) as well as in unrelated prime–target pairs (e.g., *fetish–fete*). The same authors then addressed the same question using morphologically structured nonword primes (McCormick et al. 2009). To this end, they created nonword primes with a missing <e> at the morpheme boundary (e.g., *adorage-adore*) and compared it to orthographically related prime-target pairs (e.g., *blunana-blunt*). The observed data showed that morphologically structured nonword primes facilitated the recognition of their stem targets, and that the magnitude of these priming effects was significantly larger than for orthographic control pairs. They interpreted this result as supporting their previous conclusions on word primes (2008) according to which stems that regularly lose their final <e> may be represented in an underspecified manner (i.e., absent or marked as optional). But far to call the decomposition mechanism into question, they claimed that the process of morphological decomposition was robust to regular orthographic alterations that occur in morphologically complex words.

According to me, these results could be interpreted on the contrary as being totally incompatible with the hypothesis of a mandatory decomposition process based on the surface morphology because this mechanism is only based on a minimalist condition of having two surface morphemes. If not, the decompositionalist approach needs to explain to how/on which criteria these words are actually decomposed. So far, the decompositionalists only proposed the idea of fast acting morphological effects (see Diependaele et al. 2013) without specifying on what visual/perceptual base these effects could actually operate. Recently, Giraudo & Dal Maso (2016) discussed this issue through the notion of morphological salience and its implications for theories and models of morphological processing. More precisely, the impact of the salience of complex words and their constituent parts on lexical access was questioned in light of the benchmark effects reported in the literature and the way they have been unilaterally interpreted. The issue of the relative prominence of the whole word and its morphological components has been indeed overshadowed by the fact that psycholinguistic research has progressively focused on purely formal and superficial features of words, drawing researchers' attention away from what morphology really is: systematic mappings between form and meaning. While I do not deny that formal features can play a role in word processing, an account of the general mechanisms of lexical access also needs to consider the perceptual and functional salience of lexical and morphological items. Consequently, the existence of morphemes is then recognized, but we claimed that it corresponds to secondary and derivative units of description. I hold that results obtained on the basis of masked priming are in line with holistic models of lexical architecture in which morphology emerges from the systematic overlap between forms and meanings (Baayen et al. 2011)[3] and for which the lexeme is the first unit analysis for the cognitive system. In such models, salience is not only a matter of internal structure, but also results from the organization of words in morphological families and series. As a consequence, not only syntagmatic,

---

[3]And also to abstractive approaches assuming that "the lexicon consists in the main of full forms, from which recurrent parts are abstracted" (Blevins 2006: 537).

but also paradigmatic relationships contribute to morphological salience. Certainly, the notion of salience refers primarily to formal aspects, because the perceptual body of the morpheme is necessarily the starting point of the processing mechanism. However, the notion of salience makes sense for complex word processing only if the form it refers to is associated with a meaning or function. Salience, in other words, is a property of the morpheme (i.e., a stable association of form and meaning), not simply of a phonetic or graphemic chain.

5 The final sound [o] in French

Focusing on salience from a mere formal point of view leads to consider how a decompositional hypothesis could deal with some phonological endings whose graphemic transcriptions are various.

I present a distributional study of the final sound [o] in French suggesting that paradigmatic relationships are more suitable to guide morphological processing than morphological parsing. The data have selected from Lexique 3 database (New 2006).

In French, the final sound [o] can be written in 9 different ways:

(1) *-au* as in:
     noyau, préau,   tuyau, bestiau
     'core', 'courtyard', 'pipe', 'cattle'

(2) *-aud* as in:
     noiraud,   rougeaud, crapaud, nigaud
     'black+aud', 'red+aud', 'toad',   'idiot'

(3) *-aut*, as in:
     quartaut
     'quarter+aut

(4) *-eau* as in:
     poireau, grumeau, tableau, drapeau
     'leek',   'lump',   'board', 'flag'

(5) *-od* as in:
     pernod
     'pernod'

(6) *-op* as in:
     galop,   sirop,   trop
     'gallop', 'syrup', 'too much'

(7) *-os* as in:
     gros, dos,   enclos, chaos
     'big', 'back', 'pen',   'chaos'

(8) *-ot* as in:
     bistrot, cachot,   chiot,   jeunot
     'pub', 'dungeon', 'puppy', 'youngster'

(9) *-o* as in:
     auto, ado,   mécano,   fluo
     'car', 'teenager', 'mechanic', 'fluo'

Among these words, one can distinguish semantically transparent complex words (e.g., *drap-eau*) M+, semantically opaque complex words (e.g., *crap-aud*) M−, simple words (e.g., *trop*) O and clippings (e.g., *ado* from *adolescent*) C, whose distributions in terms of size, i.e., number of different words sharing the same ending (N) and cumulative frequencies of these words (F) are sometimes very heterogeneous. Tables 1 and 2 present these different distributions.

Table 1: Number of different words having the same ending.

Ending	Distribution of the Size				
	Transparent complex words (M+)	Opaque complex words (M−)	Simple words (O)	Clippings (C)	Total N (M+, M−, O)
-au	2	3	13	5	18
-aud	20	15	11	35	46
-aut	0	1	22	1	23
-eau	74	47	74	121	195
-od	0	0	1	0	1
-op	0	0	4	0	4
-os	0	0	179	0	0
-ot	43	46	130	89	221
-o	18	8	430	26	581
Total	157	120	864	277	1089

As one can see above, among the 9 possible transcriptions of the sound [o], 6 can correspond to suffixes (i.e., *-au, -aud, -aut, -eau, -ot, -o*). It means that 66% of these endings can correspond to a suffix. Moreover endings in [o] are globally carried by a larger number of simple words (864 for O vs. 277 for M), and these simple words are much more frequent than complex words (13280 occ./million for O vs. 870 occ./million for M).

If we examine the size distributions of the different transcriptions, it appears that *-o* represents more than a half of the overall endings (581 words in *-o* for a total of 1089 words ending in [o]). The ending *-eau* dominates among the other endings (121/277 = .44) and only *-eau* (121 complex words for 74 simple words) and *-aud* (35 complex words for 11 simple words) show a morphological probability higher than an orthographic probability (p(M-*eau*) = 121/195 = .62; p(M-*aud*) = 35/46 = .76). All the other endings are dominated by

Table 2: Cumulative frequencies of words having the same ending.

Ending	Distribution of the cumulative frequency				
	Transparent complex words (M+)	Opaque complex words (M−)	Simple words (O)	Clippings (C)	Total N (M+, M−, O)
-au	6.55	34.86	5350.20	41.41	5391.61
-aud	2050.73	67.31	184.53	108.04	292.57
-aut	0	0	2009.41	0.20	2009.61
-eau	169.35	300.23	1559.39	427.10	1986.49
-od	0	0	4.73	0	4.73
-op	0	0	868.94	0	868.94
-os	0	0	1596.89	0	1596.89
-ot	1.80	4.05	1002.49	263.69	1493.18
-o	1.17	1.05	703.47	29.48	1037.39
Total	2229.60	407.50	13280.05	869.92	14681.41

simple words. This means that even if 66% of [o] endings can function as suffixes, their morphological probability is very low (p(M) = 227/1084 = .21). Therefore, morphological decomposition would conduct to a procedural deadlock in 81% of the cases. Finally, when the N distributions of M+ words are compared to M− words, we can see that M+ globally dominates M− (157 vs. 120) but when each ending is examined it appears that except for *-eau* (74 vs. 47) and *-o* (18 vs. 8) it is more a 50/50 ratio than a clear dominance. It suggests than even when the cognitive system encounters a complex word, morphological decomposition is semantically useless in 50% of the cases.

If one turns now to the details of frequency distributions, the cumulative frequencies of simple words are systematically higher than those of complex words, the highest value being associated with simple words ending in *-au* (5350 occurrences per million). As for the N distributions, the cumulated frequencies of the suffixed words ending in *-eau* dominates the other suffixed words (427 occ./million for a total of 870 occ./million). M+ words are much more frequent than M− words (2230 occ./million vs. 407 occ./million) but this dominance is explained by the cumulated frequencies of M+ suffixed words in *-aud* (2051 occ./million). When the data of *-aud* are removed, the cumulated frequency of M− words (340 occ./million) becomes almost twice as high as the one of M+ words (179 occ./million). Altogether, this suggests that simple words and semantically opaque complex words ending in [o] should be accessed more rapidly than the semantically transparent complex ones.

To sum up, the reported study of the 9 possible transcriptions of [o] according to the size and the cumulative frequency reveals that the probability for this phonological ending to correspond to a suffix is low. More importantly, the cumulative frequency of

suffixed words bearing a semantically transparent construction is weak relative to the non-suffixed words. Consequently, a decomposition hypothesis according to which any item bearing a structured morphological surface is first decomposed into morphemic constituents would lead to numerous useless prelexical mechanisms.

6 Something is rotten in the state of the decomposition hypothesis

In the present paper, I reviewed results from masked morphological priming reported in the literature and I highlight the shortcuts made by the decompositionalist to interpret some data, in particular those related to formal effects, forgetting the semantic and the paradigmatic aspects of morphology. Although I do not deny that morphology plays a role during lexical access, I doubt that fast morphological effect can operate under masked priming conditions (i.e., within a window of a 50–60 ms). In addition, I propose an alternative interpretation of its role within the mental lexical

Recently, Giraudo & Voga (2014) proposed a revised version of the supralexical model. This new model is sensitive to both lexical (e.g. frequency) and exo-lexical characteristics of the stimuli (e.g., family size) and capable to cope with various effects induced by true morphological relatives (e.g., *singer–sing*) and pseudo relatives (e.g., *corner–corn*). According to the model, morphological relationships are coded according to two different dimensions: syntagmatic and paradigmatic. The first level captures the perceptive regularity and the salience of morphemes within the language. It contains stems and affixes that have been extracted during word acquisition. Accordingly, during language acquisition, the most salient perceptive units (i.e., recurrent and regular) will be caught and coded by the cognitive system as lexical entries. At this very early level of processing, morphologically complex words, pseudo-derived words and nonwords whose surface structure can be divided into (at least two) distinct morphemes are equally processed. As a consequence, this level cannot properly be considered to be a morphological level, but rather as a level containing morcemes (from French *morceau* 'piece'). Morcemes correspond to word pieces standing as access units that speed up word identification each time an input stimulus activates one of them. Therefore, there is no need to assume, at this stage, a process of morphological decomposition; this would be unnecessary.

Contrary to the first level, the second level deals with the internal structure of words, their formation according to morphological rules. This level contains base lexemes, units abstract enough to tolerate orthographic and phonological variations produced by the processes of derivation and inflection. Base lexeme representations are connected to morphologically related word representations and these connections are determined by the degree of semantic transparency between wordforms and base lexemes. Semantically transparent morphologically complex words are connected both with their morphemes and their base lexeme. Words with a semantically opaque structure, as for example, *fauvette* 'warbler' (not related anymore to its free-standing stem *fauve* 'tawny') or with an illusory structure, as for example *baguette* 'stick' in which *bagu-* is not a stem and has

nothing to do with *bague* 'ring', are not connected with a base lexeme. These two types of items are only connected with their surface morphemes situated at the morceme level. Indeed, the model makes the fundamental assumption that base lexeme representations are created in long-term memory according to a rule that poses family clustering as an organisational principle of the mental lexicon. This rule stipulates that as soon as two words share form and meaning, a common abstract representation emerges; all the incoming forms respecting this principle then feed this representation. In the course of language acquisition and learning, family size grows and links are continually being strengthened.

Finally, if we turn back to priming effects, the model postulates that it depends on the kind of relationships the prime entertains with the target (formal and/or semantic) and consequently, on the number of excitation sources that target recognition triggers: a) when the prime is semantically transparent and complex M+O+S+ (like in the pairs *banker–bank* or *hatched–hat*), its perception gives birth to three sources of excitation, from morcemes, wordforms and base lexemes; b) when the prime is semantically transparent, complex but not decomposable M+O–S+ (like in the prime target pair *fell–fall*), it activates two sources of excitation, from wordforms and base lexemes; c) when the prime is semantically opaque M+O+S– (it concerns complex or pseudo-complex words like *apartment–apart* or *corner–corn*), its recognition triggers two sources of excitation, from morcemes and wordforms; d) when the prime is not complex and not decomposable MO–S– (like *freeze–free*), it gives raise to only one source of excitation, from wordforms.

In our view, much work still needs to be done on morphological processing, but within the framework of a lexical network that codes word representations as the result of both syntagmatic and the paradigmatic influences. Separating form from meaning, words from their family and series within experimental paradigms like the masked priming paradigm that exclusively focuses the attention of the readers on visual formal aspects, leads to a confirmation bias and reduces the notion of morphology to form only. It is indeed very important to consider that masked priming effects do not only correspond to the early processes of lexical access as suggested by numerous authors, but to a picture of lexical access that takes place at a given time within an ocean of complex relationships.

References

Amenta, Simona & Davide Crepaldi. 2012. Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology* 3(232).

Andoni Dunabeitia, Jon, Manuel Perea & Manuel Carreiras. 2008. Does *darkness* lead to *happiness*? masked suffix priming effects. *Language and Cognitive Processes* 23(7-8). 1002–1020.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes.* Cambridge: MIT Press.

Baayen, R. Harald, Petar Milin, Dušica Filipović Đurđević, Peter Hendriks & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118. 438–481.

Beyersmann, Elisabeth, Max Coltheart & Anne Castles. 2012. Parallel processing of whole words and morphemes in visual word recognition. *The Quarterly Journal of Experimental Psychology* 65(9). 1798–1819.

Beyersmann, Elisabeth, Samantha F. McCormick & Kathleen Rastle. 2013. Letter transpositions within morphemes and across morpheme boundaries. *The Quarterly Journal of Experimental Psychology* 66(12). 2389–2410.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.

Burani, Cristina & Alfonso Caramazza. 1987. Representation and processing of derived words. *Language and Cognitive Processes* 2(3-4). 217–227.

Burani, Cristina & Anna M. Thornton. 2003. The interplay of root, suffix and whole-word frequency in processing derived words. *Trends in Linguistics Studies and Monographs* 151. 157–208.

Caramazza, Alfonso, Alessandro Laudanna & Cristina Romani. 1988. Lexical access and inflectional morphology. *Cognition* 28(3). 297–332.

Christianson, Kiel, Rebecca L. Johnson & Keith Rayner. 2005. Letter transpositions within and across morphemes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(6). 1327.

Colé, Pascale, Cécile Beauvillain & Juan Segui. 1989. On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and language* 28(1). 1–13.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique.* Tübingen: Max Niemeyer Verlag.

Crepaldi, Davide, Lara Hemsworth, Colin J. Davis & Kathleen Rastle. 2016. Masked suffix priming and morpheme positional constraints. *The Quarterly Journal of Experimental Psychology* 69(1). 113–128.

Crepaldi, Davide, Kathleen Rastle & Colin J. Davis. 2010. Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition* 38(3). 312–321.

Crepaldi, Davide, Kathleen Rastle, Colin J. Davis & Stephen J. Lupker. 2013. Seeing stems everywhere: Position-independent identification of stem morphemes. *Journal of Experimental Psychology: Human Perception and Performance* 39(2). 510.

Diependaele, Kevin, Jonathan Grainger & Dominiek Sandra. 2012. Derivational morphology and skilled reading. In Michael Spivey, Ken McRae & Marc Joanisse (eds.), *Cambridge handbook of psycholinguistics*, 311–332. Cambridge: Cambridge University Press.

Diependaele, Kevin, Joanna Morris, Raphael M. Serota, Daisy Bertrand & Jonathan Grainger. 2013. Breaking boundaries: Letter transpositions and morphological processing. *Language and Cognitive Processes* 28(7). 988–1003.

Diependaele, Kevin, Dominiek Sandra & Jonathan Grainger. 2009. Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition* 37(6). 895–908.

Duñabeitia, Jon Andoni, Manuel Perea & Manuel Carreiras. 2007. Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition* 105(3). 691–703.

Ellis, Nick C. 2002. Reflections on frequency effects in language processing. *Studies in Second Language Acquisition* 24(2). 297–339.

Forster, Kenneth I. & Chris Davis. 1984. Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10(4). 680.

Fradin, Bernard. 1996. L'identification des unités lexicales. *Sémiotiques* 11. 55–93.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie.* Paris: Presses Universitaires de France.

Fruchter, Joseph, Linnaea Stockall & Alec Marantz. 2013. MEG masked priming evidence for form-based decomposition of irregular verbs. *Frontiers in Human Neuroscience* 7.

Giraudo, Hélène & Serena Dal Maso. 2016. The salience of complex words and their parts: Which comes first? *Frontiers in Psychology* 7.

Giraudo, Hélène, Serena Dal Maso & Sabrina Piccinin. 2016. The role of stem frequency in morphological processing. In *Mediterranean morphology meetings*, vol. 10, 64–72.

Giraudo, Hélène & Jonathan Grainger. 2000. Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes* 15(4-5). 421–444.

Giraudo, Hélène & Jonathan Grainger. 2001. Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review* 8(1). 127–131.

Giraudo, Hélène & Karla Orihuela. 2015. Visual word recognition of morphologically complex words: Effects of prime word and root frequency. In Vito Pirelli, Claudia Marzi & Marcello Ferro (eds.), *Word structure and word usage. Proceedings of the NetWordS final conference*, 128–131.

Giraudo, Hélène & Madeleine Voga. 2013. Prefix units in the mental lexicon. In Nabil Hathout, Fabio Montermini & Jesse Tseng (eds.), *Morphology in Toulouse. Selected Proceedings of Décembrettes 7 (Toulouse 2–3 December 2010)*, 91–107.

Giraudo, Hélène & Madeleine Voga. 2014. Measuring morphology: The tip of the iceberg? A retrospective on 10 years of morphological processing. In Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé & Fabio Montermini (eds.), *Proceedings of the 8th Décembrettes*, vol. 22 (Carnets de grammaire), 136–167. Bordeaux.

Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in honor of Sylvain Bromberger*, 111–176. Cambridge: MIT Press.

Halle, Morris & Alec Marantz. 1994. Some key features of Distributed Morphology. *MIT working papers in linguistics* 21(275). 88.

Hathout, Nabil. 2005. Exploiter la structure analogique du lexique construit: une approche computationnelle. *Cahiers de lexicologie: Revue internationale de lexicologie et lexicographie* (87). 5–28.

Hathout, Nabil. 2008. Acquisition of the morphological structure of the lexicon based on lexical similarity and formal analogy. In *Proceedings of the COLING workshop Textgraphs-3*, 1–8. Manchester.

Kinoshita, Sachiko. 2006. Additive and interactive effects of word frequency and masked repetition in the lexical decision task. *Psychonomic Bulletin & Review* 13(4). 668–673.

Kinoshita, Sachiko & Stephen J. Lupker. 2004. *Masked priming: The state of the art*. New York: Psychology Press.

Laudanna, Alessandro, Antonella Cermele & Alfonso Caramazza. 1997. Morpho-lexical representations in naming. *Language and Cognitive Processes* 12(1). 49–66.

Longtin, Catherine-Marie & Fanny Meunier. 2005. Morphological decomposition in early visual word processing. *Journal of Memory and Language* 53(1). 26–41.

McClelland, James L. & David E. Rumelhart. 1981. An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review* 88(5). 375.

McCormick, Samantha F., Marc Brysbaert & Kathleen Rastle. 2009. Is morphological decomposition limited to low-frequency words? *The Quarterly Journal of Experimental Psychology* 62(9). 1706–1715.

McCormick, Samantha F., Kathleen Rastle & Matthew H. Davis. 2008. Is there a 'fete'in 'fetish'? Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *Journal of Memory and Language* 58(2). 307–326.

McCormick, Samantha F., Kathleen Rastle & Matthew H. Davis. 2009. Adore-able not adorable? Orthographic underspecification studied with masked repetition priming. *European Journal of Cognitive Psychology* 21(6). 813–836.

Morris, Joanna, Jonathan Grainger & Phillip J Holcomb. 2013. Tracking the consequences of morpho-orthographic decomposition using ERPs. *Brain Research* 1529. 92–104.

New, Boris. 2006. Lexique 3: Une nouvelle base de données lexicales. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN)*, 892–900.

Pastizzo, Matthew J. & Laurie B. Feldman. 2002. Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(1). 244.

Rastle, Kathleen & Matthew H. Davis. 2008. Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes* 23(7-8). 942–971.

Rastle, Kathleen, Matthew H. Davis, William D. Marslen-Wilson & Lorraine K. Tyler. 2000. Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes* 15(4-5). 507–537.

Rayner, Keith, Sarah J. White, Rebecca L. Johnson & Simon P. Liversedge. 2006. Raeding wrods with jubmled lettres there is a cost. *Psychological Science* 17(3). 192–193.

Rueckl, Jay G. & Anurag Rimzhim. 2011. On the interaction of letter transpositions and morphemic boundaries. *Language and Cognitive Processes* 26(4-6). 482–508.

Sánchez-Gutiérrez, Claudia & Kathleen Rastle. 2013. Letter transpositions within and across morphemic boundaries: Is there a cross-language difference? *Psychonomic Bulletin & Review* 20(5). 988–996.

Solomon, Richard L. & Leo Postman. 1952. Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology* 43(3). 195.

Taft, Marcus. 1979. Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior* 18(1). 21–39.

Taft, Marcus. 1994. Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes* 9(3). 271–294.

Taft, Marcus & Kenneth I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14(6). 638–647.

Taft, Marcus & Kenneth I. Forster. 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15(6). 607–620.

Voga, Madeleine & Hélène Giraudo. 2009. Pseudo-family size influences the processing of French inflections: Evidence in favor of a supralexical account. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 148–155. Somerville: Cascadilla Press.

Chapter 17

Les affixes dérivationnels ont-ils des allomorphes ? Pour une modélisation de la variation des exposants dans une morphologie à contraintes

Fabio Montermini

CLLE-ERSS, CNRS & Université de Toulouse 2 Jean Jaurès

Cet article traite des phénomènes de variation formelle en dérivation (écart entre la forme attendue et la forme réellement observée pour un lexème dérivé) qui ne peuvent pas être traités en termes de variation thématique, ce qui suggère que les exposants des constructions morphologiques peuvent à leur tour être sujets à variation. Pour modéliser cette variation des exposants, je propose d'étendre la notion de contrainte non seulement à une propriété qui est spécifique à une langue donnée, mais également à une construction donnée. Les exposants des constructions morphologiques sont alors eux-mêmes vus comme des (ensembles de) contraintes qui interagissent avec les autres contraintes en jeu dans la formation des lexèmes complexes. Chaque « allomorphe » d'un exposant est donc représenté comme une contrainte qui, en tant que telle, peut être hiérarchisée par rapport aux autres, ce qui rend compte de l'observation que certaines de ces variantes jouent un rôle de défaut, alors que d'autres émergent uniquement dans des conditions particulières. Afin d'illustrer ce modèle, je propose deux études de cas de constructions morphologiques de naissance ou développement récent. Il s'agit, d'une part, de la création de noms de locuteurs en *-phone* à partir du nom d'une langue et la création de lexèmes avec un sens génériquement appréciatif / superlatif en *-issimo*. Chacune de ces deux constructions est à son tour comparée à des constructions proches: la dérivation en *-phone* est comparée à la dérivation correspondante et cognate de lexèmes en *-fono* en italien ; la dérivation en *-issimo* est comparée à la dérivation, plus canonique, de superlatifs en *-issime* en français. Ces comparaisons mettent en lumière le fait que des constructions formellement et sémantiquement similaires et qui ont la même origine peuvent, dans des langues différentes ou dans la même langue à des époques et pour des finalités différentes, développer des spécifications phonologiques différentes, ce qui se traduit, dans le cadre adopté ici, par des ensembles de contraintes différentes et/ou agencées différemment.

1 Introduction

Un des changements majeurs qu'a connus l'étude de la morphologie dans les dernières décennies a été le glissement des modèles morphématiques, décompositionnels et combinatoires vers des modèles davantage tournés vers la description des relations existantes entre des mots plus ou moins complexes. Une des conséquences de ce changement est le fait que ces relations ne sont plus analysées en termes de règles orientées, déterministes et existant indépendamment des unités qui les incarnent, mais en ayant recours à des concepts comme celui de « patron » ou « schéma », plus souples, et qui rendent compte de la manière dont les locuteurs établissent des généralisations à partir du lexique existant. C'est ce que l'on observe, par exemple, dans la Morphologie des Constructions (Construction Morphology), élaborée principalement par Booij (2010), mais aussi dans le modèle à contraintes, élaboré par Hathout (2009) et surtout dans les travaux récents de Marc Plénat et Michel Roché (Plénat & Roché 2014, Roché & Plénat 2014, 2016). Toutes ces approches sont « output-oriented », au sens qu'elles sont moins intéressées à décrire l'ensemble de procédures qui permettent de passer d'un input à un output (un lexème (plus) complexe) qu'à rendre compte des contraintes qui pèsent sur la forme (et le sens) d'un lexème construit, ou, plus précisément, de tous les lexèmes construits qui appartiennent à la même série (c'est à dire, qui sont construits par la même opération morphologique). Parmi d'autres résultats, les approches en question ont permis de rendre compte de manière efficace de la variation allomorphique observée dans le lexique construit, en particulier en ce qui concerne la sélection du thème du lexème de base et les éventuelles modifications qu'il subit. En revanche, à quelques exceptions près (notamment Lignon & Roché 2011), la variation de forme des exposants (celle qui est appelée traditionnellement l'allomorphie affixale) a été peu discutée dans ce cadre. Une des raisons principales est certainement le fait que les approches dont il est question ci-dessus ont le plus souvent pris le parti de maximiser la complexité des représentations lexicales en simplifiant, parallèlement, l'instruction phonologique associée aux opérations morphologiques, et donc de repousser, autant que possible, l'allomorphie du côté des radicaux plutôt que du côté des affixes (Bonami et al. 2009, par exemple, sont très clairs sur ce point). Pourtant, le fait que l'allomorphie puisse toucher aussi bien les radicaux des mots construits que les affixes semble souvent aller de soi, en lexicographie, dans plusieurs cadres phonologiques (par exemple en Théorie de l'Optimalité), mais également pour la morphologie, que ce soit la morphématique traditionnelle (ce qui est normal, puisque dans ces cadres les radicaux et les affixes sont des objets de la même nature) ou la morphologie lexématique dite « classique ». Dans ce contexte, une position emblématique me semble être celle de Scalise (1999), qui, en traitant des noms déverbaux de l'italien, se demande « in *amministrazione* il suffisso sarà *-azione*, *-zione* o *-ione* ? » ('dans *amministrazione* le suffixe est-il *-azione*, *-zione* ou *-ione* ?'), en suggérant simultanément qu'il est possible (et intéressant) d'identifier une forme précise pour le suffixe dans le dérivé en question – et par conséquent d'établir une frontière nette entre le suffixe et le radical – et que celui-ci peut potentiellement se présenter sous de différentes formes.

Dans cet article je vais proposer, au contraire, qu'une question comme celle ci-dessus n'est pas une question pertinente et que, si l'on se place dans un cadre morphologique orienté vers les outputs et basé sur les contraintes, la séquence formelle qui correspond à l'exposant d'une opération morphologique résulte uniquement de l'application d'une contrainte qui, en tant que telle, interagit et peut entrer en compétition avec les autres qui pèsent sur la forme d'un mot construit. Si l'exposant d'une opération morphologique correspond lui-même à une contrainte, il n'y a plus aucune nécessité théorique à ce qu'il ait une forme définie et constante dans l'ensemble des dérivés dans lesquels il apparaît, y compris dans le cas par défaut. Au contraire, l'existence de plusieurs « allomorphes », par exemple pour un même affixe, est prévisible, et ceux-ci peuvent être hiérarchisés, puisque chacun d'entre eux permet la satisfaction d'un certain nombre de contraintes formelles, à leur tour potentiellement en concurrence. Plus généralement, j'adopte un cadre et un inventaire des contraintes qui, avec peu de modifications, sont ceux proposés par Plénat & Roché (2014) et Roché & Plénat (2014,2016). Il faut noter que le cadre dans lequel je me place, et la modélisation que je propose pour la variation des exposants des opérations morphologiques, est particulièrement adapté dans le cadre d'un modèle exemplairiste de la morphologie[1]. Les contraintes ne sont donc qu'un moyen de modéliser les préférences que les locuteurs manifestent dans leur activité de création morphologique ; de ce point de vue, intégrer aux contraintes des propriétés purement déclaratives comme la forme d'un affixe est parfaitement légitime et en ligne, je considère, avec les recherches citées, puisque cette propriété fait crucialement partie de celles que les locuteurs identifient dans les mots complexes existants et ont envie de reproduire dans ceux qu'ils construisent.

Le modèle que je propose constitue l'état actuel de réflexions sur la forme des mots complexes que je mène depuis plusieurs années, et que j'ai déjà exposées dans des publications antérieures. Si je remonte dans le temps, une des premières lectures qui m'ont poussé à réfléchir sur ce sujet est l'article de Fradin (2000) sur les mots-valises et ceux qu'il appelait « related phenomena »[2]. Cet article, qui propose une analyse et une classification d'un large spectre de constructions morphologiques qui se détachent de l'affixation canonique, contient, entre autres choses, des données comme celles en (1)[3], qui, en prenant comme modèle *pérestroïka*, désignent des réformes politico-économiques qui ont eu lieu, respectivement, en France, à Cuba et en Afrique du Sud, ainsi qu'un renouveau dans les mœurs sexuels dans l'ancienne URSS :

(1)   a.   *Béréstroïka ← (Pierre) Bérégovoy*

    b.   *Castroïka ← (Fidel) Castro*

    c.   *Prétoriastroïka ← Prétoria*

    d.   *Sextroïka*

---

[1] Par « exemplairiste », j'entends un modèle de la grammaire selon lequel les patrons (dans ce cas morphologiques) émergent dans la compétence des locuteurs à partir des lexèmes existants auxquels ils sont exposés (cf. Bybee 2006, 2013 ; Blevins & Blevins 2009 pour des aperçus récents).

[2] Article que j'ai lu avant sa parution, puisque je le citais – comme « à paraître » – dans mon mémoire de DEA de 1998.

[3] Les mêmes données sont reprises dans Fradin (2003 : 212–213).

Des données comme celles-ci sont clairement problématiques pour tout modèle qui essayerait d'appliquer mécaniquement un processus de combinaison de morphèmes. Une des formes, *Prétoriastroïka*, est clairement issue de la concaténation de deux éléments, mais les deux autres présentent différents degrés de fusion entre les éléments concernés. De plus, il semble y avoir une séquence phonologique ([stʁɔjka]) qui, en français est obligatoirement présente dans ces mots complexes, et de ce point de vue elle peut à juste titre être considérée comme l'« exposant » de la construction morphologique. Cependant, le lexème construit peut conserver une portion plus importante du matériel phonologique du mot-modèle (comme dans le cas de *Béréstroïka*), et la base peut être conservée dans sa totalité ou subir différents types de réajustements. Quelques-uns des mots de (1), notamment *Béréstroïka* et *Castroïka*, pourraient également être analysés comme des mots-valises, puisque le partage de matériel phonologique est souvent considéré comme un élément essentiel de ce type de formations (Fradin 2000 : 28-31). Cependant, dans l'article en question Fradin montre de manière convaincante, sur une base sémantique, que les formes de (1) sont bien des cas d'affixation (« sécrétive », puisque l'affixe provient de la réduction d'un lexème). À l'argument sémantique développé par Fradin on peut ajouter le fait que, à la différence des mots-valises, ces mots construisent une série, qui aurait certainement été plus importante, si les vicissitudes historiques n'avaient pas privé la pérestroïka d'une grande partie de son impact politique et médiatique, et donc réduit de manière cruciale la saillance du mot dans la conscience linguistique des locuteurs. Une notion comme celle de série dérivationnelle, qui est aujourd'hui considérée comme un élément fondamental de l'organisation morphologique du lexique, ne faisait pas partie, à la fin des années 1990, des outils théoriques disponibles. Si les mots de (1) sont bien le résultat d'un processus d'affixation, une manière relativement simple de représenter l'exposant de cette construction morphologique est d'établir une contrainte qui veut que le dérivé se termine par la séquence phonologique [stʁɔjka], qui peut être simplement agglutinée à une base (*Prétoriastroïka*), mais qui peut aussi partager des segments avec celle-ci (*Castroïka*). En plus de proposer une proposition de classification des procédés morphologiques non canoniques fondée sur une analyse très fine des propriétés formelles et sémantiques des éléments en question et sur des critères solides, l'article en question, à mon sens, a joué un rôle important sur un autre plan, à savoir l'identification des formations « mineures », marginales, apparemment étrangères au « noyau » de la langue, comme des objets légitimes non seulement pour la lexicologie ou la lexicographie, mais aussi pour une approche formelle du langage, et en particulier de la morphologie. Dans les années qui ont suivi, la prise en compte de tous les types de données, en particulier des données créées spontanément par les locuteurs dans des situations non contrôlées, est devenue une pratique consolidée, et leur intérêt théorique pour l'étude de la morphologie, surtout dérivationnelle, est admis. Ce développement est allé de pair avec l'expansion et la diffusion des ressources linguistiques, et donc l'élargissement progressif des bases de données lexicales disponibles[4]. Dans ce contexte, et à une époque où les données de morphologues étaient encore pour la plupart puisées aux sources « tra-

---

[4]La liste des travaux qui, surtout en France, ont adopté cette approche extensive à la morphologie, et des avancées théoriques qu'elle a rendues possibles irait certainement au-delà des finalités de cet article. Je me limite donc à citer quelques travaux qui proposent plutôt une réflexion métathéorique sur le processus en cours et ses conséquences, par exemple Hathout et al. (2008) ; Hathout et al. (2009) ; Dal & Namer (2012, 2016).

ditionnelles », Bernard Fradin (avec d'autres) a été un des premiers à voir l'importance des données « marginales » et à les exploiter pour nourrir la réflexion théorique. Cet article s'inscrit dans le même mouvement de morphologie extensive fondée sur l'usage. En particulier, je m'appuierai, pour justifier le modèle de l'allomorphie affixale que je propose, sur deux études de cas de procédés morphologiques du français de naissance ou de développement récents, pour lesquels les locuteurs ne disposent ni d'indications métalinguistiques (intégrées plus ou moins consciemment) sur leur fonctionnement, ni d'un nombre important de lexèmes qui font partie du lexique établi et qui peuvent servir de modèles dans la création de nouveaux mots. Il s'agit, comme on le verra, de procédés qui sont partiellement en structuration, et pour lesquels les choix des locuteurs ne sont pas toujours univoques, puisque ceux-ci peuvent se fonder, dans la création lexicale, sur plusieurs indices, en attribuant un poids différent à chacun d'entre eux. Le premier phénomène que je vais regarder est la construction de noms (ou adjectifs) qui désignent les locuteurs d'une langue et qui sont construits au moyen de l'élément *-phone* (*francophone*, *occitanophone*, *quechuaphone* / *quechuophone*, *wolophone*), que je compare aux noms correspondants en italien (*francofono*, *occitanofono*, *quechuofono*, *wolofono*) (Section 3). Le deuxième est la construction de noms ou adjectifs (souvent, mais pas exclusivement, des noms commerciaux) au moyen du suffixe *-(i)ssimo* (*Colissimo*, *Doctissimo*, *Tassimo*, *Vernissimo*), que je compare aux adjectifs (et noms) construits au moyen du suffixe, plus établi, *-issime* (Section 4). Avant ces études empiriques, cependant, je propose quelques observations sur la prise en compte de la variation des exposants des constructions morphologiques dans un modèle fondé sur les contraintes, et je montre que ce paramètre n'est pas différent, dans la substance, des autres contraintes formelles qui pèsent sur la forme des lexèmes construits (Section 2).

2   La variation des exposants dans un modèle morphologique à contraintes

Pour beaucoup de linguistes, que ce soit dans des cadres formels ou plus descriptifs, le fait que les exposants d'opérations morphologiques puissent être sujets à la variation formelle (ou, pour le dire plus simplement, l'existence de phénomènes d'allomorphie affixale) ne fait pas de doute. Ceci est même attendu dans des modèles qui n'établissent aucune distinction de nature entre les unités lexicales et les unités sublexicales (les affixes), si ce n'est dans leurs propriétés combinatoires et dans leur autonomie syntaxique. À titre d'exemple, les exposants des entrées consacrées par le *TLFi* aux suffixes qui construisent *aimable* et *amabilité* ont les formes, respectivement, « -able, -ible, -uble » et « -té, -eté, -ité ». De la même manière, dans son ouvrage qui a contribué à l'établissement de l'approche lexicaliste à la morphologie, Aronoff (1976 : 100), tout en reconnaissant que les affixes n'ont pas d'existence autonome en dehors des règles de construction de mots qui les introduisent, considère que le suffixe qui construit des noms d'action en anglais « has at least four, and possibly five, forms » : +*Ation*, +*ition*, +*ution*, +*ion*, +*tion*[5]. Dans

---

[5] « + » est le symbole utilisé par Aronoff pour indiquer un type de frontière morphologique.

de tels cas, on considère implicitement qu'un affixe, qu'il ait une existence indépendante de la règle qui l'introduit ou pas, doit pouvoir être représenté sous une forme discrète, et qu'il est donc toujours possible de tracer une frontière entre celui-ci et le radical du lexème de base, qui à son tour peut présenter ou pas une forme allomorphique. La variation phonologique observée – qui, on remarquera en passant, concerne toujours la partie censée être en contact avec la base – est parallèle à la variation allomorphique observée pour les lexèmes, et peut être traitée en faisant appel aux mêmes conditionnements morphophonologiques. Un développement récent de la morphologie basée sur les lexèmes a consisté à voir de plus en plus ces derniers comme des unités multiformes, mais structurées à leur intérieur, y compris du point de vue formel, une approche informellement nommée « morphologie thématique » (par exemple par Plénat 2008a, se référant à des travaux précédents, comme ceux de Bonami & Boyé 2003). Dans ce cadre, l'allomorphie, synchroniquement irréductible, observée pour certains lexèmes est admise comme une propriété intrinsèque de ceux-ci, encodée de façon explicite dans leur représentation lexicale. Le pendant de cet élargissement de la quantité d'information mémorisée par les locuteurs est une forte simplification des procédures morphologiques. En d'autres termes, la plus grande partie de la variation observée – et donc la plus grande complexité – est transférée du côté des bases (thèmes ou radicaux), avec une simplification des opérations morphologiques (flexionnelles ou dérivationnelles), et par conséquent de leurs exposants, qui sont, autant que possible, considérés comme uniques. L'article de Bonami et al. (2009) est un des cas dans lesquels cette approche a été illustrée de manière la plus claire et convaincante. Dans la proposition de Bonami et collègues, le suffixe qui construit des noms d'action déverbaux en français possède une forme constante ([jɔ̃]), et la variation observée est à attribuer au thème verbal sélectionné par la règle de construction de lexèmes, un thème qui peut être soit identique à un des thèmes flexionnels du verbe (*dispersion*), soit autonome (*modification*, *réduction*). Comme je l'ai observé dans l'introduction, l'attention de la plupart de travaux réalisés dans le cadre de la morphologie thématique a tout naturellement porté sur la variation formelle des bases des processus de dérivation, en s'intéressant soit à la sélection du thème et aux modifications éventuelles qu'il subit (Plénat 2008a, Roché 2010, Roché & Plénat 2014, Hathout & Namer 2014), soit aux cas de concurrence entre opérations (Lignon & Plénat 2009, Lignon 2013, Koehl & Lignon 2014, Roché & Plénat 2016, entre autres). À ma connaissance, un des rares travaux dans ce cadre à traiter explicitement la question de l'allomorphie affixale est l'article de Lignon & Roché (2011), qui, dans la construction des adjectifs de relation en français, identifient -*éen* et -*ien* comme « deux variantes d'un même suffixe -*ien* » (Lignon & Roché 2011 : 191). D'autres cas, y compris des cas traditionnellement identifiés comme relevant de phénomènes d'allomorphie affixale, sont en revanche traités de manière moins claire et univoque. Je montre, à titre d'exemple, deux cas tirés de la littérature récente sur le français, celui des semi-voyelles présuffixales dans certains dérivés (en particulier en -*eux*)[6], et celui du suffixe qui construit des noms de qualité comme *rareté* ou *amabilité*. Des formes comme *ambitieux*, *injurieux* ou *luxueux*, qui comportent

---

[6] L'étiquette de « semi-voyelle présuffixale » est inspirée de Thornton (1999), qui a consacré un article au même phénomène en italien.

une semi-voyelle ([j] ou [w]) à la jonction entre la base et l'affixe sont souvent regardées comme comportant une forme allomorphique du suffixe, dont la distribution peut être déterminée par des contraintes de type phonologique et/ou morphologique. Le *TLFi*, par exemple, liste *-ieux* et *-ueux* comme des variantes du suffixe *-eux*. Des traitements plus récents, cependant, tendent à traiter les séries de lexèmes se terminant en *-ieux* / *-ueux* soit comme des cas d'allomorphie radicale (celle-ci semble être la position exprimée par Bonami et al. 2009 : 104-105), ou bien, tout simplement, comme des sous-séries des lexèmes en *-eux* qui, puisqu'elles comportent de nombreux lexèmes (dont un grand nombre directement issu du latin) et qu'elles sont uniformes, tendent à s'enrichir encore plus (cf. Roché 2011 :86 ; Roché & Plénat 2014). Dans ce cas, l'identification de la semi-voyelle comme appartenant à un allomorphe du thème de base ou à une variante du suffixe perd une grande partie de son intérêt, puisque « [l]es divers processus qui tendent à enrichir la rime se confondent et s'interpénètrent » (Roché & Plénat 2014 : 1867). La situation est encore moins claire en ce qui concerne les noms désadjectivaux de qualité se terminant en [te]. Plénat et Roché semblent considérer *-ité* et *-(e)té* tantôt comme deux variantes du même suffixe (Roché 2011 : 80 ; Roché & Plénat 2012 : 1395) , tantôt comme deux suffixes liés (ne serait-ce que du point de vue diachronique) mais distincts (Plénat 2008a : 1617 ; Roché & Plénat 2014 : 1865, 1869), tandis que Koehl (2012 : 173) indique explicitement que « *-ité* et *-té* sont deux variantes allomorphiques d'un même suffixe noté *-Ité* ». Ces deux exemples, en soi anecdotiques mais tout de même significatifs, montrent, à mon sens, que la voie qu'a empruntée la morphologie thématique – se poser des questions différentes de « quelle est la frontière entre le radical et l'affixe dans le lexème construit X ? » – est la bonne, mais qu'elle ne s'est pas entièrement débarrassée de certains réflexes propres de la morphologie combinatoire classique (par exemple, identifier une forme discrète et si possible univoque pour un affixe). Dans ce qui suit, je voudrais contribuer à pousser davantage la morphologie sur la voie que j'ai évoquée, en développant, en particulier, trois points : i) toute la variation formelle observée en dérivation ne peut pas être uniquement attribuée à la variation thématique des bases ; il existe des cas où la variation ne peut clairement pas être attribuée à la sélection d'un thème particulier, mais relève de l'exposant ; ii) il est nécessaire de distinguer les cas dans lesquels un ensemble de lexèmes est issu de la même construction, qui présente une variation de l'exposant, des cas dans lesquels on a affaire à plusieurs ensembles de lexèmes issus de constructions différentes avec des exposants différents (qui peuvent, éventuellement, présenter une similarité formelle et/ou sémantique) ; iii) lorsqu'on a affaire à un ensemble de lexèmes issus de la même construction qui présente une variation de l'exposant, cette variation peut être décrite sous forme de contraintes hiérarchisées du même type que les autres contraintes qui pèsent sur la forme des mots construits. Aux deux premiers points est consacrée la section 2.1, au troisième la section 2.2.

2.1 La variation formelle des exposants

Comme je l'ai observé, la morphologie thématique a adopté, comme principe général, l'idée que la variation formelle rencontrée dans les mots complexes était plus avantageu-

sement traitée en termes de supplétion thématique plutôt que de variation de l'exposant. L'intérêt de ce mouvement se comprend facilement, en particulier lorsqu'on considère que ce modèle a été conçu d'abord pour traiter des phénomènes flexionnels (principalement dans les langues romanes) : l'hypothèse de l'allomorphie thématique est d'autant plus facile à maintenir que les formes fléchies présentent peu de variation dans leurs exposants (terminaisons), et dans la plupart des cas il s'agit d'allomorphies qui peuvent être ramenées à une variation de classe flexionnelle. En revanche, il existe un certain nombre de phénomènes de variation thématique qui ne peuvent être traités, synchroniquement, qu'en termes de supplétion[7]. Si postuler l'existence de supplétions thématiques, au moins à un certain degré, est donc nécessaire, il est plus économique d'alléger le dispositif de règles, en associant, autant que possible, une seule instruction formelle à chaque construction morphologique[8]. Ce modèle, toutefois, s'il est convaincant dans beaucoup de cas, ne permet pas de rendre compte de l'ensemble des variations observées. L'incertitude dont j'ai fait état ci-dessus concernant les suffixes (pour faire vite) *-eux* et *-ité* me paraît emblématique de ce fait. Il existe, en effet, de nombreux cas de dérivation pour lesquels l'hypothèse d'une variation de l'exposant est bien plus convaincante que l'hypothèse d'une supplétion thématique. Lignon & Roché (2011), par exemple, consacrent plusieurs pages à une démonstration très solide du fait que *-ien*, *-éen* et *-ain* (et même *-en*) sont autant d'« allomorphes » d'un exposant unique de construction morphologique qu'ils transcrivent -IEN. Une explication en termes de variation de l'exposant devrait être invoquée, me semble-t-il, également pour les cas de substitution de *-este* à *-esque* (*grandiloqueste*, *titaniqueste*) étudiés par Plénat, Tanguy et al. (2002). Le fait que dans ce dernier cas les deux variantes aient des origines différentes (le suffixe latin *-iscus* via l'italien dans un cas, et le suffixe *-estis* dans l'autre) importe peu en synchronie, si les deux variantes sont employées en distribution complémentaire sur la base de la forme phonologique de la base, comme le montrent Plénat et collègues. Des cas dans lesquels nous avons affaire très probablement à une variation de l'exposant plutôt que du thème de base sont également très nombreux en préfixation, en français et dans d'autres langues. C'est le cas, par exemple, des trois variantes du préfixe négatif qui est orthographié *in-* (ou *il-*, *im-*, *ir-*) et qui se présente sous les formes [in], [i] et [ɛ̃] qui sont, au moins partiellement, en distribution complémentaire (cf. Apothéloz 2003) ; c'est le cas aussi des préfixes, comme *sous-*, pour lesquels existe une variante comportant une consonne « de liaison » (*sous-alimentation*, *sous-entendre*). Dans tous ces cas, imaginer la variation observée comme supplétion thématique semblerait peu naturel, voire impossible dans certains cas comme *in-*. Certes, on pourrait soutenir, comme il a été souvent avancé, que la préfixation et la suffixation diffèrent par nature, et que la première fait intervenir des unités qui présentent une plus grande autonomie, et donc plus de variabilité. Cependant, il existe de très bons arguments pour refuser l'idée qu'il existe une différence substantielle entre ces deux procédés dérivationnels, et le cadre que j'adopte est juste-

[7] Pour un examen critique de la morphologie thématique appliquée à la flexion qui aboutit à des conclusions sensiblement semblables à celles défendues ici, cf. Bonami (2014 : 34-84) ; Bonami & Boyé (2014 : 18-22).

[8] Naturellement, sont exclus de ce raisonnement les cas dans lesquels un exposant dérivationnel possède des formes différentes dans différentes instances du même lexème (c'est-à-dire construit plusieurs thèmes à la fois), comme par exemple [jɛ̃], [jɛn], [jan] dans *italien*, *italienne*, *italianiser*.

ment un cadre dans lequel l'ensemble des procédés morphologiques constructionnels correspond à des opérations de la même nature, avec, tout au plus, un continuum déterminé par l'autonomie plus ou moins grande des éléments concernés (cf. Lasserre & Montermini 2014).

Les exemples mentionnés ci-dessus montrent bien que, dans une relation de morphologie constructionnelle, la variation (dans des termes plus traditionnels l'allomorphie) peut concerner soit les thèmes du lexème de base, soit l'exposant (éventuellement les deux à la fois), et que, dans certains cas on est bien face à des exemples d'« allomorphie affixale ». Si c'est le cas, le premier problème qui se pose est celui d'identifier, lorsque nous observons une variation formelle dans un ensemble de dérivés similaires, s'il s'agit bien d'un cas d'allomorphie de l'exposant, ou bien de deux ou plusieurs constructions différentes dont les exposants présentent des similarités formelles et/ou sémantiques. La tâche est certainement compliquée par le fait que les cas d'« échangisme affixal », dans lesquels les locuteurs choisissent, pour une base donnée, un affixe équivalent ou même moins adapté sémantiquement que celui attendu parce qu'il apparaît comme préférable du point de vue formel (cf. entre autres Lignon & Plénat 2009, Lignon 2013, Roché 2013), sont avérés et fréquents. Il me semble qu'il y a au moins deux facteurs qui peuvent être invoqués pour identifier une variation comme étant une allomorphie affixale. Premièrement, les différentes variantes doivent être assez semblables phonologiquement pour pouvoir être identifiées par les locuteurs comme relevant du même exposant de construction, par exemple en manifestant des alternances qui sont phonologiquement naturelles et/ou qui s'observent dans d'autres cas dans la langue. C'est le cas, par exemple des segments « fluctuants » que l'on observe dans les différentes variantes de -ien (mais aussi devant -eux), de l'assimilation dans *in-*, ou de l'émergence d'une consonne « latente » dans *sous-*. Naturellement, cette homogénéité formelle doit toucher toutes les formes du même exposant qui apparaissent dans les thèmes qu'il permet de construire. C'est ce dernier critère, par exemple, qui permet de rassembler -*ien*, -*éen* et -*ain* en tant que variantes de l'exposant d'une seule construction, mais de distinguer le -*in* qui construit aussi des gentilés (*alpin*, *girondin*), puisque les lexèmes qu'il permet de dériver possèdent la même finale que les suffixes ci-dessus au thème A (celui des formes du masculin), mais pas au thème B (celui des formes du féminin)[9]. Deuxièmement, le contexte d'apparition des différentes variantes doit être clairement identifiable du point de vue phonologique ou morphologique. Dans le meilleur des cas, les différentes variantes sont en distribution complémentaire parfaite ; dans la pratique, cependant, il est plus vraisemblable d'observer des préférences pour une variante ou pour une autre selon la forme phonologique de la base. Tous les travaux mentionnés ci-dessus (Lignon & Roché 2011 sur -ien, Plénat, Lignon et al. 2002 sur -*esque*, Apothéloz 2003 sur *in-*) montrent en effet en premier lieu que le choix de l'une ou de l'autre variante ne se fait jamais de façon déterministe, et que la variation est la condition normale d'existence de toutes ces constructions. En revanche, l'origine commune ou d'autres propriétés extralinguistiques ne sont évidemment pas de bons critères pour décider du statut de deux variantes comme relevant de deux constructions différentes ou de la même. Plénat (2008b) et Roché & Plénat (2016) ont par exemple

---

[9]Pour l'étiquetage des thèmes, j'utilise les mêmes conventions que Plénat (2008a) ou Roché (2010).

montré que la distribution de -*ais* ou -*ois* comme suffixe pour la construction des gentilés (que l'on pourrait être tenté de considérer comme les deux allomorphes d'un seul suffixe, puisqu'ils proviennent du même suffixe latin et ils construisent, de façon parallèle, un thème B en [z]) repose, au moins en partie, sur des critères géographico-historiques, ce qui pousse à les considérer comme les exposants de deux constructions morphologiques distinctes, bien que, évidemment, reliées du point de vue sémantico-fonctionnel.

Une fois que nous avons établi que toute la variation observée en dérivation ne peut pas être attribuée uniquement à la supplétion thématique, et qu'un certain nombre de phénomènes ne peuvent être analysés qu'en termes de variation des exposants, il nous reste à établir comment modéliser cette variation des exposants dans un cadre de morphologie thématique, et comment elle interagit avec les mécanismes de sélection des thèmes.

2.2 Les exposants morphologiques en tant que contraintes

Une façon simple et à mon sens efficace de représenter la variation des exposants dans un cadre comme celui adopté dans ce travail est de considérer les exposants eux-mêmes comme des contraintes. En d'autres termes, l'exposant d'une construction morphologique peut être envisagé comme un ensemble de contraintes formelles sur la forme de ses outputs. Plus précisément, je considère que chaque construction morphologique spécifie un ensemble de propriétés formelles, prosodiques ou segmentales, que ses dérivés doivent avoir. Dans ce cas, il s'agit donc de contraintes spécifiques à chaque construction dont la satisfaction est bien entendu conditionnée à la satisfaction d'autres contraintes, universelles ou spécifiques à chaque langue. Comme dans les modèles classiques qui emploient cet outil, les contraintes peuvent être contradictoires entre elles – et dans ce cas être hiérarchisées, de façon stable ou variable – ou, au contraire, converger, et donc se renforcer mutuellement (Plénat & Roché 2014 : 51, qui s'inpirent de Burzio 2002). L'existence de contraintes prosodiques (par exemple concernant la taille optimale d'un mot construit) a été observée et discutée depuis longtemps, en particulier sur le français (cf. Plénat 2009 pour un aperçu). Plus récemment, la structure segmentale des lexèmes dérivés, notamment dans les cas où l'on observe un écart entre la forme attendue et la forme attestée, a aussi été décrite en termes de contraintes. En particulier, Roché & Plénat (2014 : 1868) identifient deux contraintes, qu'ils nomment, respectivement, « Contrainte de famille » et « Contrainte de série », dont la finalité, globalement, est de faire en sorte qu'un lexème dérivé soit le plus semblable possible à d'autres lexèmes reliés, soit parce qu'ils appartiennent à la même famille (et donc sont construits sur le même lexème de base), soit parce qu'ils appartiennent à la même série (et donc sont construits au moyen du même procédé morphologique). La contrainte de série, en particulier, rend compte du fait que le même suffixe tend à sélectionner des thèmes de base le plus possible similaires du point de vue segmental. Ceci explique, entre autres, l'émergence, au sein de la même série dérivationnelle, de sous-séries homogènes. Des cas de cooccurrence suffixale ou la fréquence de certaines séquences avant un affixe (entre autres, -*titude*, -*inette*, -*alisme*, -*anisme*, -*ariat*, -*orat*, -*inat*, etc., cf. Plénat & Roché 2014 pour un aperçu) ont été analy-

sés en termes de contrainte de série. Globalement, la contrainte de série, donc, garantit que tous les mots dérivés par la même construction (qui appartiennent à la même série) soient les plus semblables possibles dans leur partie droite (dans le cas de la suffixation). Dans le modèle développé par Plénat et Roché, ceci peut correspondre au moins à deux types d'opérations, qui à leur tour peuvent être réparties en sous-groupes :

i) la sélection d'un thème, qui peut être :

   a) un thème du lexème de base qui apparaît aussi dans d'autres dérivés, par exemple le thème qui apparaît dans *snobinard* pour *snobinat*, construit sur *snob* (Plénat & Roché 2014, cette opération permet de satisfaire simultanément la contrainte de série et la contrainte de famille) ;

   b) le thème d'un autre lexème appartenant à la même famille morphologique, par exemple *personnal-* (thème savant de *personnel*) dans *personnalisme*, qui, sémantiquement, est construit sur *personne* (Roché 2009 : 159) ;

ii) la création ex novo d'un radical[10] à partir d'un thème du lexème de base, qui peut se faire ;

   a) par troncation, par exemple dans *végétariat* construit sur *végétarien* (Plénat & Roché 2014 : 67). Cette opération permet également de satisfaire des contraintes prosodiques sur la taille des dérivés ;

   b) par adjonction d'une séquence, par exemple dans *geekariat* construit sur *geek* (Plénat & Roché 2014 : 69) ;

   c) par manipulation du thème, par exemple dans les dérivés de *gouverneur*, *gouvernorat*, *gouvernatorat*, *gubernatorat*, etc. (Plénat & Roché 2014 : 59), qui reconstruisent des thèmes savants pour un lexème qui, en français, en est normalement dépourvu.

Toutes les opérations décrites ci-dessus ont le but d'inclure les lexèmes construits dans celles que Plénat et Roché appellent des « sous-séries lexicales », c'est-à-dire des ensembles de lexèmes dérivés par la même construction morphologique qui, du point de vue segmental, partagent plus que l'exposant de la construction en question, en l'occurrence [ina] ou [ɔʁa] pour la suffixation en -*at*, et [alism] pour la suffixation en -*isme*. Plus une sous-série est grande, plus elle sert de pôle d'attraction pour de nouveaux lexèmes, quitte à induire la sélection d'un thème non optimal du point de vue sémantique (comme dans *personnalisme*), ou bien une manipulation du thème (comme dans les cas en (ii) ci-dessus), en entraînant, dans les deux cas, une violation de la contrainte de fidélité base-dérivé. Plusieurs cas de combinaisons d'affixes du français, plus ou moins justifiées du point de vue sémantique, ont été traités dans la perspective d'une inclusion de lexèmes impliqués dans des sous-séries morphologiques (cf. Roché 2009, 2011, Namer 2013, Lignon et al. 2014). Dans d'autres cas, cependant, les segments qui permettent d'identifier une sous-série ne correspondent pas nécessairement (ou du moins ne correspondent plus

---

[10]Sur la distinction entre « thème » et « radical » cf. en particulier Roché (2010).

en synchronie) à un affixe ; c'est le cas de la sous-série -*inat* pour -*at* (cf. ci-dessus), mais également de la sous-série -*titude* pour -*itude* (Plénat & Roché 2014 : 53), -*acisme* pour -*isme* (Roché 2011 : 85), etc. Toutes ces séquences (qu'elles proviennent de suffixes synchroniquement analysables ou pas) ont uniquement une fonction formelle et lexicale, puisqu'elles permettent de réduire la dispersion à l'intérieur des séries morphologiques et contribuent, donc, à les rendre plus homogènes. À bien regarder, de ce point de vue il n'y a pas de distinction de substance entre ces séquences et les séquences que traditionnellement nous acceptons comme étant des affixes. Dans un cadre théorique qui ne reconnaît pas d'existence autonome aux affixes en dehors des opérations morphologiques dont ils sont les exposants, ceux-ci peuvent être conçus simplement comme des associations arbitraires de séquences de segments à une construction. Leur rôle n'est autre que de permettre de reconnaître qu'un lexème a été dérivé au moyen d'une construction donnée, et donc d'avoir des constructions qui, du point de vue formel, sont les plus homogènes possibles. Comme je l'ai évoqué plus haut, je propose donc de concevoir toutes les séquences formelles qui permettent d'identifier des séries ou des sous-séries morphologiques comme des contraintes, dérivant, en particulier, d'un élargissement de la contrainte de série, pour laquelle je propose la formulation suivante :

(2) **Contrainte de série** : tous les lexèmes relevant de la même série morphologique sont identiques.

La formulation ci-dessus est délibérément vague, pouvant englober aussi bien les propriétés formelles que les propriétés sémantiques des lexèmes dérivés (quel que soit le modèle sémantique auquel on se réfère). Si elle peut paraître paradoxale, elle est à mon avis suffisante pour rendre compte de l'ensemble des propriétés des lexèmes construits appartenant à la même série. D'un côté, la contrainte de série est contrecarrée par d'autres contraintes, en premier lieu par la contrainte de famille[11], qui met en relation chaque lexème avec les lexèmes construits sur la même base et, de fait, empêche que la contrainte de série ait pour effet de rendre tous les lexèmes de la même série identiques. De l'autre côté, dans les faits tous les membres de la même série morphologique partagent des éléments de forme qui sont communs et occupent toujours la même place, ce qui donne lieu à l'identification d'exposants qui, du moins en français, sont généralement des préfixes ou des suffixes. Il est possible, de plus, que dans certains cas il soit utile de considérer la contrainte de série, dans la formulation que j'en ai donnée, comme pondérable selon la fréquence et la saillance des lexèmes dans une série donnée. Puisque généralement tous les lexèmes de la même série ne partagent jamais tous leurs segments, on peut imaginer que les nouveaux lexèmes qui rentrent dans une série tendent à s'aligner, formellement, plutôt aux lexèmes les plus fréquents ou saillants de celle-ci. Dans des cas extrêmes, où une série contient un lexème qui, pour différentes raisons, joue un rôle de lexème prototype (un « leader word », selon les termes de Rainer 2003 ou Roché 2011),

---

[11]Parallèlement, on pourrait imaginer une Contrainte de famille qui stipulerait que tous les lexèmes de la même famille sont identiques. De telles contraintes, contraires et ayant la même force, auraient pour effet de s'annuler réciproquement, en empêchant, de fait, que tous les lexèmes de la même famille ou de la même série soient identiques, mais rendant compte du fait qu'ils partagent la plupart de leurs propriétés.

celui-ci constitue le modèle auquel les autres lexèmes tendent à ressembler, y compris du point de vue formel. C'est le cas des lexèmes appartenant à la série donnée en (1) dans l'introduction, dans laquelle *pérestroïka* est de loin le lexème le plus saillant, puisqu'il en est à l'origine. Dans ce cas, la forme des nouveaux lexèmes inclus dans la série (peu nombreux, au final) est évaluée, par rapport à la contrainte de série, en fonction de leur similarité principalement avec ce lexème prototype, ce qui explique que différents lexèmes (par exemple *Béréstroïka* ou *Castroïka*) aient pu retenir des portions variables dans leur exposant.

Concrètement, nous pouvons imaginer que la contrainte donnée en (2) se décline en contraintes et sous-contraintes plus spécifiques qui, pour chaque construction, définissent les segments que les mots de la série correspondante partagent et leur position. Plénat & Roché (2014 : 54) eux-mêmes évoquent l'idée qu'une construction morphologique puisse être considérée « comme une macro-contrainte résultant de la présence dans le lexique d'une série de mots ». Pour reprendre et développer le cas discuté par eux des noms en *-at* du français, leur représentation formelle peut être vue comme comportant les contraintes [Xa], [Xaʁja], [Xika], [Xɔna], [Xɔʁa], etc. (cf. la liste donnée par Plénat & Roché 2014 : 54). Le fait que les sous-contraintes [Xaʁja], [Xika], [Xɔna], [Xɔʁa] soient partiellement en contradiction les unes avec les autres n'est évidemment pas problématique, dans un cadre dans lequel la satisfaction simultanée de toutes les contraintes n'est pas indispensable. Les mêmes contraintes peuvent être considérées comme étant dans une relation de « Elsewhere Condition » avec la contrainte plus générale : celle-ci correspond au choix par défaut adopté au cas où d'autres contraintes empêcheraient les sous-contraintes plus spécifiques d'être satisfaites. L'idée que des contraintes de ce type soient dans une telle relation hiérarchique est cruciale dans ce cadre. Dans les faits, il est en effet évident que, toute chose égale par ailleurs, les lexèmes issus de la même construction tendent à présenter toujours la même forme d'exposant, qui correspond donc à sa forme par défaut. Ce cas par défaut peut, comme dans le cas général discuté ici, correspondre à une forme sous-spécifiée par rapport aux autres ([Xa]), mais il peut aussi correspondre à une forme qui a le même degré de spécification que les autres, mais qui est plus fréquente dans la série en question. Pour expliquer des formes en *-at* comme *hôtessariat*, *shérifariat*, *victimariat*, etc., Plénat & Roché (2014 : 71) observent qu'« il faut que *-ariat* soit devenu, pour certains locuteurs, la forme par défaut du suffixe ». L'existence d'une « forme par défaut » de marqueurs morphologiques a été observée dans plusieurs cas. Lignon & Roché (2011 : 191), par exemple, indiquent *-ien*, *-éen*, *-ain* et *-en* comme formes possibles pour le suffixe -ɪᴇɴ, avec la première variante qui a la forme par défaut. Dans des travaux antérieurs (Montermini 2010, 2015), j'ai soutenu une position semblable pour les suffixes cognats de l'italien. En prenant en compte des données néologiques comme celles en (3), j'ai soutenu que l'exposant en question possède une forme sous-spécifiée [Vano], dont la position V est remplie par défaut par un segment [j] lorsque la base n'est pas problématique pour la phonologie de l'italien (finale en voyelle simple non accentuée ou en consonne : *calcuttiano*, *hannoveriano*), ou par une voyelle fournie par la base, lorsque celle-ci présente une finale problématique (voyelle accentuée, hiatus, diphtongue) ; enfin, la forme [ano] non précédée par une voyelle émerge très majoritairement avec des bases qui se terminent par une voyelle [a] atone (*wojtylano*).

(3)   a.   *calcuttiano ← Calcutta*

       b.   *hannoveriano ← Hannover*

       c.   *deandreano ← (Fabrizio) De Andrè*

       d.   *murnauano ← (Friedrich) Murnau*

       e.   *pessoano ← (Fernando) Pessoa*

       f.   *wojtylano ← (Karol) Wojtyla*

Les contraintes qui correspondent aux différentes variantes d'un affixe peuvent donc être elles-mêmes dans des relations hiérarchiques, avec généralement une forme qui, par rapport aux autres, a le statut de forme par défaut. Cette relation hiérarchique peut prendre au moins deux formes : i) la forme par défaut est une forme sous-spécifiée par rapport aux autres ([Xa] vs. [Xaʁja], [Xika], [Xɔna], [Xɔʁa]) ; ii) la forme par défaut a le même degré de spécification que les autres formes, mais est plus fréquente dans la série correspondante (-*ien* vs. -*éen*, -*ain*, -*en*), voire est plus spécifiée ([jano] vs. [Vano] en italien). Naturellement, les formes qui ne correspondent pas au défaut peuvent elles-mêmes être dans une relation hiérarchique. Ainsi, dans le cas des noms en -*at* du français, selon ce que disent Plénat & Roché (2014), [Xaʁja] semble fonctionner comme un défaut secondaire, plus fréquent dans la série, et donc plus disponible, que les autres variantes.

Comme je l'ai observé plus haut, les contraintes qui correspondent à la forme phonologique des exposants des constructions morphologiques (que je considère, je le rappelle, comme autant de sous-contraintes d'une contrainte de série plus générale qui a la forme en (2)), interagissent naturellement avec les autres contraintes formelles qui pèsent sur les mots construits. Par exemple, les contraintes relatives à la structure segmentale des lexèmes construits en -*at* du français, indiquées ci-dessus, sont associées à une contrainte plus générale du français qui demande qu'un lexème construit comporte, préférentiellement, trois syllabes. De même, ces contraintes segmentales entrent en relation avec des contraintes généralement considérées comme universelles, comme des contraintes phonologiques anti-marque, ou une contrainte de fidélité base-dérivé. Quelques-uns des lexèmes de (3) exemplifient ce fait. Une forme comme *wojtylano*, par exemple, respecte la contrainte de fidélité base-dérivé, ainsi qu'une contrainte phonologique générale qui défavorise les séquences de voyelles identiques (qui serait violée par *wojtylaano), mais viole partiellement la contrainte segmentale [Vano]. La forme alternative *wojtyliano*, également attestée, au contraire, respecte cette dernière contrainte (et même la hiérarchie qui indique [jano] comme forme par défaut), mais peut être considérée comme moins optimale du point de vue de la fidélité base-dérivé, puisque la voyelle finale de la base est effacée. De son côté, *deandreiano* respecte la contrainte de fidélité base-dérivé (tous les segments de la base s'y retrouvent) et respecte aussi la contrainte segmentale [Vano], même si elle favorise une variante du suffixe moins haute dans la hiérarchie des formes possibles.

À partir de ce qui est dit ci-dessus, il est évident qu'une question comme « qu'est-ce qui appartient à la base et qu'est-ce qui appartient à l'affixe ? » n'est plus une question pertinente. Si nous voulons à tout prix voir les choses dans ces termes, dans *deandreano* le segment [e] « appartient » à la fois à la base et à l'affixe. Dans des termes plus appropriés

pour le modèle défendu ici, l'émergence du segment [e] permet de satisfaire plusieurs contraintes formelles à la fois. Il est évident, donc, que dans ce cadre une notion théorique comme celle de « frontière morphologique », qui a été une notion importante dans plusieurs modèles théoriques (par exemple la Phonologie Lexicale ou la Morphologie Naturelle) ne joue plus aucun rôle. Dans les exemples en question, il n'y a pas de « frontière », puisqu'il n'y a pas deux éléments accolés l'un à l'autre, mais plutôt l'application d'un ensemble de contraintes formelles à une forme (un thème). Comme on le voit, ce pas est particulièrement cohérent avec le mouvement progressif de « déréification » des exposants morphologiques que la recherche en morphologie a mis en œuvre dans les dernières décennies.

Avant de conclure, observons que les contraintes segmentales sur la forme des lexèmes construits, qui correspondent à leurs exposants, sont des contraintes d'un type particulier. Alors que les contraintes, au sens classique, sont censées capter des propriétés générales, voire universelles, des langues, ici il s'agit de contraintes hautement spécifiées et dont le domaine d'application est fortement restreint. Cependant, le modèle de morphologie à contraintes dont je m'inspire combine déjà des contraintes universelles avec des contraintes spécifiques à une langue donnée (dans ce cas le français), et même des contraintes spécifiques à une sous-partie de la langue à un stade d'évolution donné et limitées à une de ses modalités (par exemple la « Contrainte de fidélité phonographique », Roché & Plénat 2014 : 1873). S'il est légitime d'avoir de telles contraintes non seulement non universelles, mais limitées à des secteurs de la langue, il me semble que rien n'empêche, du point de vue conceptuel, d'avoir des contraintes limitées à des constructions particulières, d'autant plus que les contraintes sur la forme des dérivés identifiées ci-dessus sont issues d'une contrainte plus générale, la contrainte de série qui, elle, peut prétendre au statut de contrainte universelle de la morphologie.

Pour conclure cette section, avant de passer à l'illustration des cas concrets étudiés dans la section 3, je récapitule les différents éléments de la proposition que j'ai avancée pour rendre compte de la forme des outputs des constructions morphologiques. Tout d'abord, la forme d'un lexème construit est régie, entre autres, par une contrainte de série qui stipule qu'il doit être le plus semblable possible, y compris du point de vue segmental, aux autres lexèmes de la même série. Pour chaque construction individuelle, cette contrainte prend la forme de contraintes plus spécifiques qui stipulent les segments qu'un dérivé de la série doit contenir pour être considéré comme tel, et leur position (ce qui correspond à l'affixe au sens traditionnel). Ces contraintes plus spécifiques peuvent être multiples, ce qui rend compte de la variation observée pour les exposants morphologiques ; elles peuvent être en contradiction les unes avec les autres ou se renforcer mutuellement, et peuvent être hiérarchisées, avec, dans le cas le plus courant, une des variantes qui fonctionne comme le défaut. La forme des lexèmes construits réellement observée est déterminée par l'interaction de ces contraintes segmentales avec les autres contraintes formelles, en particulier la contrainte de famille et celles qui sont responsables pour la sélection du thème du lexème de base. Roché & Plénat (2014) ont montré plusieurs exemples dans lesquels la sélection du thème de base (ou sa manipulation) a pour but de satisfaire la contrainte de série et/ou la contrainte de famille. Dans la section

qui suit, je discuterai des cas dans lesquels cette sélection interagit également avec la hiérarchie des contraintes segmentales qui correspondent à la forme de l'exposant des constructions. Parfois, un thème spécifique est sélectionné en vertu de sa compatibilité avec une des formes de l'exposant qui est haut placée dans la hiérarchie ; dans d'autres cas, c'est une forme moins haute dans la hiérarchie qui émerge parce qu'elle est plus compatible avec le thème de la base sélectionné, par exemple parce que d'autres thèmes ne sont pas disponibles.

3 Le jeu des contraintes dans l'identification de la forme des dérivés : deux études de cas

Dans cette section, j'applique le modèle esquissé dans la section 2 à trois exemples de constructions morphologiques. Je montrerai en particulier que l'exposant d'une construction possède un ensemble de formes possibles, dont l'émergence dépend de l'interaction avec les autres contraintes en jeu (en premier lieu la contrainte de fidélité base-dérivé). Dans tous les cas, j'indiquerai les exposants dans le texte avec une forme arbitrairement choisie (généralement la forme par défaut) écrite en petites majuscules (-PHONE, -ISSIMO, etc.), en suivant ainsi la convention adoptée par Lignon & Roché (2011) et celle généralement admise pour les lexèmes.

Le premier cas étudié est la construction de noms (ou adjectifs) qui désignent les locuteurs d'une langue et qui sont construits au moyen de l'élément -PHONE en français, qui sont comparés aux noms issus de la construction correspondante en italien (-FONO). Cet exemple montre comment deux constructions similaires (et cognates) dans deux langues proches peuvent présenter des propriétés formelles (et donc un jeu de contraintes segmentales) différentes. En italien, en effet, la forme de l'exposant comporte sans exception un [o] accentué (issu de l'élément de composition grec), alors qu'en français un segment de timbre /o/ est présent uniquement dans la forme par défaut de l'exposant, mais sa position peut être occupée par une autre voyelle (*quechuaphone*, *ewephone*) et même par une consonne (*ocphone*, *pularphone*), le timbre de ce segment étant corrélé à la forme du thème de la base. Les constructions de noms de locuteurs en -PHONE / -FONO ont également la particularité de sélectionner des bases de complexité variable : dans certains cas la base est un lexème qui appartient à une famille nombreuse, qui possède donc un espace thématique riche et peut par conséquent donner lieu à une grande variation des dérivés (à partir de *portugais* j'ai recensé *lusophone*, *lusitophone*, *portugaisophone*, *portugalophone*, *portugophone*) ; dans d'autres cas, la base est un nom de langue qui n'est relié à aucun autre lexème dans le lexique, qui possède parfois une structure phonologique inhabituelle en français, et de laquelle la morphologie doit s'accommoder pour obtenir un output. Nous verrons que les manipulations que les thèmes de certaines bases subissent en français (comme dans *portugophone*) ont pour but de satisfaire différentes contraintes, dont les contraintes segmentales, et que les manipulations des thèmes sont, en italien, beaucoup plus réduites et se limitent à un ou deux types. Pour terminer, cette étude de cas me donnera l'occasion de discuter la place de ladite « composition néoclassique » dans le système morphologique des deux langues en question.

Le deuxième cas étudié est la construction de noms ou adjectifs au moyen du suffixe -ISSIMO en français. La suffixation en -ISSIMO a la particularité de construire des noms pour lesquels l'apport sémantique de la construction morphologique est très faible, dans la plupart des cas ils ont simplement une teinte évaluative génériquement appréciative. Les bases possibles pour cette dérivation sont donc très peu contraintes du point de vue sémantique (et même catégoriel). La sélection se fait alors souvent sur une base surtout ou uniquement formelle, en utilisant des bases qui sont particulièrement compatibles avec les contraintes formelles auxquelles les dérivés en -ISSIMO sont sujets. Ce procédé dérivationnel sera comparé à la construction d'adjectifs et noms en -ISSIME, plus ancienne et plus proche aux procédés dérivationnels canoniques du français.

Les études présentées dans cette section se situent dans une approche extensive à la morphologie. Cette approche se fonde sur l'idée que, pour la compréhension des mécanismes qui dirigent la construction du lexique, il est nécessaire d'observer, d'une part, une quantité importante de données et, d'autre part, de prendre en compte le lexique non établi, non institutionnalisé, et donc – vraisemblablement – construit « sur le champ » par les locuteurs. Ce deuxième point, en particulier, correspond à deux sources de données possibles : soit on s'intéresse, pour les procédés morphologiques canoniques de la langue, aux formes non établies, comme les néologismes, les occasionalismes, etc. (c'est le cas de la première étude proposée), soit on s'intéresse à des procédés morphologiques non canoniques (c'est le cas de l'étude suivante). L'idée sous-jacente est que dans le lexique établi, y compris parmi les lexèmes construits, il y a trop de risques de rencontrer des mots qui ont subi des dérives formelles et/ou sémantiques étrangères à leur mode de construction morphologique, et donc que ce n'est pas le meilleur point d'observation pour la compréhension des mécanismes morphologiques tels qu'ils opèrent en synchronie et « en vrai ».

Le type d'objets auquel je m'intéresse, bien entendu, n'est pas sans poser de problèmes, puisqu'il est nécessaire de rassembler des bases de données non attestées suffisamment importantes et fiables pour pouvoir tirer des généralisations solides et prédictives. Le but de cet article n'est évidemment pas celui de discuter les problèmes liés à la morphologie extensive, qui ont déjà été largement traités en littérature (cf. les travaux cités dans la note 3). Ici, quelques remarques sur la collecte et l'exploitation des données sont suffisantes : pour tous les phénomènes étudiés j'ai essayé de rassembler des bases de données qui, sans être exhaustives, sont les plus larges possibles. Les données ont été recueillies en premier lieu à partir de corpus basés sur le Web, FrWac et ItWac[12]. Ces bases de données ont été enrichies par des recherches ciblées sur le Web et, occasionnellement, à partir d'autres sources. Les contextes d'apparition des lexèmes inclus dans les bases de données ont été vérifiés afin d'éliminer le plus possible le bruit (textes écrits par des locuteurs non natifs, fautes de frappe, etc.). Faute de pouvoir réaliser des calculs de fréquence fiables, en particulier sur le Web, les analyses présentées ici ne prennent en compte que les types de dérivés inclus dans les bases de données et non pas le nombre de leurs occurrences (tokens). Bien entendu, des calculs de fréquence des occurrences

---

[12]FrWac comporte ~1,6 milliards de tokens et ~6 millions de types ; ItWac comporte ~2 milliards de tokens et ~6,2 millions de types (sur ces deux corpus cf. en particulier Baroni et al. 2009).

seraient utiles et intéressants pour confirmer, moduler ou enrichir les analyses proposées. Toutefois, on peut proposer au moins quatre observations pour justifier le choix effectué : i) comme je l'ai indiqué, dans l'étude de la morphologie dérivationnelle l'observation des lexèmes nouvellement produits (néologismes, occasionalismes, etc.) est tout aussi intéressante que celle du lexique établi ; or ces lexèmes sont généralement très rares y compris dans des corpus de grandes dimensions ; si l'on veut privilégier la diversité des formes produites par les locuteurs, on se retrouve avec des bases de données qui comportent un grand nombre de lexèmes avec une fréquence d'emploi très faible qui, de ce point de vue, ne permet pas de toute façon de réaliser des calculs statistiques fiables ; ii) si, comme dans ce travail, on adopte un modèle de la morphologie basé sur l'idée d'une interaction de plusieurs contraintes, en dehors du lexique établi la variation des outputs des constructions morphologiques est la norme, et la fréquence d'un lexème n'est pas nécessairement corrélée à une plus ou moins grande « régularité » du point de vue de la morphologie constructionnelle ; iii) la collecte de bases de données qui, faute d'être exhaustives, sont les plus larges possibles en termes de types permet tout de même de proposer des généralisations et des prédictions sur l'application d'une construction morphologique à une base donnée ; des études encore plus larges, ou qui prennent en compte d'autres paramètres pourront confirmer ou falsifier ces prédictions ; iv) en plus d'analyses quantitatives, il est possible de proposer des analyses qualitatives, dans lesquelles les propriétés de chaque lexème dérivé et de chacune de ses variantes éventuelles sont attribuées explicitement à la prédominance d'une contrainte (ou d'un ensemble de contraintes) ou d'une autre.

3.1 ‑PHONE / ‑FONO

Pour la première étude de cas, j'ai rassemblé une base de données de 475 lexèmes (noms et/ou adjectifs) désignant, en français, les locuteurs d'une langue, qui comportent la séquence finale [fɔn] précédée, dans la grande majorité des cas, du nom d'une langue. Une base de données parallèle, comportant 237 lexèmes, a été constituée pour l'italien. Pour rassembler la base de données du français j'ai repris celle présentée dans Lasserre (2016) que j'ai enrichie, initialement, par l'extraction des formes se terminant par les séquences <phone> et <phones> dans FrWac, le nettoyage manuel de cette première liste, et ensuite par des recherches systématiques sur le Web réalisées à partir des listes de langues (liste des langues les plus parlées au monde et liste des langues officielles des pays du monde) du Wikipedia francophone. La base de données de l'italien a été constituée à partir de ItWac et de recherches systématiques sur le Web en utilisant les mêmes ressources que pour le français, ainsi qu'en appliquant ‑FONO aux noms des habitants des régions et des principales villes italiennes. Les deux bases ont été complétées par des recherches croisées des correspondants des lexèmes présents dans l'une ou dans l'autre. Le fait que la base des données du français soit beaucoup plus importante que celle de l'italien (presque deux fois plus d'entrées) est certainement dû à la saillance, dans la culture francophone, des termes *francophone* et *francophonie*. Ces mots désignent deux concepts qui se sont développés et répandus d'abord en relation à la situation linguistique canadienne (à partir

de la fin du XIXe siècle), et ensuite dans le discours politique de l'époque postcoloniale. Puisqu'un espace italophone comparable à celui du français pour nombre de locuteurs et distribution géographique n'existe pas, *italofono*, ou des termes similaires, n'ont pas la même connotation, et l'emploi des lexèmes en -FONO, en général, est plutôt limité au discours spécialisé en linguistique, dialectologie, etc.

Les lexèmes en -PHONE / -FONO sont généralement rangés parmi les composés néo-classiques, en vertu de l'origine et de la valeur sémantique supposée du deuxième élément (issu d'un lexème nominal du grec signifiant « voix »). Dans ce travail, cependant, j'adopte un modèle de la morphologie dérivationnelle qui ne prévoit pas de distinction discrète entre les différents types de constructions. Les différentes constructions (composition, composition néoclassique, affixation) se placent, au contraire, le long d'un continuum, avec de différents degrés de grammaticalisation, c'est-à-dire de conventionnalisation des propriétés (formelles, catégorielles et sémantiques) des lexèmes qu'elles servent à former[13]. Dans ce cadre, aucune différence de nature n'est établie entre les affixes au sens traditionnel et les dits « éléments de composition néoclassique » : dans tous les cas il s'agit d'exposants de constructions, qui peuvent éventuellement se distinguer pour leur degré de grammaticalisation. En aucun cas, on n'attribue d'existence, ni de signification lexicale autonome (*contra*, par exemple, Corbin 2001) à ces éléments, qui, dans le fonctionnement synchronique de la langue, restent indissociables des constructions qui les introduisent. En ce qui concerne plus particulièrement les constructions en -PHONE / -FONO en français et en italien, plusieurs propriétés les rapprochent des cas d'affixation canoniques. D'une part, les lexèmes formés au moyen de ces constructions entrent dans des paradigmes dérivationnels avec d'autres lexèmes, simples ou construits, par exemple, dans le cas des lexèmes désignant les locuteurs d'une langue (*francophone*), avec des lexèmes à sens collectif en -*phonie* (*francophonie*). Deuxièmement, la valeur sémantique supposée véhiculée par l'élément -PHONE n'est pas toujours saillante lorsque ces lexèmes sont employés en contexte. Dans certains cas, s'ils sont employés comme adjectifs (4a), leur valeur se rapproche de celle des autres adjectifs relationnels construits sur des noms ; dans d'autres cas (4b), ces mêmes lexèmes apparaissent dans des constructions syntaxiques dans lesquelles ils partagent les mêmes contextes et les mêmes valeurs d'adjectifs relationnels (dans ce cas ethniques) canoniques :

(4) a. Le mot « Rega » serait une transformation rwandophone survenue au XXème siècle, au même titre que le mot « Reka » d'origine ougandophone.
[https ://www.edilivre.com/frontwidget/preview/book/id/626357/]

b. Cette « guerre » a aggravé et renforcé les tensions communautaires préexistantes entre communautés rwandophones et congolaises d'une région peuplée où les litiges fonciers étaient omniprésents...
[http ://www.revuenouvelle.be/Plus-de-quinze-annees-de-guerre-au-Kivu-Ca-suffit]

Un troisième argument qui permet de rapprocher -PHONE / -FONO des affixes canoniques concerne précisément leur comportement phonologique dans les deux langues

---

[13]Cf. Lasserre & Montermini (2014) pour une discussion détaillée du modèle.

et la manipulation des différentes variantes possibles via les contraintes, qui, comme je le montrerai dans ce qui suit, ne se différencie pas du comportement d'autres éléments dont l'identification comme affixes est plus consensuelle.

Les noms de locuteurs ne sont pas les seuls lexèmes dans lesquels les éléments d'origine grecque -PHONE et -FONO interviennent. Pour se limiter, pour l'instant, au français, -PHONE apparaît également dans des noms d'instruments (de musique ou autre) (*xylophone*, *saxophone*) ou d'appareils sonores (*audiophone*, *téléphone*) (cf. Lasserre 2016 : 179-183). Cependant, je considère que ces différents lexèmes relèvent de constructions qui, si leurs exposants sont reliés diachroniquement, sont distinctes. Plusieurs arguments peuvent être avancés pour justifier l'idée que le -PHONE en question est l'exposant d'une construction morphologique spécifique distincte des autres qui ont des exposants (partiellement) homophones : i) les lexèmes dérivés par cet élément présentent une grande homogénéité sémantique et catégorielle ; concernant ce dernier point, en particulier, ce sont toujours des lexèmes qui sont à la fois des noms [+humain] et des adjectifs de relation (qui ne modifient pas nécessairement un nom humain) ; ii) comme je l'ai montré ci-dessus, les lexèmes désignant les locuteurs d'une langue appartiennent à des paradigmes dérivationnels homogènes et spécifiques, qui diffèrent des paradigmes dérivationnels des autres types de lexèmes. Tous les lexèmes en -PHONE peuvent en effet avoir un lexème correspondant en -PHONIE avec un sens collectif (*téléphonie*, *visiophonie*), mais les dérivés en -*iste* (*téléphoniste*, *saxophoniste*) et en -*ique* (*téléphonique*, *microphonique*) sont réservés aux noms d'instruments et appareils, ce qui s'explique par le fait que les noms de locuteurs sont déjà à la fois des noms [+humain] et des adjectifs de relation.

Du point de vue des bases sélectionnées par la construction, le cas le plus simple est celui dans lequel un lexème en -PHONE est construit directement sur un nom de langue, qui peut désigner uniquement cette dernière (5a), ou bien correspondre à un gentilé (5b) ou à un nom ethnique non construit (5c)[14]. Si la base est un lexème variable en français, la contrainte de famille est respectée et le thème sélectionné est le plus souvent le même que celui sélectionné par les autres constructions morphologiques, à savoir un thème L, qui peut être identique à un thème B (5d) ou indépendant (5e). La base peut être aussi constituée du thème qui sert également à construire des gentilés, et dans ce cas la base est formellement ambiguë, puisqu'elle correspond, phonologiquement, au nom géographique sur lequel le gentilé est construit (5f). Pour terminer, la base peut également être un radical issu de la modification (généralement une troncation) d'un thème (5g), ou un thème supplétif savant (5h).

(5)　　a.　créolophone

　　　　b.　espagnolophone

　　　　c.　bascophone

　　　　d.　catalanophone

　　　　e.　coréanophone

　　　　f.　islandophone

---

[14]Sur les noms / adjectifs ethniques et les réseaux lexicaux dans lesquels ils apparaissent, cf. en particulier Roché (2008).

g. bulgophone, lettophone

h. magyarophone, sinophone

Parfois, un dérivé peut être ambigu et relever à la fois de plusieurs des types ci-dessus ; *italophone*, par exemple, pourrait appartenir tant au type (5f) qu'au type (5h). De plus, comme je l'ai montré dans la section 2, le même lexème de base peut donner lieu à plusieurs dérivés différents, relevant de plusieurs types. Pour *portugais*, par exemple, sont présents dans la base de données les dérivés suivants : *lusophone* (5h), *lusitophone* (5h), *portugaisophone* (5d), *portugalophone* (5f, cf. ci-dessous), *portugophone* (5f).

Dans la plupart des cas le nom de base correspond à un nom de langue identifiée et reconnue, comme dans les exemples en (5). Puisque les taxinomies courantes ne correspondent pas toujours aux taxinomies scientifiques, cependant, la base peut également correspondre à un nom ethnique désignant un groupe pour lequel on identifie une langue ou une variété spécifique (*écossophone*, *marocanophone*), à une dénomination non officielle (argotique) d'un groupe ethnique (*ritalophone*, *rosbiffophone* / *rosbiphone*), à un autre nom d'humains (*rebeuophone*) ou pas (*banlieuophone*), pourvu que l'on puisse identifier une « langue » (une variété linguistique) spécifique au groupe auquel on fait référence.

Venons-en maintenant aux propriétés formelles de ces dérivés. Du point de vue prosodique, une contrainte de taille est clairement identifiée, avec 83,5% des lexèmes considérés (397) qui sont tri- ou quadrisyllabiques (respectivement 142 et 255). La Figure 1 montre la distribution précise des lexèmes dans la base de données selon le nombre de syllabes.
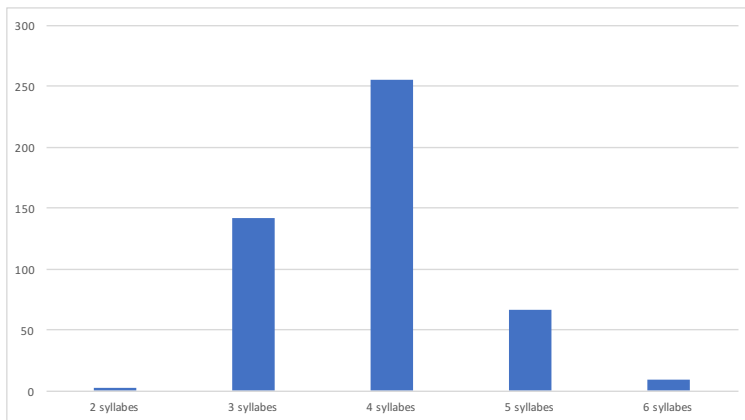
Figure 1 : Distribution des lexèmes en -phone selon le nombre de syllabes

Le fonctionnement de la contrainte de taille montre que, contrairement à ce que l'on aurait pu imaginer, le poids de *francophone* en tant que leader word de la série est limité, du moins en ce qui concerne la taille des dérivés. En effet, on aurait pu s'attendre à ce que le format trisyllabique prévale, éventuellement au prix de la réduction de bases trop

longues. Cependant, si on regarde les lexèmes en -ᴘʜᴏɴᴇ les plus fréquents dans FrWac, *francophone* vient, sans surprise, largement en tête, mais les dix premiers se répartissent de manière pratiquement équivalente entre tri- et quadrisyllabiques[15].

Parmi les 66 dérivés présents dans la base qui comportent cinq syllabes, 28 comportent également au moins une variante quadrisyllabique, la plupart du temps obtenue par troncation du thème de base (type (5g), par exemple *arménianophone* / *arménophone*, *tibéta-nophone* / *tibétophone*). Il en va de même pour 4 des 9 dérivés qui comportent 6 syllabes (*américanophone* / *américophone*). Inversement, sur 23 dérivés qui ont un radical obtenu par troncation de la base, 22 possèdent une variante « longue », généralement comportant une syllabe de plus. De la même manière, sur 91 lexèmes relevant du type (5e) (emploi du même thème que celui d'un gentilé), 79 comportent trois ou quatre syllabes. Nous pouvons donc considérer que le format tri- ou quadrisyllabique permet de satisfaire une contrainte de taille qui veut que, dans un mot construit, la base corresponde le plus fréquemment au format dissyllabique (cf. Plénat 2009) ; les troncations de thème ont principalement pour but de satisfaire cette contrainte (au détriment, bien entendu, de la contrainte de fidélité base-dérivé).

Concernant les propriétés segmentales des dérivés en correspondance de l'exposant, 79,5% des cas (378) se terminent en [ɔfɔn] et 20,5% (97) se terminent en [fɔn] précédé d'un autre segment (la plupart du temps une voyelle, cf. ci-dessous). À ce propos, il est possible d'établir une corrélation intéressante : pour le second groupe, le segment qui précède [fɔn] est déjà présent dans le thème de base dans la totalité des cas, alors que pour le premier groupe, celui se terminant en [ɔfɔn], le thème de base ne comporte un [o] final que dans 40 dérivés sur 378, répartis comme suit :

(6) a. thèmes se terminant en [o] (*espérantophone*, *lesothophone*) 17
     b. thèmes tronqués en correspondance d'un [o] (*lettophone*, *tagalophone*) 10
     c. thèmes supplétifs savants[16] (*germanophone*, *sinophone*) 13

La figure 2 résume la situation décrite (« oui » indique que le segment précédant [fɔn] est présent dans le thème de base, « non » qu'il ne l'est pas).

Pour 338 lexèmes de la base de données (71,2% du total), donc, l'opération phonologique consiste simplement en la concaténation de la séquence [ɔfɔn] à un thème, modifié ou pas ; pour 30 autres (les cas (6a) et (6c) ci-dessus), nous pouvons considérer que la présence d'un [o] dans la base n'est rien de plus que fortuite. Seuls les 10 lexèmes du type (6b) manifestent une manipulation dont l'effet est d'avoir un thème se terminant par [o] ; cependant, dans ce cas, la réduction du thème a aussi pour effet de produire un dérivé tri- ou quadrisyllabique dans la totalité des cas. On peut donc considérer qu'ici, au mieux, on assiste à une convergence entre la contrainte de taille et la contrainte qui demande que le dérivé se termine en [ɔfɔn].

---

[15] Les dix lexèmes en question sont : *francophone*, *anglophone*, *germanophone*, *arabophone*, *hispanophone*, *lusophone*, *néerlandophone*, *turcophone*, *berbérophone*, *russophone*.

[16] Je considère que les thèmes supplétifs savants comportent un [o] final, dans la mesure où ils peuvent apparaître sous cette forme, par exemple dans des composés (*germano-soviétique*, *sino-japonais*).
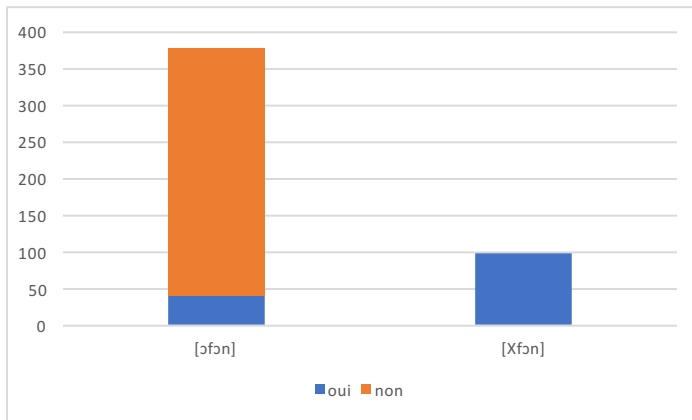
Figure 2 : Distribution des segments précédant [fɔn] présents ou non présents dans la base

Considérons maintenant les 97 cas dans lesquels le dérivé ne se termine pas par [ɔfɔn]. Tout d'abord, plus de deux tiers de ces dérivés (66) possèdent également une variante en [ɔfɔn]. De plus, comme je l'ai observé, il s'agit toujours de cas comme ceux exemplifiés en (7), dans lesquels le segment qui précède [fɔn] est toujours déjà présent dans le thème de base en tant que segment final. En (7) je donne le détail du nombre de dérivés selon la séquence finale :

(7) a. [afɔn] *aymaraphone* 34

  b. [ifɔn] *swahiliphone* 28

  c. [efɔn] *malinképhone* 9

  d. [Cfɔn] *tamoulphone* 8

  e. [wafɔn] *danoiphone* 6

  f. [ãfɔn] *flamanphone* 5

  g. [ufɔn] *ourdouphone* 4

  h. [œfɔn] *banlieuphone* 3

On pourrait être tenté d'identifier les formes [afɔn] et [ifɔn] comme des sous-défauts, vu leur prépondérance dans cette classe de dérivés. Il est probable, cependant, que leur fréquence soit surtout liée à la fréquence globale des noms de langues se terminant par [a] ou [i] par rapport aux autres segments. Notons que les 89 dérivés dans lesquels [fɔn] est précédé d'une voyelle différente de [o] constituent la majorité des outputs pour les thèmes de base se terminant en voyelle. La base de données comprend en effet 58 autres dérivés de bases en voyelles, dans lesquels soit la voyelle est effacée en faveur de [ɔfɔn] (*bambarophone*), soit, bien plus rarement (uniquement 6 exemples), la séquence [ɔfɔn] est attachée après la voyelle (presque uniquement un [i], *thaïophone*) (notons, de plus,

que dans ce cas il s'agit toujours de bases brèves, susceptibles de donner des dérivés tri- ou quadrisyllabiques).

Une interprétation des données présentées consiste à attribuer à la construction en question une forme d'exposant par défaut qui est [ɔfɔn], et une variante hiérarchiquement subordonnée, [Vfɔn] (où V représente une voyelle quelconque). L'ensemble des contraintes formelles (de série) qui pèsent sur les outputs de cette construction stipule donc qu'un dérivé doit comporter quatre (à défaut trois) syllabes et se terminer en [ɔfɔn] (à défaut en [Vfɔn]). Le reste des propriétés formelles observées pour les dérivés en question provient des autres contraintes générales qui pèsent sur la forme des mots construits, et en particulier de la contrainte de fidélité base-dérivé, qui est responsable de la forme des lexèmes en (7) et, plus en général du timbre de la voyelle qui précède [fɔn] lorsque ce n'est pas un [o]. À son tour, la contrainte de fidélité interagit avec les autres contraintes qui sont responsables pour la sélection et/ou la modification des thèmes de base, par exemple la contrainte de famille. Si une base est isolée dans sa famille lexicale (c'est le cas de la majorité des noms de langues non européennes), alors la sélection du thème n'est pas un enjeu : c'est le thème unique qui est choisi et qui est éventuellement manipulé pour satisfaire d'autres contraintes. Au contraire, si la base appartient à une famille lexicale nombreuse, le thème sélectionné peut correspondre au nom de la langue, construit ou pas (*coréanophone*, *corsophone*, *picardophone*, cela correspond, grosso modo, à la « Contrainte de fidélité à la forme libre » de Roché & Plénat 2014 : 1873), à un thème supplétif savant (*francophone*, *lusophone*, *magyarophone*), ou bien, moins préférentiellement, au thème qui apparaît devant les affixes construisant des gentilés et qui correspond, dans la plupart des cas, comme je l'ai observé, à un nom géographique de pays, région, etc. (*islandophone*, *japonophone*). Concernant ce dernier cas, la plupart des dérivés sont ambigus, comme ceux mentionnés ; cependant, il est possible que, du moins pour certains locuteurs, les deux possibilités soient disponibles. Dans certains cas, en effet, le thème de base correspond sans ambiguïté soit au thème qui précède un suffixe ethnique (*champenophone*, *néerlandophone*) soit à un nom géographique (*allemagnophone*, *portugalophone*). Ainsi, s'il existe un nom en -PHONE construit sur une base supplétive savante, qu'elle soit ambiguë (8a) ou pas (8b-c) par rapport à un autre thème, on peut rencontrer des variantes qui font prévaloir la fidélité à la forme libre du nom de la langue (souvent homophone à un ethnique) et/ou d'un nom de pays :

(8)   a.   italophone   italianophone
      b.   germanophone   allemandophone, allemagnophone
      c.   hispanophone   espagnolophone, espagnophone

Considérons maintenant les données de l'italien. Le premier fait à remarquer est que tous les lexèmes présents dans le corpus comportent, avant la séquence [fon], un [o] qui porte l'accent tonique de mot, et ont donc la structure [X'ɔfono][17]. L'exposant possède

---

[17]La hauteur des voyelles moyennes n'est pas importante dans ce contexte, puisqu'elle est phonologiquement déterminée par la place de l'accent. Pour avoir une représentation phonologique complète, j'indique la forme du masculin singulier (finale en -*o*), mais ce qui suit s'applique à toutes les formes fléchies des lexèmes en -FONO (finales en -*a*, -*i*, -*e*).

donc une forme fixe à la fois plus contrainte et plus longue qu'en français. Puisque, je le rappelle, je considère qu'un exposant est simplement une séquence phonologique associée de façon arbitraire à une construction, on ne doit pas nécessairement chercher des raisons qui expliquent la plus grande rigidité de l'italien par rapport au français dans la forme de celui-ci. Il est possible, néanmoins, qu'une des raisons réside dans le fait que l'italien tolère moins bien une variation sur une voyelle qui porte l'accent primaire de mot, même si l'on peut remarquer que cette voyelle n'est pas toujours [ɔ] lorsque le dérivé n'est pas un nom de locuteurs (*telèfono*, *vibràfono*).

Du point de vue prosodique, on observe une plus grande dispersion des formats possibles, avec une prédominance du format pentasyllabique, mais avec presque autant de lexèmes à quatre ou à six syllabes, comme le montre la figure 3.
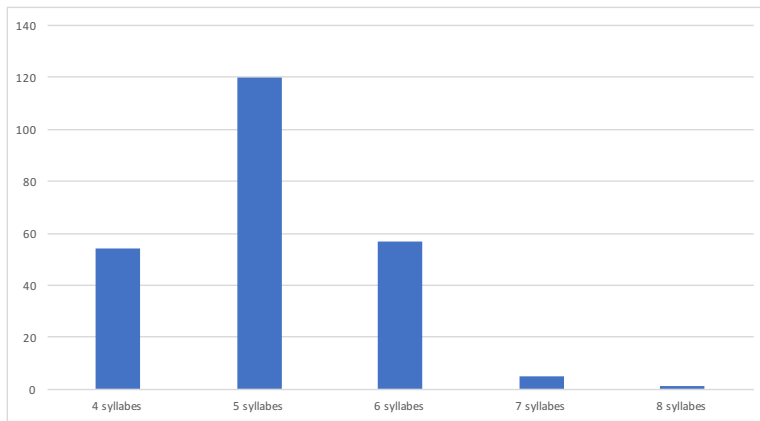
Figure 3 : Distribution des lexèmes en -FONO selon le nombre de syllabes

Comme il a déjà été observé dans d'autres cas, la contrainte de taille est donc moins forte en italien qu'en français, et elle est certainement soumise à la contrainte de fidélité base-dérivé. Concernant l'interaction entre la base et l'exposant, le cas par défaut en italien est celui dans lequel la séquence [ɔfono] est directement accolée au thème de base, si celui-ci se termine en consonne (*amazighofono*, *yiddishofono*), ou bien – plus fréquemment – la voyelle finale de la base est effacée (*bantofono*, *ligurofono*, *quechuofono*). À eux seuls, ces cas couvrent exactement deux tiers des dérivés de la base (158 sur 237), auxquels nous pouvons rajouter 22 cas dans lesquels la base est un thème supplétif d'origine savante. 75,9% des dérivés ne posent donc aucun problème particulier, ni pour le choix du thème de base, ni pour l'interaction phonologique entre ce thème et l'exposant. Concernant le phénomène d'effacement de la voyelle finale en dérivation en italien[18], deux hypothèses sont possibles, dans un cadre de morphologie thématique basée sur les contraintes : i) le thème sélectionné est un thème dépourvu de voyelle, le même que l'on retrouve dans d'autres dérivés, qui est sélectionné en respectant la contrainte de

---

[18]Cf. Montermini (2010) pour une discussion.

famille ; ii) le thème sélectionné est un thème qui contient une voyelle (par exemple un thème qui coïncide formellement avec une des formes fléchies), qui est effacée sous l'effet d'autres contraintes, par exemple une contrainte phonologique anti-hiatus. Les deux hypothèses en question ne sont pas nécessairement inconciliables. La première peut être valable pour les bases qui appartiennent à des familles lexicales nombreuses, alors que pour les autres il est plus difficile d'imaginer qu'un thème sans voyelle soit déjà présent dans le lexique. De plus, comme j'ai essayé de le montrer dans des travaux précédents (Montermini 2003, 2010), l'effacement de voyelle en dérivation est un phénomène qui, au moins en partie, est aussi influencé par la phonologie, avec des voyelles qui sont plus facilement effaçables que d'autres. Dans la base de données considérée ici on retrouve en effet deux exemples de non-effacement de voyelle, *bantuofono* et *urduofono* (qui co-existent avec les formes plus « régulières » *bantofono* et *urdofono*). Le fait que dans les deux cas la voyelle non effacée est un [u] n'est peut-être pas un hasard, puisqu'il s'agit de la voyelle qui en général résiste plus à l'effacement en italien (cf. les travaux cités ci-dessus).

En ce qui concerne le petit quart de dérivés restants, la quasi-totalité présentent des réductions du thème et peuvent être répartis en deux groupes. Les deux contiennent majoritairement des lexèmes qui sont des variantes d'autres lexèmes construits plus « régulièrement ». Le premier groupe, plus nombreux (45 lexèmes), correspond au cas déjà relevé pour le français dans lequel un lexème en -FONO est construit à partir d'un thème qui sert aussi de base à des gentilés et/ou à un nom géographique. Comme en français, on y retrouve de nombreux cas dans lesquels le thème de la base est ambigu de ce point de vue (9a), ainsi que des cas, plus rares, dans lesquels le thème est sans ambiguïté soit un thème de gentilés (9b), soit un nom géographique (9c) :

(9)   a.   islandofono, milanofono
        b.   portogofono
        c.   polonofono

Le deuxième groupe, plus restreint (5 lexèmes au total), comprend des dérivés dans lesquels le thème est réduit au format bisyllabique, indépendamment de sa structure morphologique (*albofono*, *estofono*, *lettofono*). Cette tendance, marginale, à avoir des bases bisyllabiques (et donc des dérivés quadrisyllabiques) doit très probablement être attribuée à la tendance que présentent les éléments de composition d'origine néoclassique, surtout initiaux, à être bisyllabiques en italien (cf. Thornton 2007 : 253–259). Il est possible que, pour certains locuteurs, un nom en -FONO doive encore se conformer au format d'un composé néoclassique (peut-être sur l'exemple des dérivés dans lesquels la base est un thème savant). Cependant, vu le nombre de lexèmes concernés, il s'agit d'une tendance minoritaire, voire résiduelle, ce qui peut être considéré comme une preuve indirecte du fait qu'en synchronie ces formations tendent à être manipulées par les locuteurs comme des dérivés affixaux à part entière. À la différence du français, il est difficile d'établir une corrélation précise entre ces réductions du thème de base et une quelconque contrainte prosodique, puisque, comme nous l'avons vu, les contraintes de taille sont moins importantes en italien, et probablement subordonnées aux contraintes de fidélité.

Pour conclure sur l'analyse de l'italien, les deux contraintes qui semblent prévaloir dans la construction des noms en -ꜰᴏɴᴏ sont la contrainte sur la forme des dérivés, qui unit, en réalité, plusieurs contraintes segmentales et prosodiques, et qui stipule qu'ils doivent avoir la structure [X'ɔfono], sans contrainte forte sur le nombre de syllabes, et la contrainte de fidélité base-dérivé. Ceci entraîne une tendance moins grande qu'en français à modifier les thèmes des bases pour satisfaire des contraintes prosodiques ou segmentales.

Ce que la comparaison entre les deux langues montre est que des constructions apparemment similaires, dans le processus de leur intégration aux systèmes phonologiques et morphologiques des langues en question, peuvent en réalité se développer comme des jeux de contraintes agencées de manière différente. L'italien a développé une construction dans laquelle la forme de l'exposant est fortement contrainte et la fidélité entre la base et le dérivé prime sur les autres contraintes formelles, alors que les contraintes prosodiques de taille ont moins de poids. En français, en revanche, ces contraintes jouent un rôle important, comme dans les autres procédés affixaux, ce qui, combiné à la contrainte de fidélité base-dérivé, entraîne une diversification des structures segmentales possibles pour l'exposant, qui, s'il contient toujours de préférence une voyelle étymologique de timbre /o/ à la jonction entre le thème de la base et l'exposant, admet d'autres voyelles, voire d'autres segments dans la même position.

3.2   -ɪssɪᴍᴏ et -ɪssɪᴍᴇ

La deuxième étude de cas concerne un suffixe du français qui n'a pas encore suscité, à ma connaissance, l'intérêt des linguistes et des lexicographes. Il s'agit du suffixe -ɪssɪᴍᴏ, que l'on retrouve notamment dans la construction de noms d'enseignes, événements, marques ou produits, les plus connus étant probablement *Colissimo* et *Doctissimo*[19]. Cependant, on peut également repérer des contextes dans lesquels des lexèmes en -ɪssɪᴍᴏ sont créés et employés en discours par les locuteurs, comme les suivants[20] :

(10)   a. Enfin bref je suis tout le contraire de ce qu'il aime c'est ça le plus drolissimo.
           [Twitter, 4 novembre 2013]

       b. J'ai un « torticolissimo ». C'est-à-dire que mon cou est coincé depuis 3 semaines et que personne ne sait quand la situation sera débloquée.
           [Twitter, 26 mai 2015]

---

[19]L'ensemble des lexèmes en -ɪssɪᴍᴏ cités dans cette section est donné en Annexe, avec une indication de leur signification dans les contextes dans lesquels ils ont été repérés.

[20]Il est possible que pour ces emplois de lexèmes en -ɪssɪᴍᴏ en discours les contraintes catégorielles et sémantiques pèsent plus lourd que les contraintes formelles par rapport à ceux qui servent de dénominations commerciales, en les rendant, de ce point de vue, plus proches des lexèmes en -ɪssɪᴍᴇ (et des autres lexèmes construits « canoniques »). Cependant, j'ai recensé trop peu d'exemples de ce type pour pouvoir tirer des conclusions fiables. Si cela est vrai, l'ordonnancement des contraintes serait également influencé par des paramètres externes à la morphologie liés à l'emploi pragmatique et sociolinguistique des lexèmes construits. (Je remercie le relecteur de cet article pour m'avoir fait réfléchir sur ce point).

   c. C'était vraiment énorme ! L'entrée de Médine énorme. Daniel Allouche
      (speaker), énorme. Le public havrais, énormissimo (sic). [http://www.
      lebannerofficial.com/index.php?option=com_content&task=view&id=355]

Les lexèmes ci-dessus occupent une position canonique de noms ou adjectifs, et véhi-
culent un sens génériquement appréciatif / superlatif. En ce sens, le suffixe en question
est proche du suffixe -ISSIME, qui a la même origine, mais une histoire différente. Les deux
sont issus du suffixe latin superlatif -*issimus*. Selon les dictionnaires, -ISSIME est rentré en
français via l'italien à partir du XIVᵉ siècle, d'abord via des mots d'adresse comme *séré-*
nissime (Perko 2010). En ce qui concerne -ISSIMO, son origine italienne est rendue encore
plus évidente par la voyelle [o] finale (on peut d'ailleurs considérer qu'il possède une va-
riante en [a], par exemple dans *Diorissima*, *Naturissima*, etc.). Sans en avoir la certitude,
je présume que sa disponibilité en français a été renforcée par l'existence d'un certain
nombre de mots du vocabulaire musical directement empruntés à l'italien (*fortissimo*,
pianissimo, etc.). Les premières attestations que j'ai pu documenter remontent à la se-
conde moitié des années 1960 : *Vernissimo* apparaît dans le slogan d'une annonce de
vernis pour ongles de 1966, *Parfumissimo* dans une annonce de savons de 1969 et *Ero-*
tissimo est le titre d'un film de 1969. Comme je l'ai montré ci-dessus, le suffixe, d'abord
employé dans des dénominations, a partiellement pénétré dans la langue courante. Il est
intéressant de remarquer que, si parfois il est employé dans des contextes spécifiques à
la réalité italienne (ou plus généralement « latine »), ceci n'est absolument pas systéma-
tique, comme le montre en particulier la troisième attestation de (10).

Pour cette étude, j'ai rassemblé une base de données de 294 lexèmes. Comme dans
le cas des -PHONE, la base a été recueillie en rassemblant en premier lieu les mots se
terminant par les séquences <issimo> ou <issima> dans FrWac. Ici aussi, la liste a été
nettoyée manuellement ; de plus, le contexte de chaque forme a été vérifié afin d'élimi-
ner les nombreux exemples provenant de pages écrites en italien ou en latin ramassées
par FrWac. Également, tous les mots du vocabulaire musical auxquels j'ai fait allusion
ci-dessus, ainsi que d'autres qui étaient clairement des emprunts directs (par exemple
campionissimo) ont été éliminés. Pour terminer, la liste a été complétée par des mots en
-ISSIMO provenant de différentes sources[21], et par des recherches ciblées sur le Web. Pa-
rallèlement, j'ai rassemblé une liste de 373 lexèmes en -ISSIME présents dans FrWac, que
je compare à ceux en -ISSIMO.

Concernant tout d'abord ce dernier suffixe, il s'attache principalement à des adjec-
tifs ou des noms pour former des superlatifs[22]. Du point de vue formel, Plénat (2002) a
identifié au moins quatre paramètres pour définir son comportement :

   i) il s'agit d'un suffixe « mi-savant », qui peut sélectionner tant des thèmes savants
      que des thèmes populaires (*universalissime* vs. *naturellissime*) ;

  ii) le suffixe -*ique* peut être effacé devant -ISSIME (*catholissime*, *nostalgissime*) ;

---

[21]Une liste contenant de nombreux mots en -ISSIMO m'a été fournie par ma collègue Antonella Capra, que je
remercie.

[22]Dans toute la base on ne trouve qu'un seul lexème en -ISSIME qui est indubitablement construit sur un mot
qui n'est ni un nom ni un adjectif : *obligatoirementissime*.

iii) la rime (voyelle et consonne finales) tombe si le thème de base est tri- ou quadri-syllabique et se termine par une consonne sifflante latente (*bruxellissime, rigouris-sime*) ;

iv) la rime tombe si le thème de base est tri- ou quadrisyllabique et se termine par un [i] suivi d'une consonne latente (*favorissime, interdissime*).

Les propriétés i) et ii) captent plutôt des tendances que des règles. La deuxième en particulier connaît plusieurs exceptions (*critiquissime, sympathiquissime*), pour lesquelles Plénat fait l'hypothèse que les lexèmes en question ont sélectionné un thème populaire, alors que ce sont les thèmes savants qui perdent la séquence [is] devant -ISSIME pour respecter la contrainte de dissimilation (*catholissime* vs. **catholicissime*). Les données de FrWac semblent indiquer, dans ce cas, que -ISSIME tend plutôt à sélectionner des bases populaires : sur 23 dérivés construits sur des bases qui possèdent un thème L distinct des thèmes A et B, 17 utilisent le thème A ou B (*lamentablissime, sensuelissime, supérieu-rissime*), et seulement 6 utilisent le thème L (*formidabilissime, prétenciosissime*). Concernant les deux paramètres iii) et iv), les données tirées de FrWac potentiellement concernées sont extrêmement rares, mais semblent tout de même confirmer les hypothèses formulées par Plénat. Dans l'ensemble de la base de données, on retrouve seulement trois lexèmes dans lesquels les tendances identifiées ne sont pas vérifiées : *andalousissime, pré-tenciosissime* (iii) et *favoritissime* (iv). On peut tout de même observer que, si ces lexèmes ne respectent pas les contraintes phonologiques (dissimilatives) qui sont à l'origine des principes en question, ils respectent entièrement la fidélité base-dérivé. Dans la base, on retrouve également quatre lexèmes qui correspondent à des cas de surapplication des règles ci-dessus, c'est-à-dire des effacements qui ont eu lieu là où on ne les aurait pas attendus : *Barbérissime, Optalissime* (iii), *splendissime* et *sublissime* (iv). Les deux derniers sont déjà discutés par Plénat ; concernant les deux premiers, il s'agit d'hapax construits, respectivement, sur le nom propre *Barbéris* et sur *Optalis*, qui est le nom commercial d'une série de produits financiers. À propos des cas d'effacement discutés par Plénat, cependant, il est intéressant d'observer un autre fait. L'effet des effacements en question est que le radical sur lequel le dérivé en -ISSIME est construit est presque toujours identique à des thèmes de la famille dérivationnelle de la base, qui dans la plupart des cas correspondent au thème d'un lexème autonome. C'est le cas des exemples *bruxellissime* et *nostalgissime*, et également des dérivés *prestigissime* et *ténébrissime*, présents dans Fr-Wac. Dans une perspective plus actuelle, les cas en question pourraient probablement être expliqués en termes de sélection de thème plutôt qu'en termes d'effacement. Notons tout de même, pour terminer, qu'un effacement a certainement lieu dans plusieurs cas lorsque la base se termine par une voyelle, et notamment par [e], cas où, dans les données de la base (six concernées au total), il est systématique (*branchissime, pavissime ←pavé*). Globalement, en tout cas, les modifications des thèmes des bases restent extrêmement rares dans la base de données. Au total, elles ne concernent que 32 lexèmes (moins de 10% de la base), distribués comme il suit :

(11)   a. effacement d'une rime voyelle + sifflante : 5 (*prestigissime*)

   b. effacement d'une rime [i] + consonne : 4 (*érudissime*)

    c. effacement du suffixe -*ique* : 4 (*exotissime*)

    d. effacement d'une voyelle finale : 14 (*branchissime*, *pourrissime*)

    e. épenthèse : 5 (*absolutissime*, *merveilleutissime*)

Concernant la structure prosodique des dérivés présents dans la base, la distribution est semblable à celle observée pour les dérivés italiens en -ꜰᴏɴᴏ, avec une prédominance du format quadrisyllabique, mais avec une dispersion des dérivés entre les formats tri-, quadri- et pentasyllabique. La distribution des dérivés de la base de données selon le nombre des syllabes est résumée dans la figure 4.
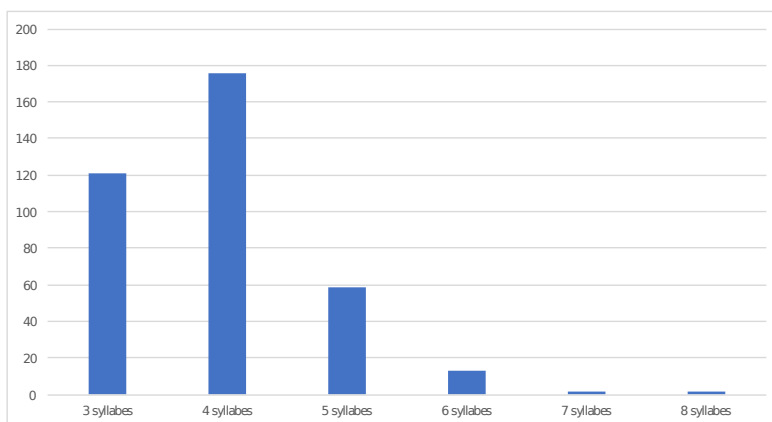
Figure 4 : Distribution des lexèmes en -ɪssɪᴍᴇ selon le nombre de syllabes

Une telle distribution peut être corrélée à la rareté des cas de manipulation des thèmes de base qui a été observée ci-dessus. La concomitance entre ces deux facteurs semble en effet suggérer que les contraintes de taille, si elles sont actives, sont subordonnées à la contrainte de fidélité base-dérivé : la taille des dérivés dépend alors plus de la taille des bases (dont la longueur en syllabes est distribuée de façon aléatoire) que de manipulations réalisées sur les thèmes.

Penchons-nous à présent sur les lexèmes en -ɪssɪᴍᴏ. La première observation que nous pouvons formuler à leur égard concerne les propriétés catégorielles et sémantiques du suffixe en question. Comme je l'ai observé plus haut, à côté des cas « canoniques », comme ceux exemplifiés en (10), -ɪssɪᴍᴏ sert souvent à construire des dénominations d'enseignes commerciales, événements, marques, produits, etc., ainsi que des occasionalismes destinés à être employés dans des slogans. Sa valeur sémantique se limite donc dans la plupart des cas à une valeur connotative superlative, voire génériquement positive. Les bases potentielles pour ce suffixe sont donc moins contraintes du point de vue sémantique, et même catégoriel ; parfois, au contraire, la sélection de la base (ou du thème de la base retenu) semble être faite plutôt à partir de sa compatibilité formelle avec la construction que de sa compatibilité sémantique. Une première conséquence de ce fait est que les bases potentielles de -ɪssɪᴍᴏ sont beaucoup plus variées que celles de

-ISSIME, y compris du point de vue catégoriel. La base de données comprend par exemple 5 lexèmes qui sont quasiment sans ambiguïté construits sur des verbes (12a), ainsi qu'au moins 13 lexèmes pour lesquels décider si la base est un verbe ou un nom (plus rarement un verbe ou un adjectif) est difficile (12b).

(12)   a. Courissimo, Jonglissimo, Repassimo[23]

         b. Agrandissimo, Investissimo, vomissimo

Si l'on voulait privilégier l'homogénéité catégorielle, on pourrait penser que, parmi les mots de (12b), *Agrandissimo* et *Investissimo* sont construits sur les noms *agrandissement* et *investissement*, et *vomissimo* sur *vomi*; au contraire, si l'on veut privilégier la transparence formelle, on peut imaginer que ces lexèmes sont construits, à partir des thèmes du verbe disponibles, sur celui qui est le plus compatible avec les contraintes imposées par la construction (dans ce cas, le Thème 1, celui se terminant en [is] pour les verbes du deuxième groupe). Dans l'analyse, j'ai choisi d'adopter cette deuxième solution, et j'ai donc décidé de considérer que les lexèmes en question (et les autres semblables) sont construits sur un verbe, dont un des thèmes est sélectionné[24]. Ce choix semble justifié par le fait que, dans d'autres cas, le radical sélectionné pour la dérivation en -ISSIMO pourrait correspondre à un des thèmes disponibles dans l'espace thématique, choisi soit en vertu de sa compatibilité phonologique avec l'exposant, soit d'autres facteurs. Des cas comme *Linguissimo*, *optimissimo* ou *scientissimo* ne peuvent, me semble-t-il être analysés que comme ça.

Une deuxième observation concerne la séquence finale de ces dérivés. À la différence de -ISSIME, dans les lexèmes dérivés par -ISSIMO la séquence [simo] peut être précédée d'une voyelle différente de [i], notamment [a], [e], [o] et [y]. Au total, 24 lexèmes de la base de données sont concernés :

(13)   a. Pizzassimo, Prépassimo

         b. Bébéssimo, Cinessimo

         c. Dodossimo, Vélossimo

         d. Revenussimo

À ce point, je pense qu'il est clair que la meilleure manière de rendre compte de cette variabilité dans le modèle adopté ici est de l'attribuer à une allomorphie de l'exposant, et que le choix de la voyelle dépend d'un segment présent dans la base. Ce point sera développé ci-dessous.

Du point de vue de la sélection du thème de base, mis à part les cas d'incertitude mentionnés ci-dessus, -ISSIMO semble se comporter, comme -ISSIME, en suffixe mi-savant, même si les données sont trop rares pour pouvoir tirer des conclusions probantes. Sur 9 lexèmes construits sur des bases qui comportent un thème L distinct des thèmes A et

---

[23]*Repassimo* est le nom d'un pressing, et est donc très vraisemblablement construit sur *repasser*.

[24]Un cas légèrement plus complexe, mais qui peut recevoir la même explication, est celui des dérivés construits sur le thème 13 d'un verbe (cf. Bonami et al. 2009), par exemple *Locatissimo*, *Nutrissimo*, *Sélectissimo*.

B, 5 utilisent le thème L (par exemple *Urbanissimo*, *Valorissimo*) et 4 utilisent un thème
A homophone du thème autonome (*formidablissimo*, *incroyablissimo*) ; 9 autres utilisent
un thème supplétif d'origine savante (*altissimo*, *Equissimo*, *Historissimo*).

Du point de vue des modifications que subissent les thèmes des bases, quasiment au-
cun exemple dans la base ne permet de confirmer les observations proposées par Plénat
(2002) pour -ɪssɪᴍᴇ (cf. (11)), mis à part 4 dérivés d'un adjectif en *-ique* où ce dernier
suffixe est, comme dans les dérivés en -ɪssɪᴍᴇ, effacé :

(14) a. Authentissima
      b. Erotissimo
      c. Olympissimo
      d. Optissimo

Lorsqu'on compare les bases de données en -ɪssɪᴍᴇ et en -ɪssɪᴍᴏ, cependant, le fait le
plus frappant est certainement la grande proportion de thèmes de bases qui ont subi une
modification dans cette dernière. Au total, en effet, 124 dérivés en -ɪssɪᴍᴏ sur 294 (42,1%)
présentent une modification de la base (presque uniquement des réductions), alors que
pour les lexèmes en -ɪssɪᴍᴏ, je le rappelle, cette proportion était de 10%. En (15) je donne
le détail des types de modifications rencontrées :

(15) a. effacement d'une rime voyelle (≠[i]) + sifflante : 11 (*dégueulassimo*,
         Promessimo, *Revenussimo*)
      b. effacement d'une rime [i] + consonne[25] : 52 (*Apéritissimo*, *Jurissimo*,
         Permissimo, *Tennissimo*)
      c. effacement d'une voyelle finale : 61 (*Bébéssimo*, *Espérantissimo*, *Pizzassimo*)

Il est notable, d'ailleurs, que pratiquement toutes les bases qui appartiennent à un des
types (15a-c) sont réduites. Les quatre seules exceptions sont *Blingissimo* (qui possède une
base monosyllabique), *Bijoutissimo* (dont le thème est employé par ailleurs, par exemple
dans *bijoutier*), *Caféissimo* et *successissimo*, qui coexistent, dans la base, avec *Caféssimo*
et *successimo*. On peut aussi remarquer qu'à la différence de ce qui a été observé par
Plénat pour -ɪssɪᴍᴇ, la longueur du thème de base ne semble pas avoir une incidence
particulière sur ses chances d'être modifié, puisque peuvent être réduits des thèmes de
longueur différente, y compris des monosyllabiques (cf. *Tassimo* ← *tasse*, nom d'une
marque de café).

Chacun des types présentés en (15) mérite d'être observé dans le détail. Parmi les bases
dont le thème comporte, en finale, une voyelle différente de [i] et une consonne (latente
ou pas), un seul (*anglissimo*) présente un exposant où apparaît la voyelle [i]. Dans tous
les autres, à l'instar de ceux exemplifiés ci-dessus, la voyelle qui précède [simo] est la
même qui apparaît dans le thème. Parmi les bases en [i] + consonne, 34 se terminent par
une sifflante (ou par la séquence [st], comme dans *Jurissimo*, qui est le nom d'un cabinet
d'avocats), et 18 se terminent par une autre séquence, presque toujours une consonne.

---

[25]Ce chiffre comprend les 4 bases en *-ique* mentionnées en (14).

Dans quelques cas, cependant, le thème de base est coupé en correspondance d'un [i] qui précède une séquence plus longue qu'une simple consonne :

(16)   a. Apprentissimo
         b. Acquissimo
         c. narcissimo
         d. Numissima
         e. Ravissimo

Les deux premiers exemples, en particulier, sont intéressants. D'un certain point de vue, ils sont parallèles aux exemples de *Agrandissimo* et *Investissimo* vus en (12b), puisqu'ils dérivent de deux noms d'action (*apprentissage*, *acquisition*), mais la base employée dans ces cas ne correspond pas à un des thèmes du verbe. En ce qui concerne le type (15c), plus de la moitié des thèmes en voyelle se terminent par [i], et les autres se distribuent comme indiqué dans le tableau 1.

Tableau 1 : Distribution des voyelles finales effacées (voir 15c)

Voyelle	Effectif
[i]	34
[o]	12
[a]	7
[e]/[ɛ]	7
[y]	1

Lorsque la voyelle finale de la base est un [i], l'exposant a évidemment toujours la forme [isimo]. Lorsqu'il s'agit d'une voyelle différente, l'exposant a également la forme [isimo] dans un tiers des cas (9 sur 27, par exemple *Espérantissimo*) et une forme où [simo] est précédé par la même voyelle que celle qui apparaît dans la base dans les deux tiers restants (*Bébéssimo*, *Pizzassimo*).

Dans le tableau 2, je détaille les chiffres présentés en (15), en donnant la distribution des thèmes réduits selon la séquence sujette à réduction :

Que suggère l'ensemble de ces données ? En premier lieu, me semble-t-il, il suggère que la forme de l'exposant ne possède pas un segment vocalique fixe comme dans le cas de -ISSIME. À l'instar de ce que j'avais proposé pour -PHONE, on peut considérer que l'exposant de la construction en -ISSIMO possède une forme par défaut [isimo] et une forme subordonnée [Vsimo], dont l'émergence dépend crucialement de la contrainte de fidélité base-dérivé. Plus précisément, du point de vue segmental, cette construction impose les deux contraintes hiérarchisées [Xisimo] > [Visimo] sur la forme de ses dérivés. Du point de vue prosodique, également, nous pouvons observer un comportement partiellement différent de celui de la construction en -ISSIME, pour laquelle j'ai argumenté que

Tableau 2 : Distribution des types de séquences finales dans les thèmes réduits

Finale	Effectif
[i]+sifflante	34
[i]	34
V≠[i]	27
[i]+consonne	18
V≠[i]+sifflante	11

les contraintes de taille jouent un rôle moindre que dans d'autres procédés construction-nels en français, et en particulier qu'elles sont subordonnées à la contrainte de fidélité base-dérivé. En ce qui concerne -ISSIMO, la distribution des dérivés selon le nombre de syllabes est celle donnée dans la figure 5.
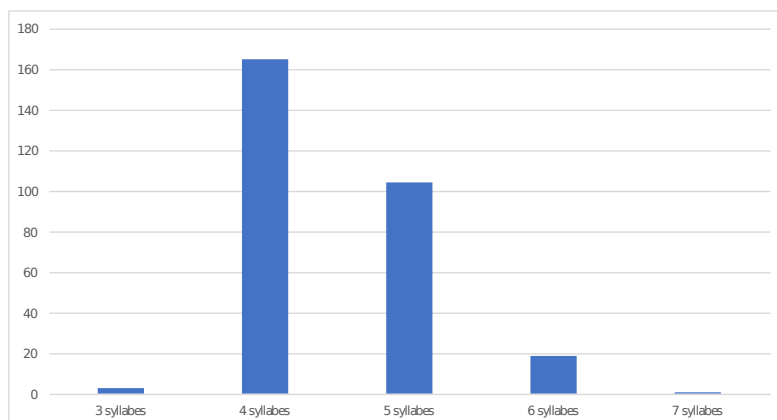
Figure 5 : Distribution des lexèmes en -ISSIMO selon le nombre de syllabes

Dans l'interprétation de ces chiffres, il faut considérer que, puisque -ISSIMO se termine par voyelle, un dérivé quadrisyllabique correspond à un dérivé trisyllabique en -ISSIME, un pentasyllabique à un quadrisyllabique, etc. En prenant en compte cette différence, les deux formats les plus fréquents pour -ISSIME (trois et quatre syllabes) représentent 79,6% des cas (cf. la figure fig :Montermini :4), alors que pour -ISSIMO les deux formats les plus fréquents (quatre et cinq syllabes) représentent 91,4% des cas. Il semble donc que la contrainte de taille soit plus forte pour -ISSIMO que pour -ISSIME, ce qui expliquerait la plus grande tendance de cette construction à modifier les thèmes de base sélectionnés en les réduisant. Cette tendance que l'on observe pour -ISSIMO a cependant, également, une autre explication, complémentaire à celle que je viens de proposer. Dans le tableau 3, je récapitule le nombre de bases qui subissent une modification (réduction) du thème pour -ISSIME et -ISSIMO, en le comparant au nombre de bases totales qui présentent les condi-tions pour une telle modification (rime en voyelle + sifflante, rime en [i] + consonne,

finale vocalique). Le premier chiffre indique le nombre de bases potentiellement modifiables, le deuxième le nombre de bases qui sont effectivement modifiées :

Tableau 3 : Nombre de bases subissant une modicication

	Bases modifiables		Bases modifiées	
-ISSIME	99	(26,5%)	28	(28,3%)
-ISSIMO	128	(43,5%)	124	(96,9%)

Ces chiffres nous disent fondamentalement deux choses : premièrement, dans la dérivation en -ISSIMO une base potentiellement modifiable est quasi systématiquement modifiée ; deuxièmement, dans cette dérivation les bases qui présentent une structure segmentale compatible avec une réduction (et donc un amalgame avec l'exposant) sont surreprésentées par rapport à celle en -ISSIME, pour laquelle nous pouvons considérer que la distribution des bases, principalement sélectionnées sur base catégorielle et sémantique, est aléatoire du point de vue phonologique[26]. Cette surreprésentation est justement due à la faible sélection qu'opère -ISSIMO sur ses bases du point de vue catégoriel et sémantique, ce qui laisse la place pour que la phonologie y joue un rôle plus important. Peu importe que -ISSIMO constitue, de ce point de vue, une construction non canonique – la plupart des affixes, en effet, privilégient les propriétés catégorielles et sémantiques dans la sélection de leurs bases. Ce que ces données, et leur interprétation, mettent en lumière, en effet, est une des voies que la morphologie peut prendre dans la conventionnalisation des propriétés (dans ce cas formelles) qui sont associées à ses constructions.

Pour conclure, on peut considérer que les contraintes formelles attachées à la construction en -ISSIMO sont les suivantes : le dérivé doit avoir la forme [Xisimo] > [XVsimo] (où les formes possibles pour l'exposant sont hiérarchisées) ; le dérivé doit comporter quatre ou cinq syllabes ou, à défaut, six syllabes ou un nombre supérieur. On peut également imaginer que les contraintes catégorielles et sémantiques de sélection de la base sont remplacées par des contraintes de sélection formelle que l'on peut formuler et ordonner ainsi :

Une base optimale pour un dérivé en -ISSIMO :

— est bi- ou trisyllabique ;

— se termine par un [i] ou par une séquence [i]+sifflante ;

— est plus que trisyllabique et contient une séquence [i]+sifflante à la deuxième ou à la troisième syllabe ;

— se termine par un [i] suivi d'une autre consonne ;

— se termine par une voyelle différente de [i], suivie ou non d'une sifflante.

---

[26]Le nombre de bases potentiellement modifiables est même surestimé dans le tableau 3 pour -ISSIME, puisqu'ici aucune distinction n'est faite entre les bases bisyllabiques et les bases plus que bisyllabiques qui sont les seules, selon Plénat, qui peuvent subir un effacement d'une rime complexe.

Comme on le voit, les contraintes qui occupent une place moins élevée dans la hiérarchie manifestent des relâchements d'une des propriétés spécifiées par les deux premières, soit sur le nombre de syllabes, soit sur le timbre de la voyelle, soit sur la nature de la consonne de la rime.

Pour rappel, j'ai considéré plus haut que les constructions en -ISSIME, de leur côté, sont soumises, du point de vue de la sélection des bases, à des contraintes catégorielles et sémantiques semblables à celles qui opèrent pour les autres constructions affixales canoniques. Du point de vue segmental, cette construction spécifie uniquement que le dérivé doit avoir la forme [Xisim] ; du point de vue prosodique, si des contraintes de taille existent, elles sont soumises aux contraintes de fidélité base-dérivé.

4 Conclusion

La prise en compte des écarts entre la forme attendue et celle réellement observée des lexèmes morphologiquement complexes est un des domaines dans lesquels la recherche en morphologie, sur le français et sur d'autres langues, a le plus évolué dans les dernières décennies. Ceci s'est traduit, d'une part, par la reconnaissance des lexèmes comme des structures complexes auxquelles peuvent correspondre, synchroniquement, plusieurs thèmes, des représentations formelles qui sont irréductibles, mais connectées entre elles et organisées. Parmi les opérations que la morphologie (dérivationnelle) met en place lors de l'application d'une règle (ou construction) morphologique, il y a la définition d'un radical, c'est-à-dire la forme à laquelle est appliquée l'opération formelle spécifiée par la règle. Cette définition passe par la sélection d'un des thèmes du lexème de base et par d'éventuelles modifications phonologiques de celui-ci. Une façon de modéliser cet ensemble d'opération est de considérer qu'elles sont régies par un ensemble de contraintes, c'est-à-dire de spécifications des propriétés qu'un lexème dérivé doit avoir. Les contraintes peuvent être spécifiques à une langue (ou même à un secteur de la langue) ou bien universelles ; elles peuvent se renforcer mutuellement, ou bien se contredire, et dans ce cas la forme réellement observée pour un dérivé sera déterminée par la tendance à satisfaire une contrainte ou une autre, avec des issues potentiellement différentes lorsqu'une opération est appliquée à la même base. Les travaux qui se sont inspirés de ce modèle, cependant, se sont principalement intéressés à la variation thématique et aux facteurs qui en sont responsables ; la variation des exposants des constructions morphologiques (qui correspond à ce qui traditionnellement était vu comme l'allomorphie affixale), en revanche, a moins suscité leur intérêt. Pourtant, j'ai proposé des arguments forts pour soutenir que certains cas de variation formelle que l'on observe en dérivation ne peuvent pas être traités en termes de variation thématique. Il faut donc admettre que les exposants des constructions morphologiques peuvent aussi être sujets à variation, une variation qui mérite d'être prise en compte et, si possible, modélisée. Cela pose, tout d'abord, le problème d'identifier clairement les cas de variation d'un exposant au sein de la même construction des cas de constructions différentes qui, éventuellement, peuvent avoir une sémantique proche et des exposants formellement semblables. Pour considérer que deux formes sont des variantes du même exposant, il faut qu'elles soient

non seulement proches et si possible liées par des relations phonologiques naturelles, mais qu'elles apparaissent dans des lexèmes dérivés qui ont des propriétés catégorielles et sémantiques semblables (c'est-à-dire qui appartiennent à la même série), et surtout qu'elles soient en distribution complémentaire ou du moins que leurs contextes d'apparition soient clairement identifiables du point de vue phonologique. Pour modéliser la variation des exposants morphologiques, j'ai proposé d'étendre la notion de contrainte non seulement à une propriété qui est spécifique à une langue donnée, mais également à une construction donnée. Les exposants des constructions morphologiques peuvent alors être vus eux-mêmes comme des contraintes, ou des ensembles de contraintes, qui interagissent avec les autres contraintes en jeu dans la formation des lexèmes complexes. Chaque « allomorphe » d'un exposant est donc une contrainte qui, en tant que telle, peut être hiérarchisée par rapport aux autres, ce qui rend compte de l'observation que certaines de ces variantes jouent un rôle de défaut, alors que d'autres émergent uniquement dans des conditions particulières[27].

Afin d'illustrer le modèle que je propose, j'ai réalisé deux études de cas de constructions morphologiques de naissance ou développement récent. Ce travail se place, en effet, dans une approche extensive de la morphologie, dans laquelle est essentielle la prise en compte d'un nombre important de données et, si possible, de données qui manifestent la pratique réelle de construction des mots par les locuteurs. C'est pour cela que l'observation des nouvelles formations, néologismes, occasionalismes, etc. est tout aussi importante, sinon plus, que l'observation du lexique établi. Les deux constructions que j'ai considérées sont la création de noms de locuteurs en -PHONE à partir du nom d'une langue et la création de lexèmes avec un sens génériquement appréciatif / superlatif en -ISSIMO. La première a la particularité de prendre comme bases aussi bien des noms de langues qui appartiennent à des réseaux lexicaux nombreux, et pour lesquels la sélection est donc un enjeu, et des noms de langues qui n'entretiennent aucun lien lexical, ou très peu, qui peuvent donc être sujets à des modifications destinées à en faire de « bons » radicaux pour la construction en question. La deuxième, à cause de sa valeur pragmatique, définit peu de contraintes catégorielles et sémantiques sur ses bases potentielles, qui sont, en revanche, plutôt sélectionnées sur une base formelle, selon leur compatibilité avec les contraintes segmentales qui en définissent l'exposant. Chacune de ces deux constructions fait également l'objet d'une comparaison. La dérivation en -PHONE est comparée à la dérivation correspondante et cognate de lexèmes en -FONO en italien ;

---

[27]Un relecteur de l'article suggère, en alternative, de considérer qu'une construction peut comporter plusieurs variantes de l'exposant, dont le choix est déterminé par des contraintes de sélection (un système, à mon sens, semblable à celui proposé par Bonet et al. 2007 pour le créole haïtien et le catalan, qui prévoit, pour certains procédés morphologiques, l'existence d'un « catalogue » de variantes hierarchisées). Il est vrai que l'efficacité des contraintes a été déjà montrée pour la sélection du thème dans les procédés morphologiques constructionnels (cf. Plénat & Roché 2014, Boyé & Plénat 2015), et une telle hypothèse permettrait d'unifier l'analyse des deux. Cependant, il me semble qu'une telle hypothèse devrait être considérée, au mieux, comme une variante de l'hypothèse principale que je défends, pour au moins deux raisons : i) dans certains cas, comme celui de -PHONE, le « catalogue » des exposants correspondrait à une simple liste de formes largement redondante (dans le cas en question [fɔn] précédé de n'importe quelle voyelle et possiblement de plusieurs consonnes) ; ii) cette hypothèse ne permettrait pas de capter l'interaction entre la forme du thème de la base et l'exposant, un élément crucial de l'analyse proposée ici.

la dérivation en -ɪssɪᴍᴏ est comparée à la dérivation, plus canonique, de superlatifs en -ɪssɪᴍᴇ en français. Ces comparaisons mettent en lumière le fait que des constructions formellement et sémantiquement similaires et qui ont la même origine peuvent, dans des langues différentes ou dans la même langue à des époques et pour des finalités différentes, développer des spécifications phonologiques différentes, ce qui se traduit, dans le cadre adopté ici, par des ensembles de contraintes différentes et/ou agencées différemment. Je prends ce constat pour une démonstration du fait que l'exposant d'une règle morphologique correspond simplement à l'association arbitraire entre un ensemble de spécifications catégorielles et sémantiques et un ensemble de contraintes formelles.

Le modèle de morphologie à contraintes ouvre de nombreuses perspectives de recherche et de connexions potentielles avec des modèles théoriques proches (par exemple la Morphologie des Constructions). S'il a été jusqu'à présent appliqué presque uniquement au français, ce modèle mériterait d'être testé sur d'autres langues et sur des ensembles de données plus variés. Le travail que j'ai présenté constitue, je l'espère, un premier pas dans cette direction.

Remerciements

Annexe

Le tableau 4 liste tous les lexèmes en -ɪssɪᴍᴏ cités dans l'article avec une indication de leur emploi, tel qu'il a pu être identifié à partir des recherches effectuées. Si le lexème n'est suivi d'aucune indication, cela signifie qu'il a été repéré dans un emploi en discours et que sa signification correspond en gros au superlatif du lexème de base.

Références

Apothéloz, Denis. 2003. Le rôle de l'iconicité constructionnelle dans le fonctionnement du préfixe *in-*. *Cahiers de Linguistique Analogique* 1(1). 35–63.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge : MIT Press.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The wacky wide web : A collection of very large linguistically processed web-crawled corpora. In *Language resources and evaluation*, t. 43, 209–226.

Blevins, James P. & Juliette Blevins. 2009. *Analogy in grammar : Form and acquisition*. Oxford : Oxford University Press.

Bonami, Olivier. 2014. *La structure fine des paradigmes de flexion. Études de morphologie descriptive, théorique et formelle*. Paris : Université Paris Diderot-Paris 7 Mémoire d'Habilitation à diriger des recherches.

Tableau 4 : Liste des lexèmes en *-issimo*

Acquissimo (type de crédit immobilier)

Agrandissimo (constructeur immobilier)

Altissimo (magasin d'équipement pour escalade)

Anglissimo

Apéritissimo (bar à apéritifs)

Apprentissimo (salon sur l'apprentissage)

Authentissima (magasin de mobilier)

Bébéssimo (site de produits pour enfants)

Bijoutissimo (site de vente de bijoux)

Blingissimo (marque de bijoux)

Caféíssimo (huile au café)

Caféssimo (site de vente de café)

Cinessimo (nom d'une carte de crédit qui comporte des réductions au cinéma)

Colissimo (service de colis de la Poste)

Courissimo (compétition de course à pied)

Dégueulassimo

Diorissima (parfum de la marque Dior)

Doctissimo (site de médecine)

Dodossimo (pyjama pour enfants)

Drolissimo

Énormissimo

Equissimo (salon de chevaux)

Erotissimo (film, 1969)

Espérantissimo (site d'espéranto)

Formidablissimo

Historissimo (librairie)

Incroyablissimo

Investissimo (agence immobilière)

Jonglissimo (festival de jonglerie)

Jurissimo (cabinet d'avocats)

Linguissimo (concours de langues)

Locatissimo (site de location d'appartements)

Narcissimo

Numissima (entreprise de rachat de précieux)

Naturissima (salon sur la nature)

Nutrissimo (jeu de société sur la nutrition)

Olympissimo (jeu de société sur les Jeux Olympiques)

Optimissimo

Optissimo (chaîne d'opticiens)

Parfumissimo (slogan publicitaire, 1969)

Permissimo (site pour la récupération des points du permis de conduire)

Pizzassimo (sauce tomate pour pizza)

Prépassimo (école préparatoire)

Promessimo (type de contrat immobilier)

Ravissimo (machine à découper les raviolis)

Repassimo (pressing)

Revenussimo (site de conseils financiers)

Scientissimo (site d'activités scientifiques)

Sélectissimo (club d'affaires)

Successimo

Successissimo

Tassimo (machine à café)

Tennissimo (centre sportif)

Torticolissimo

Urbanissimo (site d'urbanisme)

Valorissimo (agence immobilière)

Vélossimo (association cycliste)

Vernissimo (slogan publicitaire, 1966)

Vomissimo

Bonami, Olivier & Gilles Boyé. 2003. Supplétion et classes flexionnelles dans la conjugaison du français. *Langages* 152. 102–126.

Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. In Florence Villoing, Sarah Leroy & Sophie David (éds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 17–45. Presses Universitaires de Paris Ouest.

Bonami, Olivier, Gilles Boyé & Françoise Kerleroux. 2009. L'allomorphie radicale et la relation dérivation-construction. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 103–125. Saint-Denis : Presses Universitaires de Vincennes.

Bonet, Eulália, Maria-Rosa Lloret & Joan Mascaró. 2007. Allomorph selection and lexical preferences : Two case studies. *Lingua* 117(6). 903–927.

Booij, Geert. 2010. *Construction morphology*. Oxford : Oxford University Press.

Boyé, Gilles & Marc Plénat. 2015. L'allomorphie radicale dans les lexèmes adjectivaux en français. Le cas des adverbes en -*ment*. In Eulália Bonet, Maria-Rosa Lloret & Joan Mascaró (éds.), *Understanding allomorphy. Perspectives from optimality theory*, 70–106. London : Equinox Publishing.

Burzio, Luigi. 2002. Surface-to-surface morphology : When your representations turn into constraints. In Paul Boucher (éd.), *Many Morphologies*, 142–177. Somerville : Cascadilla Press.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Bybee, Joan. 2013. Usage-based theory and exemplar representations of constructions. In *The Oxford handbook of construction grammar*.

Corbin, Danielle. 2001. Préfixes et suffixes : du sens aux catégories. *Journal of French Language Studies* 11(1). 41–69.

Dal, Georgette & Fiammetta Namer. 2012. Faut-il brûler les dictionnaires ? Ou comment les ressources numériques ont révolutionné les recherches en morphologie. In Franck Neveu, Valélia Muni Toke, Peter Blumenthal, Thomas Klinger, Pierluigi Ligas, Sophie Prévost & Sandra Teston-Bonnard (éds.), *Actes du 3e Congrès Mondial de Linguistique Française. Lyon, 4-7 juillet 2012*, 1261–1276. Paris : Institut de Linguistique Française.

Dal, Georgette & Fiammetta Namer. 2016. À propos des occasionalismes. In Franck Neveu, Gabriel Bergounioux, Marie-Hélène Côté, Jean-Marc Fournier, Linda Hriba & Sophie Prévost (éds.), *Actes du 5e Congrès Mondial de Linguistique Française. Tours, 4-8 juillet 2016*. Paris : Institut de Linguistique Française.

Fradin, Bernard. 2000. Combining forms, blends and related phenomena. In Anna M. Thornton & Ursula Doleschal (éds.), *Extragrammatical and marginal morphology*, 11–59. München : Lincom Europa.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie*. Paris : Presses Universitaires de France.

Hathout, Nabil. 2009. *Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie*. Toulouse : Université de Toulouse le Mirail Mémoire d'Habilitation à diriger des recherches.

Hathout, Nabil, Fabio Montermini & Ludovic Tanguy. 2008. Extensive data for morphology : Using the World Wide Web. *Journal of French Language Studies* 18. 67–85.

Hathout, Nabil & Fiammetta Namer. 2014. Discrepancy between form and meaning in word formation : The case of over- and under-marking in French. In Franz Rainer, Wolfgang U. Dressler, Francesco Gardani & Hans Christian Luschützky (éds.), *Morphology and meaning*, 177–190. Amsterdam : John Benjamins.

Hathout, Nabil, Fiammetta Namer, Marc Plénat & Ludovic Tanguy. 2009. La collecte et l'utilisation des données en morphologie. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 266–287. Saint-Denis : Presses Universitaires de Vincennes.

Koehl, Aurore. 2012. *Altitude*, *négritude*, *bravitude* ou la résurgence d'une suffixation. In Franck Neveu, Valéria Muni-Toke, Peter Blumenthal, Thomas Kingler, Pierluigi Ligas, Sophie Prévost & Sandra Teston-Bonnard (éds.), *3e congrès mondial de linguistique française - lyon, 4-7 juillet 2012* (Institut de Linguistique Française), 1307–1323. Paris.

Koehl, Aurore & Stéphanie Lignon. 2014. Property nouns with *-ité* and *-itude* : Formal alternation and morphopragmatics or the sad-itude of the Aité$_N$. *Morphology* 24(4). 351–376.

Lasserre, Marine. 2016. *De l'intrusion d'un lexique allogène. L'exemple des éléments néoclassiques*. Université de Toulouse 2 Jean Jaurès Thèse de doctorat.

Lasserre, Marine & Fabio Montermini. 2014. Pour une typologie des lexèmes construits : entre composition, composition néoclassique et affixation. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberger, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1797–1812. Paris : Institut de Linguistique Française.

Lignon, Stéphanie. 2013. *-iser* and *-ifier* suffixation in French : Verify data to verize hypotheses. In Nabil Hathout, Fabio Montermini & Jesse Tseng (éds.), *Morphology in Toulouse. Selected Proceedings of Décembrettes 7 (Toulouse 2-3 December 2010)*, 109–132. München : Lincom Europa.

Lignon, Stéphanie, Fiammetta Namer & Florence Villoing. 2014. De l'agglutination à la triangulation ou comment expliquer certaines séries morphologiques. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1813–1835. Paris : Institut de Linguistique Française.

Lignon, Stéphanie & Marc Plénat. 2009. Échangisme suffixal et contraintes phonologiques. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 65–82. Saint-Denis : Presses Universitaires de Vincennes.

Lignon, Stéphanie & Michel Roché. 2011. Quelle distribution pour *-éen* vs. *-ien* ?. In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.), *Des unités morphologiques au lexique*, 191–250. Paris : Hermès Science-Lavoisier.

Montermini, Fabio. 2003. Appunti sulla cancellazione di vocale in derivazione. In Antonietta Bisetto, Claudio Iacobini & Anna M. Thornton (éds.), *Scritti di morfologia in onore di Sergio Scalise in occasione del suo 60° compleanno*, 171–188. Cesena : Caissa Italia.

Montermini, Fabio. 2010. The lexical representation of nouns and adjectives in Romance languages. *Recherches Linguistiques de Vincennes* 39. 135–162.

Montermini, Fabio. 2015. Regole (e irregolarità) nella formazione delle parole. In Nicola Grandi (éd.), *La grammatica e l'errore. Le lingue naturali tra regole, loro violazioni ed eccezioni*, 63–83. Bologna : Bononia University Press.

Namer, Fiammetta. 2013. Adjectival Bases of French *-aliser* and *-ariser* verbs : Syncretism or Under-specification ? In Nabil Hathout, Fabio Montermini & Jesse Tseng (éds.), *Morphology in Toulouse. Selected Proceedings of Décembrettes 7 (Toulouse 2-3 December 2010)*, 185–210. München : Lincom Europa.

Perko, Gregor. 2010. Le suffixe *-issime* dans le paysage dérivationnel du nom propre en français. In Maria Iliescu, Paul Danler & Heidi Siller-Runggaldier (éds.), *Actes du XXVe Congrès International de Linguistique et Philologie Romanes, Innsbruck 2007*, t. I, 465–470. Berlin : de Gruyter.

Plénat, Marc. 2002. Jean-Louis Fossat : Fossatissime. Note sur la morphophonologie des dérivés en -issime. In Lidia Rabassa (éd.), *Mélanges offerts à Jean-Louis Fossat*, t. 11/12, 229–248.

Plénat, Marc. 2008a. Le thème L de l'adjectif et du nom. In Jacques Durand, Benoît Habert & Bernard Laks (éds.), *Actes du premier Congrès Mondial de Linguistique Française*, 1613–1626. Paris : Institut de Linguistique Française.

Plénat, Marc. 2008b. Quelques considérations sur la formation des gentilés. In Bernard Fradin (éd.), *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, 155–174. Amsterdam : John Benjamins.

Plénat, Marc. 2009. Les contraintes de taille. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 47–63. Saint-Denis : Presses Universitaires de Vincennes.

Plénat, Marc, Stéphanie Lignon, Nicole Serna & Ludovic Tanguy. 2002. La conjecture de Pichon. *Corpus* 1. 105–150.

Plénat, Marc & Michel Roché. 2014. La suffixation dénominale en *-at* et la loi des (sous-)séries. In Florence Villoing, Sophie David & Sarah Leroy (éds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 47–74. Nanterre : Presses Universitaires de Paris Ouest.

Plénat, Marc, Ludovic Tanguy, Stéphanie Lignon & Nicole Serna. 2002. La conjecture de Pichon. *Corpus* 1. 105–150.

Rainer, Franz. 2003. Semantic fragmentation in word formation : The case of Spanish *-azo*. In Rajendra Singh & Stanley Starosta (éds.), *Explorations in Seamless Morphology*, 197–211. New Delhi : Sage Publications.

Roché, Michel. 2008. Structuration du lexique et principe d'économie : le cas des ethniques. In Jacques Durand, Benoît Habert & Bernard Laks (éds.), *Actes du premier Congrès Mondial de Linguistique Française*, 1571–1585. Paris : Institut de Linguistique Française.

Roché, Michel. 2009. Un ou deux suffixes ? Une ou deux suffixations ? In Bernard Fradin, Françoise Kerleroux & Marc Plénat (éds.), *Aperçus de morphologie du français*, 143–173. Saint-Denis : Presses Universitaires de Vincennes.

Roché, Michel. 2010. Base, thème, radical. *Recherches Linguistiques de Vincennes* 39. 95–134.

Roché, Michel. 2011. Quel traitement unifié pour les dérivations en *-isme* et en *-iste*? In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.), *Des unités morphologiques au lexique*, 69–143. Paris : Hermès Science-Lavoisier.

Roché, Michel. 2013. Briard, bougeoir et camionneur : dérivés aberrants, dérivés possibles. In Georgette Dal & Dany Amiot (éds.), *Repères en morphologie*, 42–50. Lille : Publication en ligne du laboratoire STL.

Roché, Michel & Marc Plénat. 2012. Tous les dérivés en *-at* sont-ils des conversions du thème 13? In Franck Neveu, Valélia Muni Toke, Peter Blumenthal, Thomas Klinger, Pierluigi Ligas, Sophie Prévost & Sandra Teston-Bonnard (éds.), *Actes du 3e Congrès Mondial de Linguistique Française. Lyon, 4-7 juillet 2012*, 1387–1405. Paris : Institut de Linguistique Française.

Roché, Michel & Marc Plénat. 2014. Le jeu des contraintes dans la sélection du thème présuffixal. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberger, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1863–1878. Paris : Institut de Linguistique Française.

Roché, Michel & Marc Plénat. 2016. De l'harmonie dans la construction des mots français. In Franck Neveu, Gabriel Bergounioux, Marie-Hélène Côté, Jean-Marc Fournier, Linda Hriba & Sophie Prévost (éds.), *Actes du 5e Congrès Mondial de Linguis- tique Française. Tours, 4-8 juillet 2016.* Paris : Institut de Linguistique Française.

Scalise, Sergio. 1999. Rappresentazione degli affissi. In Paola Benincà, Alberto M. Mioni & Laura Vanelli (éds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia*, 453–479. Roma : Bulzoni.

Thornton, Anna M. 1999. On Italian derivatives with antesuffixal glides. In Geert Booij & Jaap van Marle (éds.), *Yearbook of morphology 1998*, 103–126. Dordrecht : Kluwer Academic Publishers.

Thornton, Anna M. 2007. Phénomènes de réduction en italien. In Elisabeth Delais-Roussarie & Laurence Labrune (éds.), *Des sons et des sens. Données et modèles en phonologie et morphologie*, 241–268. Paris : Hermès.

Chapter 18

A frame-semantic approach to polysemy in affixation

Ingo Plag

Marios Andreou

Lea Kawaletz
University of Düsseldorf

One of the central problems in the semantics of derived words is polysemy. The most advanced theory of derivational semantics to date is the Lexical Semantic Framework developed by Lieber (2004 et seq.). This theory, however, does not have a straightforward answer to the question of which kinds of meaning extensions are possible and which ones should be impossible for a given derivative. This is all the more so for deverbal derivation, where Lieber explicitly leaves open exactly what the 'semantic body' of verbs, i.e. (roughly) the encyclopedic and cultural knowledge involved in interpretation, looks like Lieber (2004: 72)).

This paper tackles this problem by putting forward a new formal approach to derivational semantics, i.e. frame semantics. In frame theory (Barsalou 1992a,b, Löbner 2013), frames are complex structures which model mental representations of concepts. These representations are typed, recursive attribute-value structures, where the attributes are functional relations, assigning unique values to the concept they describe (see Petersen 2007). Using the apparatus of this framework, we hypothesize that the semantics of a derivational process is describable as its potential to perform certain operations (such as metonymic shifts) on the frames of its bases.

We propose a particular model of affixal polysemy in which attested readings of words of a given morphological category result from indexation of particular elements of the frame-semantic representation, combined with inheritance mechanisms. For deverbal nominalizations in English -*ment*, the shifts can target (syntactically) argumental and non-argumental components. Different bases thus go along with different kinds of semantic shifts in their derivatives. Given a particular verb class, possible readings of the respective derivatives are predictable.

Ingo Plag, Marios Andreou & Lea Kawaletz

1 Introduction

In many languages polysemy in word-formation is all-pervasive (e.g. Rainer 2014). Following Bauer et al. (2013), Kawaletz & Plag (2015: 291) list a number of readings of English deverbal nominalizations involving the suffixes *-ing, -ation, -ment, -ance/-ence, -th* and conversion, as given in Table 1.

Table 1: Readings of English nominalizations (Kawaletz & Plag 2015)

Semantic category	Paraphrase	Examples
Event	'the event of V-ing'	*production, training*
Result	'the outcome of V-ing'	*acceptance, alteration*
Product	'the thing that is created by V-ing'	*pavement, growth*
Instrument	'the thing that V-s'	*seasoning, advertisement*
Location	'the place of V-ing'	*dump, residence*
Agent	'people or person who V-s'	*administration, cook*
Measure	'how much is V-ed'	*pinch, deceleration*
Path	'the direction of V-ing'	*decline, direction*
Patient	'the thing affected or moved by V-ing'	*catch, acquisition*
State	'the state of V-ing or being V-ed'	*alienation, disappointment*
Instance	'an instance of V-ing'	*belch, cuddle*

For other languages, similar lists have been produced. For example, for French we find the data shown in Table 2 in Fradin (September 7, 2012) (see also, for example, Uth (2011), Fradin (2011), Fradin (2012) for French, Roßdeutscher & Kamp (2010), Roßdeutscher (2010) for German).

Table 2: Readings of French nominalizations (Fradin September 7, 2012)

Semantic category	Paraphrase	Example	Translation
Event	'action of V-ing'	*lavage*	'washing'
Product	'resulting object'	*construction*	'building'
Means	'what Vs'	*emballage*	'wrapping'
State	'fact of being Ved'	*embrouillement*	'muddle'
Manner	'manner of V-ing'	*marche*	'gait'
Location	'place where one V-s'	*garage*	'garage'
Group	'people who V'	*équipage*	'crew'
Period	'time during which one V-s'	*hivernage*	'wintering'

These facts raise a number of very general questions. Do affixes have meaning, and if so, how can we describe this meaning? Given the variety of interpretations that derivatives of a given affix can give rise to, this does not seem to be a trivial task. Which kinds

of readings or meaning extension are possible and which ones should be impossible for a given derivative? How does the meaning of the base interact with the meaning of the affix? What are the principles or mechanisms that account for this interaction? In spite of the growing number of studies in this domain the answers to these questions are still under debate and we are still facing the task of accounting "for the substantial evidence that affixes […] are frequently semantically underspecified, and subject to polysemy and meaning extensions of various sorts" (Bauer et al. 2013: 641).

The crucial question is how the different readings of a given derivative emerge, and, as a result, how the different readings of different derivatives of a particular morphological category come about. Some generalizations have been proposed that give at least partial answers to these questions. For example, authors like Bauer et al. (2013: 212) have claimed that certain base verbs evoke certain readings in the nouns derived from them, but systematic studies exploring this claim in more detail and with larger amounts of data are rare. Hence, Bauer et al. (2013: 213) only list a few potential generalizations, for example that state nominalizations frequently derive from verbs of psychological state, and that verbs with inherently spatial denotations give rise to location nominalizations.

With regard to French, Ferret (2013) and Ferret & Villoing (2015) hold that specific readings of derived nouns only arise "if very specific semantic conditions are met by the base verb" (Ferret & Villoing 2015: 480). In the case of instrument readings with nouns in *-oir* or *-age*, this reading can only occur if the base verb denotes an externally caused event which involves an instrumental semantic participant.

What is perhaps noteworthy at this point is the fact that deverbal nominalizations can not only lexicalize the event denoted by the verb or the verb's syntactic arguments, but also other entities that are part of the semantic representation of the base verb. For illustration consider (1). In (1a) we find an eventive interpretation of the converted noun *purchase*, while in (1b) there is an object argument reading ('the thing that was purchased'). Similarly, (2a) shows an eventive reading, but, as shown in (2b), also other things can be profiled. Thus an *embroidery* is not the thing that is embroidered (i.e the internal argument of the verb), but the entity that results from the activity of embroidering.

(1)   a.   [S]earching through the store to find someone to help, I completed my *purchase* and then went home feeling dismissed (COCA NEWS 1998)

     b.   Outside the store I deposited my *purchase* in a trash can. (COCA FIC 2008)

(2)   a.   Her daughter Daphne wisely made no comment and pretended to be engrossed in her *embroidery*. (COCA FIC 2000)

     b.   [T]he nails of her feet and hands matched the color of the *embroidery* of her leine. (COCA FIC 2010)

In this paper we will introduce a new approach to the formalization of the interpretation of derived words based on frames and apply this approach to the analysis of *-ment* derivatives that are based on change-of-state verbs and psych verbs.

2 The framework: Frame semantics

The approach adopted in the present paper builds on predecessors in cognitive science and artificial intelligence such as Marvin Minsky's (1975) frame theory, the schema theory of Bartlett (1932), and, specific to linguistics, Fillmore's work on situation frames (Fillmore 1982; see Busse 2012 for a historical overview of the development of frame semantics). We use the notion of 'frame' in the specific sense of Barsalou (1992a,b), Petersen (2007) and Löbner (2013). In this framework, frames are recursive attribute-value structures as known from other frameworks (e.g. HPSG, Pollard & Sag 1994). Frames are taken to be a general format of mental representations of concepts which is also applicable to linguistic phenomena. Frames can be depicted as graphs with nodes and arcs, or as attribute-value matrices, as shown for the toy example *John hit the ball* in Figure 1, with the graph on the left and the attribute-value matrix on the right.
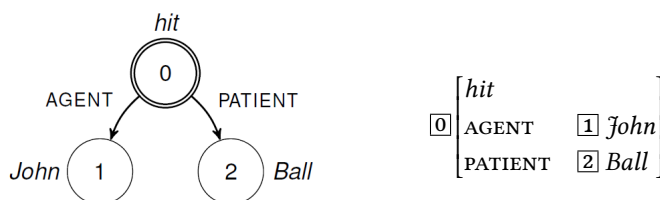
Figure 1: Two ways of depicting a frame

In both representations the referential node, which represents the event as a whole, is labeled *hit* (marked by a double circle in the graph), and this hitting event has two attributes (which, in this case, stand for the participants), an AGENT attribute with the value *John* and a PATIENT attribute with the value *ball*. Entities in graphs and matrices are often indexed for ease of reference (for example with ⓪, ① and ②, as in the attribute-value matrix).

In this approach, attributes are functional in the mathematical sense. The attribute-value structures are recursive and they allow for structure sharing (identities of attribute values). The values by which an attribute can be specified are subordinate concepts of this attribute (Barsalou 1992b: 43). In Petersen's frame approach, the resulting taxonomy is incorporated in the type signature underlying each frame (cf. Petersen 2007: Def. 8 and Fig. 9).

Returning to the problem of verbal bases, our formalism can be used to depict the semantic representation of specific verb classes. For illustration consider a class that is frequently discussed in the literature and that is also a possible base for -*ment* derivation, change-of-state verbs (e.g. Levin 1993, Levin & Rappaport Hovav 1995, Rappaport Hovav & Levin 1998, Dowty 1979, Pustejovsky 1991, Van Valin & LaPolla 1997, Alexiadou et al. 2015). According to many analyses, causation events as expressed by change-of-state verbs (such as *break*) are complex events that consist of two sub-events, a cause and an effect. In a frame semantic analysis, causation events can be formalized as in Figure 2.

$$
\begin{bmatrix}
\textit{causation event} \\
\text{ACTOR} \qquad \boxed{1} \\
\text{UNDERGOER} \quad \boxed{2} \\
\text{INSTRUMENT} \quad \boxed{3} \\[1em]
\text{CAUSE} \qquad \boxed{4}
\begin{bmatrix}
\textit{activity} \\
\text{ACTOR} \qquad \boxed{1} \\
\text{UNDERGOER} \quad \boxed{2} \\
\text{INSTRUMENT} \quad \boxed{3}
\end{bmatrix} \\[2em]
\text{EFFECT} \qquad \boxed{5}
\begin{bmatrix}
\textit{change-of-state} \\
\text{INITIAL STATE} \quad \boxed{6}
\begin{bmatrix}
\textit{state} \\
\text{PATIENT} \quad \boxed{2}
\end{bmatrix} \\[1em]
\text{RESULT STATE} \quad \boxed{7}
\begin{bmatrix}
\textit{state} \\
\text{PATIENT} \quad \boxed{2}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}_{\boxed{0}}
$$
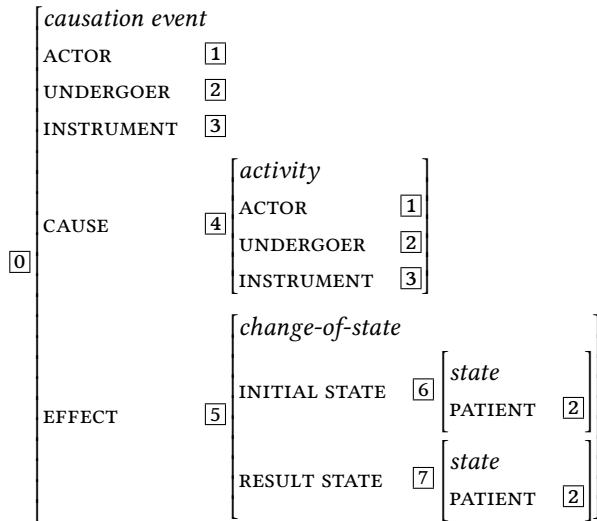
Figure 2: Change-of-state verbs

Figure 2 depicts a typical *change-of-state* verb. The representation is based on established semantic roles (e.g. ACTOR, UNDERGOER) in combination with an event frame. In other words, it combines the participants typically associated with such verbs, and embeds them in the event structure assumed for externally caused events.

A change-of-state verb has three core participants: ACTOR ([1]), UNDERGOER ([2]) and, quite often, an INSTRUMENT ([3]). One of the two sub-events, CAUSE ([4]) consists of an *activity* with the same three participants. The CAUSE sub-event is typically an *activity*, but could also be any other type of event. The activity has an EFFECT ([5]), which constitutes the second sub-event, which is a *change-of-state*. The *change-of-state* involves an INITIAL STATE ([6]) and a RESULT STATE ([7]) of a PATIENT. The PATIENT of the two states is the UNDERGOER of the event [0].

Another verb class that is very common as a base for *-ment* derivatives is that of psych verbs. The use of the term 'psych verb' is not consistent in the literature, and different authors define this class differently. We use the term in this paper as referring to so-called 'object experiencer verbs'. These are verbs (such as *amuse*) where the subject denotes the stimulus, and the object denotes the experiencer in an event in which the experiencer undergoes a change in its psychological state (see, for example, Levin (1993: 189) for discussion). Psych verbs can thus be considered a sub-class of change-of-state verbs, and they are also referred to as 'psych causation' verbs. A frame-semantic representation of such verbs is given in Figure 3.

The verb has two arguments, a STIMULUS and an EXPERIENCER. Similar to the representation of change-of-state verbs there are two sub-events, CAUSE and EFFECT. The CAUSE is an *activity* which has two participants, the ACTOR and the UNDERGOER, and the EFFECT

$$\begin{bmatrix} \textit{psych causation} \\ \text{STIMULUS} \quad \boxed{1} \\ \text{EXPERIENCER} \quad \boxed{2} \\ \\ \text{CAUSE} \quad \begin{bmatrix} \textit{activity} \\ \text{ACTOR} \quad \boxed{1} \\ \text{UNDERGOER} \quad \boxed{2} \end{bmatrix} \\ \\ \text{EFFECT} \quad \begin{bmatrix} \textit{change-of-psych-state} \\ \\ \text{INITIAL STATE} \quad \boxed{3} \quad \begin{bmatrix} \textit{psych state} \\ \text{EXPERIENCER} \quad \boxed{2} \end{bmatrix} \\ \\ \text{RESULT STATE} \quad \boxed{4} \quad \begin{bmatrix} \textit{psych state} \\ \text{EXPERIENCER} \quad \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$
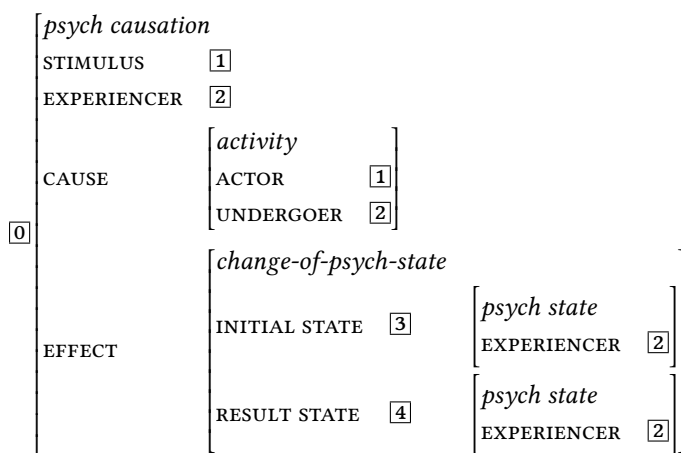
Figure 3: Psych verbs

is a *change-of-psych-state* in the EXPERIENCER entity. Note that the frames depicted here are only partial, as they omit all information that is not immediately relevant for our discussion.

In the following we will apply the frame-semantic approach to the morphological category of *-ment* derivatives in English. Kawaletz & Plag (2015) presented already a first analysis of psych verbs as bases for *-ment* derivation. We will extend this analysis to other verb classes and propose an account in which attested readings of *-ment* words result from indexation of particular elements of the frame-semantic representation, combined with inheritance mechanisms. Specific interpretations can target (syntactically) argumental and non-argumental components, and, consequently, different types of base verb go with different kinds of readings. Given a particular verb class, possible readings of the respective derivatives are predictable. As a result, the multiplicity of meaning in a particular morphological category can be expressed in an inheritance hierarchy of lexeme formation rules. Predecessors of our approach are, for example, Desmets & Villoing (2009) and Tribout (2010), who also tackle polysemy in word formation by positing (slightly different) feature structure representations of lexical semantics in inheritance hierarchies.

3 The suffix *-ment*: Data collection and attested readings

3.1 Overview

The nominalizing suffix *-ment* derives event nominals of various readings, among which Bauer et al. (2013: chapter 10) list events (*assessment*), results (*containment*), states (*contentment*), products (*pavement*), instruments (*entertainment*) and locations (*embankment*).

The suffix was very productive in earlier periods, particularly between the 15th and 17th centuries (Marchand 1969, Lindsay & Aronoff 2013), but is still moderately productive in present-day English with many "novel or low-frequency words" (Bauer et al. 2013: 199) in corpora such as the *Corpus of Contemporary American English* (COCA) (Davies 2008) or the *British National Corpus* (BNC) (Burnard 1995). The suffix mainly attaches to verbs, but adjectival (*foolishment*) and nominal bases (*illusionment*) are also attested, as well as many bound roots (*compartment*) (Bauer et al. 2013: 198).

3.2 Methodology

For the present study we were interested in new coinages, as these can be taken to best reflect the present day speakers' morphological knowledge. The investigation of old and established forms is of course also possible, but such forms are more prone to exhibiting idiosyncratic properties resulting from long-term semantic drift or other processes that accompany lexicalization. Plag (1999: 119), for example, states that "[t]he advantage of dealing primarily with neologisms is that by largely excluding lexicalized formations one has a better chance to detect the properties of possible words rather than of actual words, which may eventually lead to the correct formulation of the productive word formation rule instead of merely stating redundancies among institutionalized words."

In order to arrive at a sizeable number of forms we first extracted all pertinent neologisms of the 20th and 21st centuries from the *Oxford English Dictionary* (OED). In addition, we searched COCA for hapax legomena, i.e. words that occur only once in a corpus. Hapax legomena are not necessarily new words, but the proportion of actual neologisms is highest among hapax legomena (see, for example, Plag 2003: chapter 3.4 for discussion). We ended up with 109 deverbal *-ment* derivatives. We then categorized the base verbs according to the verb classes proposed by Levin (1993) (and extended in the VerbNet project, Kipper et al. 2008). The verbs come from 29 verb classes, with the class of psych verbs being the largest in the data set (N=23).

In order to investigate possible interpretations of the derivatives, we sampled attestations from other corpora (e.g. GloWbE, WebCorp, Google). The attestations were semantically coded using semantic categories such as STATE, EVENT, EXPERIENCER, STIMULUS, RESULT STATE, etc. (see section 3.3. for further discussion). The examples in (3) illustrate the EVENT, RESULT STATE and STIMULUS readings.

(3) a. EVENT Did you put a sound system in your car not specifically for your enjoyment but for the *perturbment* of others within three square miles? (Google BLOG 2008)

   b. RESULT STATE I know a lot of our compatriots also feel the same angst, consternation and *confoundment*. (GloWbE ART 2012)

   c. STIMULUS Here comes a *confoundment* (new word I just made up :)) for you. (Google COMM 2006)

The reader might wonder whether this way of sampling data might favor readings that necessarily deviate from the ordinary, the reason for this being that the new formations

in *-ment* may have been coined because a competing nominalization with another suffix already expressed a more expectable meaning. Two points are important in this respect. First, synonymy blocking has been shown to be an inadequate concept to explain the attested distributions of competing affixes. Very often, different affixes appear on the same base with no discernible difference in meaning (e.g. Bauer et al. 2013: section 26.4). Second, we find the full range of meanings in our data that have also been described in the literature on *-ment* (e.g. Bauer et al. 2013, Marchand 1969). We can thus safely assume that our data represent the semantic possibilities contemporary speakers and listeners of English have at their disposal when creating, using and interpreting *-ment* nominalizations.

The crucial question is which interpretations are possible and whether or how these interpretations depend on the semantics of the base verb. To answer that question the following sections will present an analysis of the attested readings couched in the frame-semantic approach sketched above, focusing on two verb classes, i.e. change-of-state verbs and psych verbs.

3.3 Results: attested readings

Our findings on change-of-state verbs are illustrated in (4).

(4) a. EVENT
Markham sets down the rules about park *befoulment*. (WebCorp BLOG 2012)

 b. INSTRUMENT
Minimal bleeding and I didn't have to have any guaze/tissue in my mouth at all to try and stop it? I'm thinking that they must have used a *congealment* or something to make it clot while I was under or something? (GloWbE COMM 2010)

 c. CAUSE (*activity*) or EVENT
Why do we as Blackpool Fans sit and take this constant *bedragglement* and farce, what is it we are scared of? (Google COMM 2013)

 d. EFFECT (*change-of-state*)
For one second she clung to her son, and then, disengaging herself, froze up like the sudden *congealment* of a spring. (Google FIC 2008)

 e. EFFECT (*result state*)
Sarcasm, Deb … trying to excuse the *bedragglement* of the hair, etc?. (Google COMM 2013)

 f. PATIENT (*in result state*)
I set down the scrap of doll's dress, a *bedragglement* of loose lace hem (COCA FIC 1999)

In (4a) we find an EVENT interpretation. This type of derivative is often referred to as 'transpositional' in the sense that the derived word preserves the sense of the base verb and merely recategorizes ('transposes') the word from verb to noun (but see Lieber

(2015) for a critique of such a view). In (4b), *congealment* denotes the INSTRUMENT, that is, the participant that is manipulated by an ACTOR, and with which an (intentional) act is performed.[1] In (4c), *bedragglement* is ambiguous between an EVENT 'transpositional' reading and a CAUSE reading. In the case of a CAUSE reading, *bedragglement* denotes the first subevent, i.e. the causing event, in the complex event, which is most frequently an activity. The nominalization *congealment* in (4d) refers to the second subevent, i.e. the *change-of-state. Bedragglement* in (4e) denotes a RESULT STATE, that is, the state that the undergoer is in after or during the event. Finally, in (4f), *bedragglement* is interpreted as the PATIENT in a result state, that is, as the participant that is affected by the event.

As far as *-ment* derivatives that are based on psych verbs are concerned, some preliminary results appeared in Kawaletz & Plag (2015). In the present paper, we build on those findings and provide new data. Example 5 lists all readings attested for this class in our data.

(5)   a.   EVENT
           Did you put a sound system in your car not specifically for your enjoyment but for the *perturbment* of others within three square miles? (Google BLOG 2008)

      b.   STIMULUS
           Here comes a *confoundment* (new word I just made up :)) for you. (Google COMM 2006)

      c.   CAUSE (*event*)
           I realize that I often awaken in mindless mid-journey getting jarred by a pothole in the road. That's a quick call-to-action, or *perturbment*. Mindfulness will curb that perturbment and make the journey all the more pleasant and fulfilling. (WebCorp COMM 2013)

      d.   EFFECT (*change-of-psych-state*), CAUSE (*activity*) or EVENT
           "[…] that being told, 'that job is not for you' is an enraging experience." In her own case, Miss Reuben said, the *enragement* began when a professor told her that it really wouldn't matter if she finished her doctoral thesis. (Google MAG 1972)

      e.   EFFECT (*result state*)
           I know a lot of our compatriots also feel the same angst, consternation and *confoundment*. (GloWbE ART 2012)

As is the case with *-ment* on change-of-state verbs, *-ment* derivatives that are based on psych verbs can denote the whole EVENT, giving rise to 'transpositional' readings as in (5a). In a similar vein, they can denote the first, causing subevent as in (5c) and the state that the undergoer is in after or during the event, as in (5e). In addition, *-ment* derivatives that are based on psych verbs can denote the STIMULUS. This finding shows that Pesetsky's claim is wrong that stimulus or event nominalizations should be impossible

---

[1]In the present paper, no claim is made with respect to the relation between *instruments* and *means*. For such a discussion, the interested reader is referred to Fradin (2012).

with psych verbs (Pesetsky 1995: 71): "*Amusement* does not refer to something amusing something, but to the state of being amused" (see also Kawaletz & Plag (2015) for this observation). In (5b), *confoundment* denotes the participant that elicits an emotional or psychological response in the experiencer. Notice that this reading is not evident in derivatives that are based on change-of-state verbs. With respect to *change-of-psych-state* readings as in (5d), it should be noted that we have found no unambiguous example of a derivative with this particular reading.

Among our neologisms RESULT STATE is the dominant reading. This is in accordance with findings in the literature (e.g. Bauer et al. 2013: 209, Pesetsky 1995). EVENT 'transpositional' readings, CAUSE readings, *change-of-(psych)-state* readings, and RESULT STATE readings are attested with both change-of-state verbs and psych verbs. INSTRUMENT and PATIENT (in result state) readings are only attested with change-of-state verbs. Finally, STIMULUS readings are only available with psych verbs.

4 Formalization

In what follows we generalize over the observations we made in the previous section. In particular, we give all referential shifts attested per verb class for *-ment* derivatives in the form of attribute-value matrices.

Figure 4 generalizes over *-ment* lexemes that are based on change-of-state verbs. The frame also contains phonological specifications.

In order to formalize possible referential shifts, we introduce the attribute REF that signals 'reference'. The value of this attribute determines the reference of the derived word. As depicted in Figure 4, the reference (REF) of a lexeme with the phonology *x-ment*, that is based on a change-of-state verb, may be identified with one of the elements of the morphological base (M-BASE). In more detail, the value of REF is [0] in the case of EVENT 'transpositional' readings, [3] when the derived word denotes the INSTRUMENT, [4] in CAUSE readings, [5] in *change-of-state* readings, [7] in RESULT STATE readings, and, finally, [2]-[7] when the derivative denotes the PATIENT in RESULT STATE.[2]

---

[2]It is not an easy task to formally define a referent that is in a particular state (of more than one possible states) in the course of a dynamic event, here to a PATIENT in RESULT STATE in a change-of-state event. The difficulty arises from the fact that dynamic elements would need to be incorporated into the – essentially static – attribute value matrix. There have been several attempts to solve this vexed issue, and the interested reader is referred to these proposals (Gamerschlag et al. 2014, Löbner 2017, Osswald submitted). Future work will have to show how a technical definition of PATIENT in RESULT STATE can be included in the frames we propose in this paper.

```
⎡ lexeme                                                                        ⎤
⎢ PHON    x̄-ment                                                                ⎥
⎢        ⎡ PHON    x̄                                                          ⎤ ⎥
⎢        ⎢        ⎡ causation event                                        ⎤ ⎥ ⎥
⎢        ⎢        ⎢ ACTOR          1                                       ⎥ ⎥ ⎥
⎢        ⎢        ⎢ UNDERGOER      2                                       ⎥ ⎥ ⎥
⎢        ⎢        ⎢ INSTRUMENT     3                                       ⎥ ⎥ ⎥
⎢        ⎢        ⎢                  ⎡ activity                        ⎤   ⎥ ⎥ ⎥
⎢        ⎢        ⎢ CAUSE        4   ⎢ ACTOR          1                ⎥   ⎥ ⎥ ⎥
⎢ M-BASE ⎢ SEM  0 ⎢                  ⎢ UNDERGOER      2                ⎥   ⎥ ⎥ ⎥
⎢        ⎢        ⎢                  ⎣ INSTRUMENT     3                ⎦   ⎥ ⎥ ⎥
⎢        ⎢        ⎢                  ⎡ change-of-state                     ⎤⎥ ⎥ ⎥
⎢        ⎢        ⎢                  ⎢                     ⎡ state       ⎤ ⎥⎥ ⎥ ⎥
⎢        ⎢        ⎢ EFFECT       5   ⎢ INITIAL STATE  6    ⎣ PATIENT  2 ⎦ ⎥⎥ ⎥ ⎥
⎢        ⎢        ⎢                  ⎢                     ⎡ state       ⎤ ⎥⎥ ⎥ ⎥
⎢        ⎣        ⎣                  ⎣ RESULT STATE   7    ⎣ PATIENT  2 ⎦ ⎦⎦ ⎦ ⎥
⎢ REF =  { 0, 3, 4, 5, 7, 2-7 }                                                 ⎥
⎣                                                                               ⎦
```
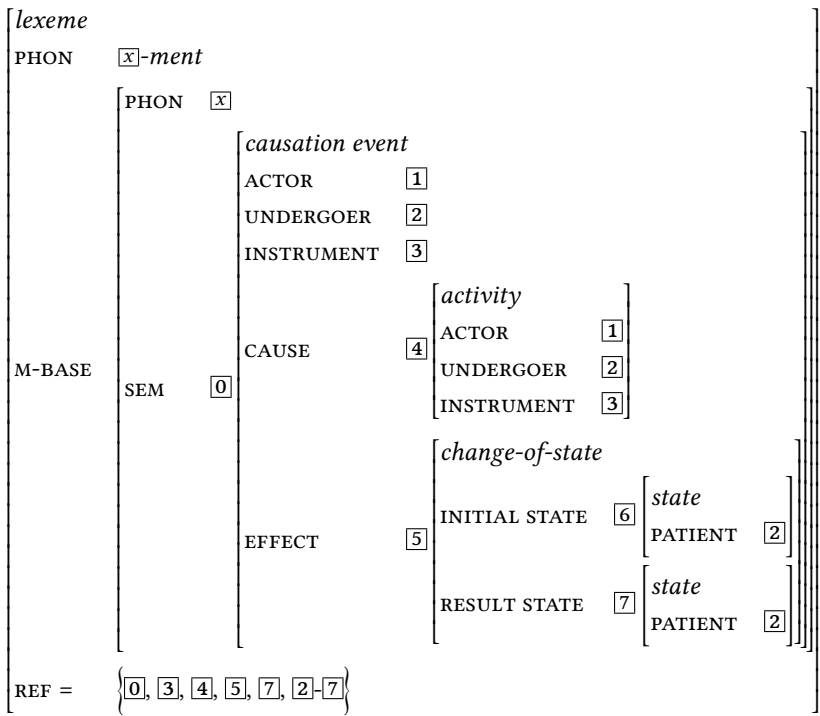
Figure 4: *-ment* on change-of-state verbs

In a similar vein, Figure 5 gives all possible referential shifts attested in *-ment* derivatives that are based on psych verbs.

Based on this figure, the reference (REF) of a lexeme with the phonology x̄-*ment* may have the value 0, 1, 3, 4, or 6. Thus, it may refer to one of the elements of the verbal base: 0 accounts for EVENT 'transpositional' readings, *-ment* derivatives with value 1 refer to the STIMULUS, 3 captures CAUSE readings, 4 accounts for *change-of-psych-state* readings, and *-ment* derivatives with REF 6 have a RESULT STATE reading.

Although Figures 4 and 5 show the range of values available for the reference of *-ment* derivatives per verb class, they collapse all possible readings under REF. In other words, REF = { 0, 1, 3, 4, 6 } and REF = { 0, 3, 4, 5, 7, 2-7 } state all possible readings for *-ment* derivatives based on psych state verbs and change-of-state verbs respectively, but do not address the mechanisms by which these readings arise. In addition, these figures establish no link between shared readings among the two verb classes. We will deal with these issues in the following section.
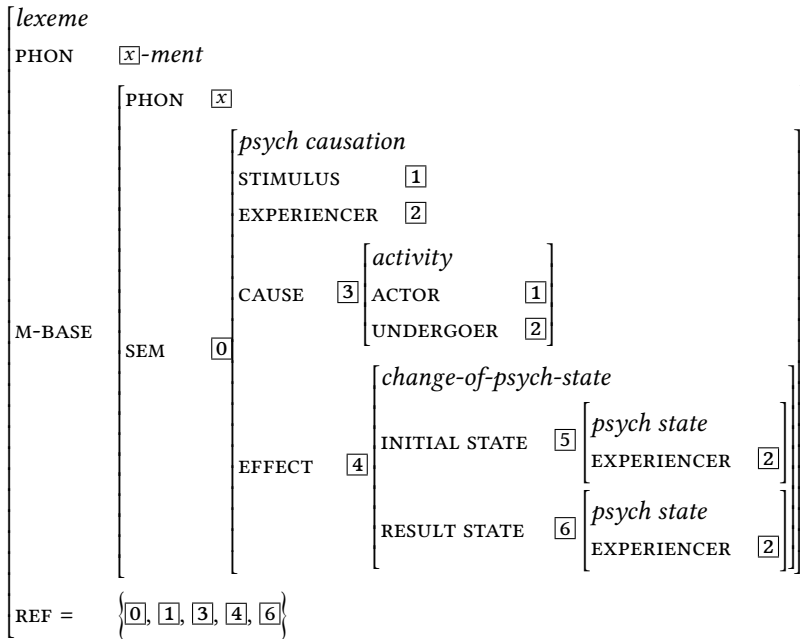
Figure 5: *-ment* on psych verbs

5 Accounting for polysemy

There are two approaches to multiplicity of meaning in derivation: monosemy and polysemy. We will first discuss the monosemy approach.

5.1 A monosemy approach to multiplicity of meaning

In the monosemy approach, multiplicity of meaning is reduced by assigning an underspecified meaning to an affix. More specific meanings of affixes derive from a general highly underspecified meaning. This is done by means of semantic extension rules and interaction between the semantics of the base and the affix. Concrete meanings of derived formations can also be attributed to contextual and encyclopedic information.

The monosemy approach figures prominently in a number of works on deverbal formations. Consider for example the discussion of *-er* nominalizations (for Dutch see Booij 1986 and for English Rappaport Hovav & Levin 1992, Plag 2003). A closer inspection of the analysis put forward by Plag (2003) illustrates the monosemy approach. According to Plag (2003: 89), *-er* derivatives often denote active or volitional participants in an event (e.g. *singer*, *writer*). Plag also mentions that *-er* is used to derive instrument nouns (e.g. *blender*, *mixer*), to denote entities associated with an activity (e.g. *diner*, *toaster*), and to derive person nouns indicating place of origin or residence (e.g. *Londoner*, *New Yorker*).

The multiplicity of meaning evident in *-er* affixation leads Plag to propose that "the semantics of *-er* should be described as rather underspecified, simply meaning something like 'person or thing having to do with X.' The more specific interpretations of individual formations would then follow from an interaction of the meanings of base and suffix and further inferences on the basis of world knowledge." (Plag 2003: 89)

Let us now apply the monosemy approach to *-ment* derivatives. In order to do so we have to reduce multiplicity of meaning by identifying meanings that are shared by all *-ment* derivatives. The results in section 3.3 suggest that *-ment* forms denote (a) eventualities (see 4a), and (b) entities (see 4f). Thus, the abstract core meaning of *-ment* seems to be 'eventuality or entity having to do with X'.

The disjunction 'eventuality or entity' illustrates the first problem that monosemy approaches are confronted with. In particular, the aim of monosemy approaches is to reduce multiplicity of meaning by postulating a unitary abstract meaning. But how abstract should this meaning be? In the case of *-er*, one could claim that *-er* derivatives denote 'an entity having to do with X'. This qualifies as a unitary meaning since all *-er* derivatives do denote an entity. Derivatives in *-ment*, however, do not always denote an entity. They may be eventualities as well. Thus, we have to resort to the disjunction 'eventuality or entity' to capture the semantics of *-ment* derivatives. This, however, shows that the desirable underspecified meaning cannot always be sensibly reduced to a single unitary meaning.

The second problem with the monosemy approach is overgeneration. Let us assume that the semantics of *-ment* derivatives could be reduced to the underspecified meaning 'eventuality or entity having to do with X'. What kind of predictions would follow from this meaning with respect to (a) already attested meanings and (b) meanings that are excluded? Although the meaning 'eventuality or entity having to do with X' is abstract enough to tackle all attested readings of *-ment* derivatives, it leads one to expect that *-ment* derivatives could in principle denote all 'entities'. This is not verified by data, however, since agentive readings are never part of the heterogeneous meanings of *-ment*. Thus, we have to conclude that the monosemy approach does not fare well with respect to which meanings are possible and which meanings are not possible, simply because it leads to massive overgeneration.

5.2 Polysemy in Frame Semantics

In this section we propose that polysemy in derivation should be treated as multiplicity of meaning in word formation patterns. As we will show, given the architecture of frame semantics, this multiplicity of meaning can be expressed in an inheritance hierarchy of lexeme formation rules.

Like some previous authors working on polysemy in word-formation (e.g. Desmets & Villoing 2009, Tribout 2010), we assume that attributes and their values are given in a type signature which can be considered as an ontology which covers world knowledge. According to Petersen & Gamerschlag (2014: 203-204), a type signature restricts the set of admissible frames, includes a hierarchy of the set of types, and states appropriateness

conditions. These conditions declare the set of all admissible attributes for a lexeme of a certain type and the values these attributes take. Appropriateness conditions are inherited by subtypes (see also Riehemann 1998, Koenig 1999, Bonami & Crysmann 2016, Andreou & Petitjean 2017). Consider, for example, the type signature in Figure 6:
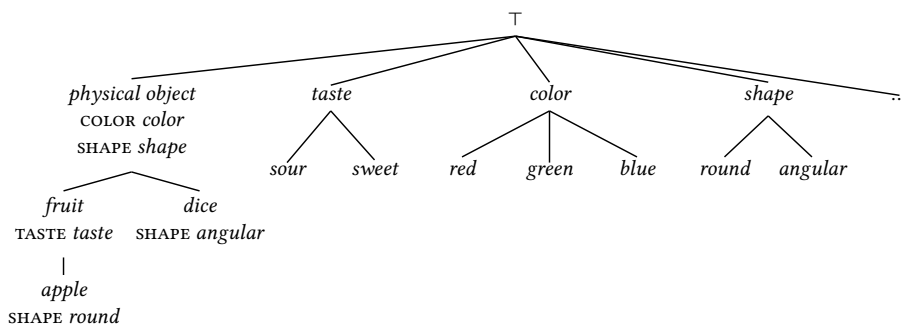
Figure 6: Example type signature (adapted from Petersen & Gamerschlag 2014: 204)

In this type signature, subtypes are given below supertypes. For example, *apple* is a *fruit*, which is itself a *physical object*. The node *physical object* meets two ACs, that is, it is characterized by the attributes COLOR and SHAPE that have the values *color, red, green, blue* and *shape, round, angular*, respectively. According to the ACs on *physical object*, TASTE does not attach to nodes of this type. Thus, not all *physical objects* have a taste. Given that ACs are inherited and further specified by subtypes, *apple* inherits the ACs on *fruit* and *physical object*. Thus, *apple* is characterized by the attributes TASTE, COLOR, and SHAPE. The value of SHAPE is *round* since subtypes not only inherit attributes from their supertypes, but also specify and further restrict the value of inherited attributes. In a similar vein, *dice* inherits the attribute SHAPE from the node *physical object* and specifies the value of SHAPE as *angular*.

The careful reader may have noticed that *color* in Figure 6 is used as an attribute label (i.e. COLOR) and as a type label (i.e. *color*). In frames, this redundancy is attributed to the ontological status of attribute concepts. These functional concepts can be interpreted both *denotationally* and *relationally* (Guarino 1992). Thus, the denotational interpretation of color covers the set of all colors (i.e. type label *color*) and the relational interpretation covers the use of color as a functional attribute that assigns a particular color (e.g. *red*) to the referent of the frame (for more on the use of functional attributes see Löbner 2015).

In the spirit of previous analyses (Riehemann 1998, Koenig 1999, Booij 2010, Bonami & Crysmann 2016) we assume that lexeme formation rules are also organized in an inheritance hierarchy. In particular, consider the following inheritance hierarchy of lexeme formation rules ('*lfr*') for deverbal nouns ('*v-n*') derived by -*ment*.

Figure 7 gives a partial hierarchy of the referential shifts attested in -*ment* affixation. It is only partial for two reasons. First, we do not model the use of -*ment* on adjectives (e.g. *foolishment*) and on nominal bases (e.g. *illusionment*). Second, due to space limitations we model only three possible readings of -*ment* derivatives, namely, event-nouns (*evt-n*),
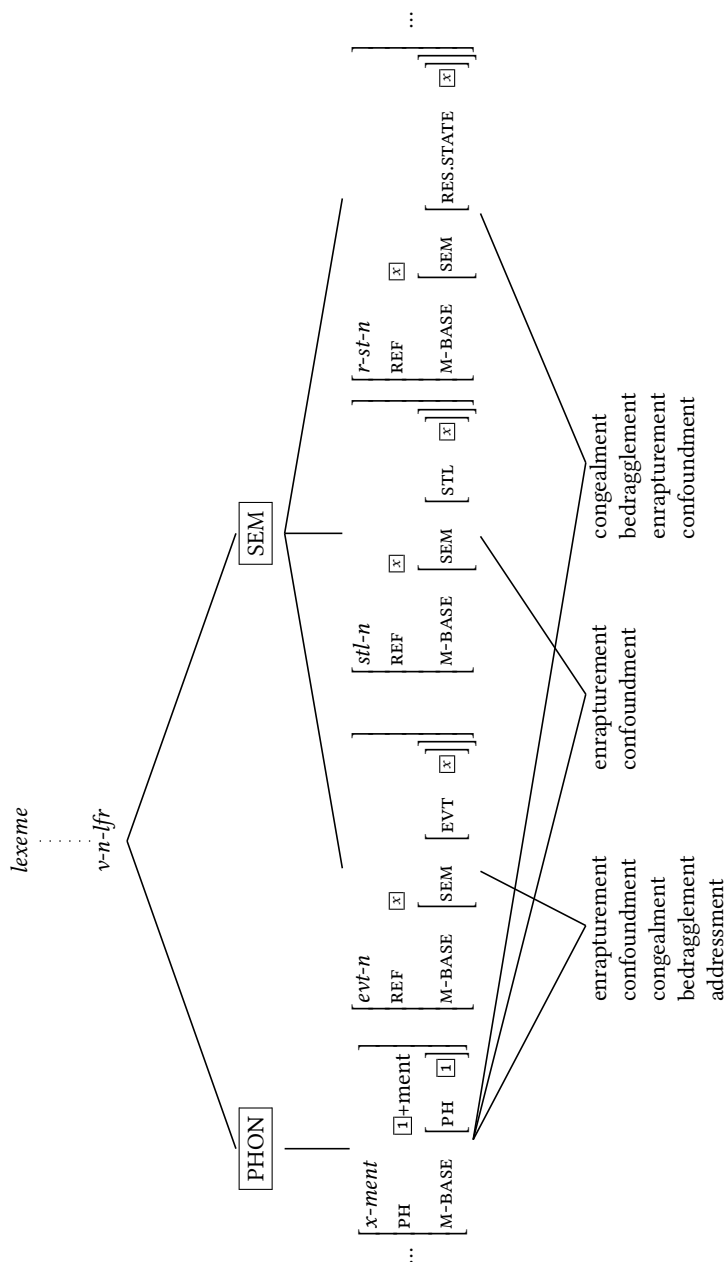
Figure 7: Partial inheritance hierarchy of lexeme formation rules for the affix
-ment

stimulus-nouns (*stim-n*), and result-state-nouns (*r-st-n*). The three dots on the right-hand side show that there are other readings which we do not model here.

The information on the left hand side provides the phonology (PHON) of *-ment* derivatives. That is, *x-ment* formations have the phonology ⎡1⎤+/ment/, where the boxed ⎡1⎤ is the phonology of the base (i.e. M-BASE). The possible readings are given on the right hand side of this figure under SEM (i.e. SEMANTICS).

In more detail, in event-nouns (*evt-n*), the event argument (EVT) of the morphological base is identified with the referential argument (REF) of the derivative. This category includes all *-ment* derivatives in which a transpositional reading is attested. As shown in Figure 7, the category of event nouns includes *enrapturement* and *confoundment* that are based on psych causation verbs, *congealment* and *bedragglement* that are based on change-of-state verbs, and *addressment* that is based on a verb of yet another class, illustrate verbs.

In the case of stimulus-nouns (*stl-n*), the reference of the noun is identified with the stimulus argument (STL) of the base. This category includes *-ment* derivatives based on psych causation verbs only (e.g. *enrapturement, confoundment*). *-ment* derivatives based on change-of-state verbs (e.g. *congealment*) are not included in this category since a stimulus argument is incompatible with change-of-state verbs (see the frame for change-of-state verbs in Figure 2).

In the case of result-state-nouns (*r-st-n*), the reference of the noun is identified with the result state argument (RESULT STATE) of the morphological base. This category includes derivatives based on both psych causation verbs (e.g. *confoundment*) and change-of-state verbs (e.g. *bedragglement*).

6 Conclusion

In this paper we have advocated a new approach to the formalization of polysemous derivational categories, based on frames as represented in attribute-value structures. The approach was illustrated with recent English neologisms derived with the suffix *-ment*, which we have shown to exhibit a wide range of possible readings.

We have argued against an approach that assumes a highly underspecified meaning of *-ment* and in favor of an analysis that assumes hierarchically structured lexical rules and inheritance mechanisms. The proposed analysis has three main characteristics. First, it links the shared readings that are attested among the various verb classes. In the case of event-nouns, for example, we need not pose different rules per verb class since all *-ment* derivatives that are based on change-of-state verbs and psych verbs can inherit the *evt-n* reading. Second, certain readings are excluded by means of appropriateness conditions that give rise to incompatibility. For instance, linking *-ment* derivatives that are based on change-of-state verbs to stimulus readings fails because the stimulus argument is incompatible with change-of-state verbs. Thus, inheritance fails. These characteristics allow us to deal with derivational polysemy without having to resort to underspecified meanings. Finally, the use of appropriateness conditions that give rise to incompatibility

is an effective step to tackle overgeneration, which is a major problem for monosemy approaches to meaning.

As a next step in our research agenda, the approach will have to be applied to more verb classes that take *-ment*, and to other affixes.

Acknowledgements

References

Alexiadou, Artemis, Elena Anagnostopoulou & Florian Schäfer. 2015. *External arguments in transitivity alternations: A layering approach.* Oxford: Oxford University Press.

Andreou, Marios & Simon Petitjean. 2017. Describing derivational polysemy with XMG. In Iris Eshkol & Jean-Yves Antoine (eds.), *Actes de TALN 2017, 24e Conférence sur le Traitement Automatique des Langues Naturelle: Volume 2*, 94–101.

Barsalou, Lawrence W. 1992a. *Cognitive psychology: An overview for cognitive scientists.* Hillsdale: Erlbaum.

Barsalou, Lawrence W. 1992b. Frames, concepts, and conceptual fields. In Adrienne Lehrer (ed.), *Frames, fields, and contrasts*, 21–74. Hillsdale: Erlbaum.

Bartlett, Frederic Charles. 1932. *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology.* Oxford: Oxford University Press.

Bonami, Olivier & Berthold Crysmann. 2016. The role of morphology in constraint-based lexicalist grammars. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge handbook of morphology.* Cambridge: Cambridge University Press.

Booij, Geert. 1986. Form and meaning in morphology: The case of Dutch agent nouns. *Linguistics* 24. 503–517.

Booij, Geert. 2010. *Construction morphology.* Oxford: Oxford University Press.

Burnard, Lou (ed.). 1995. *Users guide for the British National Corpus.* Oxford University Computing Service: British National Corpus Consortium.

Busse, Dietrich. 2012. *Frame-Semantik: Ein Kompendium.* Walter de Gruyter.

Davies, Mark. 2008. *The Corpus of Contemporary American English: 400+ million words, 1990-present*. Available online at https://corpus.byu.edu/coca/.

Desmets, Marianne & Florence Villoing. 2009. French VN lexemes: Morphological compounding in HPSG. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, Georg-August-Universität Göttingen, Germany*, 89–109. Stanford: CSLI publications.

Dowty, David R. 1979. *Word meaning and Montague Grammar*. Dordrecht: Reidel.

Ferret, Karen. 2013. Semantic propertie(s) of the base verbs and locative or instrumental interpretation of N-oir(e). Paper presented at the 9th Mediterranean Morphology Meeting, Duvrovnik, 15–17 September 2013.

Ferret, Karen & Florence Villoing. 2015. French N-age instrumentals: Semantic properties of the base verb. *Morphology* 25(4). 473–496.

Fillmore, Charles. 1982. Frame semantics. In Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin Pub. Co.

Fradin, Bernard. 2011. Remarks on state denoting nominalizations. *Recherches Linguistiques de Vincennes* 40. 73–99.

Fradin, Bernard. 2012. Les nominalisations et la lecture 'moyen'. *Lexique* 20. 125–152.

Fradin, Bernard. September 7, 2012. *Sur la corrélation existant entre les suffixes -age et -ment et les distinctions sémantiques observables dans les nominalisations du français*. Nancy.

Gamerschlag, Thomas, Wilhelm Geuder & Wiebke Petersen. 2014. Glück auf, der Steiger kommt: A frame account of extensional and intensional 'steigen'. In Doris Gerland, Christian Horn, Anja Latrouite & Albert Ortmann (eds.), *Meaning and grammar of nouns and verbs*, 115–144. Düsseldorf: Düsseldorf University Press.

Guarino, Nicola. 1992. Concepts, attributes and arbitrary relations: Some linguistic and ontological criteria for structuring knowledge bases. *Data and Knowledge Engineering* 8(3). 249–261.

Kawaletz, Lea & Ingo Plag. 2015. Predicting the semantics of English nominalizations: A frame-based analysis of *-ment* suffixation. In Laurie Bauer, Pavol Štekauer & Livia Körtvélyessy (eds.), *Semantics of Complex Words*, 289–319. Dordrecht: Springer.

Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation* 42(1). 21–40.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.

Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge: MIT Press.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle. 2015. The semantics of transposition. *Morphology* 25(4). 353–369.

Lindsay, Mark & Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Morphology in Toulouse*, 133–153. Munich: Lincom Europa.

Löbner, Sebastian. 2013. *Understanding semantics.* 2nd, revised edition. London: Arnold.

Löbner, Sebastian. 2015. Functional concepts and frames. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald & Wiebke Petersen (eds.), *Meaning, Frames, and Conceptual Representation*, 35–62. Düsseldorf: Düsseldorf University Press.

Löbner, Sebastian. 2017. Frame theory with first-order comparators: Modeling the lexical meaning of punctual verbs of change with frames. In Helle Hvid Hansen, Sarah E. Murray, Mehrnoosh Sadrzadeh & Henk Zeevat (eds.), *Proceedings of the Eleventh International Tbilisi Symposium on Language, Logic, and Information* (LNCS series). Heidelberg & New York: Springer.

Marchand, Hans. 1969. *Categories and types of Present-Day English word-formation.* Second edition. München: C.H. Beck.

Minsky, Marvin. 1975. A framework for representing knowledge. In Patrick Henry Winston & Berthold Horn (eds.), *The psychology of computer vision*, 211–277. New York: McGraw-Hill.

OED. 2013. *The Oxford English Dictionary online.* Oxford: Oxford University Press.

Osswald, Rainer. 2015. Activities, accomplishments and causation. ms.

Pesetsky, David. 1995. *Zero syntax: Experiencers and cascades.* Cambridge: MIT Press.

Petersen, Wiebke. 2007. Representation of concepts as frames. In Jurgis Skilters (ed.), *The Baltic international yearbook of cognition, logic and communication*, vol. 2, 151–170.

Petersen, Wiebke & Thomas Gamerschlag. 2014. Why chocolate eggs can taste old but not oval: A frame-theoretic analysis of inferential evidentials. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald & Wiebke Petersen (eds.), *Frames and concept types: Applications in language, cognition, and philosophy*, 199–218. Dordrecht: Springer.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation.* Berlin: Mouton de Gruyter.

Plag, Ingo. 2003. *Word-formation in English.* Cambridge: Cambridge University Press.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Pustejovsky, James. 1991. The syntax of event structure. *Cognition* 41. 47–81.

Rainer, Franz. 2014. Polysemy in derivation. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*, 338–353. Oxford: Oxford University Press.

Rappaport Hovav, Malka & Beth Levin. 1992. Er-nominals: Implications for the theory of argument structure. In Tim Stowell & Eric Wehrli (eds.), *Syntax and the lexicon*, 127–153. New York: Academic Press.

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments*, 97–134. Stanford: CSLI publications.

Riehemann, Susanne Z. 1998. Type-based derivational morphology. *The Journal of Comparative Germanic Linguistics* 2(1). 49–77.

Roßdeutscher, Antje. 2010. German *-ung*-formation: An explanation of formation and interpretation in a root-based account. *Linguistische Berichte* 17(117). 101–132.

Roßdeutscher, Antje & Hans Kamp. 2010. Syntactic and semantic constraints on the formation and interpretation of *-ung* nouns. In Monika Rathert & Artemis Alexiadou (eds.), *The semantics of nominalizations across languages and frameworks*, 169–214. Berlin: Mouton de Gruyter.

Tribout, Delphine. 2010. How many conversions from verb to noun are there in French? In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot, Paris 7, France*, 341–357. Stanford: CSLI publications.

Uth, Melanie. 2011. *Französische Ereignisnominalisierungen: Abstrakte Bedeutung und regelhafte Wortbildung*. Berlin: De Gruyter.

Van Valin, Robert D. & Randy LaPolla. 1997. *Syntax: Structure, meaning, and function*. Cambridge: Cambridge University Press.

Chapter 19

Word formation in LFG-based layered morphology and two-level semantics

Christoph Schwarze

This article treats the problem of how the semantics of word formation can be accounted for in terms of rules and representations. A comprehensive model of multilayered, LFG-based morphology is proposed. It comprises four layers of representation: phonology, constituent structure, functional feature structure and lexical semantics. The meaning of derived words is treated in the framework of two-level semantics. It is assumed that rules of word formation derive underspecified semantic forms, parting from which the actual meanings are construed by recourse to conceptual structure. The model is illustrated on the basis of three morphological processes: French é-prefixation, Italian denominal verbs of removal, and noun-to-verb conversion in French. The analyses of é-prefixation and of verbs of removal are taken from the literature; the study on noun-to-verb conversion is original work.

1 Introduction

The hypothesis that the semantics of word formation is an aspect of grammar assumes that the processes of word formation concern both form and meaning. However, actual work on this basis encounters considerable challenges. The data available for the study of a given process of word formation never seem to show a perfect parallelism between form and meaning: forms that stem from a given generative process often have meanings on which it seems to be impossible to form a descriptive generalization. It is the aim of this paper to show how challenges to the semantics of word formation can be dealt with.

I will first address the question of how morphological processes and structures can comprehensively be represented. I will then present three hypotheses concerning the semantics of word formation, namely

    i.   The semantic output of morphological rules is underspecified.

    ii.   The meanings of the words that constitute the data arise from a sequence of steps.

iii.   These steps are a. the morphological rule defines an underspecified semantic form, b. semantic form is turned into a specified meaning on the basis of conceptual knowledge, c. the derived word enters the lexicon, and d. the lexicalized word may have its own development, independently from morphology, and its original meaning may thus be changed and its morphological origin be obscured.

2 LFG-based Layered Morphology

LFG-based Layered Morphology (LLM) integrates essential properties of Construction Grammar Morphology,[1] which does justice to the multi-layered nature of the lexicon, and HPSG-based morphology, which elaborates on the features that syntax receives from morphology.[2]

Notice that LLM is a model, not a theory or a hypothesis. Unlike theories and hypotheses, which can be empirically evaluated with reference to observable data, models can only be evaluated with respect to their usefulness for the progression of knowledge. This kind of usefulness cannot be measured, it can only be shown by actual work on specific phenomena. That is what I will try to do in this study.

Lexicalist models of grammar commonly assume that words are linguistic objects with layered representations, phonological, syntactic and semantic. Accordingly, morphological processes operate simultaneously at various layers or levels of representation.[3] In accordance with Lexical Functional Grammar (LFG) the LLM model makes a distinction between the level of constituents, called the c-structure level, and a level of functional features, called the f-structure level.[4] The latter contains features concerning agreement, tense, mood, inflectional class etc. It also contains grammatical functions and, importantly, predicate features, which are labels of lexical meanings and encode grammatical functions, the syntactic reflex of argument structure.

In addition to these two "syntactic" levels, morphological representations need to comprise a phonological level to account for non-concatenative morphological processes, like German *Umlaut*; cf. Germ. *krank* /krank/ 'sick' + −*lich* /lɪχ/ 'ly' → *kränklich* /krɛnklɪχ/ 'sickly'.

And, of course, there is a semantic level, where the lexical meanings encoded in the lexicon are represented and processed. Resuming, morphological representations and processes are located at

- The level of constituent structure (the c-structure level)

---

[1]See Booij (2010), Booij & Audring (2017).

[2]For work on French, see Fradin (2005) and Tribout (2010b).

[3]I fully agree with Aurnague & Plénat (2008: 1)when they say: "Une lexie est un n-tuplet de représentations reliées entre elles, mais relevant chacune d'un niveau linguistique (phonologique, syntaxique, sémantique, etc.) distinct. La description d'un mode de formation lexical productif suppose par conséquent que soient relevées et expliquées les régularités apparaissant à chacun de ces niveaux."

[4]LLM was first presented in a seminar held by the author at the University of Padova in 2008 and subsequently applied to the formation of Italian past and passive participles in Schwarze (2011).

- The functional level (the f-structure level)

- The phonological level (the p-structure level)

- The semantic level (the s-structure level)

Unlike syntax, morphology may manipulate predicates, thus deriving new predicates.

3 A sample analysis: French *é*-prefixation

I will illustrate LLM on the basis of Namer & Jacquey's (2003)[5] article on the French *é*-prefixation, which endevours to give a formalized version of the findings of Aurnague & Plénat (1997). In one of its modalities, *é*-prefixation turns nouns into transitive verbs that denote events where the referent of the base noun is distanced, removed, or separated from the referent of the direct object, as in (1):[6]

(1)   FR.

   a. *é + branche → ébrancher (un arbre)*
          'branch'   'to prune a tree'

   b. *é + feuille → effeuiller (x)*
          'leaf'   'to strip the leaves or petals from x'

   c. *é + gorge   → égorger (x)*
          'throat'   'to cut x's throat'

   d. *é + pou   → épouiller (x)*
          'louse'   'to delouse x'

Moreover, as has been shown by Aurnague & Plénat (1997, 2007, 2008), the relation that holds between the two dissociated entities must be "usual" and "natural",[7] or, as Namer & Jacquey put it:

> [D]escribing the process consisting in clearing a tree of e.g. the magpies (*pie*) or the cats (*chat*) that colonize it cannot be performed by processes referred to by the ?*épier*[8] or ?*échatter* impossible derived verbs. (Namer & Jacquey 2003: 2)

Table 1 gives the rule that generates verbs like *ébrancher*, *effeuiller*, *égorger* or *épouiller* in the LLM notation.

The c-structure change as formulated in Table 1 should be self-explanatory, whereas a few comments on the f- and s-structure part of the rule will be useful.

---

[5]In a subsequent article, Namer & Jacquey (2012) proposed a modelization of the N>V vs. V>N derivations within the framework of the Generative Lexicon.

[6]Changes like adding /j/ to the stem as in *épouiller* are idiosyncratic and must be accounted for in the lexicon.

[7]"[…] les dérivés en *é*- expriment la dissociation […] par un agent intentionnel […] d'une relation d'attachement habituel […] créée naturellement […] et à laquelle il s'oppose […]" (Aurnague & Plénat 2008: 28).

[8]Not to be confounded with existing *épier qu.* 'to spy on someone'.

Table 1: The LLM rule for *é*-prefixation

c-structure	$[\acute{e}]_{V\_prefix} + N_{stem} \rightarrow [\acute{e}N_{stem}]_{V\_stem}$
f-structure	$(\uparrow \text{PRED1})=\text{'P'} \rightarrow (\uparrow \text{PRED2})=\text{'DISSOCIATE} (\uparrow \text{SUBJ}),(\uparrow \text{OBJ})\text{'}$
p-structure	\<no morphologically relevant change\>
s-structure	$\lambda y \, \mathsf{p}(y) \rightarrow \lambda x \, \lambda y \, \exists z$
	$\mathtt{dissociate}(x, y, z) \wedge \mathtt{natural\_relationship}(y, z) \wedge \mathtt{agent}(x) \wedge \mathtt{theme}(y)$

PRED is a feature attribute, whose value identifies a word's lexical meaning and argument structure. The input, $(\uparrow \text{PRED1})=\text{'P'}$, contains a predicate variable, *p*, which ranges over the nominal predicates associated with constituent N_{stem}. The up-arrow is an abbreviation for a function that projects the feature to the dominant c-structure node. The output of the semantic change is a new predicate PRED2, which is defined by the rule. It has two arguments, an agent and a theme, realized as the subject and the object respectively. Notice that the prefix, in accordance with Namer & Jacquey (2003), has no direct functional representation, because it has no referential meaning.[9] As to the s-structure level, the derived predicate, 'dissociate', has three semantic arguments: *x*, which is the subject and refers to the agent, *y*, which is the object and refers to the theme, and *z*, which is incorporated in the verb's meaning and refers to the entity which is dissociated from *y* . The additional predication, repeated as (2), is needed to constrain the range of *y* and *z*:

(2) $\mathtt{natural\_relationship}(y, z)$

This part of the representation expresses the fact that the relation between *y* and *z* must not be a merely accidental one, as reported above.

Notice that the change in s-structure as expressed in Table 1 does not predict the full actual meanings of the verbs derived by *é*-prefixation: the derived representation is underspecified.[10] In the following section I will give some background for such an assumption.

---

[9]"Our purpose is … to represent the verb class obtained by the *é*-prefix derivation. To achieve this, two basic ways are provided: (i) representing the prefix itself or (ii) representing an abstract, parametrized lexical unit describing the output (verbal) class. The motivation for the first choice would be the fact that the affix can be seen as some kind of predicate, operating on and controlling two arguments, the base and the derived word, from a structural, categorial and above all semantic points of view. However, the nature itself of the affix is a counterargument: according to the morphological theory defended here, an affix does not belong to any of the major categories. In addition, we have seen that it bears no referential meaning: consequently, it is not foreseeable to modelize its semantic content, as it has no proper semantic content" (Namer & Jacquey 2003). I follow this argumentation, with the exception that not belonging to a major category does not generally imply the lack of functional or semantic information.

[10]This assumption is quite common in the literature, see the survey in Tribout (2010b: 282–284).

4 Two-level semantics

If we assume that lexical morphology is a generative subsystem that feeds the lexicon, then its semantics is part of lexical semantics. Now, the fundamental question is to which extent lexical semantics is an affair of grammar. According to the conception known as two-level-semantics, lexical meaning is represented at two distinct levels, semantic form (sf) and conceptual structure (cs).[11]

Semantic form is linguistic knowledge. sfs "are systematically connected to, and hence covered by, lexical items and their combinatorial potential to form more complex expressions" (Lang & Maienborn 2011: 711). They "form an integral part of the information cluster represented by the lexical entries of a given language" (Lang & Maienborn 2011: 711). They are "accessibly stored in long-term memory" (ib.). They are underspecified with respect to cs representations (Lang & Maienborn 2011: 713). And, importantly, sf is the level at which two-level semantics endeavors to represent the compositionality of lexical meaning and the grammatical role of lexical decomposition (Lang & Maienborn 2011: 723).

Returning to the semantics of word formation, it is an aspect of grammar, as far as semantic form is concerned. Most of the characteristics of sf that hold for ordinary lexical semantics also hold for the semantics of word formation, with one exception: compositionality is not a general feature of lexical morphology. In fact, non-concatenative processes may be absolutely regular, but cannot be compositional, since compositionality presupposes concatenation.

In order to see whether the semantics of lexical morphology can reach out to phenomena that are situated beyond sf, let us see what two-level semantics means by conceptual structure.

Conceptual structure can be said to be world knowledge (Lang & Maienborn 2011: 711). That does not mean, however, that it has nothing to do with language, actually, it is closely related to sf: cs representations are built upon and enrich sf representations. Thus, semantic representations typically contain both, cs and sf features. This happens in such a way that, for the representation of a given lexical meaning, the cs features specify and enrich sf representations, thus enabling words to denote their referents.[12,13,14]

---

[11]For a critical state-of-the-art overview, see Lang & Maienborn (2011).

[12]In Lang and Maienborn's words: "…for every linguistic expression *e* in language *L* there is a cs representation *c* assignable to it via sf(e), but not vice versa" (Lang & Maienborn 2011: 711); "… cs representations are taken to belong to, or at least to be rooted in, the non-linguistic mental systems based on which linguistic expressions are interpreted and related to their denotations."

[13]This conception has an important consequence: if the features retrieved from cs are combined with or replace sf features, doing lexical semantics does not mean to represent the entire bulk of knowledge and beliefs that we have about the referents of the lexemes under investigation.

[14]As to the mental status and processing of cs representations, they are assumed to be "activated and compiled in working memory", contrarily to sf representations, which, as has been said above, are stored in long-term memory (Lang & Maienborn 2011: 712). I am not sure about the mental status of cs: it may safely be assumed that concepts, once they are lexicalized as meaning components, are as stable as sfs.

5 A second sample analysis: Italian denominal verbs of removal

It will be useful to illustrate underspecification and its resolution with an example from derivational morphology. I will briefly present the analysis of the denominal verbs derived by *s-* prefixation in Italian as proposed by von Heusinger & Schwarze (2006).[15]

The morphological process generates verb stems from noun stems by prefixing the constituent *s-* to the noun stem;[16] cf. (3) and (4):

(3) a. crem(a)$_N$ $\longrightarrow$ screma(re)$_V$
         'cream'   'to skim'

     b. carcer(e)$_N$ $\longrightarrow$ scarcera(re)$_V$
         'prison'   'to release from prison'

(4) a. La mattina, la nonna   scremava   il latte.
         the morning the grandmother skimmed$_{IMPERFECT}$ the milk

         'In the morning, Grandmother used to skim the milk'

     b. Il giudice ha scarcerato   Giovanni Rossi.
         the judge   has released-from-prison Giovanni Rossi

         'The judge released Giovanni Rossi from prison.'

Both verbs, *scremare* and *scarcerare*, mean 'x removes y from z'. However, they differ with respect to the role of the nominal base in the verbs' meaning. In terms of Leonard Talmy's (1985) lexical typology of motion events, the entity denoted by the base noun may be the Figure or the Ground. In (4a) the cream (*crema*) is the Figure; it is removed from the milk (*latte*), which is the Ground. Inversely, in (3b) the prison (*carcere*) is the Ground, from which Giovanni Rossi, the Figure, is released. Thus the speaker needs to decide on the assignment of Figure and Ground for every single verb generated by N→V *s*-prefixation. In a two-level semantics, SF will only state that the verbs under discussion denote caused motion, the role of the incorporated noun being left open. The general semantic form of these verbs may thus be written as (5):[17]

(5) $\texttt{cause}(x, \texttt{become}(\neg\texttt{located}(y, z)))$ & $[\texttt{N}(y)\lor\texttt{N}(z)]$

The first part of representation (5), $\texttt{cause}(x, \texttt{become}(\neg\texttt{located}(y, z)))$, is the lexical decomposition of the main feature of all verbs of removal, $\texttt{remove}(x, y, z)$. The second part,

---

[15]Giuseppina Todaro (2017) applies the von Heusinger & Schwarze (2006) approach to prefixed deadjectival verbs in Italian.

[16]Notice that Italian also has a V→V *s*-prefixation, which derives verbs of reversal, see Mayo et al. (1995: 932), among others. This is a different morphological process, which I do not discuss here.

[17]In von Heusinger & Schwarze (2006) the representation given here as (5) is not the final version, which uses indices in order to account for the correlation between ambiguity of role assignment and the alternative of quantification. In fact, if the predicate of the base noun is incorporated in the verb, it is only existentially bound by ∃. If it becomes the direct object, it is bound by the λ operator. In (5), quantification is omitted for the sake of easier reading.

N(y)∨N(z), expresses the underspecification of *s*-prefixed verbs of removal by a disjunction, where N is the predicate of the base noun.

The ambiguity expressed by this disjunction is resolved at the cs level. According to von Heusinger & Schwarze's (2006) analysis, the resolution of the underspecification passes through the following phases: the concepts associated with the base noun predicates are looked up in cs and checked regarding their aptitude to be a Figure or a Ground in a motion event. Objects that may contain something, are apt to take the role of Ground, objects that may easily perform or undergo motion are apt to be the Figure. Some objects, such as a sheet of paper, may meet both criteria and may consequently motivate derived verbs with two alternative fully specified meanings. Italian *scartare*, derived from *carta* 'paper', is such a case: it may be used as both a Ground verb or a Figure verb; cf. (6):

(6) Mario scarta   il   regalo.
    Mario *s*-paper₃ₛₘ.ᵢₙd.ₚᵣₑₛ the gift

    'Mario takes the gift out of the paper'   *Ground verb*

    'Mario takes the paper off the gift'   *Figure verb*

Table 2 gives the rule that derives denominal *s*-prefixed verbs of removal in the LLM format, with the semantic layer formulated in such a way as to generate underspecified SF representations.[18],[19]

Table 2: The rule for deriving Italian denominal *s*-prefixed verbs

c-structure	$[s]_{V\_prefix} + N_{stem} \rightarrow [sN_{stem}]_{V\_stem}$
f-structure	$(\uparrow$ PRED1$)=$'P' $\rightarrow$ $(\uparrow$ PRED2$)=$'REMOVE $(\uparrow$ SUBJ$),(\uparrow$ OBJ$),(\uparrow$ OBL$)$'
p-structure	\<no morphologically relevant change\>
s-structure	$\pi(y) \rightarrow$ cause$(x,$ become$(\neg$located$(y,z)))$ & $[N(y) \lor N(z)]$

6 A third sample analysis: French N→V conversion

I will now present a case study of French N→V conversion, as exemplified by the pairs in (7):

(7) a. amidon – amidonner
       'starch' 'to starch'

    b. archives – archiver
       'archives' 'to archive'

---

[18]For easier reading, I do not express here the case-marking of the Oblique, which must be *ne* if its predicate is 'PRO' and must be marked by preposition *da* elsewhere.

[19]The [s] vs. [z] realization of the prefix is a matter of post-lexical phonology, hence it is not expressed in the morphological rule.

c. bêche   – bêcher
   'spade'   'to dig'

d. sucre   – sucrer
   'sugar'   'to sugar'

The relation between the nouns and the respective verbs in (7) is clearly directed, which does not hold for other noun-verb pairs, as those given in (8):

(8)   a. chant – chanter
       'song'   'to sing'

    b. gel   – geler
       'frost'   'to freeze'

    c. prêt   – prêter
       'loan'   'to lend'

    d. vent   – venter
       'wind'   'to be windy'

The difference between (7) and (8) is due to the ontological class of the nouns' predicates: whereas the nouns in (7) denote objects or substances and thus are clearly distinct from the respective verbs, those given in (8) denote events or results of events and thus are not clearly distinct from the verbs they relate to. The derivational direction in (7) clearly is N→V, because event predicates may be built upon object or substance predicates, but not inversely.[20] On the contrary, the conversion in (8) may be the opposite, V→N,[21] or non-directional, N↔V, because the nouns' and the verbs' predicates are identical or very closely related.

As for the semantics of N→V conversion, I assume that the rule defines an underspecified semantic form, from which full meanings are derived by a retrieval of conceptual structure.[22] To account for actual meanings that are not predicted on this basis, post-morphological processes are taken into account. It is also assumed that there are certain verbs that look like N→V converts, but are idiosyncratic items not derived by the rule.

---

[20]Cf. the more explicit formulation by Tribout (2010b: 140): "… le recours aux propriétés sémantiques des deux lexèmes pour déterminer l'orientation de la conversion repose, par exemple, sur l'idée que le lexème dérivé est nécessairement défini par le biais de son lexème base, tandis que le lexème base est sémantiquement indépendant de son lexème dérivé. Ainsi pour la paire CLOU∼CLOUER, CLOUER est nécessairement défini relativement à CLOU comme 'faire quelque chose avec des clous' tandis que CLOU est défini comme un petit objet pointu, indépendamment de clouer. Cette asymétrie dans la relation sémantique entre les deux lexèmes permet de prédire une orientation de la conversion de nom à verbe."

[21]For a state-of-the-art discussion on the direction of the French N→V vs. V→N conversion see Tribout (2010a: 348–356).

[22]Tribout (2010b: 284–290) criticizes the underspecification approach; instead she proposes and spells out a fully specified semantics, based upon a classification of the output verbs. I am trying to show that an underspecification-based analysis of the French N→V conversion is an achievable goal.

6.1 A database

As a descriptive basis for the study, I established a database of 170 verbs that clearly are N→V converts. 19 of these verbs are prefixed and have no lexicalized unprefixed counterpart, such as *emprisonner* 'to imprison'.

I consider including prefixed verbs of this kind as legitimate, because the prefixes involved, *en-*, *dé-* and *re-*, require a verbal base. *Emprisonner*, e.g., thus has the derivational history shown by (9):

(9) $prison_N \rightarrow prisonner_V \rightarrow emprisonner_V$

In addition to the verbs and their base nouns, the database contains the following information:

- The verb's underspecified semantic form (SF), if there is one

- The specified semantic representation (SR)

- The conceptual class of the base noun

- The presence of a prefix, if there is one

- Remarks on formal and semantic properties of the derived verbs

6.2 The underspecified semantic forms

Underspecified semantic forms could be construed for 142 of the 170 verbs. The predominant one, which holds for 136 of the 170 verbs contained in the database, states the following:

- The verb denotes an event, which is an action

- It involves an agent and a theme

- The denotation of the noun from which the verb is derived is a salient component of the action

For an illustration, see example (10):

(10)   Le secrétaire a   archivé la   correspondence.
       the secretary has archived the correspondence
       'The secretary archived the correspondence.'

The SF underlying (10) states that the sentence describes an action. The denotation of the noun *archives* is a salient component of that action. The verb, *archivé*, has two arguments, *le secrétaire* and *la correspondence*, whose roles are agent and theme respectively.

In addition to the predominant SF, two more SFs have been identified; they are closely related to the predominant one, see examples (11) and (12). (11) describes an action, but

unlike (10), the verb has no argument in the role of theme. (12), where the reflexive pronoun is the operator of the middle voice, describes a process, the verb's only argument is in the role of theme.

(11) Nous avons passé l'après-midi à magasiner.   (Canadian Fr.)
      we   have   spent the afternoon to window-shopping

      'We spent the afternoon window-shopping.'

(12) Leurs genoux se   sont ankylosés.
      their knews Reflexive_pronoun are ankylosed

      'Their knees have become stiff.'

All SFs assumed for the verbs contained in the database are shown in Table 3, which also shows the forms of the semantic predicates involved, the mapping of the arguments onto grammatical functions and the number of verbs for each SF.[23]

Table 3: Underspecified semantic forms of converted denominal verbs

	Semantic predicate	Grammatical functions	SF	verbs
SF1	$P(e, x, y)$ agent(x) theme(y)	P (SUBJ), (OBJ) $\quad\mid\qquad\mid$ $\quad$x$\qquad$y	x accomplishes an action on y, N is salient in that action.	136
SF2	$P(e, x)$ agent(x)	P (SUBJ) $\quad\mid$ $\quad$x	x accomplishes an action. N is salient in that action.	4
SF3	$P(e, x, y)$ theme(x)	P (SUBJ) $\quad\mid$ $\quad$x	x undergoes a process. N is salient in that process.	4

We can now formulate the rule for French N→V conversion, see Table 4.[24] At the semantic layer, only the predominant SF is given.

---

[23]There are two questions that I cannot address here in detail. First, how productive is the process analyzed here? French is a language that overwhelmingly prefers affixation to conversion. I assume that N→V is fully productive, but that much of its output is blocked by the output of competing rules of affixation. Second, can the non-dominant SFs be derived from the predominent one? Further research is needed here.

[24]Except the selection of alternative lexicalized stem variants, see fn. 8. In the table I omit quantification again in order to make reading easier.

Table 4: The layered rule for N→V conversion

c-structure	$N_{stem} \rightarrow V_{stem}$, 1st inflectional class
f-structure	(↑ PRED)='P1' → (↑ PRED)='P2 (↑ SUBJ),(↑ OBJ)'
p-structure	<no morphologically relevant change>
s-structure	p1(z) → p2(e, x, y) ∧ agent(x) ∧ theme(y) ∧ salient_component_of(e) = p1(y)

7 Resolving the underspecified semantic forms

As has already been pointed out, the underspecified semantic forms cannot be used in discourse, because they are unable to refer to the specific actions denoted by the verbs. Hence the underspecification needs to be resolved. This happens by accessing the conceptual knowledge associated to the base nouns. Regarding N→V conversion, I assume that the speaker or hearer looks up the conceptual knowledge associated with the noun, inspects the event types in which the noun's denotation is typically involved, and finally creates a new semantic predicate in which one of these event types is, so to speak, incorporated. The noun's meaning is then turned into a feature of the new predicate, a feature that becomes visible by lexical decomposition. I will try to illustrate this idea by means of two examples, the first is (13):

(13) L' orfèvre   a   ciselé   leurs noms sur les alliances.
      the goldsmith has chiseled their names on the wedding_rings

      'The goldsmith engraved their names on the wedding rings.'

The verb contained in (13) has the general, underspecified semantic form listed as SF1 in Table 3, and repeated here as (14):

(14) X accomplishes an action on y; N is salient in that action.

For *ciseler* 'to chisel, to engrave' we replace N with "a chisel", getting (15):

(15) X accomplishes an action on y, a chisel (Fr. *ciseau*) is salient in that action.

The conceptual knowledge associated with *ciseau* contains, among others, the information given under (16):

(16) A chisel is a tool, used for cutting wood, stone or metals.

The predicate cut(x, y) is the semantic counterpart of the concept of cutting. Going back from conceptual structure to semantic form, the speaker inserts it into the decomposed semantic representation of the new predicate created by the conversion rule. The meaning of the new predicate also contains chisel(x), taken from the base

noun. Since, according to (16), a chisel is a tool, i.e. an instrument, the feature will be `instrument_used`(e, x, y)=`chisel`(z). Notice that z is not an argument of the new predicate and will not be realized in the sentence. (17) is the assumed semantic representation of *ciseler*, after the resolution of underspecification.

(17) $\exists e \; \lambda x \; \lambda y$ `chisel`(e, x, y)[25]
    `event_type`(e) = `action`(e)
    `action_type`(e) = `cut`(x, y)
    `agent`(e) = x
    `theme`(e) = y
    `instrument_used`(e) = `chisel`(z)

The first line of (17) gives the semantic representation of *ciseler* in the standard notation. The remaining lines give its decomposed meaning in terms of features, written as equations, in the tradition of unification grammars. (This notation mainly shows its usefulness when larger sections of the lexicon are analyzed: it makes it easy to express feature inheritance, and it helps to control the consistency of the features declared.)

The second example I give for the resolution of underspecification is (18):

(18) Les chasseurs ont   huilé leurs fusils.
    the hunters   have oiled their shotguns

    'The hunters oiled their shotguns.'

The verb *huiler* 'to oil' has the same sf as *ciseler*. Applied to the base noun *huile* 'oil' it reads:

(19) X accomplishes an action on y, oil (Fr. *huile*) is salient in that action.

Accessing the conceptual knowledge associated with *huile,* the speaker gets, among others, the information given under (20):

(20) Oil is a substance used to lubricate a mechanism.

The predicate `lubricate`(x, y) is the semantic counterpart of the concept of lubricating. The speaker inserts it into the decomposed semantic representation; the meaning of the new predicate also contains `oil`(x), taken from the base noun. Since, according to (20), oil is a substance, the feature will be `substance_used`(e, x, y) = `oil`(z). (21) is the assumed semantic representation of *huiler*:

(21) $\exists e \; \lambda x \; \lambda y$ `oil`(e, x, y)
    `event_type`(e) = `action`(e)
    `action_type`(e) = `lubricate`(x, y)
    `agent`(e) = x
    `theme`(e) = y
    `substance_used`(e) = `oil`(z)

---

[25]For readers not familiar with the French language, I use English to name semantic features, even though this may make the analysis somewhat inaccurate.

7.1 Polysemy in lexical morphology

The conceptual categorization of 'oil' I assumed for the above sample analysis, i.e. that 'oil' is a substance used to lubricate a mechanism, is far from being the only one.[26] As we know, oil also is used to preserve wood or iron, to cook and season food, it also is a fuel, and an ingredient of oil paint. As linguists, we do not have scientific methods to find out to what extent knowledge of this kind is contained in the conceptual structure and we have no precise knowledge of how conceptual structure is processed during the resolution of semantic underspecification. However, we can look at the lexicon and see those elements of conceptual structure that show up in the lexical meanings of a given language. Thus we can observe that, in the meaning variation of the French verb *huiler* 'to oil' the following bits of information clearly play a role:

   i. Oil is a lubricant (22), repeated from (18).

   ii. Oil is a preservative (23).

(22) Les chasseurs ont   huilé leurs fusils.
      the hunters   have oiled their shotguns

      'The hunters oiled their shotguns.'

(23) Cette table a   besoin d' être huilée.
      this   table has need   to be   oiled

      'This table needs to be oiled.'

As to using oil for preparing or seasoning food, the situation is less clear. According to the reviewer of this article, whom I believe to be a native speaker of French, *huiler* cannot mean 'to season with oil'. I briefly searched the Internet and found out that there were zero hits for *huiler la viande* (*viande* means 'meat') and *huiler les steaks*. There were several hits for *huiler la salade*, but only two of them were from real text (24) and (25), the others being citations from dictionaries.

(24) J'aime faire des vinaigrettes qui ne font pas qu'assaisonner ou huiler la salade mais qui apportent plutôt une valeur ajoutée.[27]

      'I like to make vinaigrettes that do not only season or oil the salad but rather bring an additional value.'

(25) Ne pas huiler la salade, car ainsi suivant son goût chacun fera sa propre vinaigrette, et puis s'il reste de la salade, elle se conservera plus facilement sans vinaigrette.[28]

      'Don't oil the salad, because that's how everyone will make their own vinaigrette to their taste, and then, if some salad is left over, it will be preserved more easily without vinaigrette.'

---

[26]I inserted this section as a response to a comment I received from an anonymous reviewer. For the analysis of polysemy in lexical morphology, also see Schwarze (2012).

[27]http://brutalimentation.ca/2017/01/14/salade-festive-vinaigrette-digestive [2017-08-29].

[28]http://ilovecuisine.blogspot.ch/2013/09/ma-salade-de-lete-la-salade-nicoise.html [2017-08-29].

The remaining known uses of oil do not seem to play a role in the meaning variation of French *huiler*. Instead of speculating about why this should be so, let us pass on to a question that immediately arises from what we could observe.

Assuming that the accessible conceptual structure offers competing information for the resolution of the underspecified meaning generated by the morphological process, the full meaning of *huiler* shows the following variants:

(26) a. 'To lubricate with oil'

  b. 'To preserve with oil'

  c. 'To prepare or season with oil'

The question now is: How do speakers pick out the convenient reading in producing or parsing utterances? This is a very general question, not specific to the semantics of word formation. In the case of transitive verbs such as *huiler*, a sort of semantic agreement is at work, which checks the compatibility of the verb's reading with the conceptual class of the direct object.

Regarding the avoidance of *huiler* with a direct object denoting meat, there may be practical reasons or no reason at all; there are phenomena in verbal behavior that are beyond the reach of linguistic analysis.

8 Restrictions on the input

It can easily be seen that many nouns are not fit to be a base in the French N→V conversion. In a list of the first 100 non-eventual nouns contained in the *Petit Larousse*, only two are a base of N→V converts, and only one of them, *acier* 'steel', is the stem of a verb with a transparent meaning, *aciérer* 'to cover with steel'.[29] Notice, however, that this finding rests on a very weak empirical basis. The nouns considered are very few, and the data are limited to strongly lexicalized items. More research is needed to get reliable quantitative results. So I will just characterize the database with respect to the 143 nouns that are the base of verbs with a transparent meaning. Turning these observations into well-founded constraints and disentangling grammatical constraints on the input and conditions for use and lexicalization of the output must be left to further research.

The following semantic characteristics of the base nouns can be gathered from the database:

- Most base nouns denote an instrument (42 items),[30] a substance (36 items), a container (seven items), or a body part (nine items); see Tables 8 to 11 in the Appendix.

---

[29]The other, *abîme* 'abyss', has *abîmer* 'to damage' as a convert, but that verb has a meaning that does not seem to be derived in a straightforward way from the noun's meaning.

[30]Cf. "Les verbes converts instrumentaux sont parmi les plus nombreux. Ils sont mentionnés dans toutes les études portant sur la conversion et sont généralement définis comme signifiant 'utiliser N', selon le schéma … X utiliser Nb" (Tribout 2010b: 263).

- Only one noun, *enfant* 'child', denotes a human being. The derived verb, *enfanter* 'to give birth to', is infrequent and strongly marked as belonging to the literary register.

- Only two nouns denote an animal, *raton* 'young rat', and *zèbre* 'zebra'. *Ratonner* 'to commit a racial attack (*ratonnade*) on North-African immigrants' has no transparent semantic relationship to its base. As to *zèbre*, the derived verb, *zébrer* 'to stripe', is only weakly transparent: rather than to the animal, it refers to a visual pattern, black stripes upon a white ground.

Regarding the formal properties of the base nouns, short words are preferred: most of them are mono- or disyllabic, only three (*ankylose* 'ankylosis', *courbature* 'ache, stiffness', and *magasin* 'store') have three and only one (*photographie* 'photography') has four syllables.

Nouns consisting of one morpheme only are clearly preferred; only *tambourin* 'tambourine' and *photographie* 'photography' may be segmented into morphemes. There are no agent nouns in *–(at)eur* and no quality nouns in *–(i)té* in the stems of derived verbs.

9   Reduced or lacking transparency – construed lexemes in time

The database contains several verbs whose relationship with the base noun is not fully transparent or not transparent at all. For none less than 25 of the 170 verbs, no underspecified semantic form could be identified, which means that the meaning of the base noun is not a feature of the derived verb, see the examples in (27):

(27)   a.   fourrager   fourrage
            'to rummage through'   'forage'
       b.   fronder   fronde
            'to satirize'   'slingshot; revolt'
       c.   gueuler   gueule
            'to yell, to bawl'   'mouth'

Ten verbs can be analyzed as having undergone some post-morphological change along one of the familiar paths of semantic change or variation, such as narrowing or widening an original meaning. Examples are shown in Table 5:

A particular kind of incomplete semantic transparency of the converted verb is due the fact that, rather than the verb, the base noun underwent a change after the derived verb entered the mental lexicon. Examples are *échafauder* 'to put up scaffolding' and *mitrailler* 'to machine-gun'. The base noun of *échafauder*, *échafaud*, does not mean 'scaffolding' any longer, it means 'executioner's platform' in modern-day French. The verb's meaning came about when *échafaud* still meant 'scaffolding'. Likewise, *mitrailler* 'to

Table 5: Post-morphological change along some familiar paths

N	English	V	English	kind of change
fer	'iron'	*ferrer*	'to shoe (a horse)', 'to strike (a fish)'	narrowing
jardin	'garden'	*jardiner*	'to do some gardening'	narrowing
mur	'wall'	*emmurer*	'to wall (a prisoner)'	narrowing
ombre	'shadow'	*ombrer*	'to shade, to hatch'	narrowing
peau	'skin'	*peler*	'to peel'	narrowing
piste	'trace'	*dépister*	'to track down (a game)'	narrowing
plume	'feather'	*plumer*	'to pluck (a bird)'	narrowing
tapis	'carpet'	*tapisser*	'to decorate (a wall and similar)'	widening

machine-gun' was created when the noun, *mitraille*, still meant 'machine gun'. Its meaning changed to 'hail of bullets', which lessened the semantic transparency of the derived verb.

The formal transparency may also be obscured, i.e. the noun's stem may differ to some extent from the derived verb's stem.[31] The variation in such cases mostly is due to morphologization of a phonological variation existing at an earlier stage of the language and may be made less opaque by the existence, in modern French, of other examples that exhibit the same lexical variation. The variation between /o/ and /ɛl/ or /əl/ as in *peau* /po/ 'skin' – /pɛl/ 'peels' and *peler* /pəle/ 'to peel' is such a case. Its transparency is improved by the presence of numerous items like those given in (28):

(28) a. nouveau
          /nuvo/
          'new.MAS'

      b. nouvelle
          /nuvɛl/
          'new.FEM'

      c. renouveler
          /rənuvəle/
          'to renew'

      d. niveau — niveler
          /nivo/   /nivəle/
          'level'   'to level'

---

[31] For a complete list of the kinds of allomorphy involved in N→V conversion see Tribout (2010b: 114f). She argues that even totally opaque pairs such as *pierre* 'stone' and *lapider* 'to stone' may be analyzed as cases of conversion, because they are related by suppletion (Tribout 2010b: 110, 118).

But this is not always the case. See the right-most column in Table 6.

Table 6: Stem variation in N-V pairs

N	English	V	English	remarks
ciseau /sizo/	'chisle'	*ciseler* /sizəle/	'to chisle'	see (25)
faux /fo/	'scythe'	*faucher* /foʃe/	'to scythe'	isolated variation
grain /gʁɛ̃/	'grain'	*engrener* /ɑ̃gʁəne/	'to engage with'	no transparency
hiver /ivɛʁ/	'winter'	*hiverner* /ivɛʁne/	'to winter'	cf. *jour -journée*
marteau /maʁto/	'hammer'	*marteler* /maʁtəle/	'to hammer	see (26)
nœud /nø/	'knot'	*nouer* /nue/	'to knot'	cf. *jeu-jouer*
poil /pwal/	'hair'	*peler* /pəle/	'to peel'	cf. *moi-me*
poil /pwal/	'hair'	*épiler* /epile/	'to depilate'	native vs. borrowed
sang /sɑ̃/	blood'	*saigner* /sɛɲe/	'to bleed'	isolated variation

Most of these cases of reduced or lacking transparency have originated from the development of the grammar combined with the effects of lexicalization. N→V conversion has been a persistent rule in a changing grammar. It was present at the Latin stage of the language (see Table 7), and endured throughout the centuries up to the present day, while there happened important changes elsewhere in the grammar.

Table 7: N→V conversion in Latin

N	English	V	English
cor cordis	heart	*recordor*	to call to mind, to remember
glacies	ice	*glaciō*	to freeze
navigium	vessel, ship	*navigō*	to navigate, to sail
onus oneris	cargo, burden, load	*onerō*	to load, to burden
pignus pignoris	bet, stake, pledge	*pignorō*	to pledge
pilum	hair	*pilō*	to depilate
pugnus	fist	*pugnō*	to fight
sal	salt	*salō*	to salt
vēlum	curtain, sail, covering	*vēlō*	to enfold, envelop, veil

When speakers found it useful for communication, the output of the rule entered into usage and was lexicalized. This happened at various periods, when the meaning of the base noun could be different from today's, and when there was a regular phonological variation given up later. But the original forms and meanings could remain in the lexicon.

Moreover, once a construed word has entered the mental lexicon, its meaning may develop freely, which leads to reduced or lacking transparency with respect to the original meaning, founded on some sf and its conceptual resolution.

What does that mean for the morphological process as a part of mental grammar? Remember that word formation rules are thought to have a double purpose: they create possible words, and they analyze existent words. Hence the N→V conversion rule will not create opaque or semi-transparent forms. However, as a means of learning and understanding construed lexemes, it will also cope with semi-transparent forms, to the extent that suitable variation patterns are present in the lexicon. Thus speakers will presumably be able to relate *ciseler* /sizəle/ to *ciseau* /sizo/ or *marteler* /maʁtəle/ to *marteau* /maʁto/, because these pairs show a variation pattern that is also present elsewhere in the lexicon. In addition, a clear semantic relationship between the noun and the verb certainly is a strong support to transparency. It would be interesting to see experimental research on this point.

Acknowledgments

Appendix

The Appendix contains some tables that would have disturbed the reading process of the main text.

Table 8: Verbs derived from nouns that denote a body part

V	English	N	English
ciller	to blink	*cil*	eyelash
enculer	to sodomize	*cul*	arse
doigter	to use one's fingers correctly on a piano and similar	*doigt*	finger
griffer	to scratch	*griffe*	claw
gueuler	to yell	*gueule*	mouth
manier	to handle	*main*	hand
peler	to peel	*peau*	skin
plumer	to pluck (a bird)	*plume*	feather
dépiler	to depilate	*poil*	hair
sourciller	to raise one's eyebrows	*sourcil*	eyebrow
talonner	to follow someone's heels	*talon*	heel
zyeuter	to take a look at	*yeux*	eyes

Table 9: Verbs derived from nouns that denote an instrument

V	English	N	English
ancrer	to anchor	*ancre*	anchor
arquer	to curve	*arc*	bow
basculer	to topple over	*bascule*	seesaw
bêcher	to dig with a spade	*bêche*	spade
boulonner	to bolt	*boulon*	bolt
brosser	to brush	*brosse*	brush
ceinturer	to surround	*ceinture*	belt
ciseler	to chisle	*ciseau*	chisle
claironner	to shout from the rouftops	*clairon*	bugle
clouer	to nail	*clou*	nail
cravacher	to whip	*cravache*	whip
crocheter	to pick (a door, a lock)	*crochet*	picklock
chaîner	to put on snow chains	*chaîne*	chain
faucher	to scythe	*faux*	scythe
filtrer	to filter	*filtre*	filter
flinguer	to blow away, to shoot	*flingue*	gun
flûter	to produce a flute-like sound	*flûte*	flute
fouetter	to flog, to whip	*fouet*	whip
fourcher	to split	*fourche*	fork
freiner	to brake	*frein*	brake
fronder	to satirize	*fronde*	sling, revolt
fusiller	to shoot (a condemned person)	*fusil*	rifle
hacher	to chop	*hache*	ax
griller	to grill	*gril*	grill
limer	to file	*lime*	file
marteler	to beat, to pound	*marteau*	hammer
menotter	to handcuff	*menottes*	handcuffs
miner	to mine, to sap	*mine*	mine
mitrailler	to machine-gun	*mitraille*	hail of bullets
peigner	to comb	*peigne*	comb
photographier	to photograph	*photographie*	photography
pilonner	to bombard, to grind	*pilon*	pestle
poignarder	to stab	*poignard*	dagger
raboter	to plane	*rabot*	plane
sabrer	to cut down	*sabre*	sword
scier	to saw	*scie*	saw
tambouriner	to hammer, to drum	*tambourin*	tambourin
tamiser	to sieve	*tamis*	sieve
téléphoner	to phone	*téléphone*	phone
se tirebouchonner	to be twisted, to be wrinkled	*tirebouchon*	corkscrew
visser	to screw on	*vis*	screw
vriller	to bore, to pierce	*vrille*	spiral

Table 10: Verbs derived from nouns that denote a substance

V	English	N	English
aérer	to air	*air*	air
amidonner	to starch	*amidon*	starch
argenter	to silver	*argent*	silver
bétonner	to concrete	*béton*	concrete
beurrer	to butter	*beurre*	butter
bitumer	to asphalt, to tarmac	*bitume*	bitumen
charbonner	to blacken	*charbon*	coal
chiffonner	to crumple	*chiffon*	mousseline, rag
cimenter	to cement	*ciment*	cement
cirer	to polish (shoes, the floor)	*cire*	wax
crotter	to muddy	*crotte*	dropping
cuivrer	to bronze, to copper	*cuivre*	copper
émailler	to enamel	*émail*	enamel
fariner	to flour	*farine*	flour
ferrer	to shoe (a horse)	*fer*	iron
feutrer	to felt	*feutre*	felt
enfieller	to fill with bile	*fiel*	bile, venom
enfumer	to fill with smoke	*fumée*	smoke
gazer	to gas	*gaz*	gas
givrer	to frost over	*givre*	frost
goudronner	to tarmac	*goudron*	tar
graisser	to grease	*graisse*	grease
huiler	to oil	*huile*	oil
larder	to lard	*lard*	fat streaky bacon
pimenter	to put chillies in	*piment*	hot pepper
plastiquer	to carry out a bomb attack on	*plastic*	plastic explosive
plâtrer	to plaster	*plâtre*	plaster
plomber	to fill (a tooth), to seal	*plomb*	lead
poivrer	to pepper	*poivre*	pepper
poudrer	to powder	*poudre*	powder
rouiller	to rust	*rouille*	rust
sabler	to sandblast	*sable*	sand
saigner	to bleed	*sang*	blood
savonner	to rub soap on	*savon*	soap
saler	to salt	*sel*	salt
sucrer	to put sugar in	*sucre*	sugar

Table 11: Verbs derived from nouns that denote a container

V	English	N	English
archiver	to archive	*archives*	archive
cuver	to ferment	*cuve*	tank
engainer	to put into its sheath	*gaine*	sheath
engranger	to gather in, to store	*grange*	barn
emmagasiner	to store	*magasin*	store
emprisonner	to imprison	*prison*	prison
enregistrer	to register	*registre*	register

References

Aurnague, Michel & Marc Plénat. 1997. Manifestations morphologiques de la relation d'attachement habituel. In Danielle Corbin, Bernard Fradin, Benoît Habert, Françoise Kerleroux, Marc Plénat & SILEX (eds.), *Mots possibles et mots existants* (Silexicales), 15–24. Villeneuve d'Ascq: URA 382 - CNRS & Université de Lille 3.

Aurnague, Michel & Marc Plénat. 2007. *Contraintes sémantiques et dérivation en é-: attachement habituel, naturalité et dissociation intentionnelle.* Carnets de Grammaire 16. CLLE-ERSS.

Aurnague, Michel & Marc Plénat. 2008. Sémantique de l'espace et morphologie : le cas de la préfixation en *é-*. *Bulletin de la Société de Linguistique de Paris* CIII. 201–236.

Booij, Geert. 2010. *Construction morphology.* Oxford: Oxford University Press.

Booij, Geert & Jenny Audring. 2017. Construction morphology and the parallel architecture of grammar. *Cognitive science* 41(S2). 277–302.

Fradin, Bernard. 2005. On a semantically grounded difference between derivation and compounding. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (eds.), *Morphology and its demarcations*, 161–182. Amsterdam: John Benjamins.

Lang, Ewald & Claudia Maienborn. 2011. Two-level semantics: Semantic form and conceptual structure. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 709–740. Berlin: De Gruyter Mouton.

Mayo, Bruce, Marie-Therese Schepping, Christoph Schwarze & Angela Zaffanella. 1995. Semantic in the derivational morphology of Italian: Implications for the structure of the lexicon. *Linguistics* 33. 883–938.

Namer, Fiammetta & Evelyne Jacquey. 2003. Lexical semantics and derivational morphology: The case of the popular '*é-*' prefixation in French. In *2nd International Workshop on Generative Approaches to the Lexicon*, 115–123.

Namer, Fiammetta & Evelyne Jacquey. 2012. Word formation rules and the generative lexicon: Representing noun-to-verb versus verb-to-noun conversion. *Advances in Generative Lexicon Theory*. 385–414.

Schwarze, Christoph. 2011. Wie syntaktisch ist die Morphologie? Der vierte Verb-stamm des Italienischen in einem Schichtenmodell. In *Was grammatische Kategorien miteinander machen: Form und Funktion in romanischen Sprachen von Morphosyntax bis Pragmatik: Festschriftø für Ulrich Wandruszka*, 57–78. Tübingen: Narr.

Schwarze, Christoph. 2012. *Le vert*, *les jeunes*, *l'important*: Aspects sémantiques de la conversion adjectif-nom. *Scolia* 26.

Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure and lexical form. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3, 57–150. Cambridge: Cambridge University Press.

Todaro, Giuseppina. 2017. Nomi (e aggettivi) che diventano verbi tramite prefissazione: quel che resta della parasintesi. Università Roma Tre & Université Toulouse Jean-Jaurès dissertation.

Tribout, Delphine. 2010a. How many conversions from verb to noun are there in French? In *Proceedings of the HPSG 2010 conference*, 341–357. Stanford: CSLI Publications.

Tribout, Delphine. 2010b. *Les conversions de nom à verbe et de verbe à nom en français.* Université Paris Diderot - Paris 7 Thèse de doctorat.

von Heusinger, Klaus & Christoph Schwarze. 2006. Underspecification in the semantics of word-formation: The case of denominal verbs of removal in Italian. *Linguistics* 44(6). 1165–1194.

Chapter 20

Lexeme equivalence or rivalry of lexemes?

Jana Strnadová

Google, inc.

This paper deals with the purported interchangeability between nouns and adjectives derived from nouns in French. The question of equivalence or rivalry between a morphologically complex adjective and a syntactic construction containing a morphologically-related noun links a field of studies on rivalry between inflected word forms, derivational suffixes or different syntactic constructions to express the same meaning. This paper then presents a corpus-based study of the relative distribution of nominal or adjectival realizations of a modifier of the same head noun and discusses some motivations that play a role in the choice of one or the other strategy.

1 Introduction

Both in syntax and in morphology, the same content can be expressed by different structural means.

In syntax, this may take the form of valency alternations such as the English dative alternation (e.g. *Mary gave a watch to me* vs. *Mary gave me a watch*) or of word order alternations such as exemplified by the position of French attributive adjectives with respect to their governing noun. Such alternations have been the focus of much attention in the recent literature which focuses on establishing the interplay of various non-categorical factors (see e.g. Bresnan et al. 2007 on the dative alternations, Thuilier 2012 on French adjectives).

In morphology, the consensus has long been that such alternations are inexistent or unexpected: in inflection, a unique form was assumed to fill each cell of a lexeme's paradigm (Anderson 1992, Stump 2001), in word formation, rivalry between affixes was taken to be resolved by blocking (Aronoff 1976). This consensus has progressively collapsed in the last two decades. Under the impulsion of Thornton (2012), the phenomenon of overabundace, where multiple forms fill a paradigm cell, has become a central issue in inflectional morphology (see e.g. Bermel & Knittl 2012 for Czech noun declension,

Stump 2016, Bonami & Crysmann this volume, Thornton this volume). Likewise, situations of non-categorical competition between derivational processes have moved from the fringes (Rainer 1988, Plag 1999) to the center of attention for derivational morphologists (Lindsay & Aronoff 2013, Villoing 2009, Tribout 2010, Fradin 2012, Koehl 2012, Namer 2013, Strnadová 2014).

In this paper, I focus on situations of alternation between the morphological or syntactic expression of some content. This is familiar in the context of inflection where overabundance between synthetic and periphrastic expression of paradigm cells is well-documented (Aronoff & Lindsay 2014, Bonami 2015). For example, *friendlier* and *more friendly* are both realizations of the comparative degree of the lexeme FRIENDLY. Situations in which a syntactic construction and a derivational process led to the expression of the same content have been comparatively less studied.[1] Here I will specifically examine the expression of nominal modification by a prepositional phrase containing some noun N or a denominal adjective derived from that same noun. This is illustrated in (1): the adjective *grammaticale* in (1a) and the noun *grammaire* introduced by the preposition *de* in (1b) roughly make the same contribution.

(1) a. faute grammaticale
       'grammatical mistake'
    b. faute de grammaire
       'grammar mistake'

The central questions that arise in view of such examples are 1) to what extent can the adjective and the prepositional phrase be taken to be semantically equivalent and 2) whether the two constructions should be taken to be paradigmatic alternatives in the same way as *friendlier* and *more friendly* are.

2 Background and methodology

The proximity between a denominal adjective and a prepositional phrase containing a morphologically related noun was observed as early as Dumarsais (1769: 413): "When there is a simple preposition *de*, without an article, the preposition and its complement are considered adjectively. *Un palais de roi*, is equivalent to *palais royal* 'royal palace'; *une valeur de héros* equals to *une valeur héroïque* 'heroic value'."[2] Bally (1944) used the term *transpositions* and Tesnière (1969) called this kind of adjectivisation *translations*.

The idea of equivalence between the two constructions was discussed later for example by Bosredon (1988) or Bartning & Noailly (1993), or in a more semantic approach, by

---

[1]In French, for example, the topic of possible competition between morphologically complex words and syntactic phrases has been studied for causative verbs (Dal & Namer 2003).

[2]Orig. "Lorsqu'il n'y a qu'une simple préposition *de*, sans l'article, la préposition et son complément sont pris adjectivement. *Un palais de roi*, est équivalent à un *palais royal*; *une valeur de héros* équivaut à *une valeur héroïque*."

Nowakowska (2004) and Roché (2006: 380), who insists on the equivalence by describing "the adjectivized noun lexically as it can be syntactically with the preposition de".[3]

Functional and semantic equivalence between a denominal adjective and its base noun used in a prepositional phrase is thus considered as one of the characteristics of denominal adjectives. The examples (1)-(3) show the possibility to substitute a derived adjective with a prepositional phrase.

(2) a. le climat social
         'social climate'

     b. le climat de la société
         'climate of the society'

(3) a. secret de famille
         'family secret'

     b. secret familial
         'family secret'

The question is then to what extent are prepositional phrases functionally and semantically equivalent to denominal adjectives in French? This question was of central importance in the 1980s and 1990s. At that time, the interest focused on the argument realization of the head noun with the goal of defining the syntactic and semantic relations within a noun phrase (Bartning 1980, Pinchon 1980, Monceaux 1993, etc.). These works showed that adjectives and prepositional phrases are not equivalent and are not interchangeable without any restriction.

More recently, Deléger & Cartoni (2010) studied the use of an adjective or of its corresponding prepositional phrase in specialized or general medical corpora and showed that there is a preference for the use of adjectives in specialized texts, while corresponding prepositional phrases are more frequent in non-specialized texts (4).

(4) a. rythme cardiaque
         'cardiac rhythm'

     b. rythme du cœur
         'heart rhythm'

Finally, Boleda et al. (2012) provided some statistical evidence supporting the claim that an ethnic adjective, which is in a certain way a denominal adjective, cannot be interpreted as the argument of the noun as in (5). The adjective acts as a simple modifier. In their study, the modified noun is a predicative noun.

(5) a. French agreement
     b. agreement by France

---

[3]Orig. "le nom adjectivé lexicalement comme il peut l'être syntaxiquement par la préposition de".

All these studies have one thing in common: they do not differentiate between cases where the prepositional phrase contains a fully determined NP and those where it contains just a bare noun. In (6), the adjective *gouvernementale* is in competition with the prepositional phrase containing a definite noun phrase (*le gouvernement*[4]), while in (7), the preposition governs a bare noun (*publicité*). Semantically, in (6), the noun phrase within the PP refers to the cabinet, while the noun phrase in (7) doesn't refer to an advertisement.

(6) a. décision gouvernementale
         'governmental decision'

    b. décision du gouvernement
         'the government's decision'

(7) a. campagne publicitaire
         'advertising campaign'

    b. campagne de publicité
         'advertising campaign'

Contrary to these previous studies, I examine denominal adjectives and their syntactic equivalents with the restriction on prepositional phrases containing a bare noun introduced by the preposition *de*. In such cases, the noun does not head a referential expression. Note that this restriction entails that the investigation be limited to cases where the adjective is derived from a common noun, as exemplified in (8). Adjectives derived from proper names are excluded since the proper names being definite noun phrases are referential expressions.

(8) campagne de publicité / publicitaire

Three situations must be distinguished concerning the availability of a denominal adjective corresponding to a French noun: (i) there is an adjective regularly derived from a noun (9); (ii) there is an adjective with a formal mismatch in comparison with the noun (10); (iii) there is no adjective (11) and hence a prepositional phrase is the only possible realization of the modifier (12).

(9) PUBLICITÉ 'advertisement' → PUBLICIT-AIRE 'advertising'

(10) LANGUE 'language' ~ LINGUISTIQUE / *LANGUIQUE

(11) a. DÉCOLLAGE 'take-off' → ?

    b. ARRIVÉE 'arrival' → ?

    c. SECOURS 'emergency' → ?

(12) a. piste de décollage
         'runway'

---

[4]The definite article *le* is merged with the preposition *de* which results in *du gouvernement*.

b. hall d'arrivée
   'arrival hall'

c. issue de secours
   'emergency exit'

It is notable that languages differ in this respect. As Table 1 shows, Czech tends to have available denominal adjectives where French does not. English has the same gap as French but uses compounding rather than PP modifications as an alternative strategy.

Table 1: Comparison between French, Czech and English noun phrases

| French | | Czech | | English |
N_1	deN_2	A_d	N_1	N_2N_1
hall	*d'arrivée*	*příjezdová*	*hala*	*arrival hall*
issue	*de secours*	*nouzový*	*východ*	*emergency exit*
carte	*de crédit*	*kreditní*	*karta*	*credit card*
piste	*de décollage*	*vzletová*	*dráha*	*runway*

To study the rivalry between denominal adjectives and prepositional phrases, the following resources were used:

1. A lexicon of noun-adjective pairs from *DenALex* (Strnadová & Sagot 2011) and *Lexique3* (New 2006). 5,888 noun-adjective pairs with regularly derived adjectives and 234 noun-adjective pairs with a formal mismatch were obtained in this way.

2. The corpus *Est républicain* which covers three years of a local newspaper (1999, 2002, 2003) and contains 119.5 million word tokens with morphosyntactic annotation (Seddah et al. 2012).

Table 2 illustrates the diversity of denominal adjectives contained in the lexicon.

The following methodology was applied: search in the corpus for all combinations where a noun is followed by an adjective from the lexicon or by a prepositional phrase with *de* containing a noun from the lexicon (13).

(13)   a. lexicon entry: PUBLICITÉ - PUBLICITAIRE 'advertisment - advertising'

b. corpus search$_1$: X_N publicitaire

c. corpus search$_2$: X_N de publicité

d. search result: campagne de publicité, campagne publicitaire, etc.

The vocabulary used throughout this article can be defined as follows: N_1 is the modified noun or the head noun. A_d is the modifying denominal adjective. N_2 is the noun morphologically related to the adjective A_d. The term *combination* stands for the search results N_1A_d and N_1deN_2. In each combination, deN stands for the nominal realization and A_d for the adjectival realization of the modifying concept N.

Table 2: Sample of French Denominal Adjectives

Suffix	Noun	Adjective
-aire	CELLULE 'cell'	CELLULAIRE 'cellular'
-al	PARENT 'parent'	PARENTAL 'parental'
-el	CULTURE 'culture'	CULTUREL 'cultural'
-esque	CARNAVAL 'carnival'	CARNAVALESQUE 'of carnival'
-eux	ANGINE 'angina'	ANGINEUX 'anginal'
-ien	MICROBE 'microb'	MICROBIEN 'microbial'
-ier	CÔTE 'coast'	CÔTIER 'coastal'
-ique	MÉTHODE 'method'	MÉTHODIQUE 'methodical'
-u	FEUILLE 'leaf'	FEUILLU 'leafy'

For each triple $\langle N_1, A_d, N_2 \rangle$, I computed the frequency F_1 of the $N_1 A_d$ of the noun-adjective sequence, the frequency F_2 of the $N_1 de N_2$ sequence, their sum frequency $SumFreq = F_1 + F_2$ and the relative frequency of the $N_1 A_d$ sequence, $Rfreq = \frac{F_1}{SumFreq}$. For instance, for the triple ⟨campagne, publicitaire, publicité⟩, the corpus contains 40 occurrences of *campagne publicitaire* and 27 occurrences of *campagne de publicité*; hence $SumFreq = 67$ and $Rfreq = \frac{40}{40+27} \approx 0.6$.

3 Corpus-based results

A first study focused on the pairs containing a regular denominal adjective, i.e. there is no formal mismatch between the noun and the adjective except for the suffix. 139,838 types of combinations (out of 1,137,137 occurrences) were collected. 45% of nouns (2,686 lexemes) from the lexicon were attested in the corpus. Likewise, 30% of adjectives (1,708 lexemes) were attested. Incomplete attestation was to be expected, since the lexicon contains many scientific terms which are not found in a journalistic corpus and many types have a very low frequency anyway.

The data distribution is presented in Table 3.

There is an inverse correlation between the token frequency of the triple (*SumFreq*) and the proportion of cases where both strategies are attested. In particular, whereas only 4% of triples are attested in both strategies overall, this proportion rises to 26% for triples with a *SumFreq* above 1,000.

For the rest of the study, only the types with a sum frequency above 10 were taken into account. At this threshold, there are 17% of cases which can be realized either as an adjective or as a prepositional modifier and which are then possible rivals. This corresponds to 937 different nouns covering 16% of the lexicon and 659 adjectives corresponding to 11% of the lexicon.

Table 3: Type counts of $N_1 A_d$ and $N_1 de N_2$ combinations by sum token frequency of the triple

SumFreq	All	Only $N_1 A_d$	Only $N_1 de N_2$	Both	%
≥0	139,838	70,876	63,145	5,817	=4%
≥10	13,422	6,175	4,986	2,261	=17%
≥100	1,586	687	535	364	=23%
≥1000	100	45	29	26	=26%

46% of cases only have an adjectival realization for the same head noun and 37% of combinations only have the nominal realization. This leads to a U-shaped distribution with many cases at the edges and few cases in the middle of the distribution, what Zuraw (2016) calls a "polarized distribution". If denominal adjectives and prepositional phrases were in free variation, then many more cases would be expected in the middle of the distribution.

Table 4 shows the number of types in each interval of the distribution. As can be seen, many cases have a strong preference for one or the other realization. There are only 154 types with a relative frequency between 0.4 and 0.6, which could be described as real cases of free variation. I will call pairs having such a distribution *strong rivals*.

Table 4: Distribution of relative frequencies of triples $\langle N_1, A_d, N_2 \rangle$ with *SumFreq* ≥10

Rfreq interval	% of data	# of types	
0 < Rfreq < 1	17%	2,261 types	
0.2 < Rfreq < 0.8	5%	580 types	
0.4 < Rfreq < 0.6	1%	154 types	→ Strong rivals?

The U-shaped distribution of relative frequencies for triples is shown in Figure 1. In order to make the figure readable, only data points with *SumFreq* ≥ 20 and 0 < Rfreq < 1 are shown. If no threshold was used, the edges would be much higher as most of the cases prefer one or the other realization.

Table 5 presents examples for the whole spectrum of relative frequencies, ranging from a strong preference for the adjectival realization at the top (*Rfreq* = 0.93 for the triple ⟨spectacle, musical, musique⟩) to a strong preference for the nominal realization at the bottom (*Rfreq* = 0.06 for the triple ⟨commission, disciplinaire, discipline⟩).
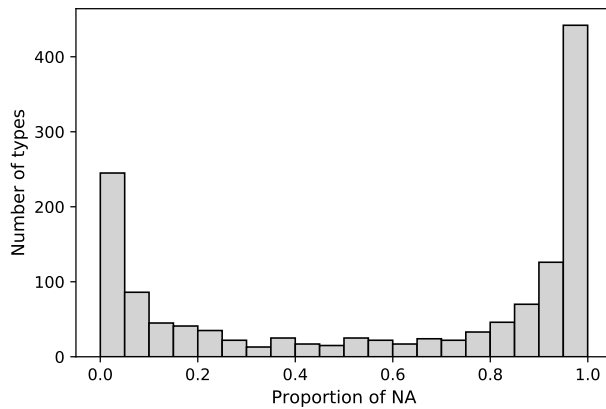
Figure 1: Distribution of relative frequencies of triples $\langle N_1, A_d, N_2 \rangle$

Table 5: Examples of $N_1 A_d / N_1 de N_2$ combinations with their frequencies

N_1	$A_d / de N_2$	Freq	Rfreq	Translation
spectacle	*musical*	409	0.93	'music show'
	de musique	31		
musée	*archéologique*	640	0.9	'archaeological museum'
	d'archéologie	69		
réseau	*électrique*	333	0.79	'electrical grid'
	d'électricité	87		
troupe	*théâtrale*	347	0.5	'theatrical troupe'
	de théâtre	347		
situation	*critique*	47	0.37	'critical situation'
	de crise	78		
soleil	*automnal*	21	0.16	'automn sun'
	d'automne	109		
commission	*disciplinaire*	15	0.06	'disciplinary committee'
	de discipline	226		

Table 6 shows some examples which could be considered in free variation between A_d and $de N_2$ since the relative frequency is situated between 0.4 and 0.6. For triples such as ⟨fête, familial, famille⟩ or ⟨troupe, theâtral, theâtre⟩, adjectival and nominal realizations are equivalent.

These strong rivals are distributed across all suffixes, as shown in Table 7 which contains a couple of adjectives which compete with their corresponding nouns introduced by *de*.

Table 6: Examples of strong rivals ($0,4 <$ Rfreq $< 0,6$)

N_1	A_d/deN_2	Freq	Rfreq	Translation
fête	familiale	102	0.42	'family party'
	de famille	143		
exposition	photographique	185	0.42	'photography exhibition'
	de photographie	258		
musicien	talentueux	38	0.44	'talented musician'
	de talent	48		
troupe	théâtrale	347	0.5	'theatrical troupe'
	de théâtre	347		
campagne	publicitaire	40	0.6	'advertising campaign'
	de publicité	27		
politique	sécuritaire	19	0.44	'security policy'
	de sécurité	24		

Table 7: Examples of strong rival adjectives sorted by suffix

Suffix	Examples of adjectives	Count
-aire	BUDGÉTAIRE 'budgetary', SÉCURITAIRE 'security'	21
-al	ARCHITECTURAL 'architectural', AUTOMNAL 'autumnal' MUSICAL 'musical' , FAMILIAL 'family'	51
-el	CONCURRENTIEL 'competitive', PROMOTIONNEL 'promotional'	20
-esque	CARNAVALESQUE 'carnaval'	1
-eux	ARGILEUX 'clay', ORAGEUX 'stormy' GLORIEUX 'glorious', PRESTIGIEUX 'prestigious'	14
-ier	LÉGUMIER 'vegetable', PRINTANIER 'spring'	11
-ique	ARCHÉOLOGIQUE 'archaeological', INFORMATIQUE 'information', TOURISTIQUE 'touristic'	36

As has been shown, the number of cases where both realizations receive the same preference is rather low.

Remember that we focused for now on cases where the formal relationship between the denominal adjective and its base noun is straightforward. One might expect to find different results where the relationship is more opaque. This is not what we found with the lexicon containing 234 noun-adjective pairs with a formal mismatch. Table 8 presents the distribution of rivals in this category according to the type frequency and Table 9 gives some examples of combinations with their frequencies. The results on this data set present a similar U-shaped distribution as we have seen in Figure 1.

Table 8: Absolute frequencies of triples $\langle N_1, A_d, N_2 \rangle$ in the corpus where A_d has an idiosyncratic form

SumFreq	# of types	Both realizations	
f≥0	29,884	1713	=6%
f≥10	3,641	673	=18%
f≥100	582	140	=24%
f≥1000	52	19	=36%

Table 9: Examples of $N_1 A_d$ / $N_1 de N_2$ with absolute and relative frequencies where A_d has an idiosyncratic form

N_1	$A_d/de N_2$	*Freq*	*Rfreq*	Translation
eau	*pluviale*	435	0.75	'rain water'
	de pluie	149		
éclipse	*solaire*	128	0.66	'solar eclipse'
	de soleil	79		
stage	*linguistique*	17	0.49	'language course'
	de langues	18		
loisir	*estival*	13	0.14	'summer leisure'
	d'été	81		

Overall, there are not many cases where the adjective and the noun are used to modify the same noun: We are far from a situation of interchangeability between the two.

4 Discussion

4.1 Grammar conditions

The low number of strong rivals is certainly due at least in part to grammatical or semantic constraints. For example, the acceptability of the $N_1 de N_2$ realization is reduced

where N_1 is a deverbal noun. A likely explanation is that prepositional complements of deverbal nouns tend to be interpreted as realizing an argument of the noun (14a), while adjectives can act as simple modifiers (14b). The same deN_2 is fine if the head noun is not deverbal (14c).

(14) a. ?visite d'archéologie
          'visit of archaeology'

    b. visite archéologique
        'archaeological visit'

    c. laboratoire d'archéologie
        'archaeological laboratory'

Another example of such constraints, but this time in favor of free variation, is represented by quality nouns such as *exception* 'exception', *prestige* 'prestige', *talent* 'talent', *etc.* and derived qualifying adjectives, such as *talentueux* 'talented', *prestigieux* 'prestigious', *etc.* In this case, both the PP and the adjective can be functionally equivalent as shown in (15).

(15) a. musicien de talent
          'talented musician'

    b. musicien talentueux
        'talented musician'

With this being said, there is a large residue of examples with preference for one or the other type of modifier without any clear grammatical motivation. I consider these to be a matter of usage-based conventionalization. Therefore, in (16), the very strong preference for the given alternative —383 *versus* 1 for (16a) and 62 *versus* 5 for (16b) — is only a matter of pure convention. In certain cases, a partial semantic specialization can be observed. This is the case for the "false rivals" in (17) which do not have the same meaning.

(16) a. fourniture scolaire 'school supplies' / f = 383
    b. sac d'école 'school bag' / f = 62

(17) a. sortie scolaire 'school outing' / f = 65
    b. sortie d'école 'end of the school day' / f = 73

In conclusion, denominal adjectives and prepositional phrases with *de* are not in free variation. Some cases can be explained by grammar, but conventionalization seems to be an important factor which should be studied more in detail.

4.2 Lexical conditions

Looking at the distributions of modifiers, the choice between adjectival or prepositional modifiers seems notably conditioned by the lexical identity of the modifying concept.

Thus, if your modifier denotes 'security', there is a clear preference to use a PP *de sécurité*, while if your modifier denotes 'region', then the preferred modifier will be the adjective RÉGIONAL, as shown in Figure 2.
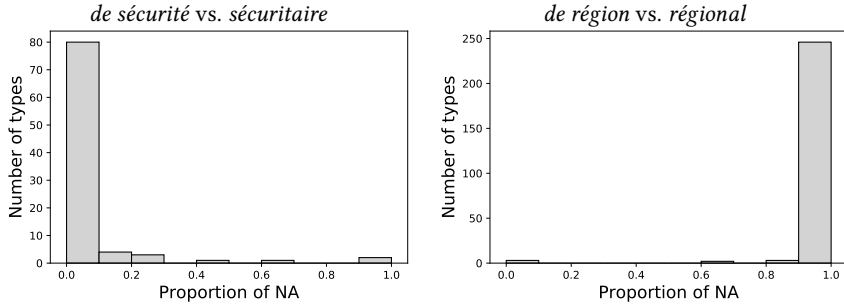
Figure 2: Distribution of relative frequencies of triples $\langle N_1, A_d, N_2 \rangle$ where N_2 = sécurité / région

Each strategy has its own distribution. For example for the pairs THÉÂTRE 'theater' / THÉÂTRAL 'theatrical' and MUSIQUE 'music' / MUSICAL 'musical', there is a real rivalry between the adjectival and the prepositional realization, as illustrated in Figure 3.
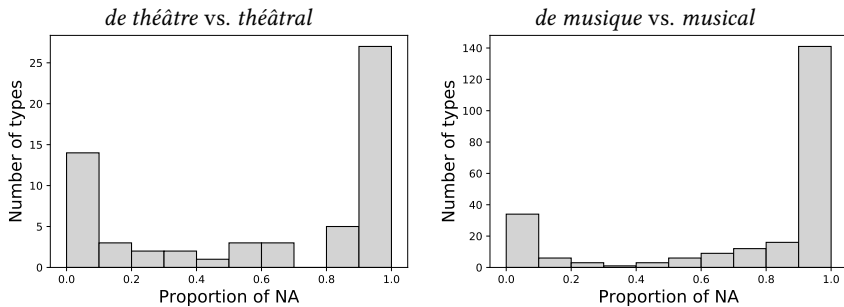
Figure 3: Distribution of relative frequencies of triples $\langle N_1, A_d, N_2 \rangle$ where N_2 = théâtre / musique

The four seasons, such as the example (18), can be presented as another good example: as shown in Figure 4, the use of a PP is much more frequent than the use of denominal adjectives which are commonly used only in a poetic register.

(18)   balade d'automne / automnale 'autumn walk'

Thus, register can also play a role in the choice of the realization. This observation corresponds to the conclusion of Deléger & Cartoni (2010) on medical texts where adjectives are more frequent in specialized texts than in more general texts.
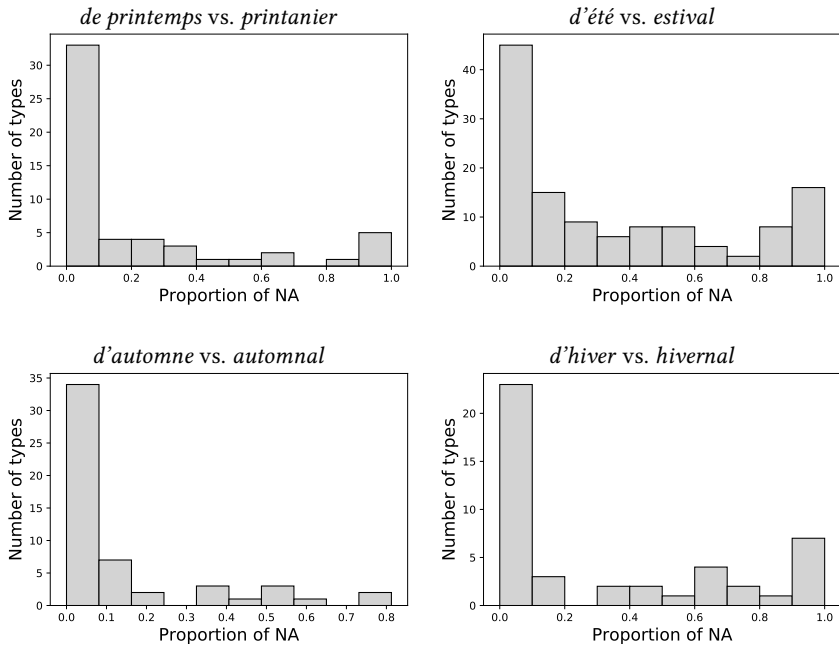
Figure 4: Distribution of relative frequencies of triples $\langle N_1, A_d, N_2 \rangle$ where N_2 is a season

Another question is to know to which extent the choice of one or the other alternative is conditioned by the identity of the head noun. For example, the nouns *zone* 'zone' and *concours* 'competition' have equally distributed adjectival and prepositional modifiers, as presented in Figure 5. This would need to be assessed against the whole dataset taking into account the semantic relationship between the head noun and the modifying concept, for example by relying on the principles of distributional semantics.
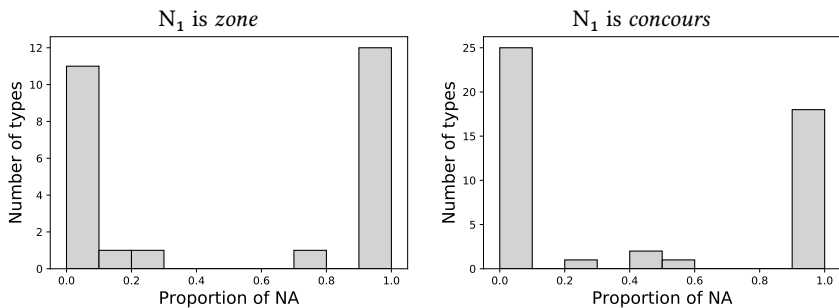
Figure 5: Distribution of relative frequencies of triples $\langle N_1, A_d, N_2 \rangle$ where N_1 = zone / concours

This section has shown that denominal adjectives and prepositional phrases are seldom equivalent. First, free variation is rare. Then, the lexical identity of the pair noun ~ adjective is decisive for the choice of the preferred realization. Finally, in many cases, this preference is purely conventional and cannot be explained in terms of grammar alone.

5 Conclusion

This paper questioned the purported equivalence between French denominal adjectives and morphologically related nouns embedded in prepositional phrases introduced by *de*. This idea has been present in the literature since at least the 18th century. In the same way as two word forms can fill the same cell of a lexeme's inflectional paradigm or two dative constructions can alternate or a synthetic and a periphrastic form can compete for degree realization on adjectives and adverbs, there are two ways to express noun modification —with a denominal adjective or with a prepositional phrase introduced by *de*. Of course, there are other linguistic means that can be used for modification, but they have not been taken for granted to the same extent.

I have shown that denominal adjectives and prepositional phrases are not in free variation (*sortie scolaire / sortie d'école*). Instead, they have a U-shaped distribution with a majority of cases favoring one or the other strategy and only few cases in the middle of the distribution. In general, there is some usage-based conventionalization which is not written in any grammar rules but learned implicitly when learning the language. Some language register preference may also play a role.

This paper presents a certain phenomenology of the question and the overview of what kinds of factors need to be taken into account and studied in more detail with respect to the choice between adjectival and nominal realization. Moreover, not only is it important to look into the rivals, but one also needs to look into the edges of the distribution: are there any specific constructions where the use of one or the other strategy can be predicted? A quick look at the data reveals that, for example, in combinations which favor nominal realization, there are cases where N_1 is a deverbal noun and the noun embedded in the prepositional phrase saturates its argument structure (*demande de soutien* 'request for support', *abandon de chien* 'dog abandonment') or cases where N_2 is a deverbal noun and there is no adjective derived from it (*horaire d'ouverture* 'opening hours', *issue de secours* 'emergency exit'). Another group that favors $N_1 \, de N_2$ are combinations where N_1 is a quantity or a measure noun (*vingtaine de commerçants* 'twenty of shopkeepers', *tonne d'acier* 'ton of steel').

To conclude, both denominal adjectives and nouns embedded in prepositional phrases with *de* can be used as modifiers, but they usually do not have the same distribution or the same meaning. This brings us to a more theoretical question: could a prepositional phrase be considered as a possible candidate for the *modifier* cell of a derivational paradigm? As could be seen, especially nouns for which there is no corresponding derived adjective would have this cell empty for a synthetic form, but they could have it filled with a prepositional phrase. This could be considered as a sort of periphrasis, in a very similar

way as inflectional paradigms contain synthetic and periphrastic forms. The results of our corpus study suggest that extending this possibility to all lexemes would bring many new challenges.

References

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: MIT Press.

Aronoff, Mark & Mark Lindsay. 2014. Partial organization in languages: La langue est un système où la plupart se tient. In Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé & Fabio Montermini (eds.), *Proceedings of Décembrettes 8*, 1–14. Bordeaux: CLLE-ERSS.

Bally, Charles. 1944. *Linguistique générale et linguistique française.* Paris: PUF.

Bartning, Inge. 1980. *Remarques sur la syntaxe et la sémantique des pseudo-adjectifs dénominaux en français*. Stockholm: Acta Universitatis Stockholmiensis.

Bartning, Inge & Michèle Noailly. 1993. Du relationnel au qualificatif: flux et reflux. *L'Information grammaticale* 58. 27–32.

Bermel, Neil & Luděk Knittl. 2012. Morphosyntactic variation and syntactic constructions in Czech nominal declension: Corpus frequency and native-speaker judgments. *Russian Linguistics* 36. 91–119.

Boleda, Gemma, Stefan Evert, Berit Gehrke & Louise McNally. 2012. Adjectives as saturators vs. modifiers: Statistical evidence. In Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit Weidman Sassoon, Katrin Schulz & Matthijs Westera (eds.), *Logic, language and meaning*, 112–121. Dordrecht: Springer.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25. 63–110.

Bosredon, Alain. 1988. Un adjectif de trop: l'adjectif de relation. *L'Information grammaticale* 37. 3–7.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Dal, Georgette & Fiammetta Namer. 2003. Complex Words vs Phrases: The Case of Causative Verbs in French. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting*. 129–148.

Deléger, Louise & Bruno Cartoni. 2010. Adjectifs relationnels et langue de spécialité: vérification d'une hypothèse linguistique en corpus comparable médical. In *Actes de la 17e conférence sur le Traitement Automatique des Langues Naturelles, Montréal*.

Dumarsais, César Chesneau. 1769. *Logique et principe de grammaire*. Paris: Briasson, Le Breton et Hérissant fils.

Fradin, Bernard. 2012. L'interprétation des nominalisations. *Lexique* 20. 129–156.

Koehl, Aurore. 2012. *La construction morphologique des noms désadjectivaux suffixés en français*. Université de Lorraine dissertation.

Lindsay, Mark & Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Morphology in Toulouse*, 133–153. Munich: Lincom Europa.

Monceaux, Anne. 1993. Adjectivation d'un complément de nom. *Lingvisticae Investigationes* 17(2). 375–404.

Namer, Fiammetta. 2013. Adjectival Bases of French *-aliser* and *-ariser* verbs: Syncretism or Under-specification? In Nabil Hathout, Fabio Montermini & Jesse Tseng (eds.), *Morphology in Toulouse. Selected Proceedings of Décembrettes 7 (Toulouse 2-3 December 2010)*, 185–210. München: Lincom Europa.

New, Boris. 2006. Lexique 3: Une nouvelle base de données lexicales. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN)*, 892–900.

Nowakowska, Małgorzata. 2004. *Les adjectifs de relation employés attributivement*. Kraków: Wydawnictwo Naukowe Akademii Pedagogicznej.

Pinchon, Jacqueline. 1980. Syntagme prépositionnel et adjectif de relation. *Cahiers de lexicologie* 37. 91–100.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin: Mouton de Gruyter.

Rainer, Franz. 1988. Towards a theory of blocking. *Yearbook of Morphology*. 155–185.

Roché, Michel. 2006. Comment les adjectifs sont sémantiquement construits. *Cahiers de grammaire* 30. 373–387.

Seddah, Djamé, Marie Candito, Benoit Crabbé & Enrique Henestroza Anguiano. 2012. Ubiquitous Usage of a French Large Corpus: Processing the Est Republicain corpus. In *Proceedings of LREC 2012*, 3249–3254. Istanbul.

Strnadová, Jana. 2014. *Les réseaux adjectivaux: sur la grammaire des adjectifs dénominaux en français*. Université Paris Diderot & Univerzita Karlova v Praze Thèse de doctorat.

Strnadová, Jana & Benoît Sagot. 2011. Construction d'un lexique des adjectifs dénominaux. In *Actes de la 18e conférence sur le Traitement Automatique des Langues Naturelles, Montpellier*, vol. 2, 69–74.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.

Tesnière, Lucien. 1969. *Eléments de syntaxe structurale*. Paris: Klincksieck.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5(2). 183–207.

Thuilier, Juliette. 2012. *Contraintes préférentielles et ordre des mots en français*. Université Paris Diderot dissertation.

Tribout, Delphine. 2010. *Les conversions de nom à verbe et de verbe à nom en français*. Université Paris Diderot - Paris 7 Thèse de doctorat.

Villoing, Florence. 2009. Les mots composés VN. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 175–198. Saint-Denis: Presses Universitaires de Vincennes.

Zuraw, Kie. 2016. Polarized variation. *Catalan Journal of Linguistics* (15). 145–171.

Name index

Language index

Subject index

Did you like this book?

This book was brought to you for free

language
science
press
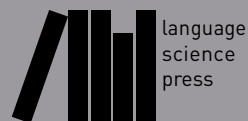
The lexeme in descriptive and theoretical morphology

Since the 1970s, the notion of a lexeme, an abstract lexical unit identifying what is common to a set of words belonging to the same inflectional paradigm, has become a cornerstone of theoretical thinking on morphology and a standard tool for description. The present volume collects papers that crucially use, discuss or question the lexeme in the context of contemporary morphology, with particular emphasis on its place in the description of word formation through the concept of a *Lexeme Formation Rule*. It will be of interest to any descriptive linguist, theoretical linguist, or psycholinguist with an interest in morphology and its interface with syntax and lexical semantics.