Complex Lexical Units

Konvergenz und Divergenz

Sprachvergleichende Studien zum Deutschen

Herausgegeben von
Eva Breindl und Lutz Gunkel

Im Auftrag des Instituts für Deutsche Sprache

Band 9

IDS INSTITUT FÜR DEUTSCHE SPRACHE

Complex Lexical Units

Compounds and Multi-Word Expressions

Edited by Barbara Schlücker

DE GRUYTER

Die Open-Access-Publikation dieses Bandes wurde gefördert vom Institut für Deutsche Sprache, Mannheim.

Redaktion: Dr. Anja Steinhauer

ISBN 978-3-11-063242-2
e-ISBN (PDF) 978-3-11-063244-6
e-ISBN (EPUB) 978-3-11-063253-8

Library of Congress Cataloging-in-Publication Data 2018964353

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Typesetting: Annett Patzschewitz
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Inhalt

Rita Finkbeiner/Barbara Schlücker

Compounds and multi-word expressions in the languages of Europe

1 Introduction

This volume deals with compounds (e.g., *boat house*, *softball*) and multi-word expressions (*piece of cake*, *dry cough*) in European languages.[1] Compounds and multi-word expressions (henceforth MWEs) are similar as they are both lexical units and complex, made up of at least two constituents. The most basic difference between compounds and MWEs seems to be that the former are the product of a morphological operation and the latter result from syntactic processes. This is, admittedly, a very vague distinction. However, as soon as one takes into account more than one specific language (or language family), it seems that this is the closest one may come to a definition that is more or less applicable to the European languages. In fact, in light of Romance examples such as French *glace au chocolat*, Spanish *helado de chocolate* 'chocolate ice cream' which have often been analyzed as compounds although they contain syntactic relational markers, even the morphological criterion for compoundhood seems to be questionable. Further complicating matters, whereas in many languages compounds are regarded as being opposed to MWEs, in other languages, and particularly in English, compounds are often regarded as a kind of MWE. In addition, for languages that are assumed to have an opposition between compounds and MWEs, the question arises of whether compounds and MWEs act in competition or complementation with regard to the formation of new lexical units.

Given this background, the aim of the volume is to present an overview of compounds and MWEs in a sample of European languages. Central questions that are discussed for each language concern the formal distinction between compounds and MWEs (in particular prosodic, morphological, and syntactic properties), the relation between compounding and MWE formation as well as the conclusions concerning the theory of grammar and the lexicon that follow from these observations. Although several comprehensive volumes on compounding and phraseology have appeared in recent (and not so recent) years (cf.

1 We would like to thank Kristel Van Goethem and Carmen Scherer for very valuable comments on an earlier version of this chapter.

https://doi.org/10.1515/9783110632446-001

Scalise (ed.) 1992; Burger et al. (eds.) 2007; Lieber/Štekauer (eds.) 2009a; Gaeta/Grossmann (eds.) 2009; Scalise/Vogel (eds.) 2010; Gaeta/Schlücker (eds.) 2012), the relationship between compounds and MWEs with respect to their status in lexicon and grammar has received comparatively little attention (cf. Hüning/Schlücker 2015 for an overview). For this reason, this relationship constitutes the central focus of this volume.

The aim of the present chapter is to review the language-specific properties, bring them together and compare them against German. German is well-known for its propensity for (nominal) compounding, as compared to, e.g., French. Also, there is a rather clear demarcation line between compounds and MWEs in German, in contrast to English, for instance. Taking German as a reference point may help to shed more light on some of the crucial questions with respect to the compound-MWE relationship in the various European languages such as, for instance, the potential competition between the two processes, or their demarcation line. By way of language comparison, the differences and commonalities between languages – both within language families and across these borders – become clearer, ultimately revealing that a cross-linguistically valid definition of compounds and the demarcation from MWEs may be impossible, given that languages vary greatly in their defining properties and in the number and productivity of compound and MWE subpatterns.

The volume contains chapters on English, German, Dutch, French, Italian, Spanish, Greek, Russian, Polish, Finnish, and Hungarian. Although this sample is neither complete nor representative of "the" languages of Europe, it nevertheless provides thorough analyses of a large set of central European languages. Importantly, it should be noted that the selection here is mostly due to various practical reasons, rather than an assessment of the relevance of languages. In addition to the languages mentioned, the present chapter also comprises an overview of the North Germanic languages.

The structure of this chapter is as follows: Section 2 starts with general considerations about the lexicon and the lexicon-syntax interface and discusses basic notions such as morphological vs. syntactic lexical unit, lexicalization, and the problem of correspondence. Section 3 discusses compounds and MWEs against the background of German, sorted by language families. The chapter ends with a brief conclusion in Section 4.

2 Theoretical considerations

At the outset of our overview, a short remark on the notion of MWE is in order. It is widely known that different research traditions within this field have focused on different types of MWEs, applying an extremely diverse terminology. In the early Anglo-American structuralist tradition (e.g., Weinreich 1969; Newmeyer 1974), the focus was on idioms as semantically and/or syntactically irregular MWEs. Idioms – a notorious example being *kick the bucket* – were mainly discussed under the assumption that they posed a problem to rule-based grammar. Traditional German phraseology, on the other hand, which is influenced by the Soviet tradition, has been investigating idioms in their own right, as a core phenomenon of the linguistic subfield of phraseology (Häusermann 1977; Fleischer 1982; Burger et al. 2007). This tradition has put much effort into issues of classification, studying not only idioms, but also other types of MWEs which need not be idiomatic, for instance collocations such as *starker Raucher* (lit. strong smoker, 'heavy smoker') or routine formulae such as *Kein Problem* ('no problem') (e.g., Burger 1998). However, under the growing influence of theories such as Construction Grammar (Fillmore/Kay/O'Connor 1988; Goldberg 2006; Hoffmann/Trousdale (eds.) 2013), and insights from applied linguistics, such as research in foreign language learning (Pawley/Syder 1983; Wray 2002), and with the advent of new technologies within quantitative linguistics and corpus linguistics (Sinclair 1991; Gries 2008), the notion of MWE has broadened dramatically in the last decades. In particular, it has become increasingly accepted that there is a large inventory of lexically partially fixed patterns in the lexicon such as [N *by* N] (*page by page*, *year by year*, *country by country*, cf. Jackendoff 2008) that may or may not be fully compositional, and that may be used productively to create new instances. Under such a broad view, MWEs are "co-occurrence phenomena at the syntax-lexis interface" (Gries 2008: 8) that may be defined as syntactic patterns consisting of at least two words, the combination of which may be more or less fixed, more or less idiomatic, and more or less productive. Crucially, as idiomaticity is not a defining feature of all of these patterns, their status as stored MWEs hinges on sufficient frequency and on their function as a lexical unit; hence the term 'phrasal lexical unit', which is regularly employed throughout the volume and the remainder of this introduction. To decide whether or not a frequent syntactic pattern is a lexical unit, a well-defined notion of lexical unit, and of the lexicon, is required.

2.1 The notion of the lexicon

It is a widely held assumption that the lexicon is a repository of stored linguistic knowledge, in particular, a repository of words. In fact, it may seem that under the last 50 years of linguistic research, this assumption has hardly been challenged, compared to the lively and ongoing debate about what the most adequate theory of grammar is (cf. Wunderlich 2006: 1). However, it is clear that our theory of the lexicon crucially depends on our theory of grammar. For example, whether the lexicon is viewed as a repository of only words or also of affixes depends on whether morphology is conceptualized as a subcomponent of the lexicon or as part of syntax. Under a mainstream view, linguistic knowledge comprises two components:

> One is a finite list of structural elements that are available to be combined. This list is traditionally called the "lexicon", and its elements are called "lexical items". [...] The other component is a finite set of combinatorial principles, or a grammar. (Jackendoff 2002: 39)

This view entails the idea that lexical items have to be learned, as they are not predictable. By contrast, grammar – which is often equated with syntax – is viewed as the domain of rules, or principles, that enable speakers of a language to productively generate new sentences. For example, it is an idiosyncrasy of English that the word *squirrel* (and not, say, the word *dog*, or the word *hamburger*) refers to the concept SQUIRREL. Speakers of English have to learn this word with its specific phonological, categorial and semantic features. However, they do not have to learn the sentence *The squirrel is eating nuts*, as they can productively generate it by combining the respective words according to the rules of grammar. Therefore, the dichotomy between lexicon and grammar also tends to be conceptualized as a dichotomy between words and phrases, and between idiosyncrasies and rules (Engelberg/Holler/Proost 2011: 1).

However, it has long been recognized that there are a considerable number of phenomena in the languages of the world that pose a serious problem to the view of a strict lexicon/grammar divide. Compounds and MWEs are a pertinent case in point. As to compounds, Jackendoff (2009: 108) points out that on the one hand, speakers must store thousands of lexicalized compounds, e.g., *peanut butter*, but on the other hand, they may build compounds "on the fly", e.g., *bike girl* for a girl who left her bike in the vestibule. Thus, compounds arguably are part of the lexicon, but at the same time, compounding is a productive, and therefore rule-based process. For this reason, it is necessary to distinguish between the properties of being morphological, and of being lexical (Gaeta/Ricca 2009).

As to MWEs such as *kick the bucket*, it is obvious that on the one hand, they are phrasal units, often showing a fully regular syntactic behavior, but on the

other hand, they must be part of the lexicon, as their meaning is non-compositional and has to be learned (Nunberg/Sag/Wasow 1994; Gries 2008). What is more, there is ample evidence by now that not all MWEs are isolated units that have to be learned one by one, but that there must be something like MWEs "on the fly", as well. That is, there seem to be abstract patterns in the lexicon that can be used by speakers to create new MWEs (Fillmore/Kay/O'Connor 1988). For example, speakers might newly coin the potential, but unattested phrasal simile *heavy as a truck* on the basis of the lexicalized pattern [(*as*) A *as* NP], which comprises established examples such as *strong as a horse* or *dead as a doornail* (Finkbeiner 2008).

This raises the more general question of the interrelation between the lexicon and the two "rule-based" components of grammar, morphology and syntax. If both morphology and syntax may feed the lexicon, as is evidenced by compounds and MWEs, how is the interaction of morphology and syntax with the lexicon to be represented in our theory of grammar?

2.2 Lexicon-syntax interface

In early conceptions of Generative Grammar, the lexicon was conceived of as a passive repository of morphemes, which would be concatenated in the transformational component of syntax. Only the later stage of lexicalism, initiated by Chomsky's *Remarks on Nominalization* (1970), led to the recognition of the dual status of the lexicon as both a repository of words and an active component of the grammar (Giegerich 2009). Thus, in a lexicalist theory, morphology is acknowledged as an autonomous component of grammar that is part of the lexicon.[2] However, there is still a sharp dividing line between the lexicon, including morphology, and syntax. This divide is captured by the principle of lexical integrity, which says that syntactic processes can manipulate members of lexical categories, but not their morphological components (Di Sciullo/Williams 1987; Scalise/Guevara 2005). Behind this is the idea that the lexicon (including morphology) is a 'pre-syntactic' component that feeds syntax, but not vice versa. Thus, lexical items, with or without internal morphological structure, are taken from the lexicon and inserted into a syntactic tree. The resulting syntactic structures are later 'spelled out' in phonology as well as in semantics.

2 A weaker form of lexicalism assumes that inflectional morphology is more closely related to syntax, while word-formation is more closely related to the lexicon (e.g., Anderson 1982; cf. also Giegerich 2009).

Under such a conception, one may account for the fact that compounds are part of the lexicon, while at the same time being the output of a productive morphological component. However, the model does not account for the difference between listed compounds and novel ones, as morphosyntactically they look exactly the same. Even more importantly, lexicalism predicts a lexicon free of syntactic phrases. Thus, not only do MWEs such as idioms and collocations pose a serious problem to lexicalism, but also phenomena like phrasal compounds, i.e. compounds with a phrasal modifier constituent (e.g., Pafel 2015; Trips/Kornfilt 2015), and particle verbs, i.e. verbs with a separable particle (e.g., Lüdeling 2001; Zeller 2001).

The linear view of the lexicon/syntax relation is abandoned in Jackendoff's (1997) Parallel Architecture. At the heart of this approach is the hypothesis of representational modularity, which states that grammar is organized into three autonomous and generative components: viz. phonological structure, syntactic structure, and conceptual structure. Each domain generates representations of its own. The interaction between the components is established by separate interface modules between the systems that contain correspondence rules. In this model, a lexical entry is exactly such a (small-scale) correspondence rule. It links a small chunk of phonology with a small chunk of syntax and a small chunk of semantics. Instead of lexical insertion, there is lexical licensing, in that a lexical item licenses its chunks of information as the result of three independent processes. As Jackendoff (2009) puts it:

> A word therefore is to be thought of not as a passive unit to be pushed around in a derivation, but as a part of the interface components. It is a long-term memory linkage of a piece of phonology, a piece of syntax, and a piece of semantics, stipulating that these three pieces can be correlated as part of a well-formed sentence. (ibid.: 107)

The crucial point is that this model allows for including into the lexicon all kinds of units, not only simplex and complex words, but also phrases of different kinds. That is, MWEs can be listed in the lexicon as correspondence rules like every other lexical item. The only difference is that in an MWE such as *kick the bucket*, the three syntactic words are associated with three phonological words, but only with one element in semantics ('to die'). Complex words, such as compounds, are treated as instantiations of more abstract morphosyntactic schemata that contain variables at the three representational levels. Thus, morphology is not a separate component in Jackendoff's model. There is no difference between words and rules, but both are conceived of as declarative schemata that have the status of (more or less abstract, and more or less productive) lexical units.

The Parallel Architecture has much in common with Construction Grammar and Construction Morphology (Booij 2010). In a way, one can say that Construc-

tion Grammar, or at least certain variants of it, are realizations of the Parallel Architecture. At the heart of Construction Grammar is the insight that linguistic knowledge largely consists of stored knowledge of constructional schemata, from morphological schemata via lexical, phrasal, and even discourse schemata. Both the Parallel Architecture and Construction Grammar thus argue for a continuity between lexicon and grammar.[3] In Construction Grammar, this continuum view culminates in the notion of the 'constructicon', which replaces older views of a lexicon/grammar dichotomy. The constructicon is conceived of as a large structured inventory of constructions of all levels of abstraction. Under this approach, compounds and MWEs can easily be treated as on a par with each other, both being complex constructions sharing certain conceptual or functional features.

2.3 Lexicalization

The continuum view of the syntax/lexicon relationship may lay the ground for an integrated and systematic treatment of both compounds and MWEs as the output of productive or semi-productive schemata localized in the lexicon. Still, it does not say anything about the differences in the lexical status between, e.g., the compounds *grass frog* vs. *grass slug*, or the VPs *hit the road* vs. *hit the dog*. While *grass frog* is a lexicalized compound, *grass slug* is not, and while *hit the road* is a lexicalized MWE, *hit the dog* is not. Obviously, some outputs of schemata, or rules, have the status of established lexical items listed in the lexicon, while others have not (Hohenhaus 2005; Bauer 2006; Gaeta/Ricca 2009). In order to account for these differences, one needs a concept of lexicalization.

According to Hohenhaus (2005: 356), the term lexicalization denotes both the process of listing and the state of listedness, that is, the property of some element to be a lexical item of a language. The main rationale behind the joint investigation of compounds and MWEs is precisely their common status as complex lexical items. In order to delimitate the field of investigation, it is therefore crucial to

3 One difference between the Parallel Architecture approach and Construction Grammar lies in the conceptualization of productivity. While Jackendoff (2009, 2013) clearly differentiates between productive and semi-productive phenomena, Construction Grammar is somewhat less explicit in this respect, assuming a flexible continuum of productivity of constructions. Another difference lies in the conceptualization of the contents of constructions. While in a homogeneous approach (e.g., Goldberg 1995, 2006), all linguistic units are taken to be meaningful constructions – there being no autonomous syntactic principles – a heterogeneous approach takes meaningful constructions as only one kind of stored structure, assuming that the grammar can also contain independent principles of syntactic form or semantic structure (Jackendoff 2013: 78f.).

properly define the notion of 'lexical item'. That is, while we want to include *grass frog* and *hit the road* into our field of investigation, we would like to exclude *grass slug* and *hit the dog*. In particular, the following two criteria seem to be crucial in this respect.

Firstly, a lexical item functions as a semantic, or conceptual unit. For example, *grass frog* refers to a unitary concept, a certain species, and *hit the road* refers to a specific kind of activity. Both are concepts that speakers of the language have stored together with the respective items. By contrast, while speakers of English will be able to assign an interpretation to *grass slug*, they do not have stored it as a unit together with a certain conventional concept, or stable referent. Similarly, speakers will be able to interpret the phrase *hit the dog*, but they do so on compositional grounds, and not because they have learned this phrase together with a certain concept.

Secondly, for an element to have the status of a lexical item, it must occur with significant frequency in the language. This criterion has received increasing attention with the growing influence of usage-based approaches and rapidly developing quantitative methods in corpus linguistics. It is closely related to the first criterion, because high frequency makes it more likely that an item is becoming listed with a certain meaning. For example, if during a rainy summer a plague of slugs that eat all the grass in people's gardens were to sweep over a country, and everybody started talking about the nasty grass slugs, it might be that after a while, this compound would get stored in the English lexicon as a label of this specific concept ('certain kind of nasty grass-eating slug').

2.4 Compounds and multi-word expressions in the lexicon

The criterion of lexicalization, i.e. the property of being a (complex) lexical unit, thus allows us – at least, theoretically – to distinguish between those instances of morpho-syntactic schemata that are listed in the lexicon, and those that are not. However, we also need a good criterion to distinguish, within the class of complex lexical units, between compounds, on the one hand, and MWEs, on the other. This criterion, obviously, must be found in their internal structure.

Compounds are the output of morphology, while MWEs are the output of syntax. Accordingly, Gaeta/Ricca (2009: 38) suggest a quadripartite typology which is based on the idea that one has to strictly distinguish between the properties of being morphological, and of being lexical. The property of being morphological implies that an item is the output of some morphological schema or rule, which is different from a syntactic schema or rule. The property of being lexical implies that an item is lexicalized in the above-mentioned sense, i.e., that it refers to a

stable concept and occurs with sufficient frequency in the language. Cross-classifying the two properties results in the following matrix (ibid.):

(a) [+morphological], [+lexical]
(b) [+morphological], [–lexical]
(c) [–morphological], [+lexical]
(d) [–morphological], [–lexical]

Of these four options, (a) represents the prototypical instance of a lexicalized compound, i.e., an item that is the output of a morphological process and that is listed in the lexicon with a stable meaning, e.g., *grass frog*, *play list*, or *milkshake*. Option (b), by contrast, represents an item that is the output of a morphological process, but is not listed, e.g., *bike girl*, *grass slug*, or *Trump problem*.[4] Option (c) is represented by MWEs, that is, phrasal, not morphological items for which it is plausible to assume listedness, either because of semantic idiomaticity or sufficient frequency, or both, e.g., *hit the road*, *heavy smoker*, or *by and large*.[5] Finally, option (d) represents the prototypical syntactic phrase, i.e., a VP such as *hit the dog* that is formed according to a syntactic rule, or schema, and whose meaning is compositional, therefore not requiring separate storage in the lexicon.[6] The quadripartite typology makes it very clear that, contrary to traditional views, morphological units do *not need* to be lexical units, while syntactic units *may* be lexical units.

Against this background, we may now attempt to pin down the defining criteria of compounds, and of MWEs. In this we do not aim for more than a rough approximation, as it is clear that the respective criteria are not only in part language-specific, but also a matter of controversial theoretical debate. Generally, we take it for granted that compounds have the features [+morph], [+lex], whereas MWEs have the features [–morph], [+lex]. Compounds may be defined, following Bauer (2009a), on both phonological, morphological, and syntactic grounds (cf. also Lieber/Štekauer (eds.) 2009a; Giegerich 2015; Bauer 2017). First, compounds

4 These items are also called occasionalisms. They may become listed in the lexicon at a later stage, but not all of them will. Hohenhaus (2005) discusses the question whether there are occasionalisms that are not listable (non-lexicalizable) in principle.

5 Booij (2010: 190) uses the term "lexical phrasal constructions" to refer to these units.

6 While Gaeta/Ricca (2009) focus on the delimitation between compounds and MWEs, i.e., *complex* lexical units, it is clear that the feature combination [–morph], [+lex] also applies to established simplex words, such as *grass*. Likewise, the combination [–morph], [–lex] also applies to inexistent simplex words, such as the nonce verb *to gorp* from a textbook sentence on language acquisition ("The duck is gorping the bunny"), cf. Saxton (2010).

usually behave like single words phonologically. For example, the stress pattern in English compounds is more like the stress pattern in single words than the stress pattern in phrases, e.g., *gréen card* (compound; 'residence permit') vs. *green cárd* (phrase; 'green card', e.g., in a game of cards).

Second, compounds are marked as word-like units morphologically. While the prototypical case is that a compound is made up of two unmarked lexemes, in languages with inflection, the non-head may carry an inflection-like element (e.g., the element *-s* in German *Liebe+s+brief* 'love letter'). Crucially, though, this inflection-like element does not vary as a function of the compound's role in the matrix sentence (Bauer 2009a: 346). What carries the inflection for the compound as a whole, according to its role in the matrix sentence, is the head (ibid.). For example, the linking element *-s* in the German compound *Liebe+s+brief* is carried by the non-head, while inflection according to the compound's role in a matrix sentence goes to the end of the head, e.g., *in den Liebe+s+brief+en* ('in the$_{\text{DAT.PL}}$ love letters$_{\text{DAT.PL}}$').

Third, compounds can be defined according to syntactic criteria, most importantly syntactic inseparability and an inability to modify the non-head. For example, one cannot insert an element in between the two constituents of the German compound *Alt+bau* 'old building', cf. **dieser Alt teure Bau* (lit. this old expensive building), and the non-head (the first constituent) cannot be modified: **dieser sehr Alt+bau* (lit. this very old building).

As for MWEs, scholars like Nunberg/Sag/Wasow (1994) and Gries (2008) make use of syntactic, semantic, and frequency criteria to arrive at a definition. As outlined in the beginning of this chapter, in modern phraseological research, most scholars hold a rather broad view of the notion of MWEs, including many different types of phrasal units. Syntactically, MWEs are required to consist of more than two syntactic elements, which may be of different natures. For example, the collocation *heavy smoker* consists of two words. In other MWEs, a word tends to co-occur with a particular grammatical pattern, for instance, the verb *to hem* tends to co-occur with the passive. In this case, the MWE consists of a word and a syntactic frame (Gries 2008: 5). MWEs often are syntactically more or less fixed, but there are also fully flexible MWEs. For instance, the MWE *by and large* is completely fixed (e.g., the reverse order **large and by* would be ungrammatical), while *run amok* is rather flexible (e.g., it allows for different tenses).

Semantically, it is usually required that MWEs be semantic units, i.e. that they have a meaning just like a single word or morpheme. For example, *hit the road* roughly means 'leave'. While many MWEs tend to have a non-compositional semantics, non-compositionality is not a necessary criterion. For example, while *kick the bucket* is semantically non-compositional, *too much to ask* is fully compo-

sitional. Both can be regarded as semantic units, however. As to the frequency criterion, for something to count as an MWE, it is required that the observed frequency of the joint occurrence of the constituents be larger than the expected frequency of joint occurrence. More generally, the degree of frequency of an MWE can be related to its degree of cognitive fixedness, or "entrenchment". Naturally, the frequency criterion can only be employed on empirical grounds.

2.5 Problem of correspondence

While MWEs such as *kick the bucket* and compounds such as *blackbird* do not seem to have much in common except their being complex lexical units, it has been pointed out repeatedly in the literature that there are certain subsets of compound words and MWEs that closely correspond to each other. For example, in German, as in many other languages, there are adjectival compounds, e.g., *butter+weich* 'butter soft', that have corresponding phrasal similes, e.g., *weich wie Butter* 'as soft as butter'. These expressions share lexical material and have a very similar meaning. Another case in point are A+N combinations such as *schwarzer Tee* vs. *Schwarz+tee* 'black tea' (cf. Schlücker 2014; Hüning/Schlücker 2015). As both the morphological and the syntactic pattern are stored lexical units, they pose a problem to the principle of synonymy blocking in the lexicon, suggesting that this principle might not be as strong as often assumed. For such cases, potential tasks for the researcher are to find out how much the two competing processes overlap, if the overlap is systematic or only applies to a subset of the respective patterns, whether one is dealing with real doublets, or whether there are more specific differences in meaning or usage (cf. Masini, this volume; Schlücker, this volume). For example, Hüning/Schlücker (2015) point out that the morphological and the phrasal pattern in similes such as *butter+weich*/*weich wie Butter* are competitive only with regard to a relatively small subset of all possible similes. This can be shown by pairs such as **brot+dumm*/*dumm wie Brot* (lit. dumb as bread, 'very dumb'), where one of the two patterns is ruled out. Theoretically, the interesting question is what underlying principles guide the choice of strategy that is employed in a given language, or in a given context. For German A+N sequences, for instance, the choice between the morphological and the phrasal pattern seems to be sensitive to type frequency effects (cf. Schlücker/Plag 2011).

While all contributions to this volume discuss the compound-MWE relationship, some of them focus explicitly on corresponding patterns, while others look at the issue from a broader perspective. What can be said more generally for the different languages and language families of Europe is that the potential corre-

spondence between compounds and MWEs cannot be described in a uniform way, since it is multifaceted and manifests itself in very different ways.

An interesting aspect, from a semantic point of view, is the observation that in German, compounds such as *Rot+kraut* 'red cabbage', in contrast to their phrasal counterparts (*rotes Kraut*), seem to be more inclined to adopt a kind reading. Thus, *Rot+kraut* denotes a specific kind of cabbage, and not just a cabbage that is red. Härtl (2016) argues that this semantic specialization of compounds is not, as is often assumed, an effect of lexicalization, but can also be observed with novel compounds such as *Rot+dach* ('specific kind of roof') vs. *rotes Dach* ('red roof'), and is therefore "somehow active 'right from the beginning' in the life of a compound" (Härtl 2016: 66; cf. also Lipka 1977).[7] From a contrastive point of view, an interesting question is whether this presupposition of kind reference is true for compounds in other languages as well. Furthermore, given that Romance languages employ compounding to a far lesser extent than Germanic languages (cf. Section 3.3), one may ask whether similar effects in French are connected to the difference between the ubiquitous, determinerless [N *de* N] pattern and the 'regular' pattern with definite article [N *du*/*de la* N]. Similarly, for Swedish, one might speculate that the systematic difference between the 'regular' pattern with double determination on the one hand (*det röda kors+et* 'the red cross'), and the reduced pattern with single determination, i.e. with suffixed determiner only (*röda kors+et* 'the Red Cross') (cf. Section 3.2) on the other hand, might be functional. If this were the case, then one would expect that a novel combination with double determination such as *den stora mur+en* 'the big wall' would be less inclined to adopt a kind reading or a naming function when compared to the combination with single determination, *stora mur+en* (which should be inclined to denote a specific type of wall, e.g., the prospective wall between the United States and Mexico).

In the next section, we are going to take a more detailed contrastive look into the compound/MWE relationship in the different languages and language families of Europe as compared with German.

7 This is not to say that German phrasal patterns cannot adopt a kind reading, which is clearly not the case (e.g., *schwarzes Brett* 'bulletin board'). The point in Härtl (2016: 66) is that "right from the beginning", a compound is semantically more specialized, or more restricted than its corresponding phrase, which may, but must not adopt a kind reading. Potential counterexamples to this hypothesis are pairs such as *Warmwasser* vs. *warmes Wasser* ('warm water'), or *Blondhaar* vs. *blondes Haar* ('blond hair'), where the compound does not seem to be semantically more restricted than the phrase; cf. Schlücker (2014).

3 A contrastive overview

The second part of this chapter is devoted to the comparison of German with the West Germanic, North Germanic, Romance, Slavic, Greek, and Finno-Ugric language families in terms of the relationship between compounds and MWEs. It strives to illustrate the similarities and differences between these languages and to sketch some more general tendencies of the respective language families with respect to this relationship. The languages discussed in the following overview are restricted to those represented in the various chapters of the volume, except the North Germanic languages, which lack their own chapter and which have been added to this overview to complete the picture. When relevant, compound boundaries are marked by "+" in the following.

3.1 West Germanic languages and German

As German is a West Germanic language, more similarities than differences with other West Germanic languages are to be expected. In fact, German and the other major members, English and Dutch, are characterized by several common properties. First of all, there is no doubt whatsoever about the existence and productivity of the morphological pattern of compounding in these languages. Second, these languages have both nominal and adjectival compounding, with the former being unanimously regarded as the most frequent and productive subpattern and N+N compounding particularly apparent. Verbal compounding, on the other hand, is regarded as either scarce or non-existent. Regarding English, Bauer (this volume) and Bauer (2017: 136–140) provide sporadic examples of verbal compounding such as *dry-burn* or *mock-whisper*. Similarly, there are a few coordinate V+V compounds in German, such as *brenn+härten* 'flame-harden', *press+ polieren* 'press-polish'. They are, however, very rare and mainly belong to technical terminology. In general, it seems clear that most forms that look like verbal compounds on the surface are in fact the result of either back-formation or conversion (e.g., German *frühstücken* 'to have breakfast', < *Früh+stück*, lit. early piece, 'breakfast'). Then again there are also separable complex verbs, such as particle verbs (e.g., English *drink up*) and quasi-noun incorporation (e.g., Dutch *piano spelen* 'play the piano' (cf. Booij, this volume)) whose morphological/compound status is highly problematic given the fact that they are separable. Thirdly, English, Dutch and German all have MWEs, both those that in principle correspond to compounds and those that do not, such as proverbs or routine formulas. MWEs corresponding to compounds are those that share the basic naming function of compounds and possibly also share lexical material. For instance, there are various

kinds of nominal phrasal constructions with a naming function and which are therefore on a par with nominal compounds. Thus, they are lexical noun phrases, sometimes also termed 'phrasal nouns'. Patterns of lexical noun phrases are easily found in all these languages, e.g., close apposition (German *Prinzip Hoffnung* 'principle of hope'), genitive (or possessive) constructions (English *baby's chair*, German *Ei des Kolumbus* 'egg of columbus'), constructions with prepositional phrases (Dutch *restaurant met tuin* 'garden café'), binomials (English *fish and chips*), or A+N phrases, often with a relational adjective (Dutch *stalen zenuwen* 'nerves of steel').[8]

However, in addition to these similarities, there are also differences within West Germanic. In particular, there is one fundamental contrast that distinguishes English from the other two. Overall, in German and Dutch, compounds can be very clearly distinguished from phrasal constructions on the basis of formal criteria, primarily stress and inflection. This distinction is reflected in spelling, with compounds displaying solid spelling and MWEs being written in two (or more) orthographic words. There are only very few patterns that resist a clear classification as either morphological or phrasal, at least at first view, such as phrasal (particle) verbs. In fact, German and Dutch seem to pattern very much alike with regard to the (number of) types of compound and MWEs patterns that exist in both languages.

Leaving aside various minor differences and specific characteristics of each language, the major difference between German and Dutch seems to lie in the often noted observation that – at least in the nominal domain – Dutch seems to use phrasal patterns more often than German, which in contrast opts for compounding more frequently, although both patterns are in principle available in both languages, e.g., German *Tag+es+gespräch*, Dutch *gesprek van de dag* (lit. talk of the day, 'nine days' wonder'), German *Stumm+film*, Dutch *stomme film* ('silent film') (cf. van Haeringen 1956; De Caluwe 1990; Booij 2002; Hüning 2010; Hüning/Schlücker 2010, among others).

In English, on the other hand, the formal distinction between (nominal) compounds and phrases is notoriously difficult. First of all, the criterion of inflection is inapplicable in English. Secondly, the (formerly often invoked) criterion of stress has been shown in a number of works (cf., for instance, Plag 2006; Kunter 2011) to be incapable for drawing this distinction because although the vast majority of (NN) compounds have forestress, as predicted, there are also numerous exceptions, as can be seen from classical examples such as *'apple*

8 *Stalen* is a relational adjective derived from the noun *staal* 'steel'.

cake vs. *apple 'pie*, *'Madison Street* vs. *Madison 'Avenue*. Thirdly, the distinctive force of other tests which refer to the idea that compounds, being words, should be subject to lexical integrity (contrary to phrases), such as the pro-*one* test, internal modification, or coordination, have been proven weak in works such as Bauer (1998), Giegerich (2015) and Bauer (to appear). Also, the forms that evolve as either morphological or phrasal on basis of the stress criterion do not necessarily coincide with the outcomes of the other tests. For this reason, very divergent opinions on the definition of compounds in English and the demarcation from phrases can be found in the literature. A literature survey is beyond the scope of the present paper (but see, for instance, Olsen 2000; Lieber/Štekauer 2009b). Generally speaking, in addition to uniform analyses that assume that the constructions in question are either all morphological, and thus compounds, or all syntactic, and thus phrases, it has also been suggested that some of them are morphological whereas others are syntactic, depending on how the above-mentioned criteria are weighted (e.g., Giegerich 2004). Finally, it has been advocated that the inconclusive data are an indication of the fact that the compound-phrase distinction does not exist and that there is either a continuum or an overlap between syntax and the lexicon (e.g., Giegerich 2015; Bauer, this volume). Another problematic case is the 'descriptive' or 'classifying' genitives, e.g., *lawyer's fee*, *mother's milk*. Regardless of their obvious phrasal form, they are alike compounds in that the genitive dependent has a classifying rather than a determinative function, that it is immediately adjacent to the head noun, and that the constituents cannot be separated, e.g., by another modifier. For this reason, they have often been treated as compounds in the literature (cf. Rosenbach 2006: 82–89 for a literature survey).

In sum, the major difference between German (and Dutch) on the one hand and English on the other is that in English, due to the apparent impossibility of distinguishing clearly between morphological and syntactic N+N and A+N sequences, compounds are often regarded as just one kind of MWE, cf., for instance, Ramisch (2015), Bauer (this volume),[9] whereas in German and Dutch, compounds and MWEs are clearly opposed and there are only few patterns that elude immediate classifications as either compound or MWE. Apart from that, the West Germanic languages pattern very much alike regarding the existence of various specific subtypes of compounds and MWEs. This similarity becomes particularly obvious when German is compared to other languages and language families.

9 Moon (2015), on the other hand, explicitly excludes compounds from the set of MWEs.

3.2 North Germanic languages and German

The North Germanic languages comprise the continental Scandinavian languages Swedish, Danish, and Norwegian, as well as the insular Scandinavian languages Icelandic and Faroese.[10] As there is no separate chapter on complex lexical units in a North Germanic language in this volume, a short general description of the language family is in order. Generally, North Germanic languages are very similar to West Germanic languages in many respects. Distinctive features common to all North Germanic languages that are lacking in West Germanic languages include the suffixed definite article;[11] the agreement of the adjective in gender and number not only in attributive, but also in predicative position;[12] and the existence of a synthetic passive (termed *s*-passive or medio-passive,[13] cf. Torp 2002). As to word order, North Germanic languages share with Dutch and German V2 in declarative sentences, where English dominantly has SV.[14] On the other hand, North Germanic shares with English the predominant VO-pattern, where Dutch and German have OV.[15] Within the North Germanic languages, the insular Scandinavian languages differ from the continental Scandinavian languages most notably in their rich inflectional morphology. While Swedish, Norwegian and Danish have a rather reduced inflectional morphology, Icelandic has, of all modern Germanic languages, the most differentiated inflection in the nominal, adjectival and verbal domain (Braunmüller 2007: 248), comparable to that of Ancient Greek or Latin, but with additional combinatorial phonological changes.

Compounding is a highly productive morphological process in all North Germanic languages, as in German. Generally, compounds in North Germanic languages, as in German, are right-headed, with inflectional endings attaching to the word-final element. Also, compounding in North Germanic languages is recur-

10 In this overview, we will concentrate on examples from Swedish, Danish, and Icelandic.

11 E.g., Swedish *bil+en* (lit. car+the, 'the car').

12 E.g., Swedish *en stor bil* (common gender, 'a big car'), *ett stort hus* (neuter, 'a big house'); *bilen är stor* (common gender, 'The car is big'), *huset är stort* (neuter, 'The house is big').

13 E.g., Swedish *dörren öppnade-s* 'the door was opened', with the *-s*-suffix marking passive.

14 Cf. Swedish *Där kommer hon*, German *Da kommt sie* (Adv V S) (both lit. there comes she), but English *There she comes* (Adv S V).

15 This is reflected not only in subordinate clauses, but also in main clauses if one takes into account the position of infinite verbal parts. Cf. for main clauses Swedish *Hon har sett huset*, English *She has seen the house* (V_{fin} V_{infin} O) (both 'She has seen the house'), but German *Sie hat das Haus gesehen* (V_{fin} O V_{infin}) (lit. she has the house seen); for subordinate clauses Swedish [*Jag vet att*] *hon har sett huset*, English [*I know that*] *she has seen the house* (V_{fin} V_{infin} O) (both 'I know that she has seen the house'), but German [*Ich weiß, dass*] *sie das Haus gesehen hat* (O V_{infin} V_{fin}) (lit. I know that she the house seen has).

sive (e.g., Danish *Kilde+skatte+direktorat+et* 'internal revenue service' (lit. source tax directorate), cf. Haberland 1994, Icelandic *Norð+austur+atlant+s+haf+s+fisk+veiði+nefndin* 'The North East Atlantic Ocean Fisheries (lit. Fish-Catching) Committee', cf. Bjarnadóttir 2017).[16] North Germanic compounds normally display solid spelling and carry stress on the first constituent. Nominal compounding (N+N, A+N, V+N) is by far the most common process, with N+N being the most productive pattern, approximately as in German (cf. Thráinsson 1994; Teleman 2005; Bauer 2009b). Some examples are Swedish *ång+båt*, Danish *damp+skib*, Icelandic *gufu+bátur* 'steam boat' (N+N); Swedish *lill+finger*, Danish *lille+finger*, Icelandic *litli+fingur* 'little finger', 'pinkie' (A+N); Swedish *skriv+bord*, Danish *skrive+bord*, Icelandic *skrif+ borð*, lit. write table, 'desk' (V+N).

One difference concerning V+N compounding in the three languages is that V+N compounds in Swedish and Icelandic use the verbal stem as first constituent (*skriv-*, *skrif-*), while Danish V+N uses the infinitive of the verb ([*at*] *skrive*). This feature of Danish V+N compounds is distinct from German, which is also interesting from a theoretical point of view. If one takes infinitival endings as inflectional endings, the question arises whether Danish V+N compounds should be regarded as cases of compound-internal inflection. However, it is clear that infinitival nonheads are to be distinguished from cases where a non-head exhibits agreement features with the head. Only the latter case may pose a serious problem to the delimitation between compounds and syntactic phrases, since in cases with compound-internal agreement there is a potential overlap between compound and syntactic phrase.

A highly particular feature of Icelandic compounds, in contrast with all other Germanic languages, is that they systematically exhibit compound-internal inflection (cf. Bjarnadóttir 2017).[17] This pertains both to a subclass of Icelandic N+N compounds, i.e. those with a genitive (or sometimes also a dative) non-head, as well as to all A+N compounds. As to N+N compounds, Bjarnadóttir (ibid.: 18) distinguishes between compounds with a stem as non-head (e.g., *fjár+hús* 'sheep house'); compounds with a genitive as non-head (e.g., *vegar+endi*, 'end of road', with *vegar* being one of two possible genitive forms of the noun *vegur* 'way'); and a very small class of compounds with a special stem form or a linking element (cf.

16 Note that compounds in North Germanic languages do not exhibit regular capitalization (in contrast to German). The examples *Kildeskattedirektoratet* and *Norðausturatlantshafsfiskveiðinefndin* exhibit upper case because they function as proper names.

17 Note that internal inflection, more generally, pertains to all nouns with a suffixed definite article in Icelandic. Thus, in definite nouns, the noun and suffixed definite article *both* inflect, e.g., *hestur* 'horse', *hestur-inn* 'the$_{NOM}$ horse$_{NOM}$', *hesti-num* 'the$_{DAT}$ horse$_{DAT}$'.

also Thráinsson 1994). While she acknowledges that the nature of the genitives in Icelandic compounds and the question of whether these are true inflectional forms or linking elements are matters of debate, Bjarnadóttir (2017: 19) argues for a genitive/inflectional analysis of these forms. One of her arguments in favor of this analysis is that the inflected forms of the non-head nouns are always the "correct" genitives, in spite of the complexity of the inflectional patterns. This stands in contrast with German, where forms such as *Liebe+s+brief* ('love letter') are paradigmatically incorrect, the expected genitive feminine being *Liebe*, not **Liebes*. Internal inflection is also found in the adjectival non-heads of A+N compounds, where agreement of gender, case, and number "is exactly the same within the compounds as in syntax" (ibid.: 28f.). For example, in *litli+fingur* 'pinkie' (lit. small finger), the ending *-i* in *litli* 'small' is a marker for masculine, singular, nominative, definite. In the accusative case, the compound form would be *litla+fingur*, with the ending *-a* in *litla* marking masculine, singular, accusative, definite. Thus, on purely inflectional grounds, it is not possible in Icelandic to differentiate between a definite noun phrase *litli fingurinn* 'the small finger' and a compound word in definite form, *litli+fingurinn* 'the pinkie'. This distinction can be made only with the help of word stress (and spelling, though this is not a very robust criterion), with compound words carrying primary stress on the first constituent, and secondary stress on the second constituent in a binary compound.

In Swedish and Danish, on the other hand, compounds can be distinguished from phrasal constructions based on prosodic, morphological, and syntactic criteria. Swedish and Danish compounds, as in Icelandic, carry primary stress on the first constituent and secondary stress on the last constituent (cf. Teleman 2005; Bauer 2009b). According to the Swedish tonal system, which differentiates between accent 1 ("acute") and accent 2 ("grave"), compounds carry accent 2, which is characteristic for polysyllabic words with primary stress on the first syllable (^{2}*sport*+$_{|}$*bil* 'sports car', ^{2}*läs*+*glas*+$_{|}$*ögon*, lit. read glass eyes, 'reading glasses').[18] The difference between accent 1 and accent 2 is distinctive in pairs such as ^{2}*ande+n* ('spirit+definite') and ^{1}*and+en* ('duck+plural'). For these pairs, accent 2 is a lexical accent differentiating lexical words from inflected word forms. This is specific for Swedish and contrasts with German. In German, there is a difference between lexical stress and phrasal stress, but not between lexical stress in words vs. word forms.

Moreover, in Swedish and Danish, compounds may be distinguished from phrases on formal grounds. Generally, in contrast to Icelandic, Swedish and Dan-

18 The exponent $^{1/2}$ replaces the primary stress sign.

ish compounds do not exhibit internal inflection. While in A+N phrases, the adjective must carry inflection (e.g., Danish *et stort køb* 'a big purchase', *det store køb* 'the big purchase'), in A+N compounds, it is uninflected (e.g., Danish *et stor+køb*, *stor+køb+et* 'a wholesale'). Syntactically, while the adjective in an A+N phrase may be modified (e.g., *et meget stort køb* 'a very big purchase', *det største køb* 'the biggest purchase'), in an A+N compound, it may not (e.g., **meget stor+køb*, lit. very big purchase, **størst+køb*, lit. biggest purchase). Further evidence for the compoundhood of A+N compounds comes from definiteness inflection on the noun. While a single noun in Swedish and Danish takes a postposed definite article (*hus+et* 'the house'), a premodified noun takes a preposed definite article (cf. Swedish *det stora hus+et*, Danish *det store hus* 'the big house'). Thus, the correct definite form of the Danish compound *hvid+vin* 'white wine' is *hvid+vin+en*, but not **den hvid+vin*, as would be expected of a phrase (Bauer 2009b).

In Swedish and Danish N+N compounds, the non-head may be changed morphologically in various ways. However, these forms are normally regarded not as inflection, but rather as linking elements, as in German (Niemi, S. 2009; Bauer 2009b). Swedish compounds may display vowel deletion (*flicka* > *flick+skola* 'girl school'), vowel addition (*tjänst* > *tjänst+e+man* 'service man', 'clerk'), or the addition of *-s* (*stol* > *stol+s+ben* 'chair leg') (cf. Josefsson 1997; Teleman 2005). Danish compounds may display an *s*-link (*træning+s+bane* 'training ground'), an *e*-link (*jul+e+dag* 'christmas day'), an *er*-link (*blomst+er+bed* 'flower bed') or an (*e*)*n*-link (*rose+n+gaard* 'rose garden'). In general, this picture is consistent with West Germanic languages such as German and Dutch (but not English, which lacks linking elements).

Apart from compounds on the one hand and regular syntactic phrases on the other, a large stock of MWEs can be found in North Germanic languages, both those that in principle correspond to compounds and those that do not. For example, in Swedish, there are A+N phrases with a naming function such as *röda hund* 'measles' and *hög hatt* 'top hat'; collocations such as *ymnig grönska* 'lush greenery' and *duka bordet* 'lay the table'; complex verbs incorporating a non-referential noun such as *knipa käft* (lit. shut mouth) 'keep one's trap shut' and *vålla storm*, lit. cause storm, 'to cause a great stir'; idioms such as *tala i skägget* 'to express oneself in an obscure way'; and speech act formulae such as *Tack för senast* 'thanks for the other day'. As Koptjevskaja-Tamm (2009: 134) observes, lexicalized A+N phrases in Swedish may contain both indefinite (*hög hatt* 'top hat') and definite adjectives (*röda hund*, lit. red dog, 'measles'), with definite adjectives combining with either unmarked nouns (*röda hund* 'measles') or nouns with the suffixed definite article (*röda korset* 'the Red Cross'). However, what is avoided, according to Koptjevskaja-Tamm (2009), are lexicalizations of the nor-

mal pattern with preposed determiners, definite adjectives and nouns with suffixed article (as in *den gula hatten* 'the yellow hat') (cf. Section 2.5). A specific feature of Swedish MWEs is their connective prosody (cf. Anward/Linell 1976), whereby all stressed syllables in the MWE become deaccentuated, except for the last one. This can be taken as a distinctive feature for telling apart phrasal lexical units from phrasal syntactic units. In this respect, Swedish clearly differs from German, which does not distinguish lexical phrases from non-lexicalized phrases on prosodic grounds.

Generally speaking, the North Germanic MWE systems are very similar to the MWE system of German. Thus, there are overall commonalities both as to the number and the types of MWEs, including rather specific idioms such as German *auf keinen grünen Zweig kommen*, which directly corresponds to Swedish *ej komma på grön kvist* (lit. to not come onto a green branch, 'to get nowhere'). However, there are also many language-specific differences in lexicalization which can be easily demonstrated, e.g., for the case of collocations. For example, in Swedish, there are several collocations with the verb *torka* 'to dry (sth.)', e.g., *torka bordet* ('wipe the table'), *torka golvet* ('wipe/clean the floor'), *torka disken* ('dry the dishes'). While German has a direct verbal equivalent, *trocknen* 'to dry (sth.)', it uses three different verbs in combination with the respective nouns: *den Tisch abwischen*/**trocknen* 'wipe the table', *den Boden wischen*/**trocknen* 'wipe/clean the floor', *das Geschirr abtrocknen*/**trocknen* 'dry the dishes'.

An interesting question is whether there are any tendencies in the North Germanic languages as to the use of compounds compared to their corresponding MWEs. It is well-known that Dutch, relative to German, tends to prefer MWEs over compounds, while German, relative to Dutch, tends to prefer compounds over MWEs (cf. Section 3.1). As to North Germanic languages, as far as we can see, comprehensive studies on this issue are lacking. There is some evidence, though, that Swedish tends to use compounds more frequently than corresponding MWEs compared to other languages. For example, Dura/Gawronska (2007), in a parallel corpus study on novel expressions, found that legislative concepts such as 'quality control' were realized in the Swedish corpus as compound nouns (*kvalitet+s+kontroll*), whereas the Polish parallel corpora used nominal phrases (*kontrola jakośki*). Combinations with 'animal food' were realized as compound nouns (*djur+foder* 'animal food', *fisk+foder* 'fish food') in the Swedish corpus, but as lexical noun phrases containing prepositional phrases (*karma dla zwierzat* 'animal food', *karma dla ryb* 'fish food') in the Polish corpus. Inghult (1991), in an investigation of the principles of lexical innovations in German and Swedish, found that only 3% of all new formations in dictionaries of neologisms were phrases, while 97% were word formations. Moreover, he found that Swedish often has compounds where German has MWEs, for instance, German *kupferne*

Kanne vs. Swedish *koppar+kanna*, 'copper pot'. However, these somewhat outdated results from dictionaries should be treated with caution and are in need of confirmation by corpus-driven studies.

Comparing German and Danish, Farø (2015) finds that Danish tends to have MWEs where German has compounds, e.g., Danish *røget laks* vs. German *Räucher+lachs* ('smoked salmon'), Danish *stor begivenhed* vs. German *Groß+ereignis* ('major event'). However, there are also reverse pairs such as Danish *spanskrør* vs. German *Spanisches Rohr* ('cane'). For a comparison of Dutch and Danish, Haberland (1994: 347) remarks that where Dutch would use derivational processes, Danish would use compounds, cf. Danish *vel+smagende* 'well tasting', 'tasty' vs. Dutch *smakelijk*. While more comprehensive studies on this issue are lacking, these observations suggest, overall, that the North Germanic languages tend to pattern with German with respect to the utilization of the two competing processes.

An interesting commonality between the North Germanic languages, German, and Dutch, which clearly sets them apart from English, is put forward by Klinge (2006). Klinge investigates the [N *de* N] construction, which is well-known from French (e.g., *prisonnier de guerre*). Interestingly, this is also a productive pattern of formation in English (e.g., *prisoner of war*), yet not or only marginally in other West Germanic languages (German, Dutch) or indeed in North Germanic languages such as Danish or Icelandic. Thus, where English has *bird of prey*, German has *Raub+vogel*, Danish *rov+fugl*, and Icelandic *rán+fugl*. The hypothesis put forward by Klinge is that this may be explained as a language contact phenomenon. Thus, the originally Romance [N *de* N] pattern was adopted in English from Norman French. This would explain why it does exist in English, but not in Dutch, German, Danish, or Icelandic. Importantly, Klinge argues that MWEs such as *weapons of mass destruction* in English are not the result of some isolated lexicalization of a syntactic phrase, but instead reflect the presence of a lexical formation pattern [N *de* N] in English which instantiates such structures directly as lexical units.

In sum, one can say that the North Germanic languages largely pattern with German with respect to the availability and utilization of the processes of compounding and MWE formation. The most significant differences between North Germanic and German are to be found in the Icelandic possibility of compound-internal inflection, which makes Icelandic compounds look more "syntactic" than German compounds. However, in many other respects, the commonalities outweigh the differences.

3.3 Romance languages and German

In Romance, morphological compounding is much more restricted than in German. Verbal compounding does not exist (or is very marginal) just as in German, but the number and productivity of nominal and adjectival compound patterns is lower than in German. In general, the notion of compound has often been used also to include phrases, and thus MWEs, for instance nominal constructions containing a preposition or an inflected adjective, e.g., French *moulin à vent* 'wind mill', Italian *macchina da scrivere* (lit. machine to write) 'typewriter', Spanish *casa de campo* 'country house', French *guerre froid-e* 'cold war', Spanish *mal-a suerte* 'bad luck'. Obviously, the key reason for classifying such forms as compounds is their semantic-functional property of serving as a conventional naming entity for a unitary concept. It seems safe to say that in comparison to German such "syntagmatic/syntactic/improper compounds" (as they are often termed in the literature) are much more frequent in Romance. This can also be illustrated by the fact that the German counterparts of all of the above-mentioned examples are compounds, except for the last two, which are a lexical phrase (*kalter Krieg*) and a simplex word (*Pech*). Just as in German, these MWEs either have a fully regular syntactic structure (e.g., French *homme de la rue*, lit. man from the street, 'average person') or are syntactically deficient, for example in that the determiner is missing, e.g., French *château d'eau* (lit. palace of water, 'water tower') (e.g., Gunkel/Zifonun 2011; Gunkel et al. 2017: 1625).

Turning to "proper", morphological compounds, it is striking that for each of the languages under discussion there is no general agreement in the literature as to precisely which constructions should be classified as such. Obviously, the main reason for this is the difficulty in providing generally valid properties of morphological compounding. This problem is illustrated by the definition given in Fradin (2009: 417): "Compounds may not be built by syntax (they are morphological constructs)." Thus, compounds are defined only negatively as non-syntactic, yet this leaves open the exact nature of morphological constructs. The problem is that many of the criteria that can be positively established for compounds in other languages, and in particular in German, are not available in Romance. The first one is the absence of a unitary compound stress rule in Romance (Rainer/Varela 1992; Arnaud 2015; Fernández-Domínguez, this volume). Thus, compounds and MWEs are basically stressed in the same way, contrary to German where compounds can clearly be distinguished from phrases on the basis of stress (modifier vs. head stress). (Native) linking elements, another common property of German (N+N and V+N) compounds, do not exist in French and Italian. However, the native linking vowel *-i-* is found regularly in some adjectival and nominal compound patterns of Spanish, e.g., *roj+i+blanco* (lit. red white,

'red and white').[19] Regarding headedness, French and Italian compounds are generally left-headed, e.g., French *stylo-bille* (lit. pen ball, 'ball pen'), Italian *pesce-spada* (lit. fish sword, 'sword fish'). However, Spanish, in addition to left-headed compounds, e.g., *célula madre* (lit. cell mother, 'stem cell'), also has some right-headed compound patterns (cf., e.g., Guevara 2012; Rainer 2016), both adjectival and nominal ones (cf. Fernández-Domínguez, this volume), e.g., *drog+adicto* (lit. drug addict, 'addicted to drugs'). For Italian, on the other hand, Masini/Scalise (2012) argue that the existence of right-headed compounds does not provide evidence against the assumption that Italian compounding is generally left-headed because these cases are either neoclassical formations, Latin relics, or English calques, such as *scuolabus* 'school bus'. Another frequently mentioned property of compounds, which is again particularly valid for compounding in German (although not for all compound subpatterns) is recursivity. In general, compounding is not considered to be recursive in the Romance languages under discussion (cf., for instance, Scalise 1992 on Italian), with the exception of coordinate (or: copulative) compounds (e.g., Arnaud 2015 on French). Also, solid spelling – which is often said to be indicative of the compound (versus phrase) status in German – is often found with morphological compounds, as well as hyphenated spelling.[20] At the same time, however, there are also compounds with an unstable spelling (cf. Fernández-Domínguez, this volume) as well as MWEs written as one word (e.g., Van Goethem 2009; Van Goethem/Amiot, this volume).

So far, this brief overview has shown that in contrast to German it seems much more difficult to provide clear criteria for morphological compounds as opposed to MWEs in French, Spanish, and Italian. However, two important criteria are still missing. They are among those that have been established by Lieber/Štekauer (2009b: 8) as more general, cross-linguistic criteria of compounding, namely (in addition to stress) (a) syntactic impenetrability, inseparability, and unalterability, and (b) inflection. The first criterion is difficult to assess. On the one hand, it is a basic criterion for distinguishing compounds from phrases (cf., for instance, Fernández-Domínguez, this volume, Van Goethem/Amiot, this volume). On the other hand, however, it is well-known that it also applies to some, though not all kinds of lexicalized phrases (cf., for instance, Gunkel/Zifonun

19 In addition, Latinate and Greek linking elements are found in neoclassical compounding of all three languages, cf., for instance, Villoing (2012) on French.

20 It goes without saying that spelling is subject to conventional norms and possible changes of normative rules and, for these reasons, cannot be regarded as evidence for the grammatical status of forms. However, in particular non-normative writing tendencies might be indicative of the writer's assessment of a form as a conceptual unit.

2011; Arnaud 2015). Thus, syntactic impenetrability, inseparability, and unalterability can be regarded as a necessary criterion of morphological compounds but not as a sufficient one that distinguishes compounds from MWEs. The second criterion, inflection, very clearly distinguishes compounds from phrases in German, as compounds contain only stems and not inflected constituents. This criterion is highly problematic for the Romance languages which has, among other things, to do with the fact that Romance compounds are generally left-headed. Thus, if in a nominal compound the plural is marked on the head, this results in word-internal inflection, e.g., French *poisson*$_{\text{SING}}$-*scie*$_{\text{SING}}$ – *poissons*$_{\text{PL}}$-*scies*$_{\text{PL}}$ (lit. fish$_{\text{SING/PL}}$ saw$_{\text{SING/PL}}$, 'sawfish'). It seems that the three languages have both constructions with word-internal inflection and those without. Thus, plural, for instance, is sometimes marked only on one (usually the left) constituent and on both constituents in other cases. In particular, coordinate compounds seem to regularly inflect for plural on both constituents (e.g., French *auteurs*$_{\text{PL}}$-*compositeurs*$_{\text{PL}}$ 'songwriters-composers'). The question then is which conclusions can be drawn from these observations. In other words: how much value is attached to this criterion regarding the definition of compound? As expected, different positions can be found in the literature: Whereas some scholars quite naturally accept inflected forms as word-internal building-forms of compounds (e.g., Scalise 1992; Guevara 2012; Masini/Scalise 2012; Arnaud 2015), others are more restrictive (e.g., Villoing 2012, on French).

In sum, it is obvious that although in all three languages at hand there are constructions that are clearly morphological (and thus compounds) and others that are clearly syntactic (and thus MWEs) it is very difficult to draw a clear border between them. In this connection, proposals have been made in the context of constructionist frameworks which do away with the idea of a clear-cut borderline between syntax and the lexicon (cf. Masini 2009; Van Goethem/Amiot, this volume; Masini, this volume). If we compare German and the Romance languages with regard to compounding and MWEs, three differences can be noted: firstly, in contrast to the Romance languages, German does allow (with very few exceptions, cf. Schlücker, this volume) a clear-cut distinction between compounds and MWEs. Secondly, although empirical evidence cannot be provided here, it seems that both the number of clearly morphological compound patterns as well as the specific forms instantiated from these patterns are much rarer in Romance languages than in German, and for this reason, MWEs prevail in Romance. Thirdly, if we compare the morphological compound patterns of the Romance languages and German, two Romance patterns stand out from a German (or Germanic) perspective. The first one is V+N compounding, a productive pattern of exocentric compounding, consisting of a verb and a noun which functions as the direct object of that verb, e.g., French *abat-jour* (lit. weaken light, 'lampshade'),

Italian *porta+bagagli* (lit. carry luggage, 'trunk'), Spanish *cubre+cama* (lit. cover bed, 'bedspread'). The pattern is regarded as typical for the Romance languages and it does rarely exist in other Indo-European languages, and not at all in German nor in Dutch (there are sporadic English examples such as *turncoat*, *killjoy*). Although there have been debates concerning the stem form (and thus the morphological nature) of the left, verbal constituent, these constructions are relatively uniformly regarded as morphological compounds in contemporary works (see the literature cited in this section as well as Ricca 2015, who also discusses the interlinguistic differences of V+N compounds within Romance). The second pattern are coordinate compounds (A+A, N+N) which can be said to be fairly regular and productive in French, Italian, and Spanish (though with some restrictions regarding specific subpatterns in the individual languages). In comparison to German, they are interesting for two reasons: first, with regard to form, coordinate compounds often show inflectional marking on both heads and thus word-internal marking, which is impossible in German. One could therefore argue that the pattern is more morphological in German than in Romance. Second, the existence of N+N coordinate compounds has been widely discussed in the literature on German (in contrast to A+A coordinate compounds, whose existence has not been questioned). The main argument is that in many cases of alleged N+N coordinate compounds it seems hard to establish a semantic coordinate relationship and thus two semantic heads; instead, a determinative interpretation is available in equal measure or even preferred. There are only very few clearly nominal coordinate compounds in German (with an additive meaning) such as toponyms, for instance the names of federal states that consist of two regions, e.g., *Nordrhein-Westfalen* 'North Rhine-Westphalia', or technical terms such as *Sprecherschreiber* 'speaker-writer' which is however restricted to linguistic terminology. Thus, although in general German seems to be much more prone to morphological compounding than the Romance languages which in contrast make much more use of MWEs, there are at least these two patterns of compounding that constitute an exception from this general distribution of use of forms.

3.4 Modern Greek and German

Regarding compounds and MWEs, Modern Greek and German display many similarities. Compounding in Greek is, just as in German, a very productive device of word-formation, and both languages have various MWE patterns. As in German, compounds can be distinguished clearly from syntactic phrases (both MWEs and common ones) in Greek.

Starting with compounding proper, it is remarkable that virtually all properties that have been identified for German compounds can also be found in Greek (for comprehensive descriptions cf. Ralli 1992, 2009, 2013a, 2013b, 2016). The vast majority of Greek compounds is endocentric and right-headed, e.g., *domat+o+saláta* 'tomato salad'. Also, Greek compounds have lexical stress. More precisely, they are single-stressed and therefore form one phonological word, contrary to phrases (e.g., compound stress on the antepenultimate syllable in *kapnoxórafo* 'tobacco field' < *kapn*(*ós*) 'tobacco' *xoráf*(*i*) 'field'). In contrast to German compounds, however, which (in simple compounds) always have stress on the first constituent, compound stress in Greek compounds is more variable, depending largely on the phonological properties of the second constituent. Thus, there are several single-stressed compound patterns (e.g., Ralli 2013b: 186f.). Greek compounds consist of either stem or word constituents (most frequently, the left constituent is a stem with the right one either a stem or a word). In any case, they clearly do not have word-internal inflection. Another important point relates to linking elements. There is only one linking element, *-o-*, e.g., *kapn+o+xórafo* 'tobacco field', which is almost compulsory in Greek compounds (there are only a few phonologically conditioned exceptions). For this reason, Ralli (2008) treats linking elements in Greek and in general as compound markers. The occurrence of Greek *-o-* is much more systematic compared to linking elements in German compounds, which are restricted to particular compound subpatterns and display a broad variety of forms, including the zero form. Finally, Greek compounds display solid spelling, contrary to phrases, just as in German.

As to the differences, it seems that recursiveness – which is usually considered a typical property of German N+N compounds – is possible in Greek, too (cf. Ralli 2009), but much rarer (cf. Koliopoulou, this volume). More importantly, while German does not have verbal compounding, it is a productive pattern in Greek, with either verbs, nouns or adverbs as left constituents, e.g., N+V: *xaropalévo* 'fight (with) death' (*xár*(*os*) 'death', *palévo* 'fight'), Adv+V: *kakopernó* 'live badly' (*kak*(*á*) 'badly', *pernó* 'pass, live') (e.g., Ralli 1992). Meanwhile phrasal compounds, that is, compounds with a phrasal modifier constituent, do not exist in Greek, in contrast to German.

Greek compounds can clearly be distinguished from phrases on the basis of stress, the linking element *-o-* and the absence of inflectional markers. Furthermore, morphological compounds are subject to lexical integrity and thus the usual tests (as known from the literature on English and other languages) can be applied: inseparability and the inability to modify the non-head constituent and refer pronominally to the individual constituents (a comprehensive overview of the diagnostics for the compound-phrase distinction is given in Bağrıaçık/Ralli 2015, for instance).

As for MWEs, two particularly interesting constructions that relate to the present issue have been discussed in detail in the literature on Greek, namely [A N] and [N N_{GEN}] sequences. Classical examples are *psixrós pólemos* 'Cold War' ([A N]) and *zóni asfalías* (lit. belt safety, 'safety belt'), i.e. an [N N_{GEN}] sequence with the second, non-head constituent assigned genitive case. These [A N] and [N N_{GEN}] sequences are lexical units with a stable conventional meaning, many of them being scientific terms. They are phrasal, and thus syntactic entities, sharing some features with (morphological) compounds and are inaccessible for the syntactic operations that phrases normally allow. Thus, they are hybrid constructions and have, for this reason, been termed phrasal compounds (not to be confused with compounds containing a phrasal modifier constituent), syntactic compounds or loose multi-word compounds (cf., e.g., Ralli 1992; Ralli/Stavrou 1998; Bağrıaçık/Ralli 2015; Ralli 2016; Koliopoulou, this volume). They are phrasal in that they exhibit full inflectional marking as well as phrasal stress, thus they have two distinct prosodic domains. Also, the [N N_{GEN}] sequences are left-headed. On the other hand, they behave unlike syntactic phrases, and like morphological compounds in that they are inseparable and do not allow modification of the non-head constituent, e.g., **métria psixrós pólemos* (lit. moderately cold war). Also, the [A N] sequences do not allow doubling of the definite article (and neither do A+N compounds), which is a usual constellation in common A N phrases, e.g., **o psixrós o pólemos* (lit. the cold the war, 'the Cold War'), but *o meyálos o pólemos* (lit. the big the war, 'the big war') (Ralli 2016: 3147). Also, many of these lexical phrases do not have a compositional meaning, just like compounds (e.g., Ralli 1992). In addition to these two patterns that have been described in detail in the works of Ralli (and colleagues), there are also several other lexical syntactic patterns, consisting of two inflected nouns, cf. Gavriilidou (2013), Ralli (2013a), Koliopoulou (this volume).

In sum, these sequences are clearly both lexical units and syntactic entities, and thus MWEs. For this reason, they pose a challenge as to their exact grammatical status, given the fact that they combine syntactic and morphological properties (which is obviously not necessarily the case for all MWEs). This is particularly important because they are not individual instances of lexicalization but rather the result of productive patterns for creating new lexical units, just as with compounding.[21] In the light of this, Booij (2009, 2010) offers a formal analysis for

21 Contrary to compounding, though, they seem to be a rather recent pattern. According to Ralli (2013a), they have been observed only in the last two centuries and have most probably emerged under the influence of French and English. Also, they are almost always restricted to specific registers.

Greek lexical [A N] sequences as syntactic compounds (N_0) within the framework of Construction Morphology; similarly, a constructional analysis for various lexical [N N] sequences is proposed in Gavriilidou (2013).

3.5 Slavic languages and German

Slavic languages, exemplified here by Russian and Polish, differ clearly from German with respect to the formation of new lexical items and in particular compounding. Although compounding, and in particular nominal compounding, is a productive word-formation process in both languages, it is a less important means for expanding the lexicon than it is in German (and other languages, such as English), particularly since derivation is highly productive (Uluhanov 2016).[22]

As in German, nominal compounds, in particular N+N compounds, are the predominant compound type both in Russian and Polish, e.g., Polish *gwiazd+o+zbiór* (lit. starset, 'constellation'), Russian *gaz+o+snabženie* ('gas supply'), followed by adjectival compounds, e.g., Polish *ciemn+o+niebieski* ('darkblue'), Russian *tëmn+o+sinij* ('darkblue'). Verbal compounding is considered unproductive in Polish (cf. Szymanek 2009) and only marginally productive in Russian (cf. Benigni/Masini 2009), although both languages have a rather small inventory of (older) verbal compounds. Generally, there are neither compounds with verbal modifiers (V+X) (cf. Ohnheiser 2015: 761) nor phrasal modifiers (XP+X) (cf. Bağrıaçık/Ralli 2015: 344; Szymanek 2017) in Slavic, in contrast to German. Compounding is mostly right-headed, although there are also some (minor) left-headed subpatterns. Compounds proper have a linking element, mostly *-o-*, as in the above-mentioned examples or, less frequently, *-e-*, *-i-*, *-u-*, and they are written in one word (or with a hyphen). Compounds in Polish display lexical stress on the penultimate syllable which clearly sets them apart from phrases. Finally, Polish and Russian compounds are hardly recursive; compounds with more than two constituents are only found with adjectival coordinate compounds, e.g., Polish *polsko-rosyjsko-ukraińskie* ('Polish-Russian-Ukrainian').

22 In the following, only the most important and basic properties of compounding in Russian and Polish are described. For further details, such as the difference between proper compounds and solid compounds, the various kinds of input elements, neoclassical compounding, the gender class shift etc. the reader is referred to the contributions by Ohnheiser (on Russian) and Cetnarowska (on Polish) in this volume, as well as Szymanek (2009), Benigni/Masini (2009), Ohnheiser (2015), Uluhanov (2016), Nagórko (2016).

In addition to the formation of endocentric and coordinate compounds of various kinds which are familiar from the German(ic) perspective, Polish and Russian (and Slavic in general) have another frequent and productive type of compounding, namely synthetic (or parasynthetic) compounding (cf., e.g., Benigni/Masini 2009; Melloni/Bisetto 2010; Ohnheiser 2015). This pattern is rare in German(ic), e.g., English *blue-eyed*, German *blauäuig*. In this case, a suffix is added to the compound simultaneously with the combination of the two constituents, e.g., Polish *nos+o+rożec* ('rhinoceros'), with *nos* 'nose', *róg* 'horn', the linking element *-o-* and the nominal suffix *-ec*, or Russian *rabot+o+da+tel'* ('employer'), with *rabota* 'work', *dat'* 'give', the linking element *-o-* and the nominal suffix *-tel'*. Importantly, the linking element and the suffix necessarily co-occur and enter the structure at the same time. For this reason, they are referred to as co-formatives (cf., e.g., Szymanek 2009; Nagórko 2016). Another interesting recent phenomenon in Russian are N+N compounds that are modelled on Germanic N+N compounds and which contain stems borrowed from English or German, e.g., *press-diskussija* ('press discussion'), *eskort-uslugi* ('escort service'), cf. Kapatsinski/Vakareliyska (2013). The authors suggest that they are not instances of borrowing of individual lexemes but rather a specific compound pattern that has been developed on the basis of the individual forms.

The observation that Russian and Polish make only limited use of compounding compared to German and English has often been attributed to the fact that – in addition to the high productivity of derivational processes – these languages have various productive MWE patterns, in particular nominal ones. For instance, Szymanek (2009: 465f.) notes that the equivalents of English N+N compounds such as *telephone number*, *toothpaste*, and *computer paper* are realized in Polish either as a noun phrase with an inflected noun modifier, usually in the genitive (e.g., *numer telefonu* 'telephone number'), a noun phrase with a PP modifier (e.g., *pasta do zębów* 'toothpaste'), or a noun phrase with a relational adjective as modifier (e.g., *papier komputerowy* 'computer paper'). Other patterns for the formation of lexical noun phrases (or: phrasal nouns) mentioned in the literature are N+N sequences with a noun modifier case other than genitive, N CONJ N (binomials), and A+N patterns. Thus, there are both N+A and A+N patterns, the adjective being often but not necessarily relational (cf. Masini/Benigni 2012; Cetnarowska 2015; Nagórko 2016; Cetnarowska 2018; Cetnarowska, this volume; Ohnheiser, this volume). Masini/Benigni (2012) stress that of all these patterns the A+N pattern is by far the most productive one in Russian.

In a similar way as has been discussed for the various phrasal lexical units in other languages in the preceding sections, Russian and Polish phrasal lexical units, or more specifically, lexical noun phrases, can be distinguished both from free syntactic phrases and from compounds on formal grounds. At the same time,

they also share properties with free syntactic phrases and compounds. For instance, these lexical noun phrases are inseparable. That is, they cannot be interrupted by intervening material, e.g., Russian *sotovyj telefon* ('mobile phone'), but **sotovyj služebnyj telefon* (lit. cellular official telephone). Also, the individual constituents cannot be modified internally, e.g., Russian *posobie po bezrabotice* 'unemployment benefit', but **posobie po ženskoj bezrabotice* (lit. benefit by female unemployment). These are properties typical of morphological entities and unlike free syntactic phrases; also, the function of lexical noun phrases as lexical naming unit equals that of compounds. On the other hand, lexical noun phrases display inflectional markers, like free syntactic phrases and unlike compounds, and some patterns contain relational elements, thus prepositions and conjunctions (as *po* in the last example), again like free syntactic phrases and unlike compounds (for a more detailed discussion of the tests employed including (apparent) counterexamples cf. Masini/Benigni 2012; Cetnarowska 2015; Ohnheiser 2015; Cetnarowska 2018; Cetnarowska, this volume; Ohnheiser, this volume). Thus, again it can be shown that these lexical noun phrases are lexical entities on the interface of syntax and the lexicon, i.e. lexical entities that are created in syntax. Building on works by Booij (2009, 2010) on A+N phrases in Dutch and Greek, among others, constructionist analyses have been proposed for these Russian and Polish lexical noun phrases in Masini/Benigni (2012), Cetnarowska (2018), Cetnarowska (this volume).

MWEs, and lexical noun phrases in particular, are also known from German, although it seems likely that these (or comparable patterns) are less productive in German than they are in Slavic, given the predominance of compounding in German. There is, however, another process for the formation of lexical items which stands in a close relationship to MWEs and MWE formation. This process is specific for Slavic and without a real equivalent in German. It is a process of shortening phrasal items to a single morphological lexeme.[23] More precisely, there are several shortening processes, among them ellipsis, truncation, clipping and de-suffixation (cf. Masini/Benigni 2012 on Russian; Martincová 2015 on Slavic in general), e.g., Polish *rzut karny* ('penalty throw') > *karny* (lit. penal), Russian *mineral'naja voda* ('mineral water') > *mineralka*. These processes are referred to either as shortening, condensation or univerbation, although the latter term is somewhat misleading as univerbation is often understood elsewhere

23 There are, obviously, also shortening processes such as clipping and contamination in German. They are much less systematic in nature than the shortening processes in Slavic however. Also, they occur only sporadically and are much less frequent. Finally, they do not require MWEs as input structures.

(i.e. in the non-Slavic literature) as the fixation of a phrasal item as a single word, without any shortening or change of the form. These shortenings produce forms that are synonymous to the input phrases. However, they belong to a different register as they are usually considered to be much more colloquial, expressive and informal than the corresponding phrases. They are considered to be very productive which, according to Martincová (2015), might also be related to the lower productivity of compounding in Slavic. Importantly, it is usually assumed that the input of these shortenings are not phrases in general but rather MWEs (for discussion on this point cf. Masini/Benigni 2012; Martincová 2015). If this is the case, then the productivity of shortenings presupposes productivity of MWE formation processes and ultimately, MWE formation not only creates phrasal lexical units but also systematically underlies the formation of non-phrasal lexemes in Slavic.

3.6 Finno-Ugric languages and German

Compounding is a productive word-formation pattern both in Finnish and Hungarian and can, according to Niemi, J. (2009) and Pitkänen-Heikkilä (2016), even be considered the most productive word-formation device in Finnish. The output classes are nominal and adjectival compounds, with N+N compounds being particularly productive, e.g., Hungarian *vér+nyomás* 'blood pressure', Finnish *tee+kuppi* 'tea cup'. According to Kiefer (2016: 3310), the productivity of nominal compounding in Hungarian has been considerably increased through loan-translations of thousands of German nominal compounds at the beginning of the 19th century. There are very few (apparent) verbal compounds (less than 1% in Finnish according to Kolehmainen/Savolainen 2007) and these forms are regarded either as backformations, univerbations from verbal phrases, or derivates rather than as compounds proper (cf. Kolehmainen/Savolainen 2007; Kiefer 2009, 2016; Pitkänen-Heikkilä 2016). Compounding in Finnish and Hungarian is also similar to German(ic) in that the vast majority of compounds is endocentric and right-headed. Furthermore, Finnish and Hungarian compounds display lexical stress which distinguishes them from phrases. (N+N) compounds are recursive, e.g., Hungarian [[*vér+nyomás*]+ *mérő*] 'blood-pressure measuring', [[[*vér+nyomás*]+ *mérő*]+ *készülék*] 'blood-pressure measuring apparatus' (Kiefer 2009). Finally, compounds are written as one word.

However, there are also clear differences between compounding in German and in Finnish and Hungarian. The first one is that compounds in Finnish and Hungarian never have linking elements. The second, and more important one, is that in addition to uninflected adjectival and nominal modifiers, e.g., Hungarian

feketes+zoftver (A+N) 'black/illegal software', Finnish *kylmä+varasto* (A+N) 'cool storage',[24] Finnish and Hungarian compounds also include inflected modifier constituents, thus, word-internal inflection, e.g., Hungarian *bolond-ok+ház-a* (N+N) 'mad house', with the plural suffix *-ok* (and the possessive suffix *-a*), Finnish *käde-n+sija* (N+N) (lit. hand's place) 'handle', with the genitive suffix *-n*. Regarding Hungarian, Kiefer (2009: 539) argues that they are not productive and morphologically formed compound patterns but rather univerbations from verb phrases or possessive noun phrases. Accordingly, there are no compounds with word-internal inflection in Hungarian. In Finnish, on the other hand, sequences with an inflected adjectival or nominal modifier constituent (mostly in the genitive, but also in other cases) are considered compounds proper (e.g., Niemi, J. 2009; Karlsson 2015; Pitkänen-Heikkilä 2016; Hyvärinen, this volume), similar to Icelandic (cf. Section 3.2). They form a considerable part of all cases: according to a corpus study by Niemi, J. (2009), about 14% of all nominal modifiers are inflected, about 20% of all adjectival modifiers and 22% of the verbal modifiers. Interestingly, compounds with a genitive modifier often have a possessive interpretation, e.g., *tuoli-n+jalka* (lit. chair's leg) 'leg of a chair' (cf. Pitkänen-Heikkilä 2016: 3214), which seems to suggest that these sequences are univerbated possessive phrases rather than compounds. However, possessive relationships are also found with non-genitive modifiers and genitive modifiers may also express other meaning relations (cf. Hyvärinen, this volume). Niemi, J. (2009: 239f.) claims that with respect to syntactic islandhood and lexical integrity, respectively, which underlie the debarment of word-internal inflection, Finnish compounds differ from other Standard European languages. Although from a morphological perspective, the sequences in question are phrases rather than compounds, they are regarded as compounds due to their lexical stress pattern which distinguishes them clearly from phrases.[25] Thus, the prosodic structure is regarded as decisive (and more important than the morphological one) for the classification as compound in the Finnish literature; again as in Icelandic. Recall from the previous sections that, in contrast, sequences classified as lexical noun phrases (phrasal nouns) in various languages retain phrasal stress, which is one property that distinguishes them from morphological compounds.

24 More precisely, in the Finnish literature, this is regarded as the nominative case, as the nominative equals the base form without any inflectional suffixes (cf., e.g., Hyvärinen, this volume).

25 An additional, morphosyntactic criterion is that possessive suffixes and clitics that can be added to Finnish nouns are not allowed inside compounds (Niemi, J. 2009: 241f.).

Lexical noun phrases, their morphosyntactic properties and the question of whether there are productive patterns of the formation of lexical noun phrases have to our knowledge not yet been discussed in the literature, at least not in the non-Finnish and non-Hungarian speaking one.[26] However, lexical noun phrases obviously exist, e.g., Hungarian *nyári szünet* (A_{REL} N) 'summer holidays', *állatok világa* (N_{PL} N_{POSS}) 'animal kingdom', Finnish *valkoinen valhe* (A N) 'white lie', also in terminology, e.g., Finnish *jätteiden poltto* ($N_{PL.GEN}$ N) 'waste combustion', *batyaalinen vyöhyke* (A_{REL} N) 'bathyal zone' (cf. Liimatainen 2008).

In the verbal domain, meanwhile, several interesting phenomena with respect to the lexicon-syntax interface have been discussed (cf. Hyvärinen, this volume). For Hungarian, Kiefer (1990, 1992, 2009) and Kiefer/Németh (this volume) describe the phenomenon of quasi-noun incorporation, that is, combinations of a bare noun and a verb, e.g., *levelet ír* (lit. letter write) 'to do letter writing', *zenét hallgat* (lit. music listen) 'to do music listening', in contrast to 'writing a letter' or 'listening to a (particular) piece of music'. These complex verbs always denote institutionalized activities. They are similar to compounds in that they exhibit compound stress. Also, the non-head cannot be modified, pluralized and is non-referential, just as a compound modifier. On the other hand, the noun and the verb can be separated, e.g., by the negative particle *nem* 'not' (cf. Kiefer 1992: 76) which indicates their phrasal nature. Thus, these complex verbs can be regarded as verbal MWEs. Quasi noun-incorporation also exists in German and Dutch (as well as in Danish, Norwegian, and Swedish, cf. Section 3.2). For a constructionist analysis for quasi-noun incorporation in Dutch, see Booij (this volume).

3.7 Discussion

This section is devoted to a discussion and summary of the preceding sections. The first, very simple observation is that all languages examined here have morphological compounds. However, it turned out that the compounds in these languages do not all share the same defining properties. While lexical (compound) stress, headedness (either right or left), inseparability and debarment of word-internal inflection, recursiveness, and linking elements are generally considered essential criteria for the definition of compound, in particular from a German(ic)

26 There are, however, numerous studies on MWEs in Finnish and Hungarian in the more traditional sense, written in German or English, in particular on verbal idioms, including various Hungarian-German and Finnish-German contrastive studies. For an overview on Finnish cf. Hyvärinen (2007).

perspective, all of them also emerged as problematic in at least one language, or as non-existent.[27] Thus, it seems that there is no universal definition of compound. Rather, as pointed out by Ralli (2013b: 184):

> What makes a compound morphological should be defined on a language-specific basis, since languages vary with respect to the realization of their morphological features and the use of morphologically-proper units.

Although it is ultimately impossible to weigh the various criteria against each other, it seems that compounding in German is – hardly surprisingly given the genetical relation – particularly similar to Dutch as well as the continental North Germanic languages, but also to Greek.

In addition to the defining criteria, also the number of compound subpatterns and the productivity of these patterns vary considerably between the languages. Verbal compounding, for instance, is regarded as either unproductive or only marginally productive in most languages, in contrast however to Greek which has several productive verbal compound patterns. What all languages discussed here have in common is that nominal compounding, and in particular N+N compounding, is considered the most frequent and probably also the most productive compound type (cf. likewise Guevara/Scalise 2009 for a much larger language sample).

A second observation is that all languages under discussion have MWEs and in particular MWEs that correspond or equal functionally to compounds.[28] Notably, it has been observed that all languages have various productive patterns that instantiate these phrasal lexical units.[29]

In the literature, the existence and productivity of MWE patterns is usually explained in relation to the existence and productivity of corresponding compound patterns and other word-formation processes, in particular derivation. Thus, for instance, compounding has been deemed comparatively limited in

27 Guevara/Scalise (2009) correctly point out that defining criteria of compounding such as those mentioned here usually reflect the Germanic perspective, given the huge amount of studies on compounding in Germanic, but cannot do justice to compounding from a broader perspective.

28 In this overview, more attention has been given to nominal MWEs than to verbal ones. This is due to reasons of space as well as to the fact that the starting point of this study is German, and that German compounding is predominantly nominal. Verbal lexical phrasal units are studied in detail in the chapters on Dutch, Finnish, and Hungarian (this volume).

29 As noted in Section 3.6, as far as we are aware there is no English or German speaking literature on lexical noun phrases and, in particular, on the respective lexical patterns in Finnish and Hungarian so far. There are, however, studies on complex verbal lexical units.

Slavic due to the productivity of both derivation and MWE formation, whereas the high productivity of nominal compounding in German has often been used as an explanation for the fact that the number and productivity of nominal German MWEs seem to be lower than in other languages.

Comparing the MWE patterns, it turned out that all languages have, among others, productive patterns for the formation of [A N] phrasal units (or [N A], in left-headed configurations). Among these, units with a relational adjective play an important role. In addition, some languages (among which German, Dutch, Danish, Swedish, Polish, and Greek) also have morphological A+N compounds, which raises – for each language – the question of synonymy and synonymy blocking. Another phrasal pattern that can be observed cross-linguistically are so-called phrasal similes, i.e. comparative adjectival phrases of the type [(*as*) A *as* NP], e.g., *as red as blood* (cf. Section 2.5). Phrasal similes are attested in the West and North Germanic languages as well as in Finnish and Italian,[30] e.g., Swedish *mjuk som silke*, German *weich wie Seide*, both 'soft as silk', Italian *rosso come il sangue* 'red as blood' (note that not all comparisons make sense in their literal meaning, e.g., Danish *dum som en dør* 'as stupid as a door'). They are particularly interesting with respect to the question of synonymy and synonymy blocking since all these languages also have an equivalent A+N compound pattern with a comparative meaning, e.g., *blood-red*, Swedish *silkesmjuk* 'silky smooth'. It can be observed that in some cases the existence of a phrasal or morphological form blocks the other (e.g., Swedish *mjuk som smör* 'soft as butter', but **smörmjuk* (lit. butter-soft)), but in other cases the phrasal and morphological form co-exist (e.g., Danish *dum som snot* (lit. stupid as snot, 'very stupid'), *snotdum* (id.)). The principles that underlie the (non-)blocking in the various cases, both within single languages and cross-linguistically, are however not yet fully understood. While both phrasal A+N units and phrasal similes seem to arise quite naturally from the usual syntactic patterns of the various languages, it is very interesting to see that phrasal patterns that are more specific in that they violate the syntactic rules can also be found in various languages. A case in point is the [*an* N_1 *of an* N_2] pattern (e.g., *a hell of a guy*), again a comparative pattern. It expresses a comparison of N_2 to the reference value provided by N_1. Hence, there is a mismatch between the semantic head of the construction (N_2) and the syntactic one (N_1), referred to as 'dependency reversal' in Rijkhoff (2009: 76). This pattern exists not only in Germanic, as for instance in English (*a hell of a guy*), German (*ein Idiot von* (*einem*) *Arzt* 'an idiot of (a) doctor'), Danish (*en klovn av en statsråd* 'a clown of privy council'), Swedish (*en kretin till*

30 More detailed studies on phrasal similes can be found in this volume in the chapters on German, Dutch, and Italian.

polisprefekt 'an imbecile of police chief') and Dutch (where it is well-known in the linguistic literature in connection with the famous example *schat van een kind* (lit. sweetheart of a child, 'very sweet child'), cf. Paardekoper 1956), but also in Italian, French, and Spanish (e.g., *esta maravilla de niño* 'this wonder of a child') (cf. Gunkel et al. 2017: 1627ff.).

One has to add that while in the present context attention is given only to the formal side, i.e. the morphosyntactic and possibly phonological properties of patterns such as [(*as*) A *as* NP] and [*an* N_1 *of an* N_2], cross-linguistic similarities (and differences) have also been studied with respect to the semantic side, in particular themes and images that feed imagery and metaphors in phrasal patterns and that re-occur cross-linguistically, due to cultural links and other factors (cf. Piirainen 2012, among many others). Thus, from this perspective, it is not unexpected that similar patterns occur cross-culturally in different, even genetically unrelated languages.

4 Summary

In this chapter, we sought to present an introductory overview of compound and MWE formation in a sample of European languages. We started with some general considerations about the notion of complex lexical unit, the lexicon, and the lexicon-syntax interface, and provided some preliminary criteria for the distinction between compounds and MWEs. In the second part of the chapter, we reviewed the language-specific properties of compounds and MWEs in West Germanic, North Germanic, Romance, Greek, Slavic, and Finno-Ugric languages, comparing them to German. Central questions that were discussed for each language family included the formal distinction between compounds and MWEs (in particular prosodic, morphological, and syntactic properties), the relationship between compounding and MWE formation as well as the conclusions concerning the theory of grammar and the lexicon following from these observations. One major finding is that while there are great similarities as well as differences regarding compound and MWE formation in the languages of Europe, a cross-linguistically valid definition of compounds and MWEs is hard to establish, because the languages differ greatly with respect to both the compound criteria that can be relevantly applied to them, and the relevant types of compound and MWE patterns and their degree of productivity. The various chapters of this volume provide in-depth analyses of the situation in the respective languages and language families, also discussing in more detail the relevant implications for the theory of the lexicon-grammar interface.

References

Anderson, Stephen R. (1982): Where's morphology? In: Linguistic Inquiry 13. 571–612.

Anward, Jan/Linell, Per (1976): Om lexikaliserade fraser i svenskan. In: Nysvenska studier 55/56. 76–119.

Arnaud, Pierre J. L. (2015): Noun-noun compounds in French. In: Müller, Peter O. et al. (eds.). 673–687.

Bağrıaçık, Metin/Ralli, Angela (2015): Phrasal vs. morphological compounds: Insights from Modern Greek and Turkish. In: STUF – Language Typology and Universals 68, 3. 323–357.

Bandle, Oscar et al. (eds.) (2005): The Nordic Languages. An International Handbook of the History of the North Germanic Languages. Vol. 2. (= Handbooks of Linguistics and Communication Science (HSK) 22.2). Berlin/New York: De Gruyter.

Bauer, Laurie (1998): When is a sequence of two nouns a compound in English? In: English Language and Linguistics 2. 65–86.

Bauer, Laurie (2006): Morphological productivity. (= Cambridge Studies in Linguistics 95). Cambridge, UK: Cambridge University Press.

Bauer, Laurie (2009a): Typology of compounds. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 343–356.

Bauer, Laurie (2009b): IE, Germanic: Danish. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 400–416.

Bauer, Laurie (2017): Compounds and Compounding. (= Cambridge Studies in Linguistics 155). Cambridge, UK: Cambridge University Press.

Bauer, Laurie (to appear): Compounds. In: Aarts, Bas/Bowie, Jillian/Popova, Geri (eds.): Oxford Handbook of English Grammar. Oxford: Oxford University Press.

Benigni, Valentina/Masini, Francesca (2009): Compounds in Russian. In: Lingue e Linguaggio 8, 2. 171–193.

Bjarnadóttir, Kristín (2017): Phrasal compounds in Modern Icelandic with reference to Icelandic word formation in general. In: Trips, Carola/Kornfilt, Jaklin (eds.). 13–48.

Booij, Geert (2002): The morphology of Dutch. Oxford: Oxford University Press.

Booij, Geert (2009): Phrasal names. A constructionist analysis. In: Word Structure 2, 2. 219–240.

Booij, Geert (2010): Construction morphology. Oxford/New York: Oxford University Press.

Braunmüller, Kurt (2007): Die skandinavischen Sprachen im Überblick. 3rd ed. Tübingen/Basel: Francke.

Burger, Harald (1998): Phraseologie. Eine Einführung am Beispiel des Deutschen. Berlin: Schmidt.

Burger, Harald et al. (eds.) (2007): Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research. (= Handbooks of Linguistics and Communication Science (HSK) 28). Berlin/New York: De Gruyter.

Cetnarowska, Bożena (2015): The lexical/phrasal status of Polish Noun+Adjective or Noun+Noun combinations and the relevance of coordination as a diagnostic test. In: SKASE Journal of Theoretical Linguistics 12, 3. 142–170.

Cetnarowska, Bożena (2018): Phrasal names in Polish: A+N, N+A and N+N units. In: Booij, Geert (ed.): The Construction of Words. Advances in Construction Morphology. (= Studies in Morphology Series 4). Cham: Springer. 287–313.

Chomsky, Noam (1970): Remarks on Nominalization. In: Jacobs, Roderick A./Rosenbaum, Peter S. (eds.): Readings in English Transformational Grammar. Waltham, MA: Ginn & Co. 184–221.

De Caluwe, Johan (1990): Complementariteit tussen morfologische en in oorsprong syntactische benoemingsprocédés. In: De Caluwe, Johan (ed.): Betekenis en productiviteit: Gentse bijdragen tot de studie van de Nederlandse woordvorming. Gent: Seminarie voor Duitse Taalkunde. 9–23.

Di Sciullo, Anna-Maria/Williams, Edwin (1987): On the definition of word. Cambridge, MA: MIT Press.

Dura, Elsbieta/Gawronska, Barbara (2007): Novelty extraction from special and parallel corpora. In: Vetulani, Zygmunt/Uszkoreit, Hans (eds.): Third Language and Technology Conference (LTC 2007), Poznan, Poland, October 5–7, 2007. (= Lecture Notes in Artificial Intelligence 5603). Berlin/Heidelberg: Springer. 291–302.

Engelberg, Stefan/Holler, Anke/Proost, Kristel (2011): Zwischenräume. Phänomene, Methoden und Modellierung im Bereich zwischen Lexikon und Grammatik. In: Engelberg, Stefan/Holler, Anke/Proost, Kristel (eds.): Sprachliches Wissen zwischen Lexikon und Grammatik. Berlin/Boston: De Gruyter. 1–35.

Farø, Ken (2015): Feste Wortgruppen/Phraseologie II: Phraseme. In: Haß, Ulrike/Storjohann, Petra (eds.): Handbuch Wort und Wortschatz. Berlin/Boston: De Gruyter. 226–247.

Fillmore, Charles J./Kay, Paul/O'Connor, Mary C. (1988): Regularity and idiomaticity in grammatical constructions: The case of *let alone*. In: Language 64. 501–538.

Finkbeiner, Rita (2008): Zur Produktivität idiomatischer Konstruktionsmuster. Interpretierbarkeit und Produzierbarkeit idiomatischer Sätze im Test. In: Linguistische Berichte 216. 391–430.

Fleischer, Wolfgang (1982): Phraseologie der deutschen Gegenwartssprache. Leipzig: VEB Bibliographisches Institut.

Fradin, Bernhard (2009): IE, Romance: French. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 417–435.

Gaeta, Livio/Grossmann, Maria (eds.) (2009): Compounds between syntax and lexicon. Special Issue of Italian Journal of Linguistics/Rivista di Linguistica 21, 1.

Gaeta, Livio/Ricca, Davide (2009): Composita solvantur: Compounds as lexical units or morphological objects? In: Gaeta, Livio/Grossmann, Maria (eds.). 35–70.

Gaeta, Livio/Schlücker, Barbara (eds.) (2012): Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte. (= Linguistik – Impulse & Tendenzen 46). Berlin/New York: De Gruyter.

Gavriilidou, Zoe (2013): NN Combinations in Greek. In: Journal of Greek Linguistics 13, 1. 5–29.

Giegerich, Heinz J. (2004): Compound or phrase? English noun-plus-noun constructions and the stress criterion. In: English Language and Linguistics 8, 1. 1–24.

Giegerich, Heinz J. (2009): Compounding and Lexicalism. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 178–200.

Giegerich, Heinz J. (2015): Lexical structures: compounding and the modules of grammar. (= Edinburgh Studies in Theoretical Linguistics 1). Edinburgh: Edinburgh University Press.

Goldberg, Adele E. (1995): Constructions. A construction grammar approach to argument structure. Oxford: Oxford University Press.

Goldberg, Adele E. (2006): Constructions at Work. The nature of generalization in language. Oxford: Oxford University Press.

Gries, Stefan Th. (2008): Phraseology and linguistic theory: a brief survey. In: Granger, Sylviane/Meunier, Fanny (eds.): Phraseology. An interdisciplinary perspective. Amsterdam: Benjamins. 3–25.
Guevara, Emiliano/Scalise, Sergio (2009): Searching for Universals in Compounding. In: Scalise, Sergio/Bisetto, Antonietta (eds.): Universals of Language Today. Dordrecht: Springer. 101–128.
Guevara, Emiliano R. (2012): Spanish compounds. In: Probus 24, 1. 175–195.
Gunkel, Lutz/Zifonun, Gisela (2011): Klassifikatorische Modifikation im Deutschen und Französischen. In: Lavric, Eva/Pöckl, Wolfgang/Schallhart, Florian (eds.): Comparatio delectat: Akten der VI. Internationalen Arbeitstagung zum romanisch-deutschen und innerromanischen Sprachvergleich, Innsbruck, 3.–5. September 2008. Frankfurt a.M.: Lang. 549–562.
Gunkel, Lutz et al. (2017): Grammatik des Deutschen im europäischen Vergleich: das Nominal. Vol. 2: Nominalflexion, Nominale Syntagmen. (= Schriften des Instituts für Deutsche Sprache 14). Berlin/Boston: De Gruyter.
Haberland, Hartmut (1994): Danish. In: König, Ekkehard/Auwera, Johan van der (eds.). 313–348.
Haeringen, Coenraad Bernardus van (1956): Nederlands tussen Duits en Engels. 2de druk. Den Haag: Servire.
Härtl, Holden (2016): Normality at the boundary between word-formation and syntax. In: d'Avis, Franz/Lohnstein, Horst (eds.): Normalität in der Sprache. In: Linguistische Berichte Sonderheft 22. Hamburg: Buske. 65–92.
Häusermann, Jürg (1977): Phraseologie. Hauptprobleme der deutschen Phraseologie auf Basis sowjetischer Forschungsergebnisse. Tübingen: Niemeyer.
Hoffmann, Thomas/Trousdale, Graeme (eds.) (2013): The Oxford Handbook of Construction Grammar. Oxford: Oxford University Press.
Hohenhaus, Peter (2005): Lexicalization and institutionalization. In: Štekauer, Pavol/Lieber, Rochelle (eds.). 353–373.
Hüning, Matthias (2010): Adjective + Noun constructions between syntax and word formation in Dutch and German. In: Onysko, Alexander/Michel, Sascha (eds.): Cognitive Perspectives on Word Formation. (= Trends in Linguistics. Studies and Monographs 221). Berlin/New York: De Gruyter. 195–215.
Hüning, Matthias/Schlücker, Barbara (2010): Konvergenz und Divergenz in der Wortbildung. Komposition im Niederländischen und im Deutschen. In: Dammel, Antje/Kürschner, Sebastian/Nübling, Damaris (eds.): Kontrastive Germanistische Linguistik (= Germanistische Linguistik 206–209). Vol. 2. Hildesheim i.a.: Olms. 783–825.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.). 450–467.
Hyvärinen, Irma (2007): Phraseologie des Finnischen. In: Burger, Harald et al. (eds.). Vol. 2. 737–752.
Inghult, Göran (1991): Lexikalische Innovationen in Wortgruppenform. Zu einer Untersuchung über die Erweiterung des Lexembestandes im Deutschen und Schwedischen. In: Palm, Christine (ed.): Europhras 90. Akten der internationalen Tagung zur germanistischen Phraseologieforschung Aske/Schweden, 12.–15. Juni 1990. Uppsala: Almqvist & Wiksell. 101–114.
Jackendoff, Ray (1997): The architecture of the language faculty. Cambridge, MA: MIT Press.
Jackendoff, Ray (2002): Foundations of language. Oxford: Oxford University Press.

Jackendoff, Ray (2008): Construction after construction and its theoretical challenges. In: Language 84. 8–28.
Jackendoff, Ray (2009): Compounding in the Parallel Architecture and Conceptual Semantics. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 105–128.
Jackendoff, Ray (2013): Constructions in the Parallel Architecture. In: Hoffmann, Thomas/Trousdale, Graeme (eds.). 70–92.
Josefsson, Gunlog (1997): On the principles of word formation in Swedish. Lundastudier i nordisk språkvetenskap 51. Lund: Lund University Press.
Kapatsinski, Vsevolod/Vakareliyska, Cynthia (2013): [N[N]] compounds in Russian: A growing family of constructions. In: Constructions and Frames 5, 1. 69–87.
Karlsson, Fred (2015): Finnish: an essential grammar. 3rd edition. Milton Park i.a.: Routledge. [Trans. by Andrew Chesterman].
Kiefer, Ferenc (1990): Noun incorporation in Hungarian. In: Acta Linguistica Hungarica 40, 1–2. 149–177.
Kiefer, Ferenc (1992): Compounding in Hungarian. Rivista di Linguistica 4. 61–78.
Kiefer, Ferenc (2009): Uralic, Finno-Ugric: Hungarian. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 527–541.
Kiefer, Ferenc (2016): Hungarian. In: Müller, Peter O. et al. (eds.). 3308–3326.
Klinge, Alex (2006): The Origin of Weapons of Mass Destruction. Investigating Traces of Lexical Formation Patterns in the (Lingustic) History of Europe. In: Nølke, Henning (ed.): Grammatica: Festschrift in honour of Michael Herslund/Hommage à Michael Herslund. Frankfurt a.M./New York: Lang. 233–248.
Kolehmainen, Leena/Savolainen, Tiina (2007): Deverbale Verbbildung im Deutschen und im Finnischen: ein Überblick. Würzburg. (Internet: https://opus.bibliothek.uni-wuerzburg.de/opus4-wuerzburg/frontdoor/index/index/docId/1937, last access: 30.5.2018).
König, Ekkehard/Auwera, Johan van der (eds.) (1994): The Germanic Languages. London/New York: Routledge.
Koptjevskaja-Tamm, Maria (2009): Proper-name nominal compounds in Swedish between syntax and lexicon. In: Rivista di Linguistica 21, 1. 119–148.
Kunter, Gero (2011): Compound stress in English: the phonetics and phonology of prosodic prominence. (= Linguistische Arbeiten 539). Berlin: De Gruyter.
Lieber, Rochelle/Štekauer, Pavol (eds.) (2009a): The Oxford handbook of compounding. Oxford: Oxford University Press.
Lieber, Rochelle/Štekauer, Pavol (2009b): Introduction: Status and definition of compounding. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 3–18.
Liimatainen, Annikki (2008): Untersuchungen zur Fachsprache der Ökologie und des Umweltschutzes im Deutschen und Finnischen: Bezeichnungsvarianten unter einem geschichtlichen, lexikografischen, morphologischen und linguistisch-pragmatischen Aspekt. (= Finnische Beiträge zur Germanistik 22). Frankfurt a.M.: Lang.
Lipka, Leonhard (1977): Lexikalisierung, Idiomatisierung und Hypostasierung als Probleme einer synchronischen Wortbildungslehre. In: Brekle, Herbert E./Kastovsky, Dieter (eds.): Perspektiven der Wortbildungsforschung. Beiträge zum Wuppertaler Wortbildungskolloquium vom 9.–10. Juli 1976. Bonn: Bouvier. 155–164.
Lüdeling, Anke (2001): On particle verbs and similar constructions in German. Stanford: CSLI Publications.
Martincová, Olga (2015): Multi-word expressions and univerbation in Slavic. In: Müller, Peter O. et al. (eds.). 742–757.

Masini, Francesca (2009): Phrasal lexemes, compounds and phrases: A constructionist perspective. Word Structure 2, 2. 254–271.
Masini, Francesca/Benigni, Valentina (2012): Phrasal lexemes and shortening strategies in Russian: the case for constructions. In: Morphology 22, 3. 417–451.
Masini, Francesca/Scalise, Sergio (2012): Italian compounds. In: Probus 24, 1. 61–91.
Melloni, Chiara/Bisetto, Antonietta (2010): Parasynthetic compounds. Data and theory. In: Scalise, Sergio/Vogel, Irene (eds.): Cross-Disciplinary Issues in Compounding. Amsterdam/Philadelphia: Benjamins. 199–217.
Moon, Rosamund (2015): Multi-word items. In: Taylor, John R. (ed.): The Oxford Handbook of The Word. 1st ed. Oxford: Oxford Univesity Press. 120–140.
Müller, Peter O. et al. (eds.) (2015–2016): Word-formation. An international handbook of the languages of Europe. (= Handbooks of Linguistics and Communication Science (HSK) 40). Berlin/Boston: De Gruyter.
Nagórko, Alicja (2016): Polish. In: Müller, Peter O. et al. (eds.). 2831–2852.
Newmeyer, Frederick J. (1974): The regularity of idiom behavior. In: Lingua 34. 327–42.
Niemi, Jussi (2009): Compounds in Finnish. In: Lingue e Linguaggio 8, 2. 237–256.
Niemi, Sinikka (2009): Compounds in Swedish. In: Lingue e Linguaggio 8, 2. 257–269.
Nunberg, Geoffrey/Sag, Ivan/Wasow, Thomas (1994): Idioms. In: Language 70, 3. 491–538.
Ohnheiser, Ingeborg (2015): Compounds and multi-word expressions in Slavic. In: Müller, Peter O. et al. (eds.). 757–779.
Olsen, Susan (2000): Composition. In: Booij, Geert/Lehmann, Christian/Mugdan, Joachim (eds.): Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung/ Morphology. An international handbook on inflection and word-formation. Berlin: De Gruyter. 897–916.
Paardekoper, Petrus C. (1956): Een schat van een kind. In: De Nieuwe Taalgids 49. 93–99.
Pafel, Jürgen (2015): Phrasal compounds are compatible with lexical integrity. In: Language typology and universals (STUF) 68, 3. 263–280.
Pawley, Andrew/Syder, Frances H. (1983): Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In: Richards, Jack C./Schmidt, Richard W. (eds.): Language and Communication. London/New York: Routledge. 191–226.
Piirainen, Elisabeth (2012): Widespread idioms in Europe and beyond: toward a lexicon of common figurative units. (= International Folkloristics 5). New York i.a.: Lang.
Pitkänen-Heikkilä, Kaarina (2016): Finnish. In: Müller, Peter O. et al. (eds.). 3209–3228.
Plag, Ingo (2006): The variability of compound stress in English: structural, semantic, and analogical factors. In: English Language and Linguistics 10, 1. 143–172.
Rainer, Franz (2016): Spanish. In: Müller, Peter O. et al. (eds.). 2620–2640.
Rainer, Franz/Varela, Soledad (1992): Compounding in Spanish. In: Rivista di Linguistica 4, 1. 117–142.
Ralli, Angela (1992): Compounds in Modern Greek. In: Rivista di Linguistica 4. 143–174.
Ralli, Angela (2008): Compound markers and parametric variation. In: STUF – Language Typology and Universals 61, 1. 19–38.
Ralli, Angela (2009): IE, Hellenic: Modern Greek. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 453–463.
Ralli, Angela (2013a): Compounding in modern Greek. (= Studies in Morphology 2). Dordrecht: Springer.
Ralli, Angela (2013b): Compounding and its locus of realization: Evidence from Greek and Turkish. In: Word Structure 6, 2. 181–200.

Ralli, Angela (2016): Greek. In: Müller, Peter O. et al. (eds.). 3138–3156.
Ralli, Angela/Stavrou, Melita (1998): Morphology-syntax interface: A-N compounds vs. A-N constructs in Modern Greek. In: Booij, Geert E./Marle, Jaap Van (eds.): Yearbook of Morphology 1997. Dordrecht: Springer. 243–264.
Ramisch, Carlos (2015): Multiword expressions acquisition. A generic and open framework. New York i.a.: Springer.
Ricca, Davide (2015): Verb-noun compounds in Romance. In: Müller, Peter O. et al. (eds.). 688–707.
Rijkhoff, Jan (2009): On the co-variation between form and function of adnominal possessive modifiers in Dutch and English. In: McGregor, William (ed.): The expression of possession. Berlin/New York: De Gruyter. 51–106.
Rosenbach, Anette (2006): Descriptive genitives in English. A case study on constructional gradience. In: English Language and Linguistics 10. 77–118.
Saxton, Matthew (2010): Child Language. Acquisition and Development. London: SAGE.
Scalise, Sergio (1992): Compounding in Italian. In: Rivista di Linguistica 4. 175–199.
Scalise, Sergio (ed.) (1992): The Morphology of Compounding. Special issue of Rivista di Linguistica 4, 1.
Scalise, Sergio/Guevara, Emiliano (2005): The lexicalist approach to word formation and the notion of the lexicon. In: Štekauer, Pavol/Lieber, Rochelle (eds.). 147–187.
Scalise, Sergio/Vogel, Irene (eds.) (2010): Cross-disciplinary issues in compounding. Amsterdam/Philadelphia: Benjamins.
Schlücker, Barbara (2014): Grammatik im Lexikon. Adjektiv+Nomen-Verbindungen im Deutschen und Niederländischen. (= Linguistische Arbeiten 553). Berlin/Boston: De Gruyter.
Schlücker, Barbara/Plag, Ingo (2011): Compound or Phrase? Analogy in Naming. In: Lingua 121. 1539–1551.
Sinclair, John (1991): Corpus, Concordance, Collocation. Oxford: Oxford University Press.
Štekauer, Pavol/Lieber, Rochelle (eds.) (2005): Handbook of Word-Formation. Dordrecht: Springer.
Szymanek, Bogdan (2009): IE, Slavonic: Polish. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 464–477.
Szymanek, Bogdan (2017): Compounding in Polish and the absence of phrasal compounding. In: Trips, Carola/Kornfilt, Jaklin (eds.). 49–79.
Teleman, Ulf (2005): The Standard languages and their systems in the 20th century: Swedish. In: Bandle, Oscar et al. (eds.). 1603–1626.
Thráinsson, Höskuldur (1994): Icelandic. In: König, Ekkehard/Auwera, Johan van der (eds.). 142–190.
Torp, Arne (2002): The Nordic Languages in a Germanic Perspective. In: Bandle, Oscar et al. (eds.). 13–24.
Trips, Carola/Kornfilt, Jaklin (2015): Typological aspects of phrasal compounds in German, English, Turkish and Turkic. In: Language typology and universals (STUF) 68, 3. 281–321.
Trips, Carola/Kornfilt, Jaklin (eds.) (2017): Further investigations into the nature of phrasal compounding. (= Morphological Investigations 1). Berlin: Language Science Press.
Uluhanov, Igor' S. (2016): Russian. In: Müller, Peter O. et al. (eds.). 2953–2978.
Van Goethem, Kristel (2009): Choosing between A+N compounds and lexicalized A+N phrases: The position of French in comparison to Germanic languages. In: Word Structure 2, 2. 241–253.

Villoing, Florence (2012): French compounds. In: Probus 24, 1. 29–60.
Weinreich, Uriel (1969): Problems in the analysis of idioms. In: Puhvel, Jaan (ed.): Substance and Structure of Language. Berkeley: University of California Press. 23–76.
Wray, Alice (2002): Formulaic Language and the Lexicon. Cambridge, UK: Cambridge University Press.
Wunderlich, Dieter (2006): Introduction: What the theory of the lexicon is about. In: Wunderlich, Dieter (ed.): Advances in the Theory of the Lexicon. Berlin/New York: De Gruyter. 1–26.
Zeller, Jochen (2001): Particle verbs and local domains. Amsterdam: Benjamins.

Laurie Bauer

Compounds and multi-word expressions in English

1 Introduction

Compounds are traditionally defined as being, in the words of Lieber (2010: 43), "words that are composed of two (or more) bases, roots, or stems". Multi-word expressions (also known as multi-word units or items, henceforth MWEs) can be defined as "lexical items which consist of more than one 'word' and have some kind of unitary semantic or pragmatic function" (Moon 2015: 120). Since all words (in the sense of 'lexeme', which is what I assume Lieber to mean in the cited passage) are lexical items, the first thing to note is that these two definitions overlap (*pace* ibid.: 121). Things called compounds, if they have 'some kind of unitary semantic or pragmatic function', which they can be argued always to have, are MWEs, although not all MWEs are compounds.

In this chapter, it will be argued that this fuzzy borderline between compounds and MWEs is real, that there is no generally accepted way of dividing compounds from MWEs, and that much of this derives from their common function as lexical items. Furthermore, there is no generally accepted way of dividing compounds from syntactic phrases, so that it follows that there is no generally accepted way of dividing MWEs from syntactic phrases. This situation arises partly from the data, and partly from the varying views of different scholars, who have tried to draw dividing lines in different places, thus illustrating the lack of commonality of opinion. Because this chapter focusses on the situation in English, the arguments affect English specifically, and may not all transfer to other languages. No attempt is made here to generalise to other languages; that is left for another chapter. The effect is, however, a claim that there is no agreed definition of a compound in English (and possibly not of an MWE, as noted ibid.).

2 The notion of 'word'

Word must be one of the least well-defined technical terms in linguistics. There are innumerable discussions of why this is the case, and there is little point in adding to them here (cf., e.g., Bauer 2000; Dixon/Aikhenvald 2002; Hippisley 2015; Wray 2015). In some languages, some criterion or set of criteria can be used

https://doi.org/10.1515/9783110632446-002

to define 'word' sufficiently well to allow a definition of a compound as a word to be meaningful. In English, it is less clear that this is true. Consider just three potential criteria, which are often used in other languages.

The first of these is stress. In other Germanic languages, stress is often used as a criterion for compoundhood. Any discussion of English in these terms, however, falls foul of examples like those in (1).

(1)

Forestress	End-stress
apple cake	*apple pie*
glass cupboard	*glass cupboard*
('cupboard in which glassware is kept')	('cupboard made of glass')
toy factory	*toy factory*
('factory that makes toys')	('factory which is itself a toy')
York Street	*York Avenue*

In addition, Bauer (1983b) finds that speakers are inconsistent in assigning stress to (at least some) such expressions, and also notes (Bauer to appear) variable usage of stress in the speech of newsreaders. Kunter (2011) finds a reasonable minority of such forms show variable stress. While most authorities now see stress as not being a reliable guide to the status of such items as compounds (Giegerich 2004), this has not always been the case, so that some such expressions have seemed to be changing category from compound to non-compound in an apparently random fashion. Chomsky/Halle (1968), for example, use stress as definitional for compounds.

Spelling is, to some extent, linked with stress: *railway* is written as one word and has forestress, *iron bar* is written as two and has end-stress. Other factors are also involved, however: *schoolgirl* tends to be written as one word, while *university student* has to be written as two, despite parallel stress and semantic readings. Some of the examples in (1) equally show a distinction between stress and orthography. It is also well-known that English orthography is inconsistent when it comes to writing some compounds: *rainforest*, *rain-forest* and *rain forest* can all be found in dictionaries. Matters as difficult to quantify as house-style and fashion can influence such spellings. Spelling cannot be criterial for word status in English. Nonetheless, some scholars use it in this way, either by default (cf. Hall 1964: 134) or to make dealing with the computational analysis of written text possible (McEnery/Xiao/Tono 2006: 147).

As a third criterion, consider the notion that words allow for global inflection, but not for inflection which is internal and applies to some element within the word. If we consider a compound verb like *badge-flash* (see (2)), we can see how this works.

(2) *I badge-flashed my way to the scene*.[1]

In (2) we see that *badge-flash* can take a past tense which affects the entire entity *badge-flash*. However, even if several members of the police made their way to the scene in this way, we could not change this to **We badges-flashed our way to the scene*. Global inflection is possible, but not internal inflection. There are two problematic constructions in English in relation to this criterion. The first is illustrated by *jobs growth*, where the first element of the compound has an apparent plural. Pinker (1999) sees this as sufficient evidence to say that such constructions are phrasal, not words, others include these as compounds (and, hence, as words). The other awkward construction in this regard is the classifying genitive as in *cat's eye* ('reflecting road marker' or 'semi-precious stone'). Even if we ignore the question as to whether the *s*-genitive in English is inflectional or a clitic (cf. Bauer/Lieber/Plag 2013: 141f. for a brief summary), it is not clear whether such constructions count as single words. They are compound-like in many ways (Rosenbach 2006), though most scholars exclude them from the set of compounds. Corresponding expressions in other Germanic languages are generally thought of as compounds, although 'uneigentlich' ('non-genuine, false') compounds in Grimm's terminology.

Other criteria for wordhood are frequently used in attempting to determine whether given constructions are words (compounds) or not. These include the fixed order of constructions, non-interruptibility of elements, lack of modification of internal elements, lack of coordination of internal elements, impossibility of referring back to individual elements by pronouns, including *one*, and listedness. Not only do such criteria not define a coherent set of items as words (Bauer 1998), they are often broken in derivatives, whose wordhood is not usually queried. These criteria will be referred to below, as required. The point here is that not only do the criteria for wordhood not fit compounds particularly well (cf. also Giegerich 2015; Bauer 2017), they do not allow agreement on what is or is not a compound in English. The border of compounding is vague partly because the border of wordhood is vague.

In what follows, a number of constructions will be considered in varying detail. Some of these constructions will be ones which some scholars see as compounds, others will be MWEs more loosely defined. The borderline between these two groups of construction will be shown to be non-principled, with different theoreticians making different decisions as to what is or is not a compound.

1 Karp, Marshall (2006): *The rabbit factory*. San Francisco: MacAdam, 179.

3 Formal constructions

3.1 N+N

There are several classes of N+N constructions in English, and while some of them are regularly considered to be compounds, many of them are equally regularly considered to be excluded from the category. We can illustrate some of the classes as in (3).

(3a) *Doctor Johnson*, *Miss Havisham*, *King George*
(3b) *Elizabeth Taylor*, *John Lennon*
(3c) *President Donald Trump*, *Prime Minister Theresa May*
(3d) *beef Wellington*, *chicken Kiev*
(3e) *the category adjective*, *letter A*, *number nine*, *Model T*
(3f) *bank-box*, *bus-driver*, *car park*, *windmill*
(3g) *Oxford college*, *cutlery box*
(3h) *iron bar*, *copper wire*, *stone wall*
(3i) *the film 'Jaws'*, *the year 1952*
(3j) *egg head*, *hatchback*
(3k) *father-daughter*, *hand-eye*
(3l) *Nelson-Marlborough*, *Daimler-Benz*
(3m) *murder-suicide*, *mind/brain*
(3n) *singer-songwriter*, *lawyer-poet*
(3o) *elm tree*, *tuna fish*
(3p) *salad-salad*

Names are not usually counted as being compounds. Those in (3c) are generally seen as instances of apposition, and frequently have a pause and intonation break between the title and the name (unlike those in (3a) which would otherwise be parallel). Apposition is usually considered a syntactic construction rather than a lexical one, and so the examples in (3a–c) and also the examples in (3i) with common nouns, are excluded from compounds. However, at least those in (3a) and (3b) must be listed, since they denote individuals and have little semantic transparency. Those in (3a) appear to be left-headed (Doctor Johnson is a member of the class of people with title of doctor), while those in (3b) may not be – it is not clear whether it even makes sense to ask whether Elizabeth Taylor is a member of the set of Elizabeths or the set of Taylors, especially since asking such a question changes the category of both *Elizabeth* and *Taylor* from proper noun to common noun. The examples in (3e) may also be instances

of apposition, but it is less clear: *Model T* is something which deserves at least an encyclopedic entry, if not a lexical entry, and acts as a label for a class of objects in much the same way as the noun *T-junction* does. *Model T*, though, is left-headed. While headedness is not usually given as one of the criteria for compoundhood in English (though cf. Bauer/Lieber/Plag 2013), most of the items that are seen as clear cases of compounds in English are right-headed. The examples in (3d) are also left-headed, but here it seems even less likely that apposition is involved. These items are names of dishes and synchronically at least have little to do with any semantic content that might be derived from their second elements. They certainly fit the definition of compound given in Section 1 above, and they are listed.

The items in (3f) are the central examples of compounds (though including examples from rather different subsets), and those from (3h) are examples which are often thought of as syntactic, but for different reasons. For Giegerich (2015) the first word in these constructions is an adjective, for others they are syntactic because their orthography, stress and behaviour under coordination shows them to be so: *copper and aluminium wire* and *copper wire and cable* are both unexceptional. The items in (3g) provide an intermediate step. For some scholars they are compounds, for others (e.g. Payne/Huddleston 2002) they are syntactic, because they fail at least one of the criteria for being words. For example, *Oxford and Cambridge colleges* is perfectly acceptable, as is *four Oxford and three Cambridge colleges* and *cutlery and wine-glass boxes* and *assorted silver cutlery box*.[2] Note that Payne and Huddleston have an overarching principle that any trace of syntactic behaviour makes something a syntactic structure rather than a principle that any hint of lexical behaviour makes something non-syntactic.

The items in (3i), as mentioned above, are appositional, and are usually excluded from the set of compounds, but they contrast with the set in (3o) which are usually included. Even so, they do not easily allow interruption, though they do allow coordination, where relevant, as in *the movie and book* Jaws,[3] and they certainly allow submodification of just one element, as in *the thrilling film "Jaws"* or *the thrilling film, the notorious "Jaws"* (but note the necessity for the determiner in this last example, which may change the construction).

The items in (3j) are usually considered compounds, but exocentric compounds, often thought of as unheaded (cf. Carstairs-McCarthy 2002). The particular items listed here fit into the Sanskrit category of bahuvrihi compounds, which

2 Internet: www.spotlightstores.com/party/party-decorator/room-table/decorating-accessories/amscan-assorted-silver-cutlery-box/p/BP80402188 (last access: 17 Nov 2017).

3 Internet: https://en.wikipedia.org/wiki/Frank_Mundus (last access: 17 Nov 2017).

others see as regular endocentric compounds interpreted through the figure of speech synecdoche (sometimes considered to be a type of metonymy) (Bauer 2016).

The types in (3k–p) are various kinds of coordinative compounds. Adams (2001: 3) excludes all of these from the set of compounds, apparently because they are unheaded. It should be noted that some of these are exocentric (for different reasons): *hand-eye* (in *hand-eye coordination*) is exocentric because it is used exclusively as a premodifier (and thus, possibly, an adjective), while *Nelson-Marlborough* is neither a hyponym of *Nelson* nor a hyponym of *Marlborough*. On the other hand, a *singer songwriter* is both a *singer* and a *songwriter*, and an *elm tree* is both an *elm* and a *tree*. We have already seen that examples like *elm tree* bear some resemblance to instances of apposition, another potential reason for not including them as compounds. Some scholars include items like that in (3p) as compounds, while others might see it as reduplication or even just repetition (a *salad-salad* is one which contains things typically found in a salad like lettuce and cucumber, as opposed, say, to a pasta salad).

3.2 A+N

Again, we can find many classes of construction involving adjectives and nouns. A+N compounds are usually distinguished from syntactic constructions by their stress (forestress) and, correspondingly, their orthographic unity, by the fact that the adjective cannot be submodified or graded, and by the fact that the adjective can be denied without contradiction. Thus *blackbird* is a compound by virtue of its stress, its orthography, the fact that we cannot have *a blackerbird* or *a very-blackbird*, and because *This blackbird is brown* is not a contradiction. The syntactic construction *black bird* differs from the compound *blackbird* in all of these respects. This distinction arises because *black* in *black bird* describes, while *black* in *blackbird* categorises.

If we look at intersective adjectives like *black*, *heavy*, *silly* etc. where *a black bird* represents the intersection of black things and birds, we discover that they are not always intersective. *A red book* may illustrate an intersective use of *red*, but *a red squirrel* does not: *red squirrel* behaves semantically like *blackbird*, not like *black bird*, despite different stress and orthography. Bauer (2004) points out that there is a difference in frequency between the forestressed words and the end-stressed expressions: the forestressed words are more frequent. This would seem to indicate that there are intersective adjectives used descriptively, intersective adjectives used to categorise, and intersective adjectives used with forestress to categorise. Most authorities distinguish the compounds with forestress from

the other two types, but we might equally distinguish the descriptive adjectives from the categorising ones.

If we now turn to relational adjectives like *canine*, *dental*, *parental*, *vernal*, they are not intersective. *A canine tooth* is not the intersection of canine things and teeth, but a kind of tooth related in some way to dogs (for fuller discussion cf. Giegerich 2015). Relational adjectives are rarely descriptive unless they are figurative or used predicatively, as in *his movements were feline*, *her attitude was vaguely parental*. The precise relationship between the adjective and the noun has to be discovered by considering the individual example, just as the relationship between nouns in N+N compounds has to be discovered by considering the individual example: a *windmill* uses wind power, but a *flour mill* grinds flour. Part of the result of this is that relational adjectives are by default categorising. Nevertheless, there are instances when they, too, can take forestress: consider for instance *dramatic society*, *mental hospital*, *primary school*. The reason for the forestress here is not clear. Neither is it clear whether things like *mental hospital* are compounds. Scholars disagree on whether A+N constructions with relational adjectives are compounds or not, but they certainly seem to fulfil a similar purpose. In some cases there are pairs with a modifying noun and a modifying adjective which may be nearly synonymous (*atom bomb*, *atomic bomb*; *language description*, *linguistic description*), while in other instances they contrast in meaning (*a civic centre* is not the same as *a town centre*). Speakers must know that it is *solar flare* but *sunspot*; there does not seem to be a way to predict such distinctions.

Expressions such as *attorney general*, *court martial*, *heir apparent*, where the adjective follows the noun it modifies, are usually of French origin, and follow the French order of noun and adjective. A few such as *postmaster general* are formed in English on a French pattern. The pattern does not seem to be productive, so in principle a full list of these can be given. There seems to be little reason to include such expressions among compounds, particularly since they are left-headed though most compounds are right-headed, but they are certainly MWEs.

3.3 Other word-classes +N

Examples of potential compounds formed with other word-classes in the modifying position are given in (4).

(4a) *spoilsport*, *dreadnought*
(4b) *call girl*, *show-room*
(4c) *uptown*, *downdraught*

(4d) *go-go dancer*, *pass-fail test*, *yes-no question*
(4e) *the … if-there's-any-sort-of-difficulty-ask-William-and-he'll-fix-it-for-you person*,[4] *our fear-of-terrorist-atrocity society*,[5] *after-tax profits*
(4f) *linesman*, *salesman*, *letters column*, *jobs programme*
(4g) *cat's-eye*, *women's magazine*

The examples in (4a) probably imitate a Romance pattern which is no longer productive in modern English. However, a similar type is found with the order of the elements reversed: *prick-tease*, for example. The type in (4b) has verbs in modifying position, but is endocentric (*show room* is a hyponym of *room*). It is often the case that modifying verbs in forestressed constructions take the *-ing* form: *dining room*, *shooting party*, *walking stick*. These are then usually considered to have nominal first elements. The type in (4c) shows adverbs/prepositions/particles in modifying position. Things like *through-put* may also belong here formally, though they are probably nominalisations of phrasal verbs. Reverse ordered forms like *put-down* are also found. The type in (4d) shows alternatives in modifying position, the alternatives being, in these instances, verbs or adverbs. The type in (4e) shows apparently unlimited syntactic constructions in initial position. These expressions do not have to be idiomatic or even familiar. If these are compounds, though, and most scholars accept that they are, they allow syntactic structure within word-structure. The types in (4f–g) have already been mentioned, with plurals or genitives in the first element. In both cases there is often an alternative with an unmarked noun in the first position (*lineman* and *linesman* are synonymous; according to the OED *tailor's tack* and *tailor tack* are synonymous). At the same time, a genitive first element can contrast with an unmarked first element, as illustrated in (5) (data from the OED).

(5)

dog-tooth 'check pattern'	*dog's tooth* 'architectural feature'
dog show 'event'	*dog's show* (Aust) 'no chance'
dog collar 'clerical garb'	*dog's collar* 'collar of a dog'
duck-foot 'having webbed feet'	*duck's foot* 'plant sp.'

4 Meynell, Lawrence (1978): Papersnake. London: Macmillan, 10.
5 Francis, Dick (2006): Under orders. London: Michael Joseph, 87.

3.4 Adjectival compounds

Adjectival compounds are common, with examples like *crime-prone*, *grass-green*, *sky-blue*, *word-final*, *work-shy*, and coordinative compounds are also found: *philosophical-historic*, *spicy-mild*. It can be argued (Bell 2014) that there are a number of exocentric compound adjectives in English which are exocentric by virtue of not containing an adjectival head: words like *day-to-day*, *fly-by-wire*, *overhead*, *through and through*, *pass-fail*. Some of these may look more like non-compound MWEs, but recall the definition given in Section 1 above, that compounds are 'words that are composed of two (or more) bases', and it can be seen that all of these fit the definition. If this just indicates that the definition is incomplete, then that is part of the message of this contribution. It must be noted, though, that corresponding structures in related languages would not be considered adjectives, and the question of their status arises peculiarly in English.

There is a set of adjectives which appears to arise from the participle-form of phrasal verbs: *down-sized*, *up-graded*, *out-grown*. Whether these are viewed as compounds may well depend on whether phrasal verbs are viewed as compounds (see below). They have a form made up of two bases, but those two bases are not independent at the point of adjective-formation. The same point can be made with relation to the corresponding denominal forms like *black-hearted*, *green-eyed*, which are not strictly formed as compounds in English, since their structure is [[black heart]ed], so that they are derivatives based on phrasal structures.

3.5 Verbal compounds

Verbal compounds are something of a discussion point in English word-formation, following Marchand's (1969: 100) definitive declaration that "[v]erbal composition does not exist in Present-Day English". The point is that many of the things that look like compounds, and that we might want to term compounds, are actually formed by back-formation (*to baby-sit*, *to horror strike*) or conversion (*to breath test*, *to cold shoulder*). The argument that these are not compounds follows the pattern of the argument on *hard-hearted* in the last section. Nevertheless, it is clear that there is an increasing number of genuine verbal compounds which are not formed by these means (Bauer/Renouf 2001; Bauer 2017). Recent examples are *air-quote*, *dry-burn*, and coordinative examples like *to blow dry*, *to stir-fry* (these are controversial examples of coordinative compounds, though some authorities included them).

English does have a number of V+V constructions which might be viewed as compounds or as serial verbs (and, if the latter, probably of syntactic not lexical

origin). These are most commonly found with verbs of motion as the first verb (*go see*, *come buy*) but go beyond that (*I hope see you soon*), especially in US English. Some such constructions can be the base of further derivation, which seems to imply listedness, if not other features of words (consider *go-getter*, *jump-starter*).

3.6 Compounds in minor word-classes

Whether there are compound prepositions is a matter of definition. Things like *into*, *onto*, *throughout* are written as one word, and are probably instances of frozen syntax. Instances like *away from*, *because of*, *except for*, *off of* (esp. US English), *out of* are certainly common collocations in text, but whether they are compounds or not is not clear.

3.7 Binomials

Binomials are pairs of words linked usually by *and*, occasionally by *or*. They are normally called binominals only if they are fixed collocations. Thus *Monday or Tuesday* would not be considered a binominal, but the examples in (6) would be.

(6) *Abbot and Costello*, *bacon and eggs*, *bread and butter*, *cat and mouse*, *chalk and cheese*, *fish and chips*, *gin and it*, *kit and caboodle*, *kith and kin*, *life or death*, *milk and honey*, *salt and pepper*, *slap and tickle*, *sun and sand*, *whisky and soda*; *do or die*, *kiss and tell*, *make or break*, *put up or shut up*, *wine and dine*; *black and blue*, *free and easy*, *neat and tidy*, *sick and tired*, *spick and span*; *as and when*, *back and forth*, *far and away* 'by a wide margin', *far and near* 'everywhere', *now and again*

There is quite a large literature on the order of the elements in binomials (for a good summary cf. Benor/Levy 2006), and the fixedness of the order. Binomials vary in the degree to which each element presupposes the other. In *spick and span* we cannot have either element without the other; *black* can easily occur without *blue*, but *black and blue* is a fixed expression whose implications go beyond the colours involved; *chalk and cheese* collocate only when illustrating how different two things can be; *Abbott and Costello* illustrates a collocation which was originally purely arbitrary, but became more fixed as the team became more established. They also differ in how easily they can be interrupted: *bread and manuka honey* is perfectly possible, but *sick and really tired* is no longer an example of the relevant collocation. Again, they differ in how easily the coordi-

nated items can be reversed. *Eggs and bacon* or *bacon and eggs* seem to be equally good (and *scrambled* can be added to *eggs* in either ordering), *jam and bread* is possible, if slightly unusual (it is found in a song in *The Sound of Music*, for instance), *chips and fish* is mainly used when the chips and the fish are referred to separately rather than as a single dish. Those binomials that have a figurative reading cannot in general be interrupted or reversed: *bread and butter* 'main source of income', *salt and pepper* 'colour term', *far and away*.

3.8 N+P+N constructions

N+P+N constructions like *lady-in-waiting* are frequently established MWEs, even though there are many N+P+N constructions which appear to be perfectly freely syntactic, as in *piece of cheese*. The problem of description is exacerbated in comparison with a language like French, where N+P+N constructions are often the translational equivalent of Germanic compounds. For instance, French *chemin-de-fer*, lit. way of iron, 'railway' is equivalent to Danish *jernbane*, lit. iron way, 'railway' (compare also German, Italian and other European languages), and French *jus de fruits* 'juice of fruits' is equivalent to English *fruit juice*. The French expressions are sometimes called 'compounds' (Spence 1969 calls them 'prepositional compounds'), while an opposing view sees them as syntactic constructions that may become fixed (Bauer 2001). The English construction is not as widespread as the French one is (because English has more compounds), but there are plenty of examples (cf. (7)).

(7) *lady-in-waiting*, *line-of-sight*, *man-about-town*, *man-at-arms*, *man-of-war*, *mother-of-pearl*, *pay-per-view*, *sense of humour*, *son-in-law*, *stock-in-trade*, *trial by jury*

Part of the question here (and, incidentally, also in French) is the status of items with internal determiners, such as those in (8). Are they a different construction by virtue of having an NP (or DP) in second position, or are they a variant of the same construction?

(8) *belle of the ball*, *birds of a feather*, *two bites of the cherry*, *a Jack of all trades*, *the man in the moon*, *the man of the moment*, *a pain in the neck*, *the time of your life*, *will of the wisp*

We might also ask whether toponyms such as *Burton-in-Lonsdale*, *Gatehouse-of-Fleet*, *Moreton-in-Marsh*, *Newcastle-under-Lyme*, *Newcastle-upon-Tyne*, *Walton-*

on-Thames, *Weston-super-Mare* are part of the same construction type. Again, there is a variant with determiners: *Stow-on-the-Wold*, *Stanford-in-the-Vale*, *Widecombe-in-the-Moor*.

To the extent that DPs can form part of the construction, these forms look more syntactic. But even then, we do not appear to find random DPs: adjectival modification within that DP does not appear to occur in established constructions of this form, though forms like *cat-in-the-new-moon* or *Marton-in-the-Blue-Mountains* might appear to be possible. On the other hand, with non-established examples, such as *by the light of the new moon*, there is no problem with adjectival modification. A fortiori, post-nominal modification does not occur in established examples.

Klinge (2005: 366) claims that only the preposition *of* is particularly productive in such phrases. He uses this as an argument for the lexical nature of these constructions. This is hard to establish, since other prepositions are clearly in use in the more syntactic phrases, and it seems unlikely that the rules of production for the more syntactic and more lexical types are completely independent. A more likely explanation is that only relatively non-specific forms are frequent enough to become established in usage, and that *of* is the most frequent preposition.

Overall, the descriptive problem here seems to be similar to the descriptive problem with genitive first elements: the formal description of the construction includes expressions which are clearly listed (sometimes idiomatic) and others which appear to be produced productively, possibly by syntactic rules. Perhaps equivalently, this means that some such expressions are more word-like than others.

3.9 Phrasal verbs

Phrasal verbs are usually taken to be syntactic units in English, though many of them are figurative or idiomatic. *Look up* is literal when it means 'raise your eyes towards the sky', but idiomatic when it means 'refer to' (as in *look up a word in the dictionary*) or even 'improve' as in *business is looking up*. *Put up* is literal in *put your hand up the pipe*, figurative in *to put someone's back up* ('annoy') and idiomatic in *I can put you up in our spare room* ('accommodate'). Note that some phrasal verbs have two particles, as *put up with* 'tolerate', *look up to* 'admire', but this construction too can be literal, as in *fall out of*. Phrasal verbs have syntax-like behaviour in being interrupted by their direct objects, but are lexical to the extent that their meaning is not predictable from their elements.

3.10 Phrases as words

It might be claimed that some of the items mentioned above are simply syntactic phrases that have become more word-like, by a process usually called univerbation. Since univerbation is a diachronic process that proceeds by degrees, and since there are a number of different univerbation processes, there are many different kinds of expression which, even if they started out containing two or more words, are currently considered to be single words. Some examples are given in (9).

(9) *altogether*, *attorney general*, *bullseye*, *dyed-in-the-wool*, *forget-me-not*, *thank you*, *touch and go*, *wannabe*

Because these fit the rough definition of a compound given in Section 1, they are sometimes considered to be compounds. To the extent that the constituent words are transparent, they might be considered to be MWEs. (Note that *bullseye* fits into the type illustrated in (4g) except that it is written as a single word and the genitive is not overtly marked.) They might also be considered to be single unanalysable words, as is implied in the term univerbation. Such items span the borders of MWEs.

4 Functional categories

The last section looked at categories that are more or less formally defined; in this section other types of category are considered, including formation-types that lead to MWEs. These are grouped together as 'functional' categories, in the sense that they are not formal, but they are nonetheless a heterogeneous group. In particular, the first section below scarcely seems to be a category at all, but contrasts with other categories discussed later.

4.1 Literal interpretation

It may seem trivial that literal interpretations of such constructions exist. For example, *Kim is good at music and maths* contains a N+*and*+N construction whose interpretation follows from the construction in which the coordinated pair occurs and the meanings of the words involved. Such examples are typically non-word-like. Discussion of such types is frequently carried out under the heading of

'semantic compositionality' or 'semantic transparency', which may or may not be equivalent. It is clear that semantic transparency is a matter of degree rather than a matter of yes/no; it is less clear – despite a large literature – just what is compositional (cf. Wisniewski/Wu 2012 for a useful discussion). It must be made explicit, however, that even listed items may appear perfectly transparent. Consider such examples as *copper wire*, *singer-songwriter*, *elm tree*, *whisky and soda*, *Burton-in-Lonsdale*. Whether that is sufficient to make them compositional is partly a matter of definition. Some of these show some evidence of word-like behaviour: for instance, *whisky and soda* is not reversible to *soda and whisky*, *singer-songwriter* is not easily interrupted to give, for instance, *singer-sad song writer*, *singer-incompetent songwriter*.

4.2 Figurative interpretation

An expression may also be interpreted figuratively. This is not the place to discuss the various possible figures of speech, or the distinctions between them. Suffice it to say that a figurative interpretation is a pragmatic interpretation based on the literal meaning, but providing an interpretation which is not literal. Consider the established metaphor *a dog's breakfast*. We could interpret that as 'a morning meal for a dog', that is literally, but its established meaning is 'a mess', and that involves pragmatically inferring that where a dog has eaten, things are not tidy. *A king's ransom* means 'a lot of money', which is pragmatically inferred from the amount that would be required to ransom a king. *To be on the ropes* is a metaphor from boxing and means 'to be in a desperate position'. As has been shown in a number of publications (e.g. Lakoff/Johnson 2003), figurative language is ubiquitous in everyday communication, and appears to be cognitively normal and effortless: indeed, it is often the sign of brain damage if a listener cannot interpret figures of speech.

4.3 Idiomaticity

Following Grant/Bauer (2004), a distinction is drawn here between figurative interpretation and idiomatic interpretation (called 'core idioms' by Grant/Bauer). On this reading, an idiomatic expression cannot be understood literally (it is not semantically transparent) nor in terms of the pragmatic inferences of figurative usage. The label is frequently used for a range of different structures, including examples like *red herring* 'misleading clue' (once figurative, but the figure is not recuperable in the current state of the language), *kick the bucket* 'die', *chew the fat*

'hold a conversation', *not by a long chalk* 'fall far short', *be in fine fettle* 'be fit and healthy' (*fettle* is now extremely rare except in this phrase). The important point about this, though, is that expressions of all kinds can be idiomatic, including compounds (consider *blackmail*, *yellowhammer* 'bird sp.') and phrasal verbs (consider *put up with* 'tolerate', *pan out* 'conclude' – perhaps once figurative, but not now recuperable).

A different type of idiom is the constructional idiom, a syntactic construction where the idiomatic semantics is provided by the construction, and the construction may be filled with varied lexical content (Booij 2002). An example from English is found in (10) (cf. also Philip 2008), where all the examples mean 'not to be particularly intelligent'.

(10) *to be a couple of sandwiches short of a picnic*
to be a couple of shrimps short of a barbie
to be two pennies short of the full shilling
to be several cards short of a full deck (with a variant, *not to be playing with a full deck*)
to be a few French fries short of a Happy Meal
to be a beer short of a six-pack
to be a few cakes short of a birthday party
to be a couple of bricks short of a wall

Another type of idiomaticity may be culture-bounded idiomaticity. Svensson (2008) considers this, looking at what she terms 'encyclopedic (non)compositionality', which she illustrates with expressions such as *The White House* and *to expect a baby*, which may be understood literally but which have much greater implications in our society (cf. also Sabban 2008). Examples like this show that the line between literal/transparent and non-compositionality/transparency may be more awkward than is often assumed, but also that the line between figurative and idiomatic is not necessarily easy to perceive.

4.4 Quotations, proverbs and the like

Any language will have a large number of recognised expressions which, in some way, acknowledge the wisdom of past speakers of the language. Some of these are quotations (from traditional tales, from literary works, from songs, movies or TV shows, from religious sources) others are proverbial or even family sayings. Their length and structure is infinitely variable: in principle, an actor or literary scholar might know the whole of *Hamlet* by heart and quote from it freely. Quotations are

often abbreviated, mis-quoted or even alluded to. The proverb *Too many cooks spoil the broth* may be shortened, as perhaps *It's a case of too many cooks*, or, if someone was complaining about the number of people involved in a project, someone else might conceivably ask, *So how did the broth turn out?* Quotations may often go unrecognised by hearers. Some examples are given in (11).

(11) *eye of the needle*, *fisher of men*, *the salt of the earth* (Biblical); *the goose that lays the golden egg*, *the grand old duke of York*, *white rabbits* (said on the first of the month) (folklore); *this sceptered isle*, *pound of flesh*, *star-crossed lovers*, *strange bedfellows* (Shakespeare); *dim*, *religious light*, *a modest proposal*, *a truth universally acknowledged* (other literary sources); *the curate's egg*, *famous last words*, *lies*, *damned lies and statistics* (non-literary sources); *the early bird*, *a gift horse*, *a watched pot* (proverbial)

Also included here are established similes like those in (12).

(12) *bald as a coot*
black as coal/ink/jet/night
bold as brass
clean as a whistle
cool as a cucumber (*cool* here means 'unruffled')
daft as a brush
pure as driven snow
thick as two short planks (*thick* here means 'stupid')
white as milk/snow

4.5 Abbreviations

Initialisms and acronyms deserve a marginal place in this discussion, as they are a means by which MWEs turn into single words. In initialisms, an MWE becomes an orthographic word: *FBI* is a single orthographic entity, while its origin, *Federal Bureau of Investigation* is an MWE. In acronyms, the MWE turns into a new phonological and orthographic word: the MWE *North Atlantic Treaty Organization* turns into *NATO* (/neɪtəʊ/). Although there is a rather old-fashioned spelling convention whereby some of these items may have their individual letters interrupted by full stops/periods (*N.A.T.O.*), the more modern orthography stresses the wordhood of the outcome. For the most successful acronyms, the original MWE becomes lost, and a new morpheme arises: *scuba* < *self-contained underwater breathing apparatus*.

Blends may be seen as a cross between compounds and abbreviations. In a blend, typically, the first part of the first word and the last part of the second word are telescoped together with some loss of phonological material. An example is *infotainment* < *information* + *entertainment* or *administrivia* < *administration* + *trivia*. Because blends can be seen as a type of compound, they are MWEs.

4.6 Rhyming slang

The essence of rhyming slang is that a word is replaced with a (usually two- or three-word) phrase which rhymes with the original. In this first stage, non-MWEs are deliberately replaced by MWEs. The word *kids* is replaced by *dustbin lids*, the word *stairs* is replaced with *apples and pears*. Note that there is no semantic link between the original word and the rhyming replacement, though occasional examples may be (or may be thought to be) jocularly appropriate, such as *trouble and strife* for *wife*. To make things more difficult, the rhyming word is then often deleted, so that *kids* becomes *dustbins* and *stairs* becomes *apples* and what was an MWE is now replaced by a polysemous lexeme. Although this is often termed 'Cockney rhyming slang' it is not restricted to London English. Not only is it also found, for instance, in Glasgow, Australia and New Zealand, but occasional expressions of rhyming slang creep unacknowledged in the vocabulary of the wider language community: *to do bird* (*bird lime* = *time* [in prison]), *let's have a butcher's* (*butcher's hook* = *look*), *my old china* (*china plate* = *mate*), *use your loaf* (*loaf of bread* = *head*), *rabbit on* (*rabbit and pork* = *talk*). All of these retain the distinctly informal style level of the originals, and form new idiomatic MWEs.

All the examples provided above are established examples. But rhyming slang can also be used productively. One website cites *Jar Jar Binks* for *forty winks* ('a snooze'), clearly postdating the relevant Star Wars movie, and not necessarily widely known.

4.7 Collocation

Collocations are sets of words which habitually occur together, even if they are perfectly transparent. A standard example concerns the way in which *dry* changes its meaning depending upon what it collocates with, as shown in (13).

(13) *a dry cough* (not producing catarrh)
a dry lecture (not interesting)

a dry state (where alcohol is not sold)
a dry wall (built without cement)
a dry wine (not sweet)
a dry wit (dead-pan)
dry eyes (without tears)
dry ground (not wet)
dry toast (not buttered)
dry weather (not raining)

Collocations are not always of the same strength. Sometimes the ability to predict one of the items in the collocation from the other is strong, sometimes it is weak. This can be measured in terms of the mutual information each element provides as to the identity of the other element(s) in the collocation (Xiao 2015). This may complicate the process of deciding what belongs in the lexicon in a theoretical sense, but does not interfere with the notion that more than just the individual word might have to be listed.

Note that while *dry* in *dry ground* can be submodified (*very dry ground*), and many of these expressions can be interrupted (*a dry French wine*, *dry red-rimmed eyes*) some of them seem to be more word-like (**very dry toast*, *dry battery* does not appear to allow random insertions).

A particular kind of collocation is that provided by light verbs. It is *make a difference*, *give a lecture*, *make a mistake*, *take the opportunity*, *take a shower*, *have a smoke*. There does not seem to be any straightforward semantic reason for the selection of these light verbs, and speakers (including native speakers) will often use a different one from the one expected, and say things like *do a mistake*.

Another similar case is provided by adjectives that take complements, and then collocate with fixed prepositions, as in *afraid of*, *averse to*, *different from/than/to*, *proud of*. The case of *different*, which becomes a matter of prescription, shows that the preposition is not always fixed, but generally speaking the preposition has to be seen as being chosen by the head adjective. This puts such constructions of the borderline between being lexical combinations and syntactic structures showing government.

4.8 Formulae

Formulae are the way things are said rather than the way they could be said (cf. also Sabban 2008). In many European languages, there is an expression which can be translated as 'good day' which is a greeting. In England, *good day* is a

farewell. In Australia and New Zealand, *good day* (with a phonetically very much reduced first syllable) is again a greeting. In the usage of young New Zealanders around the turn of the millennium, *spot you later*, and *laters* were farewells (Bauer/Bauer 2003). The fact that these are greetings and farewells (as opposed to other potential expressions which are not, such as *until we meet again*, *till the next time* or *soon*), with the corresponding increase in usage of these precise phrases, makes them into formulae. Corresponding to the rather old-fashioned *How do you do?* heard in England, *How are you doing?* can be heard in other parts of the English-speaking world, but as a day-to-day greeting rather than as a greeting on first introduction. *How is it going?* is an alternative possibility, but not *How does it go?* There are many perfectly grammatical possible ways of saying things that are never used, and those that are used, and their precise meaning, may be unexpected.

Formulae, then, are particular types of collocation, with high frequency in particular social environments. While they have syntactic structure (in the case of *How do you do* a rather outmoded syntactic structure), some of them may be learned as listed, fixed expressions, or have the status of words (as with *good-bye*).

4.9 Lexicalisation

Lexicalisation is the process of becoming a lexical item. It depends on semantic shift (often called idiomatisation, e.g. Lipka 1994) and formal change. Although it may be difficult or impossible to measure degrees of lexicalisation, it is a matter of more or less not either/or. At the one end, the most lexicalised items like *lord* are historically derived from elements meaning 'loaf ward', and all internal structure and the meaning of the original elements has been lost. At the other end, we have freely produced syntactic constructions which are perfectly transparent in form and meaning. The terminology of lexicalisation is very variable, and various intermediate stages have been postulated (cf., e.g., Bauer 1983a). In formal terms, we find constructions whose elements are transparent, instances where the elements have undergone some phonetic erosion (e.g. *Christmas* which phonologically contains neither *Christ* nor *mass* any more), to constructions whose elements probably cannot be perceived without formal instruction (such as *dearth*, related to *dear*). Semantically, transparent elements may have to be interpreted figuratively (e.g. *hedgehog* or *fire dog*), or, even if appearing formally transparent, be semantically totally opaque (such as *blackmail* and *woodchuck*). It will be clear from these examples that various factors influence lexicalisation, but many MWEs are, almost by definition, somewhere on the lexicalisation spectrum.

5 Discussion

While this wide range of MWEs has to be recognised (however difficult they may be to systematise), there are a number of expressions which do not appear to be sufficiently lexical to fit in the category. Any study of n-grams will come up with expressions like *in a*, which collocate not because *in a* is constituent with its own meaning, but because all members of the category preposition are typically followed by determiner phrases, typically headed by determiners like *a* in initial position. The high number of such cross-constituent collocations has thus more to do with the productivity of syntax than anything lexical. Similarly, colligations, such as the fact that the verb *construct* is transitive, is not a matter of lexis but a matter of grammar (again, perhaps, a matter of government). It is true that *construct a building* is likely to be more frequent than *construct a daisy*, but this has as much to do with the nature of the world as with the nature of lexical items. While it makes little sense to suggest that *construct* demands in its complement something with a feature [+ constructible], as has been done on occasions, it makes rather more sense to say that pragmatically the need for a sentence which contains *construct a daisy* is likely to be extremely low (although it might be possible if people were decorating a kindergarten and making flowers out of recycled material to use as decorations). As McCawley remarked many years ago (McCawley 1971), if someone says *my toothbrush is pregnant*, it is unlikely to be their grammatical competence which is at fault.

The borderline between things which happen to collocate because they are syntactically likely to arise in similar contexts and what is lexical is not necessarily an easy one to draw. I tend to think that *is like a* is on the grammatical side, but Wikberg (2008: 136f.) makes a case for it on the basis that it is a formula used to introduce similes.

Borderlines like these, and one mentioned earlier between government and lexical structure, are potentially problematic, and the entire idea that there are such borderlines is worthy of further discussion. At one extreme we find a view, which we can characterise as essentially Chomskian, that virtually everything we produce is the result of free syntactic rules in operation. The other extreme position, and one worth arguing for, would be that there is no such thing as free syntax, but that everything is lexically-driven, with MWEs, fixed phrases and strongly restrictive constructions accounting for the fact that speakers do not say many things which might appear to be grammatical. I distrust extreme views, and suspect that there is some of each involved, but that the limits of each require careful motivation. It seems to me that a line like Carroll's (1871) '*Twas brillig and the slithy toves did gyre and gimble in the wabe* shows that there must be some syntax separate from vocabulary items, while the range of MWEs discussed in the phra-

seological and constructional literature shows that much of what we say on a day-to-day basis requires minimal independent syntax to be formed into perfectly normal conversational turns. In saying that, I imply consciously that there may be a difference between written and spoken language in this regard. All of these are open questions.

Two questions have been ignored in this presentation. The first is frequency. It might seem that MWEs must be frequent enough to be recognised by speakers, but there are many constructions that are invented on the spur of the moment and yet fit (at least some of) the criteria for recognising MWEs. Consider, for instance, examples in (4e) and (10). Frequency is a correlate of lexicalisation, but low frequency does not prevent something from being an MWE.

The second point to be considered is speaker accuracy. As was pointed out in relation to light verbs, speakers are not always consistent in what they say, and what start out as errors may spread and cause language change. This seems to go beyond performance errors in the sense of Chomsky (1965). Listening to current spoken English suggests that there is huge variation in complementation patterns at the moment, something else that lies on the borderline between government and lexical collocation (if these can be fully distinguished).

In this contribution, I have presented a sketch of some of the types of MWE that can be found in English. The classification I have used is, however, not exhaustive, and the various categories I have used are not mutually exclusive, so that I consider the classification used here to be no more than an ad hoc framework for discussion and not a typology. Various alternative classifications are provided in Granger/Meunier (eds.) (2008), but while I see the value of these classifications, I do not think we are yet at a point where a typology of MWEs is possible. Partly, as I have tried to suggest above, this is because the very nature of MWEs is pluricentric. There is no simple distinction between lexical and syntactic, there is no simple distinction between compositional and non-compositional or between lexicalised and non-lexicalised. Rather there is a host of expressions which link to syntactic structure and to semantic structure (and, indeed, even to phonological structure, although I have not discussed matters such as alliteration and rhyme here) in multiple ways. Compounds are one type of MWE, which may not easily be distinguished from other MWEs, because they are part of the network and give rise to the same problems of description and interpretation that other MWEs do.

That brings us back to the starting point of this contribution. It is hard to define compounds because they overlap with other MWEs in sharing features of wordhood, they overlap with syntax in that some things which have been called compounds are viewed by others as syntactic, because some of them, at least, are semantically transparent, and because some of the things that some scholars call

compounds arise from pieces of syntactic structure being frozen. While anyone is free to define compounds as they see fit, agreement on any definition which can determine which of the structures that have been canvassed here are really compounds seems a long way off.

References

Adams, Valerie (2001): Complex words in English. Harlow: Pearson.
Bauer, Laurie (1983a): English word-formation. Cambridge, UK: Cambridge University Press.
Bauer, Laurie (1983b): Stress in compounds. A rejoinder. In: English Studies 64. 47–53.
Bauer, Laurie (1998): When is a sequence of two nouns a compound in English? In: English language and linguistics 2. 65–86.
Bauer, Laurie (2000): Word. In: Booij, Geert/Lehmann, Christian/Mugdan, Joachim (eds.): Morphology. An international handbook on inflection and word-formation. Berlin/New York: De Gruyter. 247–257.
Bauer, Laurie (2001): Compounding. In: Haspelmath, Martin et al. (eds.): Language universals and language typology. Berlin/New York: De Gruyter. 695–707.
Bauer, Laurie (2004): Adjectives, compounds and words. In: Nordic Journal of English Studies 3, 1. 7–22.
Bauer, Laurie (2016): Re-evaluating exocentricity in word-formation. In: Siddiqi, Daniel/Harley, Heidi (eds.): Morphological metatheory. Amsterdam/Philadelphia: Benjamins. 461–477.
Bauer, Laurie (2017): Compounds and compounding. Cambridge, UK: Cambridge University Press.
Bauer, Laurie (to appear): Stressing about the news. In: New Zealand English Journal.
Bauer, Laurie/Bauer, Winifred (2003): Playground talk. Wellington: Victoria University.
Bauer, Laurie/Lieber, Rochelle/Plag, Ingo (2013): The Oxford reference guide to English morphology. Oxford: Oxford University Press.
Bauer, Laurie/Renouf, Antoinette (2001): A corpus-based study of compounding in English. In: Journal of English Linguistics 29. 101–123.
Bell, Melanie J. (2014): The English noun-noun construct: a morphological and syntactic object. In: Ralli, Angela (eds.): Morphology and the architecture of grammar. 59–91. (Internet: https://geertbooij.files.wordpress.com/2014/02/mmm8_proceedings.pdf, last access: 20.4.2018).
Benor, Sarah/Levy, Roger (2006): The chicken or the egg? A probabilistic analysis of English binomials. In: Language 82. 233–278.
Booij, Geert (2002): Constructional idioms, morphology and the Dutch lexicon. In: Journal of Germanic linguistics 14. 301–329.
Carroll, Lewis (1871): Through the looking glass and what Alice found there. London: Macmillan.
Carstairs-McCarthy, Andrew (2002): An introduction to English morphology. Edinburgh: Edinburgh University Press.
Chomsky, Noam (1965): Aspects of the theory of syntax. Cambridge, MA: MIT Press.
Chomsky, Noam/Halle, Morris (1968): The sound pattern of English. New York: Harper & Row.

Dixon, R.M.W./Aikhenvald, Alexandra Y. (2002): Word: a typological framework. In: Dixon, R.M.W./Aikhenvald, Alexandra Y. (eds.): Word. A cross-linguistic typology. Cambridge, UK: Cambridge University Press. 1–41.
Giegerich, Heinz J. (2004): Compound or phrase? English noun-plus-noun constructions and the stress criterion. In: English Language and Linguistics 8. 1–24.
Giegerich, Heinz J. (2015): Lexical structures. Edinburgh: Edinburgh University Press.
Granger, Sylviane/Meunier, Fanny (eds.) (2008): Phraseology. An interdisciplinary perspective. Amsterdam: Benjamins.
Grant, Lynn/Bauer, Laurie (2004): Criteria for re-defining idioms: Are we barking up the wrong tree? In: Applied Linguistics 25. 38–61.
Hall, Robert A. Jr (1964): Introductory linguistics. Philadelphia/New York: Blackwell.
Hippisley, Andrew (2015): The word as a universal category. In: Taylor, John R. (ed.). 246–269.
Klinge, Alex (2005): The structure of English nominals. Copenhagen: Copenhagen Business School. [Doctoral dissertation].
Kunter, Gero (2011): Compound stress in English. (= Linguistische Arbeiten 539). Tübingen: Niemeyer.
Lakoff, George/Johnson, Mark (2003): Metaphors we live by. Chicago: University of Chicago Press.
Lieber, Rochelle (2010): Introducing morphology. Cambridge, UK: Cambridge University Press.
Lipka, Leonhard (1994): Lexicalization and institutionalization. In: Asher, R.E. (ed.): Encyclopedia of language and linguistics. Vol 4. Oxford: Pergamon. 2164–2167.
Marchand, Hans (1969): The categories and types of present-day English word-formation. 2nd ed. Munich: Beck.
McCawley, James D. (1971): Where do noun phrases come from? In: Steinberg, Danny D./Jakobovits, Leon A. (eds.): Semantics. Cambridge, UK: Cambridge University Press. 217–231.
McEnery, Tony/Xiao, Richard/Tono, Yukio (2006): Corpus-based language studies. London/New York: Routledge.
Moon, Rosamund (2015): Multi-word items. In: Taylor, John R. (ed.). 120–140.
Payne, John/Huddleston, Rodney (2002): Nouns and noun phrases. In: Huddleston, Rodney/Pullum, Geoffrey K. (eds.): The Cambridge grammar of the English language. Cambridge, UK: Cambridge University Press. 323–524.
Philip, Gill (2008): Reassessing the canon: 'Fixed' phrases in general reference corpora. In: Granger/Meunier (eds.). 95–108.
Pinker, Steven (1999): Words and rules. London: Phoenix.
Rosenbach, Anette (2006): Descriptive genitives in English: a case study on constructional gradience. In: English language and linguistics 10. 77–118.
Sabban, Annette (2008): Critical observations on the culture-boundness of phraseology. In: Granger/Meunier (eds.). 229–241.
Spence, Nicol C.W. (1969): Composé nominal, locution et syntagme libre. In: Linguistique 2. 5–26.
Svensson, Maria Helena (2008): A very complex criterion of fixedness: Non-compositionality. In: Granger/Meunier (eds.). 81–93.
Taylor, John R. (ed.) (2015): The Oxford handbook of the word. Oxford: Oxford University Press.
Wikberg, Kay (2008): Phrasal similes in the BNC. In: Granger/Meunier (eds.). 127–142.
Wisniewski, Edward J./Wu, Jing (2012): Emergency!!!! Challenges to compositional understanding of noun-noun combinations. In: Hinzen, Wolfram/Machery, Edouard/

Werning, Markus (eds.): The Oxford handbook of compositionality. Oxford: Oxford University Press. 403–417.
Wray, Alison (2015): Why are we so sure we know what a word is? In: Taylor, John R. (ed.). 725–750.
Xiao, Richard (2015): Collocation. In: Biber, Douglas/Reppen, Randi (eds.): The Cambridge handbook of English corpus linguistics. Cambridge, UK: Cambridge University Press. 106–124.

Barbara Schlücker

Compounds and multi-word expressions in German

1 Introduction

This chapter reviews multi-word expressions, compounds, and their mutual relation regarding their status in grammar and lexicon in contemporary German.[1] Both multi-word expressions and compounds are lexical units and morphosyntactically complex. That is, they are made up of a minimum of two words or stems,[2] which sets them apart both from simplex lexemes and from morphologically complex words derived by other word-formation processes, in particular derivation and conversion.[3] As lexical units, they have the common function of providing labels for all kinds of concepts. This apparent similarity – which becomes immediately obvious from the existence of parallel units such as *Frischluft* / *frische Luft* 'fresh air' – raises various questions concerning the status, the function, and the division of labor between multi-word expressions (henceforth: MWEs) and compounds, but also regarding the identification and demarcation of these forms. These questions will be discussed in this chapter. To start with, it has been noted time and again that the dividing line between MWEs and compounds cannot always be clearly drawn. While many of the problems that are discussed in the following – such as the theoretical considerations concerning MWE formation and the status of MWEs and compounds in the mental lexicon – have cross-linguistic implications, the question of identification and demarcation of the forms is language-specific. Therefore, we will start our overview with a brief survey of the relevant properties in German. The chapter is organized as follows: Section 2 defines the central terms in the context of the object of investi-

1 I would like to thank Geert Booij, Jesús Fernández, Rita Finkbeiner, and Katerina Stathi for very valuable comments on earlier versions of this chapter.

2 Although the notion of word is known to be notoriously problematic, it is used in most definitions of multi-word expressions, relying (usually without further discussion) on orthography as the defining criterion. In addition, one also finds other (unspecified) terms such as 'element' (Gries 2008). The term 'stem' is mentioned here because stems rather than words form the basic constituents in compounds.

3 Strictly speaking, conversions, although derived by a morphological process, are not morphologically complex.

https://doi.org/10.1515/9783110632446-003.

gation of the study, in particular the scope of the units known as MWEs. This section covers general aspects such as the relation between morphology and the lexicon, as well as MWE formation, the proportion of compounds and MWEs in the German lexicon, and the relation between both processes with respect to their function as providing lexical units. Section 3 gives a more detailed overview of German MWEs and compounds classified according to lexical category. Section 4 discusses the theoretical implications of the findings. The chapter ends with a brief conclusion in Section 5.

2 General aspects

2.1 Identifying compounds and MWEs in German

In his chapter "Idioms and other fixed expressions: Parallels between idioms and compounds", Jackendoff (1997a: 164) writes:

> Another part of the goal is to show that the theory of fixed expressions is more or less coextensive with the theory of words. Toward this end, it is useful to compare fixed expressions with derivational morphology, especially compounds, which everyone acknowledges to be lexical items.

The main reason for investigating MWEs and compounds and their interrelation is the fact that they are quite similar with respect to (i) their status as lexical units and their function of providing labels for concepts and (ii) their form, as both are morphosyntactically complex, i.e. consisting of a minimum of two words or stems. What follows from this first description is that if MWEs and compounds are similar in being both lexical units and consisting of two (or more) words/lexemes, the crucial difference lies in the way these words are combined. Gaeta/Ricca (2009) have made this point very clear, distinguishing strictly between the properties of being [± lexical] and [± morphological], where "lexical" means that a unit has a stable referent, a unitary meaning and possibly a non-negligible frequency of occurrence (ibid.: 39). While both MWEs and compounds are [+ lexical], compounds are [+ morphological] but MWEs are [– morphological]. This means that only lexical units can be regarded as compounds that are the output of the morphological operation of compounding, which in turn must clearly differ from the syntactic operations of the language in question. For this reason, we will start with a concise description of compounding.

In German, nominal and, to a more limited extent, adjectival compounding are productive word formation patterns, whereas verbal compounding is regarded

as either non-existent or highly restricted (e.g., Motsch 2004; Fleischer/Barz 2012). Compounding is generally right-headed. In direct comparison with parallel phrases, compounds can best be characterized by the following properties:

(i) Stress, which is on the left (modifier) constituent in compounds but on the head in phrases (*Fríschluft* – *frische Lúft* 'fresh air').
(ii) The stem form of the modifier, i.e. the absence of inflection (*FrischØluft* – *frische Luft*).
(iii) Inseparability, i.e. compounds cannot be interrupted by any intervening material which is perfectly possible for phrases (*frische*, *angenehme Luft* 'fresh pleasant air').
(iv) Linking elements, although they do not occur in all subkinds of compounds, for instance *Geigenbogen* 'violin bow'.[4]
(v) Spelling, as compounds are consistently written as one word (or are hyphenated), contrary to phrases.[5]

In addition, there are several properties that apply only to specific subtypes of compounding. To the extent that they are relevant to the present issue they will be discussed in Section 3.

The properties mentioned distinguish compounds not only from phrases but also from univerbations in the strict sense ("Zusammenrückung"), such as *zulasten* (lit. on burden of, 'account of'), *demzufolge* (lit. as a result of this, 'accordingly') or *Möchtegern* ('would-be, wannabe'). These lexical units are inseparable and written in one single word. They are, however, not the result of a word formation process but rather fossilized phrases. This can be seen in the fact that they can contain inflected material instead of stem forms, such as *lasten* (PL.DAT.) in *zulasten*, *dem* (DAT.SING.) in *demzufolge* or *möchte* (1./3.PERS.SING.PRES.ACT.) in *Möchtegern*. Also, they retain phrasal stress. Contrary to compounding, the formation of such units is unsystematic and cannot be predicted. Thus, they are lexical but not morphological units. Accordingly, if we rely on the properties of [± lexical] and [± morphological] only, univerbations are no different from MWEs (see below). However, due to their inseparability and solid spelling they are generally considered words.[6]

4 Linking elements in German are not inflectional elements although some of them have evolved diachronically from inflectional affixes, cf. footnote 6.

5 It can be observed that German language users sometimes write compounds as two separate words (cf. Scherer 2012, for instance) and it has been speculated that this might be an increasing tendendy due to influence from English. This breaks the official German spelling rules, however.

6 From a diachronic perspective, it can be seen that a particular type of univerbation forms a close link between MWEs and nominal compounds. In addition to compounds proper that can be

MWEs, according to this first sketch, are [+ lexical] and [– morphological] which means that they are formed syntactically. Following definitions of MWEs as advanced by Gries (2008) or Burger (2015), for instance, MWEs are characterized as syntactic patterns that consist of a minimum of two words (but not longer than a sentence), forming either a lexical or a grammatical pattern. They may but need not exhibit idiosyncratic semantic and/or syntactic properties, i.e., MWEs may but need not have a non-compositional meaning and the constituent parts may but need not be in a fixed order, immediately adjacent or syntactically deficient. For example, the MWEs in (1a) have a non-compositional meaning, but fully regular syntactic properties (that is, the VP *ein Fass aufmachen* can be inflected as with any other VP, and can be passivized or modified, e.g., *ein großes Fass aufmachen* (lit. to open a big barrel, 'make a big fuss'). The examples in (1b), on the other hand, have a fully compositional meaning. Finally, the examples in (1c) have a non-compositional meaning and they exhibit special syntactic properties, that is, the order of words is fixed and they cannot be separated, determiners are lacking and the adjective *gut* 'good' is uninflected (a historic relic) which is, according to present-day syntax, ungrammatical.

(1a) *ein Fass aufmachen* (lit. to open a barrel, 'make a fuss'),[7] *um ein Haar* (lit. by a hair, 'very nearly')
(1b) *Dank sagen* (lit. say thanks, 'thank'), *leere Menge* ('empty set'), *in Zusammenhang mit* ('in connection with')
(1c) *Knall auf Fall* (lit. bang on fall, 'suddenly'), *auf gut Glück* (lit. on good luck, 'on the off chance')

found since Old High German (and before), a second type of compounds, the so-called 'genitive compounds', or, in Grimm's terminology, "uneigentliche Komposita" ('false compounds') arise sporadically in Old High German and Middle High German times and become more frequent later on. They are univerbations of a prenominal genitive construction and, for this reason, contain genitive case marking. In Early High German, this pattern becomes productive and collapses with the older compound type. As a result, the former case markings are reanalyzed as linking elements, and the newly coined forms are no longer conceived of as univerbations, thus (former) syntactic patterns, but as word formation proper (cf. Pavlov 1983, for instance). For instance, the genitive construction *(des) menschen herz* ('(the) human's heart') is reanalyzed as a nominal compound *Menschenherz* ('human heart') and the former suffix *-en* (SING.GEN.) is reanalyzed as a linking element.

7 The German MWE is the result of folk etymology relating to the English verb *fuss*, due to the phonological similarity of English *fuss* and German *Fass* 'barrel' and the equivalence between German *(auf)machen* and English *make*.

As formal and semantic irregularities are not defining criteria of MWEs, their identification hinges crucially on (a) the function of the combination of words as a semantic unit and (b) the frequency of occurrence, which means that the frequency of occurrence of the particular combination of words is larger than expected.[8, 9]

This definition (and many similar approaches in the literature) have led to a rather broad view of MWEs that encompasses many different types of lexical phrasal units, some of which are not regarded as MWEs in older and more traditional phraseological theory. In particular, collocations which may have a fully compositional meaning, are nowadays usually regarded as MWEs, e.g., *billige Kopie* ('cheap copy'), *den Kopf schütteln* ('to shake one's head'), *eine Entscheidung treffen* (lit. to hit a decision, 'to make a decision'). Presumably, they make up a large part of all MWEs in German. Another group are partially fixed (or: lexically filled) patterns, that is, patterns that contain open slots that can be filled with various lexical items to produce new MWEs, cf. (2):

(2a) [X *um* X] 'X by X': *Stein um Stein* ('brick by brick'), *Jahr um Jahr* ('year by year')
(2b) [*Wer* X (*der*) Y] ('he who X, Y'): *Wer rastet*, *der rostet* ('He who rests, rusts'), *Wer suchet*, *der findet* ('He who seeks, finds'), *Wer schreibt*, *der bleibt* ('He who writes, remains')

The observation that some phrasal patterns are systematically and productively used to form lexical units can already be found in early traditional German phraseological research (cf. Häusermann 1977; Fleischer 1982). Quite influentially, the idea of productive syntactic patterns in the lexicon has been discussed in detail in cognitive and constructionalist frameworks (cf. Fillmore/Kay/O'Connor 1988; Jackendoff 1997b; Kay/Fillmore 1999, among many others), often in connection with the term 'constructional idiom' (Jackendoff 1997a; Jackendoff 2002; Booij 2002). Finally, and especially in connection with recent developments in usage-based and corpus linguistics and rapidly increasing corpus sizes, the idea emerged that the vast majority of MWEs are indeed realizations of abstract

8 The first criterion can serve to exclude frequently co-occuring sequences such as *and the*, which obviously do not form a semantic unit. Yet, it is not clear what exactly a semantic unit is; Gries (2008: 6), for instance, defines it as "to have a sense just like a single morpheme or word". This, however, seems too narrow given the meaning of proverbs or (some) verbal idioms such as *ein Fass aufmachen* / *make a fuss*.

9 Other properties which are in principle compatible with these properties make use of the psycholinguistic dimension, e.g., psycholinguistic stability or retrieval as a whole.

patterns, with numerous relations between these patterns (e.g. Steyer 2015, 2016). Crucially, the idea of abstract MWE patterns implies that there are also occasional MWEs, that is, nonce-MWEs that are formed ad hoc, and even potential MWEs that might be formed according to these patterns but have yet to do so, just as is the case with occasional and potential compounds. Obviously, such ideas challenge the original idea of MWEs as idiosyncratic stored items in the lexicon. We will come back to this issue in Section 4.

For the purpose of the present chapter, some constraints on the range of MWEs to be discussed are in order. First, MWEs are usually also thought to encompass proverbs, sayings, quotations, and routine formulas, e.g. *Good Morning* or *Happy Birthday*. However, these kinds of MWEs do not denote referents (either objects or events), but rather have a propositional function due their sentence character, or, in the case of routine formulas, a purely pragmatic (communicative) function. As the present discussion focuses on MWEs that parallel compounds, they are excluded in what follows. Similarly, as will be discussed in more detail in Section 3, we will not be concerned with MWE patterns that systematically lack equivalent compound forms.

2.2 Proportion of MWEs and compounds in the German lexicon

Given the potential functional overlap between MWEs and compounds, the questions of what share they hold in the (German) lexicon and whether any regularities can be observed concerning their distribution arise. Obviously, the answers are determined by various factors: first, they crucially hinge on the definition of MWEs and the question of which combinations are considered MWEs. Furthermore, we might ask how to deal with occasional and possible/potential formations, i.e. concrete patterns that might be instantiated from abstract MWE patterns.

Most remarks in the literature on the distribution of MWEs and compounds relate to lexical categories. It has often been assumed that verbal MWEs make up the largest part of German MWEs (e.g., Burger 2001: 34). Nominal MWEs are usually considered much less frequent (e.g., Barz 1996: 131, 2007: 28; Donalies 2008: 308). According to Fleischer (1996a: 152, 1997a: 17–20), MWEs are most frequent in the verbal and least frequent in the adjectival domain, with the nominal and the adverbial domain in between. Fleischer (1996b: 336) and Barz (2007: 28) relate this distribution to differences in productivity of compounding (or word-formation in general) in the respective lexical categories: whereas nominal compounding is highly productive in German, there are considerably less word-formation patterns in the verbal domain and verbal compounding in particular is consid-

ered marginal or non-existent. In addition, is has been observed that the distribution also depends on register: nominal MWEs seem to be much more frequent in terminology, e.g., medical language or professional titles, than in general usage (Möhn 1986; Fleischer 1996a; Barz 2007).

However, these assessments about the distributions among the lexical categories crucially depend on what counts as an MWE. "Classical" verbal idioms such as *jdn. auf die Palme bringen* (lit. bring so. on the palm, 'drive so. nuts') stand out for their semantic and morphosyntactic idiosyncrasies and are therefore more often perceived as MWEs. Collocations, on the other hand, in particular nominal ones, have not always been recognized as fixed units, as many of them have a fully compositional meaning. They have often not been included in dictionaries or phraseological lists. However, inclusion in such dictionaries or lists usually forms the basis for the sort of assessment mentioned above. Thus, given a broader view on MWEs like that introduced in the preceding section, it seems hard to say whether (or to what extent) a distribution of compounds and MWEs by lexical category can be established at all.

2.3 Relation between MWEs and compounds in the German lexicon: complementarity or competition?

An old and widespread idea about the lexicon is that it usually does not contain real synonyms or doublets which means that the co-existence of compounds and MWEs with identical meanings and grammatical function/distribution is not expected (for discussion cf. Haiman 1980, for instance). It has also been assumed that real doublets only exist between terminology and general vocabulary (Barz 1996: 132). However, this view is probably too strict. Obviously, there are also examples of "real" doublets within the general lexicon, some of the (often cited) examples being *Schwert des Damokles* / *Damoklesschwert* ('sword of Damocles'), *Grüntee* / *grüner Tee* ('green tea'), *schwarzer Markt* / *Schwarzmarkt* ('black market'), *halbherzig* / *mit halbem Herzen* ('half-hearted'), although in some cases there are clear differences in the frequency of use of both forms.[10] Also, there might be regional variation concerning the use of an MWE vs. compound.

As to the differences, MWEs are often assumed to be more expressive than parallel morphological units, e.g., *jdn. übers Ohr hauen* (lit. hit so. across the ear)

10 For instance, Schuster (2016: 195) shows that the distribution of *schwarzer Markt* vs. *Schwarzmarkt* ('black market') has changed considerably in the period of 1946–2009 [ZEIT corpus], with an initial proportion of the compound of about 10 % and 90 % at the end.

vs. *jdn. betrügen*, both meaning 'cheat so.'. Expressivity is often due to metaphorical meaning, e.g. *grüne Welle* (lit. green wave, 'phased traffic lights'), *blondes Gift* (lit. blond poison, 'blonde bombshell'), but it may also arise from phonological-prosodic properties, like rhyme or alliteration, as in binomial constructions such as *null und nichtig* ('null and void'), *hegen und pflegen* ('nurture') (cf. Fleischer 1997b: 164f.). However, although expressivity and imagery might be the initial driving forces for the coinage of an MWE, these properties might wear out over time and the forms are no longer perceived as particularly expressive (cf. Fleischer 1997a). Furthermore, compounds might also have a metaphorical meaning, such as *Dickmops* (lit. fat pug, 'fat person, fatty'), *Baumdiagramm* ('tree diagram'), *Kuchenhimmel* (lit. cake heaven, 'place that serves excellent cake').

The question of whether the relation between compounds and MWEs is to be characterized as complementary or competitive depends on the ideas about the status and the formation of MWEs. According to the traditional view, MWEs are not formed by abstract patterns (or rules) in the way compounds are. Rather, their emergence has been regarded as a secondary, purely semantic process of idiomatization (e.g. metaphoric or metonymic) of syntactic units, which might in turn have an effect on the morphosyntactic properties of the unit in question (e.g., Fleischer 1997a: 11; Barz 2007: 31). Barz (1996: 132, 2007: 30) regards MWEs as less economic than complex morphological units due to their complexity, i.e. the number of constituent parts, although they are often semantically more explicit since the relation between the constituents is morphosyntactically expressed, unlike with compounds. A typical example is an adjectival phrasal simile such as *so rot wie Blut* ('as red as blood') and the corresponding adjectival compound *blutrot* ('blood-red') (cf. also Section 3.4). The comparison between 'blood' and 'red' is expressed explicitly in the phrase while this relation is implicit in the compound and must be inferred by the reader. At the same time, the morphological counterpart is structurally less complex than the phrasal unit.

According to this view, MWE formation can be regarded as complementary to compounding and is employed if compounding is not available (cf. Section 2.2) or (at least in some cases) for the purposes of increasing expressivity (e.g., Fleischer 1997b). However, on a broader view on MWE formation that acknowledges – in addition to sporadic, secondary idiomatization of phrases – the (widespread) existence of more abstract MWE patterns, both with or without a compositional meaning, MWE formation is not complementary to compounding but rather competing or at least on an equal footing. If this is indeed the case, we ought to question whether more can be said about the distribution of MWEs and compounds in the lexicon than the preferences concerning lexical category. In other words: Are there more (or other) factors influencing or determining the choice between both patterns?

In recent years, several studies have approached this question for German with a focus on nominal units, both A+N and N+N. The study by Schlücker/Plag (2011) adopts an analogical approach, investigating the idea that the choice between MWEs and compounds depends on the individual lexemes involved. The study examines the formation of new A+N combinations. It shows that there are no general preferences for coining new A+N lexical units as either MWE or compound, but that the choice depends on the way the individual adjectives and nouns have been used before, i.e. either as a compound (e. g. *voll* ('full'): *Vollbart* 'full beard', *Vollmond* 'full moon') or an MWE (e. g. *offen* 'open', *offenes Geheimnis* 'open secret', *offenes Ohr* 'sympathetic ear') or both (e. g. *rot* ('red'): *Rotwein* 'red wine', *Rotkohl* 'red cabbage'; *rote Bete* 'beetroot', *rote Grütze* 'red fruit jelly').[11] Put simply, constituents that have previously been used in compounds tend to be realized as compounds when coining new combinations, and those that have previously been used in MWEs tend to be realized as MWEs. Thus, the choice between the forms is determined by the existence and number of related similar constructions in the mental lexicon of the language users. This analogical effect has been shown to be stronger for adjectives than for nouns.[12] There is also evidence for the co-existence of both patterns as well as for analogical effects from the diachronic perspective. Studying the diachronic development of German A+N sequences since 1700, Schuster (2016) shows that both patterns have continuously co-existed and that there is no clear trend towards either of the patterns or the disappearance of the other. Again, the choice for either an MWE or a compound seems to depend on individual adjectives. Thus, some adjectives consistently form A+N phrases whereas others always occur in compounds. A third group is productive in both patterns which also leads to the formation of doublets, e.g. *rotes Wild* – *Rotwild* ('red deer') which both can be found in 19th century dictionaries (cf. Schuster 2016: 278). It is only for the third group of adjectives that a diachronic tendency towards compounding can be observed, as in the case of *Rotwild* which is the only acceptable form in present-day language.[13]

11 The same holds for the noun; examples are not provided for reasons of space.

12 In addition, morphological and semantic properties also play a role in the determination of the form, cf. Schlücker/Plag (2011); Schlücker (2014). Regarding semantics, there is a complementary distribution of metaphorical and metonymic A+N combinations such that the former are always realized as MWEs (e.g., *roter Faden*: lit. red wire, 'thread') and the latter (almost) always as compounds (e.g., *Blauhelm* 'Blue helmet'). However, the bulk of A+N combinations have neither a metaphorical nor a metonymic meaning and are found in both forms.

13 To be sure, the phrase *rotes Wild* is fully grammatical, as it is formed according to the syntactic rules for a nominal phrase with an adjectival modifier in present-day German. It is however not a conventional lexical unit denoting the concept of red deer, and thus no MWE.

Presupposing the existence of doublets in present-day language, Schlücker/Hüning (2009) (on A+N combinations) and Roth (2014, 2015) (on A+N and N+N combinations) examine the factors that determine the choice of use of either of the forms.[14] Based on corpus data, these studies show that, among other things, the context may influence the choice of use of either form. For instance, if an A+N unit is preceded by another adjectival modifier, speakers prefer compounds over MWEs, obviously to avoid the immediate sequence of two syntactic adjectival modifiers (e.g., *heißer Grüntee* vs. *heißer grüner Tee* 'hot green tea'). Similarly, sequences of two postnominal genitive attributes are avoided in favor of compounds. On the other hand, in a compound the modifier cannot be specified. Specification of the modifier thus forces the speaker to use the phrase, cf. *sehr extreme Position* vs. **sehr Extremposition* ('very extreme position'), *Abbau von 500 Stellen* vs. **500 Stellenabbau* ('reduction of 500 jobs').

Furthermore, Roth (2014, 2015) also demonstrates the influence of sentence length. It is known that long sentences generally contain more long words than shorter sentences. In accordance with this idea, compounds are shown to be used more often than phrases in longer sentences. Also, compounds appear more often in the context of other long words in the same sentence. Finally, within the same text consistence of use seems to play an important role, thus speakers tend to consistently use either the compound or the MWE.

In sum, it seems that there are competing abstract patterns as well as specific doublet forms and that, in addition to factors such as expressivity or register, the actual use in a particular context as well as analogical relations are also factors determining their distribution of use.

3 Overview of German MWEs and compounds

This section provides an overview of MWEs and compounds in German, classified according to lexical/syntactic category and syntactic function, respectively. Although it is doubtful whether a reliable general assessment of the quantitative

14 Schlücker/Hüning (2009) deal for the most part with Greek- and Latin-based relational adjectives such as *sozial* 'social' and *optimal* 'optimal'. Roth's (2014, 2015) choice of comparable patterns (i.e. compounds and collocations) relies on the quantitative method of distributional semantics which determines the meaning of an expression on the basis of its context in an automatical procedure. Expressions with very similar or identical lexical constituents in the context are considered semantically equivalent, although it is obvious that subtler meaning differences cannot be detected in this way.

distribution of MWEs and compounds according to lexical category can be made (cf. Section 2.2), it seems justified to say that at least some categories differ greatly with respect to the productivity and the use of MWEs and compounds. Thus, there are categories where either compounding or MWE formation prevail, and other cases where they co-occur. Contrary to other languages, however, in most cases German allows a clear demarcation between MWEs and compounds on formal grounds.

3.1 Prepositions and conjunctions

German has various prepositional MWEs, such as *auf Grund* 'due to', *in Anbetracht* ('in consideration of'). Some of them have morphological counterparts, in particular derivatives formed by the suffix *-lich* ('belonging to X'), e.g. *in Bezug auf – bezüglich* ('pertaining to'), *in Hinsicht auf – hinsichtlich* ('regarding'). There are also morphological counterparts that resemble compounds, often consisting of a P+N sequence, e.g. *aufgrund* (lit. on$_{P}$ ground$_{N}$, 'due to'), *anhand* (lit. at$_{P}$ hand$_{N}$, 'on the basis of'). They are, however, not the output of compounding but the result of univerbation, that is, they are former phrases that have become fixed and, as a result, are now written as one word (cf. Section 2.1). This is also obvious from the phrasal stress pattern of these forms (stress on the nominal head, e.g. *aufgrúnd*), in contrast to genuine P+N compounds which have modifier stress, e.g. *Vórdach* (lit. in front of$_{P}$ roof$_{N}$, 'porch roof'). In many cases, this transition is still in progress which means that both writing norms officially co-exist, e.g. *zu Gunsten – zugunsten* ('in favor of'). They are, for the reason just mentioned, no instances of MWE/compound doublets however. The same holds for grammatical MWEs such as conjunctions, e.g. *wenn auch* ('although'). Although there are a few non-phrasal counterparts, such as *wenngleich* ('albeit'), these are not compounds but univerbations.

3.2 Adverbs and adverbials

Fleischer (1997b: 149–153) stresses that adverbial MWEs display great structural variety. Many of them contain prepositions. Some frequent patterns are given in (3). Note that these examples are diverse regarding syntactic category (so some are structurally equivalent to the prepositional MWEs in the previous section, with others equivalent to the binomials discussed in Section 3.5). The various forms are grouped together due to their common adverbial function, in order to compare them with adverbial word-formation.

(3a) Prepositional phrases: *auf Anhieb* ('straightaway'), *in der Tat* (lit. in the deed, 'indeed'), *unter vier Augen* (lit. under four eyes, 'in private'), *von Hause aus* (lit. from home out, 'by nature')

Various kinds of binomials:

(3b) Conjoined nouns: *Tag und Nacht* ('day and night'), *bei Nacht und Nebel* (lit. at night and fog, 'in secrecy')

(3c) With prepositions: *von Zeit zu Zeit* ('from time to time'), *von Kopf bis Fuß* ('from top to toe'), *von Haus zu Haus* ('from house to house')

(3d) Identical constituents (adverbs): *durch und durch* ('out and out'), *nach und nach* ('little by little')

It is obvious (and has also been discussed by Fleischer 1997b) that many of these MWEs are instantiations of partially fixed abstract patterns (cf. Section 4).

Adverbial compounding, on the other hand, is highly restricted and often not recognized as a word formation type on its own. Adverbial compounds are only found with a handful of adverbs and prepositions, in particular directional adverbs such as *hin* 'to, there' and *her* 'to, there' (cf. Fleischer/Barz 2012), e.g. *herauf* (lit. there up, 'up'), *hinüber* (lit. there over, 'over'), *dorthin* (lit. thereto, 'there'), *daneben* (lit. there next, 'alongside'). However, in some (though not all) cases these forms seem to be univerbations rather than compounds proper. Also, contrary to genuine compounds, the head cannot be clearly identified in most cases and they are not right-headed, as is usual in German. These restrictions on adverbial compounding can explain the enormous amount and structural diversity of adverbial MWEs, in particular given the fact that adverbial derivation is also restricted to a handful of affixes. For the domain of adverbs and adverbials, this supports the idea of MWE formation as a complementary device to compounding.

3.3 Complex verbs

The verbal domain is usually regarded as the most diverse and extensive domain of German MWE formation. Verbal MWEs (the classical "idioms"), either with a fully or a partially non-compositional meaning, such as *bei jdn. einen Stein im Brett haben* (lit. have a stone in so.'s plank, 'be in so.'s good books') or *den Wald vor lauter Bäumen nicht mehr sehen* ('not see the wood for the trees') have long been at the core of phraseological research. In addition, there are various abstract verbal MWE patterns and verbal collocations (mostly N+V). However, there are no corresponding verbal MWE and compound patterns, due to the absence of verbal compounding in German. We will therefore only briefly discuss some patterns, in particular in con-

nection with the question of the demarcation between syntactic and morphological verbal units.

The first one are light verb constructions. They are either $[\text{NP V}]_{\text{VP}}$ or $[\text{PP V}]_{\text{VP}}$ sequences. All of them have corresponding morphological forms, either simplex or derived, but no compounds. The correspondence is also obvious as most of them (though not all) contain a corresponding lexical item, e.g. *einen Beschluss fassen*/*beschließen* (lit. grab a decision, 'decide'), *zur Anzeige bringen* / *anzeigen* (lit. bring to record, 'report'), but *in Kenntnis setzen* / *informieren* (lit. set in knowledge, 'inform'). The phrasal and the morphological forms are equal in meaning, but often differ in argument structure. There are also differences in register as the phrasal constructions are more formal.

Another group are particle verbs. Particle (or: phrasal) verbs such as *anlächeln* 'smile at', *abschicken* 'send off', *austrinken* 'drink up' have been widely discussed for German as well as for other Germanic languages in connection with their unclear status as either morphological or phrasal entities (cf. Los et al. 2012; Dehé 2015; McIntyre 2015; Booij, this volume; a.o.). The central problem is that they are syntactically and morphologically separable in some contexts, e.g. *Er schickt den Brief ab*. ('He sends the letter off'); past participle: *abgeschickt* ('sent off'), i.e. with the *ge*-prefix in the middle of the word rather than at the beginning, as usual. Inseparability, however, is usually considered a basic property of morphological units. Interestingly, it seems that German particle verbs are mainly discussed in morphological research (often in connection with the question of whether they form a word formation pattern on their own or not) but are rarely considered in phraseological research. For English, on the other hand, they are quite naturally also included in phraseological work, cf. Gries (2008) and Ramisch (2015), for instance.[15]

Particles in German particle verbs often have prefixal counterparts, but there are also particles that are homonymous to prepositions, adverbs, adjectives, and nouns (cf. Fleischer/Barz 2012). Thus, even forms like *herumbrüllen* (lit. yell around, 'yell'), *schönreden* (lit. talk st. beautiful, 'sugarcoat') or *totarbeiten* (lit. dead work, 'work to death') that on the surface look like compounds since they involve lexical stems rather than prefixes, are in fact particle verbs since they are separable.

For this reason, particle verbs have often been regarded as problematic regarding the demarcation between MWEs and compounds/morphological units. Whereas the cases discussed so far in this chapter raise the question of the way in which (clearly) morphological and (clearly) syntactic lexical patterns relate to

15 However, Moon (1998) argues against the classification of particle verbs as verbal MWEs in English.

each other, they rather demand a solution for the fact that there are also intermediate constructions. We will come back to the issue of intermediate constructions in Section 3.5 and 4.

Finally, another unclear, intermediate group are N+V patterns of the type *Rad fahren* / *radfahren* ('ride a bike'), *brustschwimmen* ('breaststroke') or *Eis laufen* / *eislaufen* ('ice-skate'). They have been widely discussed in the literature, regarding both their orthography and their morphosyntactic properties. However, contrary to particle verbs, they do not seem to form a homogeneous group. Thus, several co-existing subtypes of these N+V patterns have been identified, with different analyses as either verbal compounds, backformations or incorporation (cf. Fuhrhop 2007, among many others).

3.4 Adjectival compounds and MWEs

Häcki Buhofer et al. (2014) list numerous adjectival collocations, mostly an adjective preceded by a modifier (adverb, adjective or other), cf. (4):

(4) *streng geheim* (lit. strictly secret, 'top secret'), *bitter nötig* ('urgently necessary'), *geradezu klassisch* ('almost classical'), *verschwindend klein* ('vanishing small'), *spielend leicht* (lit. playing easy, 'easily'), *furchtbar traurig* ('terribly sad'), *immens wichtig* ('immensely important')

The modifiers in these phrases express gradation, i.e. they either intensify or diminish the property denoted by the adjective. A gradational meaning can also be found in adjectival compounds, as those in (5).

(5) *dunkelrot* ('dark red'), *tiefrot* (lit. deep red, 'bright red'), *heilfroh* (lit. salvation glad, 'really glad'), *stinkfaul* (lit. stinking lazy, 'bone-idle'), *grundverkehrt* ('fundamentally wrong'), *hochbegabt* ('highly talented')

However, real doublets are rare, e.g., *schwerkrank* – *schwer krank* (lit. heavily ill, 'critically ill'). In addition to such compounds having a gradational meaning, adjectival compounds very often have a determinative meaning, that is, the modifier specifies the property denoted by the adjectival head, often, though not always, in a comparative way, cf. (6).

(6) *graublau* ('grey-blue, powderblue'), *hautnah* (lit. skin close, 'very close'), *schneeblind* ('snow-blind'), *butterweich* (lit. butter soft, 'beautifully soft')

Thus, the morphological and the syntactic units discussed above only partially overlap in the semantically restricted domain of gradation and cannot generally be regarded as competing patterns.[16]

In addition to the adjectival collocations as in (4), there are also partially fixed MWE patterns in the adjectival domain. One of them are adjectival phrasal similes as in (7) (cf. Burger 2015: 56f.; Hüning/Schlücker 2015).[17] It is a typical example of a partially filled MWE. The property denoted by the adjective is – by means of the comparative conjunction *wie* 'as' – compared to a reference value provided by the noun.

(7) [(*so*) A *wie* (*ein*) N] [(*as*) A *as* (*an*) N]): *so weich wie Seide* ('as soft as silk')

Interestingly, the same comparison can also be expressed by an N+A compound, as mentioned above, e.g. *seidenweich* 'silky smooth, as soft as silk' , cf. (6). Thus, it seems that these are examples of equivalent morphological and syntactic lexical patterns which in turn raises the question of the relation between the patterns and the distribution of the specific forms.[18] First, it seems that the formation of comparative compounds is more restricted than that of phrasal similes. There are plenty of phrasal comparisons, both with a compositional and a non-compositional meaning, that do not allow the formation of a corresponding compound, cf. (8).

(8) (*so*) *stumm wie ein Fisch* / **fischstumm* (lit. as mute as a fish, 'as mute as a maggot')
(*so*) *sanft wie Regen* / **regensanft* ('as soft as rain')

16 For some examples in (5) and (6) it may be a matter of debate whether they only have a gradational or a determinative meaning or both, e.g. *dunkelrot* ('dark red'). The crucial point here is, however, that the determinative meaning is not available for the syntactic pattern and that for this reason there is only partial overlap between the morphological and the syntactic pattern.

17 Adjectival binomials form another pattern, cf. (i). However, nominal, verbal and adverbial binomials seem to be much more frequent than adjectival ones. Yet another pattern is given in (ii), cf. Fleischer (1997b: 149). However, the patterns do not have a direct morphological counterpart, neither regarding form nor semantics.

(i) [A *und* A] ([A *and* A]): *fix und fertig* (lit. fix and ready, 'beat, strung out'), *still und leise* ('silent and quiet')

(ii) [*zum* + infinitive + A] ([*to the* + infinitive + A]): *zum Weinen schön* (lit. to the crying beautiful, 'movingly beautiful'), *zum Bersten voll* (lit. to the bursting full, 'full to bursting')

18 Interestingly, adjectival phrasal similes and corresponding N+A compounds do also exist in other languages (cf. Finkbeiner/Schlücker, this volume), such as Dutch (cf. Booij, this volume), Italian (cf. Masini, this volume), and Finnish (cf. Hyvarinen, this volume).

(*so*) *dumm wie Brot* / **brotdumm* (lit. as dumb as bread, 'as thick as brick')
(*so*) *frech wie Dreck* / **dreckfrech* (lit. as cheeky as dirt, 'as bold as brass')

An obvious explanation would be that the formation of the compound is blocked due to the existence of the MWE, in line with usual assumptions of non-existence of synonymy in the lexicon. This explanation is not convincing, however, given the existence of numerous doublets as those in (9):

(9) (*so*) *weich wie Seide* / *seidenweich* ('as soft as silk')
(*so*) *weiß wie Schnee* / *schneeweiß* ('snow-white')
(*so*) *hart wie Stein* / *steinhart* ('rock-hard')
(*so*) *stark wie ein Bär* / *bärenstark* (lit. as strong as a bear, 'strong as an ox')

On the other hand, there are also compounds that lack corresponding phrasal comparisons. In these cases, the phrasal expressions are not ungrammatical but are not conventionalized lexical units and therefore much rarer, as can be seen from corpus data, cf. (10).[19]

(10) *kirschrot* [586] / *rot wie eine Kirsche* [6] ('cherry-red')
zitronengelb [832] / *gelb wie eine Zitrone* [7] ('lemon yellow')
blitzschnell [8.585] / *schnell wie* {*ein*/*der*} *Blitz* [63] ('as quick as a/the flash')

The distribution of forms is also dependent on the context, as discussed for A+N sequences in Section 2.3. Whereas both patterns can be used predicatively or adverbially, only compounds can occur in attributive position. Thus, although the phrasal pattern might be more expressive, especially since it also allows nonsensical, apparently unmotivated comparisons which compounds generally do not (e.g. *frech wie Dreck*, *dumm wie Brot*, cf. (8)),[20] compounds are more versatile concerning their syntactic distribution.

Furthermore, it has been assumed that both the phrasal and the morphological pattern have developed a semantic subpattern with an intensifying rather than a comparative meaning (cf. Hüning/Booij 2014; Hüning/Schlücker 2015). Thus, in cases like *hart wie Stein* / *steinhart* ('rock-hard'), *stark wie ein Bär* / *bären-*

19 Counts are from all corpora available through www.dwds.de.
20 A counterexample of a nonsensical comparison in a compound is *rotzfrech* (lit. cheeky as snot, 'impudent').

stark (lit. as strong as a bear, 'strong as an ox') the noun does not provide an actual measure for comparison but rather functions as an intensifier ('very hard', 'very strong'). This intensifying meaning is available for both phrasal similes and compounds, although it is not entirely clear under which condition comparative patterns develop an intensifying meaning. Importantly, neither the phrasal nor the morphological pattern do always have this intensifying meaning. For instance, the adjective *weich* 'soft' occurs in numerous comparative patterns, both phrasal and morphological, and all of them have a comparative rather than an intensifying meaning. More specifically, two subgroups can be observed, one relating to the softness of the surface and the other to the softness of the substance, cf. (11)–(12).

surface

(11a) *seidenweich*, *samtweich* ('silky smooth', 'velvety')
(11b) (*so*) *weich wie* {*Seide* / *Samt*} ('as soft as {silk / velvet}')

substance

(12a) *butterweich*, *gummiweich*, *wachsweich*, *watteweich*
('buttersoft', 'as soft as rubber', 'as soft as wax', 'cotton-soft')
(12b) (*so*) *weich wie* {*Butter* / *Gummi* / *Wachs* / *Watte*}
('as soft as {butter / rubber / wax / cotton}')

In these cases, the various measures of comparison are literally present, thus *samtweich* is different from *seidenweich* in the way velvet is different from silk. In particular, the groups in (11) and (12) have clearly different meanings and cannot be used interchangeably. It might then be concluded that an intensifying meaning can only develop if only one comparative measure is conventionalized, as in the case of *hart* (*hart wie Stein* / *steinhart*) and *stark* (*stark wie ein Bär* / *bärenstark*), and not several.[21]

21 There are also intensifying modifiers in compounds that have developed into a productive intensifying pattern such that the modifier has completely lost its literal meaning, often discussed in connection with the term affixoid. A case in point is the intensifier *stock* 'stick' which first occurred in morphological and phrasal comparisons such as *stocksteif* / *steif wie ein Stock* 'as stiff as a stick' but later, after having developed an intensifying meaning, was used as an intensifier of other, totally unrelated adjectives, e.g. *stockdunkel* ('very dark'), *stockbesoffen* ('very drunk') (cf. Hüning/Booij 2014; Hüning/Schlücker 2015). The abovementioned example of *grund-* ('ground') (cf. (5)) seems to be a similar case.

These observations lead to the conclusion that phrasal similes and adjectival compounds are an example of the co-existence of corresponding phrasal and morphological lexical structures. They show that blocking as a principle controlling the lexicon does not seem to be as strong as sometimes assumed. Also, as both patterns share lexical material and semantic subgroups (comparative/intensifying) they cannot be regarded as complementary. Rather, it can be assumed that both patterns as well as their instantiations are related to each other via their constituents and their meanings. The choice between either form in the case of doublets as in (9), (11), and (12) is likely to be determined by expressivity (in favor of the phrasal structure) as well as syntactic flexibility and conciseness (favoring the compound), but also other factors determined by the actual context, e.g. sentence length (cf. Section 2.3).

3.5 Nominal compounds and MWEs

Nominal compounding, in particular N+N compounding, is without doubt the most frequent and productive type of compounding in German. However, nominal MWEs also come in a variety of forms, cf. (13) and (14) (cf. Burger 2001, for instance). Thus, contrary to what has been assumed in the literature nominal compounding and MWE formation do not seem to complement each other; in particular, it is not the case that MWE formation is poorly developed due to the obvious productivity of nominal compounding.

(13a) Postnominal genitives: *Schlaf der Gerechten* ('sleep of the just'), *Geschenk des Himmels* (lit. gift from heaven, 'godsend'), *Macht der Gewohnheit* ('force of habit')
(13b) Prenominal genitives: *des Rätsels Lösung* (lit. the puzzle's solution, 'the answer to this problem')
(13c) Prepositional constructions: *Dame von Welt* (lit. lady of world, 'sophisticated woman'), *Nerven aus Stahl* ('nerves of steel')
(13d) Close apposition: *Häufchen Elend* (lit. heap misery, 'picture of misery'), *Vater Staat* (lit. father state, 'Uncle Sam')
(13e) Binomials: *Grund und Boden* (lit. ground and soil, 'property'), *Sack und Pack* ('bag and baggage')

(14) A+N phrases:
(14a) *lahme Ente* ('lame duck'), *heißes Eisen* (lit. hot iron, 'hot potato'), *krumme Sachen* (lit. bent things, 'criminal activities')

(14b) *gelbes Trikot* ('yellow jersey'), *echte Grippe* (lit. real flu, 'influenza'), *schwarzes Brett* (lit. black board, 'notice board')

Some of them have a (fully or partially) non-compositional meaning, e.g. *lahme Ente* ('lame duck') or *Nerven aus Stahl* ('nerves of steel'). Others differ from corresponding free phrases regarding their morphosyntactic properties. For instance, nouns in binomials do not occur with determiners (never inside the construction and only rarely before), which would be ungrammatical in a normal coordinative construction. Also, their order is not interchangeable, again contrary to free coordinative structures.[22] Prenominal genitives are no longer productive in present-day language (except with proper names and kinship terms) and thus are only found with fossilized forms.

Compared to the patterns in (13) and (14), those in (15) are more specialized regarding semantics and conditions of use:

(15a) N+A constructions: *Forelle blau* (lit. trout blue, 'blue trout'), *Sonne pur* (lit. sun pure), *Rahmspinat tiefgefroren* (lit. cream spinach deep-frozen)

(15b) [*ein* N_1 *von einem*/*einer* N_2] ('[*an* N_1 *of an* N_2]'): *ein Berg von einem Mann* (lit. a mountain of a man, 'a man like a mountain'), *eine Null von einem Stürmer* (lit. a null of a striker, 'a useless striker'), *ein Arsch von einem Professor* (lit. a butt of a professor, 'an idiot of a professor')[23]

(15c) [N_1 *von* N_2] ('[N_1 *of* N_2]'): *Salat von Flusskrebsen* (lit. salad of crayfish), *Gratin von Tomaten* (lit. gratin of tomatoes), *Suppe von Spinat und Bärlauch* (lit. soup of spinach and wild garlic)

The pattern in (15a) is characterized by a postponed uninflected adjective. Its use is highly restricted and productive only in advertising catalogues, as slogans, brand names, or product descriptions (cf. Dürscheid 2002). The pattern in (15b) is productive; it has an evaluative meaning and expresses a comparison of N_2 to the reference value provided by N_1. Thus, there is a mismatch between the semantic head of the construction (N_2) and the syntactic one (N_1). Finally, the pattern in (15c) can be described as a register-specific construction for haute cuisine language. Here, N_1 must denote a dish and N_2 an ingredient. This prepositional

22 This view is somewhat simplified as there does not seem to be a strict border between conventionalized binomials and free coordinative constructions. It can be observed that nouns in occasional binomials do not occur with determiners, but their internal order is interchangeable (cf. D'Avis/Finkbeiner 2013, for instance), so they might be regarded as in-between forms.

23 I owe the last two examples to Rita Finkbeiner.

construction is used instead of the compounds *Flusskrebssalat* 'crayfish salad', *Tomatengratin* 'tomato gratin', *Spinat-Bärlauch-Suppe* 'spinach & wild garlic soup' that are the usual (and only) way of expressing these concepts in everyday language.

The examples discussed above show that some of the nominal MWE patterns differ morphosyntactically from the corresponding free phrases. This also holds for the A+N phrases in (14). More specifically, some of them form an example for the existence of intermediate constructions, that is, constructions that are neither clearly phrasal nor morphological, similar to particle verbs (cf. Section 3.3). Two groups of A+N phrases can be distinguished. The first one (cf. (14a)) consists of phrases with a metaphorical meaning (either of the modifier alone or both modifier and head), e.g. *heißes Eisen* (lit. hot iron, 'hot potato'). The special meaning of these forms requires the adjective and the noun to be unseparated. Thus, if there is an intervening adjective (e.g., *heißes gefährliches Eisen* 'hot dangerous iron') or if the adjective is used predicatively (*das Eisen ist heiß* 'the iron is hot') only the literal meaning is available. However, these phrases allow comparative forms and the modification of the adjective, just as free A+N phrases, e.g. *ein sehr heißes Eisen* ('a very hot potato'). Thus, although the meaning of the adjective is metaphorical (e.g. 'hot' standing for 'tricky', 'delicate'), it specifies the meaning of the nominal head and can as such be modified itself, just as in any regular A+N phrase. In the second group (cf. (14b)), in contrast, the adjective has a classificatory function. It does not specify a property of the nominal head but rather identifies a subclass of the concept denoted by the head. For instance, a yellow jersey is not just a shirt that is yellow but the kind of shirt worn by the leader of the Tour de France race. Importantly, it is exactly this classifying meaning that is also a general characteristic of A+N compounds, e.g. *Gelbgold* ('yellow gold'): a kind of gold that is an alloy of gold with silver, *Stummfilm* ('silent film'): a kind of film without spoken words. Due to this classifying function, the adjective in classifying A+N phrases cannot be modified. The adjective serves to the identification of the subclass. This is a categorial property that is not gradable: either something belongs to the category of yellow jersey or not. Thus, neither intensification or comparative forms are allowed. Similarly, the adjectival modifier in A+N compounds can never be modified. Thus, if we compare metaphorical A+N phrases (as in (14a)) and classifying A+N phrases (as in (14b)), it becomes obvious that the former are clearly phrasal in nature while the latter have both phrasal and morphological features. The classifying A+N phrases allow syntactic rules of agreement and case assignment of the adjective (just like in any phrase and unlike adjectives in A+N compounds), and are meanwhile inseparable, with the adjective precluding comparative forms and modification (as morphologically complex words, cf. A+N compounds). For this reason, it seems that classifying A+N

phrases constitute an intermediate construction. Following a proposal for the analysis of A+N phrases in Dutch in Booij (2010: Chapter 7), they can be analyzed as syntactic compounds, and thus as lexical items (N_0) with a complex internal syntactic structure (cf. Schlücker 2014: 173–187).[24] With this analysis comes the idea that classifying A+N phrases are instantiations of a productive abstract pattern (or schema), and thus a phrasal pattern for the formation of new lexical entities, just like the morphological pattern of (A+N) compounding. Metaphorical A+N phrases, on the other hand, are idiosyncratic forms that result from the lexicalization (including semantic specialization) of individual regular A+N phrases.

4 Theoretical implications

In the past decennia of phraseological research it has become obvious that the existence of abstract, partially fixed phrasal patterns in the lexicon is not restricted to a handful of MWE patterns, such as binomials, but rather seems to be a fundamental characteristic of MWEs more generally. Such patterns are assumed to underlie MWEs both with and without a compositional meaning and both with and without deviant phonological, morphological, or syntactic properties.

The crucial point here is that under this view, MWE patterns are syntactic patterns in the lexicon, and thus are lexical patterns on a par with morphological ones. Booij (2002, 2010) argues that constructional idioms are syntactic expressions that function as alternatives to morphological expressions. In his definition, constructional idioms are

> syntactic constructions with a (partially or fully) non-compositional meaning contributed by the construction, in which – unlike idioms in the traditional sense – only a subset (possibly empty) of the terminal elements is fixed. (Booij 2002: 302)

This definition can capture many pattern-like, partially-fixed MWEs as, for instance, in (2), (7), or (15b). In addition, it also covers other, more grammatical kinds of MWEs such as analytic causative constructions or analytic progressives (cf. Booij 2002, 2010). These are productive patterns with the same function as their morphological, synthetic counterparts and, just like these morphological counterparts, their productivity can be shown to be subject to certain restrictions.

24 For further details, including an analysis of the adjective as either A^0 or AP, cf. Booij (2010: 176ff.) and Schlücker (2014: 177ff.).

Culicover/Jackendoff/Audring (2017) point to another parallel between morphological and MWE patterns. Obviously, not all MWEs are instantiations of a partially fixed pattern, for instance verbal MWEs that share syntactic structure (e.g., $[V\ NP]_{VP}$ or $[V\ NP\ PP]_{VP}$) but do not have a common lexical element. Culicover/Jackendoff/Audring (2017) argue that many MWEs – both those with and those without fixed elements – display a fully regular syntactic behavior. So, for instance, in the case of classical verbal idioms such as *kick the bucket* or *sell* [NP] *down the river*, there are no differences concerning the morphosyntactic behavior between the idiomatic and the literal phrases except for their meaning (and, arguably, the morphosyntactic properties that result directly from this meaning, such as the non-passivizability of *kick the bucket* 'die'). Other MWEs are lexically restricted, such as for instance *go*/*drive* [NP] *nuts*/*crazy*/*bananas*/*insane*/**wild*/**demented*/**meshuga*. The (non-)admissibility here is unpredictable with regards to the meaning of the MWE, and therefore has to be stored. The authors argue that the same contrast can be found in morphological patterns which may either be morphosyntactically unrestricted and therefore fully productive, as with the *s*-plural in English, or unsystematically restricted as is the case with several derivational affixes, leading to a restricted productivity or unproductivity of these patterns. Again, these restrictions must be stored. Thus, the resemblance Culicover/Jackendoff/Audring (2017: 14) identify between morphological and MWE patterns is that of the difference between what they call "relational" and "generative" patterns: Relational patterns are stored items that are related to more general patterns in the lexicon, and, via them, to similar stored items. Generative patterns are also relational, but in addition are productive and can be used to generate new expressions. Thus, morphological and MWE patterns are of a very similar nature in that they are both determined by the co-existence of relational and generative patterns.

One consequence that follows from this line of thought is the existence of ad hoc MWEs – that is, MWEs that are occasionally coined and used but not stored – but also the existence of potential MWEs, which are MWEs that fit the morphosyntactic and semantic specifications of a particular MWE pattern but which have not yet been realized, just as is the case with occasional and potential word formations. Empirically based research (cf., for German, Finkbeiner 2008; Steyer 2015, for instance) has provided ample evidence for this idea. However, it obviously fundamentally clashes with the notion of MWEs as stored items not only in traditional phraseological research but also in older "mainstream" generative grammar which views MWEs as a residual collection of idiosyncratic expressions stored in the lexicon.

Finally, against the background of the basic similarity between morphological and MWE patterns, it is quite possible to accept the idea of intermediate con-

structions that have both phrasal and morphological properties like the syntactic compounds discussed at the end of Section 3.5 (in addition to clearly morphological and clearly syntactic patterns). They can be regarded as a link or transitional category between morphological and syntactic lexical patterns. In other words, morphological and syntactic lexical patterns form a continuum and these intermediate constructions are situated in the middle.

In sum, treating MWEs in the way advocated here has a crucial impact on ideas about the structure of the lexicon and the division of labor between morphology and syntax.

5 Conclusion

This chapter has provided an overview of German MWEs from the perspective of relating MWEs and MWE formation to compounds and compounding. It has been shown that in German, MWEs for the most part can be clearly distinguished from compounds on formal grounds. This chapter has focused on MWEs that have – or at least could have in principle – corresponding compounds with a similar meaning and function. In general, it can be seen that the proportion of compounds and MWE differs between lexical categories. These differences – or at least some of them – can be explained by the idea about the avoidance of synonymous expressions in the lexicon. On the other hand, however, it has also become clear that there are numerous parallel and thus competing abstract patterns and even doublets on the level of specific forms.

From a theoretical perspective, it has been argued that MWEs should not generally be regarded as individual and idiosyncratic formations that are derived from "regular" syntactic phrases in a secondary process of idiomatization and lexicalization. Instead – and in accordance with numerous findings in recent literature – it can be assumed that abstract patterns underlie MWE formation and that, therefore, MWE formation can be regarded as being on a par with word formation. Thus, just as there are abstract morphological patterns for the formation of lexical units there are also syntactic ones.

References

Barz, Irmhild (1996): Komposition und Kollokation. In: Knobloch/Schaeder (eds.). 127–146.
Barz, Irmhild (2007): Wortbildung und Phraseologie. In: Burger, Harald et al. (eds.): Phraseology. Vol. 1. Berlin/New York: De Gruyter. 27–36.

Booij, Geert (2002): Constructional Idioms, Morphology, and the Dutch Lexicon. In: Journal of Germanic Linguistics 14, 4. 301–329.
Booij, Geert (2010): Construction morphology. Oxford/New York: Oxford University Press.
Burger, Harald (2001): Von lahmen Enten und schwarzen Schafen. Aspekte nominaler Phraseologie. In: Häcki Buhofer, Annelies/Burger, Harald/Gautier, Laurent (eds.): Phraseologiae Amor. Aspekte europäischer Phraseologie. Festschrift für Gertrud Gréciano zum 60. Geburtstag. Hohengehren: Schneider. 33–42.
Burger, Harald (2015): Phraseologie: eine Einführung am Beispiel des Deutschen. 5th ed. (= Grundlagen der Germanistik 36). Berlin: Schmidt.
Culicover, Peter W./Jackendoff, Ray/Audring, Jenny (2017): Multiword Constructions in the Grammar. In: Topics in Cognitive Science 9, 3. 552–568.
D'Avis, Franz/Finkbeiner, Rita (2013): "Podolski hat Vertrag bis 2007, egal, ob wir in der Ersten oder Zweiten Liga spielen." Zur Frage der Akzeptabilität einer neuen Konstruktion mit artikellosem Nomen. In: Zeitschrift für germanistische Linguistik 41, 2. 212–239.
Dehé, Nicole (2015): 35. Particle verbs in Germanic. In: Müller et al. (eds.). 611–626.
Donalies, Elke (2008): *Sandstrand*, *sandy beach*, *plage de sable*, *arenile*, *piaskowy plaża*, *homokos part*. Komposita, Derivate und Phraseme des Deutschen im europäischen Vergleich. In: Deutsche Sprache 36, 4. 305–323.
Dürscheid, Christa (2002): "Polemik satt und Wahlkampf pur" – Das postnominale Adjektiv im Deutschen. In: Zeitschrift für Sprachwissenschaft 21, 1. 57–81.
Fillmore, Charles J./Kay, Paul/O'Connor, Mary Catherine (1988): Regularity and idiomaticity in grammatical constructions. The case of "let alone". In: Language 64, 3. 501–538.
Finkbeiner, Rita (2008): Zur Produktivität idiomatischer Konstruktionsmuster. Interpretierbarkeit und Produzierbarkeit idiomatischer Sätze im Test. In: Linguistische Berichte 216, 4. 391–430.
Fleischer, Wolfgang (1982): Phraseologie der deutschen Gegenwartssprache. Leipzig: VEG Bibliographisches Institut.
Fleischer, Wolfgang (1996a): Phraseologische, terminologische und onymische Wortgruppen als Nominationseinheiten. In: Knobloch/Schaeder (eds.). 147–170.
Fleischer, Wolfgang (1996b): Zum Verhältnis von Wortbildung und Phraseologie im Deutschen. In: Korhonen, Jarmo (ed.): Studien zur Phraseologie des Deutschen und des Finnischen II. Bochum: Universitätsverlag Brockmeyer. 333–344.
Fleischer, Wolfgang (1997a): Das Zusammenwirken von Wortbildung und Phraseologisierung in der Entwicklung des Wortschatzes. In: Wimmer, Rainer/Berens, Franz-Josef (eds.): Wortbildung und Phraseologie. Tübingen: Narr. 9–24.
Fleischer, Wolfgang (1997b): Phraseologie der deutschen Gegenwartssprache. 2nd ed. Tübingen: Narr.
Fleischer, Wolfgang/Barz, Irmhild (2012): Wortbildung der deutschen Gegenwartssprache. 4th ed. Berlin: Schmidt.
Fuhrhop, Nanna (2007): Verbale Komposition: Sind *brustschwimmen* und *radfahren* Komposita? In: Kauffer, Maurice/Métrich, René (eds.): Verbale Wortbildung im Spannungsfeld zwischen Wortsemantik, Syntax und Rechtschreibung. Tübingen: Narr. 49–58.
Gaeta, Livio/Ricca, Davide (2009): Composita solvantur: Compounds as lexical units or morphological objects? In: Gaeta, Livio/Grossmann, Maria (eds.): Compounds between syntax and lexicon. Special Issue of Italian Journal of Linguistics/Rivista di Linguistica 2, 1. 35–70.

Gries, Stefan Th. (2008): Phraseology and linguistic theory: a brief survey. In: Granger, Sylviane/Meunier, Fanny (eds.): Phraseology. An interdisciplinary perspective. Amsterdam/Philadelphia: Benjamins. 3–25.
Häcki Buhofer, Annelies et al. (2014): Feste Wortverbindungen des Deutschen. Kollokationenwörterbuch für den Alltag. Tübingen: Francke.
Haiman, John (1980): The Iconicity of Grammar: Isomorphism and Motivation. In: Language 56, 3. 515–540.
Häusermann, Jürg (1977): Phraseologie: Hauptprobleme der deutschen Phraseologie auf der Basis sowjetischer Forschungsergebnisse. Tübingen: Narr.
Hüning, Matthias/Booij, Geert (2014): From compounding to derivation. The emergence of derivational affixes through "constructionalization". In: Folia Linguistica 48, 2. 579–604.
Hüning, Matthias/Schlücker, Barbara (2015): 24. Multi-word expressions. In: Müller et al. 450–467.
Jackendoff, Ray (1997a): The architecture of the language faculty. Cambridge, UK: Cambridge University Press.
Jackendoff, Ray (1997b): Twistin' the night away. In: Language 73, 3. 534–559.
Jackendoff, Ray (2002): Foundations of language. Oxford: Oxford University Press.
Kay, Paul/Fillmore, Charles J. (1999): Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. In: Language 75. 1–33.
Knobloch, Clemens/Schaeder, Burkhard (eds.): Nomination, fachsprachlich und gemeinsprachlich. Opladen: Westdeutscher Verlag.
Los, Bettelou et al. (2012): Morphosyntactic change: a comparative study of particles and prefixes. (= Cambridge Studies in Linguistics 134). Cambridge, UK: Cambridge University Press.
McIntyre, Andrew (2015): 23. Particle-verb formation. In: Müller et al. 434–449.
Möhn, Dieter (1986): Determinativkomposita und Mehrwortbenennungen im deutschen Fachwortschatz. In: Jahrbuch Deutsch als Fremdsprache 12. 111–133.
Moon, Rosamund (1998): Fixed expressions and idioms in English: a corpus-based approach. Oxford/New York: Oxford University Press.
Motsch, Wolfgang (2004): Deutsche Wortbildung in Grundzügen. 2nd ed. Berlin/New York: De Gruyter.
Müller, Peter O. et al. (eds.) (2015): Word-formation. An international handbook of the languages of Europe. Vol. 1 (= Handbooks of Linguistics and Communication Science (HSK) 40.1). Berlin/Boston: De Gruyter.
Pavlov, Vladimir M. (1983): Zur Ausbildung der Norm der deutschen Literatursprache im Bereich der Wortbildung (1470–1730): Von der Wortgruppe zur substantivischen Zusammensetzung. (Zur Ausbildung der Norm der deutschen Literatursprache (1470–1730) VI). Berlin: Akademie.
Ramisch, Carlos (2015): Multiword expressions acquisition: a generic and open framework. New York: Springer.
Roth, Tobias (2014): Wortverbindungen und Verbindungen von Wörtern. Lexikografische und distributionelle Aspekte kombinatorischer Begriffsbildung zwischen Syntax und Morphologie. Tübingen: Narr.
Roth, Tobias (2015): Kompositum oder Kollokation? Konkurrenz an der Syntax-Morphologie-Schnittstelle. In: Schmidlin, Regula/Behrens, Heike/Bickel, Hans (eds.): Sprachgebrauch und Sprachbewusstsein. Berlin/Boston i.a.: De Gruyter. 155–176.

Scherer, Carmen (2012): Vom *Reisezentrum* zum *Reise Zentrum*. Variation in der Schreibung von N+N-Komposita. In: Gaeta, Livio/Schlücker, Barbara (eds.): Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte. Berlin/New York: De Gruyter. 57–81.

Schlücker, Barbara (2014): Grammatik im Lexikon. Adjektiv+Nomen-Verbindungen im Deutschen und Niederländischen. (= Linguistische Arbeiten 553). Berlin/Boston: De Gruyter.

Schlücker, Barbara/Hüning, Matthias (2009): Compounds and phrases. A functional comparison between German A+N-compounds and corresponding phrases. In: Italian Journal of Linguistics/Rivista di Linguistica 21, 1. 209–234.

Schlücker, Barbara/Plag, Ingo (2011): Compound or phrase? Analogy in naming. In: Lingua 121, 7. 1539–1551.

Schuster, Saskia (2016): Variation und Wandel. Zur Konkurrenz morphologischer und syntaktischer A+N-Verbindungen im Deutschen und Niederländischen seit 1700. (= Konvergenz und Divergenz (KuD) 4). Berlin/Boston: De Gruyter.

Steyer, Kathrin (2015): Patterns. Phraseology in a state of flux. In: International Journal of Lexicography 28, 3. 279–298.

Steyer, Kathrin (2016): Corpus-driven Description of Multi-word Patterns. In: Pastor, Gloria Corpas et al. (eds.): Workshop Proceedings “Multi-Word Units in Machine Translation and Translation Technologies” (MUMTTT2015). Genf: Editions Tradulex. 13–18.

Geert Booij

Compounds and multi-word expressions in Dutch

1 Introduction: morphological and phrasal lexical units

It is a generally accepted insight in linguistics that not only words, but also combinations of words (multi-word expressions, MWEs) may function as lexical units, and can be stored in the mental lexicon. MWEs may vary in size, from two words to a complete sentence (for instance, a proverb) (Hüning/Schlücker 2015). The existence of MWEs raises interesting questions about the organization of the grammar of natural languages, and their relationship to morphological word combinations. This is the topic of this article, with Dutch being the object language.[1]

The number of MWEs in Dutch is enormous (cf. Schutz/Permentier 2016 for a recent survey). In this article I will discuss a specific subset of MWEs in Dutch, namely phrases that function as alternatives to compounds. Compound formation in Dutch serves to expand three major word classes, nouns, adjectives and verbs. They provide names for types of entities, properties, and events respectively. I will compare these types of compound with their phrasal counterparts with a similar naming function: noun phrases, adjectival phrases, and verbal phrases. As Koefoed (1993: 3) points out: "Naming is creating a link between an expression and a concept. The expression is often a word, but can also consist of more than one word." The other function of phrases is that of description. Koefoed gives the phrase *vaderlandse geschiedenis* 'national history' as an example, it is the conventional name for a particular form of history, and may be contrasted with the phrase *de geschiedenis van het vaderland* 'the history of the native country' which is a description (Booij 2009a: 219).

1 The existence of such a wide range of MWEs also raises the psycholinguistic question which role they play in lexical processing. As far as Dutch is concerned, there are a number of psycholinguistic studies (Levelt/Meyer 2000; Sprenger 2003; Sprenger/Levelt/Kempen 2006; Nooteboom 2011) to which the reader is referred. However, this psycholinguistic dimension will be left out of consideration here.

https://doi.org/10.1515/9783110632446-004

In most cases, these two structural options for creating names complement each other, but there is also some competition. A comparison of these two options provides insight into the organization of grammar, the role of the lexicon, and the division of labour between morphological and syntactic devices.

The topic broached in this article may be qualified as a study of the relation between compounding and forms of 'periphrastic word formation'. The latter term is used in Booij (2002c) as a characterization of the function of Dutch particle verbs. Traditionally, the term 'periphrasis' is applied to word combinations that fill cells of inflectional paradigms, for instance the cells for the perfect tense forms of Dutch verbs, combinations of an auxiliary (*hebben* 'to have' or *zijn* 'to be') and a past participle. As we will see below, phrasal word combinations can be used to fill in certain gaps in the word formation system and compete with synonymous complex words. This is the idea of complementarity between morphological and phrasal lexical units.

Investigating this relationship also makes sense from a diachronic perspective, since syntactic word combinations are the historical source of the various types of compounding that we find in Germanic languages like Dutch. Hence, it is important to understand the differences and similarities between phrasal and morphological constructions, and it may not always be easy to make this distinction due to this historical source of compounds. This demarcation problem has been pointed out by Hermann Paul in chapter XIX of his *Prinzipien der Sprachgeschichte* (Paul 1898), where he argues that "[d]er Uebergang von syntaktischem Gefüge zum Kompositum ist ein so allmählicher, dass es gar keine scharfe Grenzlinie zwischen beiden gibt" (ibid.: 304). Paul's observation on the blurred boundary between phrases and compounds implies that we need to investigate in more detail how we can distinguish compounds from phrases with a similar form and function. In this article, I will therefore first discuss the formal demarcation of compounds from phrases (Section 2). In Section 3, the naming functions of various types of compounds and their phrasal counterparts are discussed in detail. Section 4 shows how syntax plays a role besides compounding in the construction of complex numeral expressions. In Section 5, it is briefly argued what these empirical findings imply for a proper theory of the organization of grammar, and why Construction Morphology (CxM) offers an insightful account of the relevant facts.

2 Demarcation of compounds and phrases

The demarcation of compounds and phrases in Dutch is based on a number of criteria: lexical integrity, orthography, phonological properties, and morphological properties. Before I discuss these criteria in detail, let me first give a number of relevant examples of compounds and their phrasal counterparts that consist of combinations of the same word classes:

(1)		compound	phrase
	N+N	*opoe+fiets*	*opoe's+fiets*
		lit. grandma+bike, 'retro-bike'	'grandma's bike'
	A+N	*rood+baars*	*rode+wijn*
		'red bass'	'red wine'
	A+A	*donker+geel*	*rijk*[2] *versierd*
		'dark-yellow'	'richly decorated'
	N+V	*raad+plegen*	*koffie+zetten*
		lit. advice+seek, 'to consult'	lit. coffee make, 'to make coffee'
	A+V	*lief+kozen*	*schoon+maken*
		lit. love+fondle, 'to caress'	lit. clean make, 'to clean'
	P+V	*over+komen*	*over+komen*
		lit. over+come, 'to happen to'	lit. over come, 'to come across'

The N+N and A+A phrases in (1) do not have a naming function, they are descriptive in nature. The A+N phrase *rode wijn* can be used as a name for a particular type of wine, or as a description. Yet, I discuss these phrases here because we are focusing on the formal differences between compounds and phrases, whether with a naming or a descriptive function.

Not all types of Dutch compounds have a counterpart in phrasal form; this applies to the following types:

(2)		
	V+N compounds	*eet+kamer* 'dining room'
	N+A compounds	*sneeuw+wit* 'snow-white'
	V+A compounds	*druip+nat* 'drip-wet, dripping wet'

In these cases we cannot find phrasal counterparts because verbs cannot modify a nominal head, and nouns and verbs cannot modify adjectives in pre-adjectival

2 In this example, *rijk* functions as an adverb.

position. Hence, for these types of word combinations there is no phrasal interpretation possible, and thus, the demarcation issue does not arise.

2.1 Lexical Integrity

The first criterion that comes to mind for the demarcation of words and phrases is that of Lexical Integrity. The criterion of Lexical Integrity can be defined as follows: 'Syntactic rules cannot manipulate parts of words'. In other words, words are islands for syntactic operations. This narrow definition of Lexical Integrity as being restricted to syntactic operations does not exclude the possibility that the internal structure of words is accessible for other purposes, such as semantic interpretation, as should be the case (cf. Booij 2009b for detailed discussion of various definitions of Lexical Integrity).

The word combinations listed as phrases in (1) all allow for syntactic splits:

(3a) *opoe's oude fiets*
'grandma's old bike'
rode en witte wijn
'red and white wine'

(3b) *rijk en kostbaar versierd*
'richly and costly decorated'

(3c) *Jan zet koffie*
'John makes coffee'
Hij maakt de kamer schoon
lit. He makes the room clean, 'He cleans the room'
Dit komt niet goed over
lit. This comes not well over, 'This does not come across well'

In the cases (3a), the nominal head can be modified additionally, and hence we get a syntactic split between the first and the second word. The same applies to the adjectival head in (3b). The three verbal phrases in (3c) are all examples of so-called separable complex verbs (cf. Section 3.3). The non-verbal part is split off from the verb in main clauses (Booij 2010; Los et al. 2012). The word combinations in (1) that are classified as compounds, on the other hand, cannot be split. In the case of compound verbs this is clear from their not being split in main clauses:

(4) **opoe-goede-fiets* 'grandma-good-bike' / *goede opoefiets* 'good grandma's bike'
**rood-grote-baars* 'red-big-bass' / *grote roodbaars* 'big red-bass'

**donker-diep-geel* 'dark-deep-yellow' / *diep donkergeel* 'deeply dark-yellow'
**Jan pleegde zijn ouders raad* / *Jan raadpleegde zijn ouders* 'Jan consulted his parents'
**Hij koost zijn vrouw vaak lief* / *Hij liefkoost zijn vrouw vaak* 'He caresses his wife often'
**Dat komt mij niet weer over* / *Dat overkomt mij niet weer* 'This will not happen to me again'

There are two cases where it seems as if parts of compounds can be split. First, Dutch features gapping of parts of words: a compound constituent can be omitted under identity with another constituent of the same prosodic form in a phrase, as in:

(5a) *land- en tuinbouw* 'agri- and horticulture'
(5b) *voor- en achterkant* 'front- and back-side'
(5c) *ere- en eerste divisie* lit. honour- and first division, 'premier and first league'
(5d) *natuurbeheerders en -beschermers* 'nature managers and -protectors'

However, as shown in Booij (1985), this kind of ellipsis is not syntactic in nature. Instead, it is a prosodic process in which one of two identical prosodic words is omitted. Both in compounds and phrases, the word constituents correspond to separate prosodic words (also referred to as 'phonological words'). That is, this type of gapping is phonological in nature. This explains why a compound constituent like *divisie* in *eredivisie* can be omitted under identity with a separate word *divisie*, as in (5c): they are identical prosodic words, although their morpho-syntactic status is different.

The second type of split is found in phrases with coordinated elative compounds (cf. Hoeksema 2012) such as:

(6) *door- en doornat* lit. through- and through-wet, 'very wet'
dood- en doodziek lit. dead- and dead-ill, 'very ill'

In elative compounds the first part functions as an intensifier. Again, this is not a case of syntactic gapping. We cannot assume underlying structures like *doornat en doornat* or *doodziek en doodziek* as the sources of the phrases in (6) since such phrases are ill-formed. Instead, what is at stake here is the repetition of an intensifier word in the left part of a compound, a case of word-internal coordination.

2.2 Orthography

A+A compounds and A+A phrases are not always that easy to distinguish. In A+A phrases, the first adjective functions as an adverb. However, Dutch adjectives can be used as adverbs without being morphologically marked as such. Hence, when we come across an A+A sequence such as *jong getrouwd* lit. 'young married' this word sequence can be interpreted either as a compound or as a phrase. The difference between compound and phrase is primarily a semantic one. When we spell *jonggetrouwd*, it is considered a compound with a naming, classifying function, and the meaning is 'recently married'. When we use the phrase *jong getrouwd*, the phrase has a descriptive function 'married at a young age'. In the latter case, we can modify the adjective *jong*:

(7) *Ze zijn nogal jong getrouwd*
lit. They are rather young married
'They have married at a rather young age'

The orthography thus expresses a primarily semantic distinction here. Lexicalized word combinations may be felt as one word (the process of univerbation), have lost their syntactic flexibility, and are therefore spelled as one word. Thus, spelling may reflect lexicalization and univerbation.

However, orthography is not always revealing when we try to determine the status of Dutch word combinations. This is the case for separable complex words: the two parts of a separable complex verb are spelled as one word, without internal space, when they are adjacent:

(8) *Matthias was de kamer aan het schoonmaken* 'Matthias was cleaning the room'
Ik merkte dat de boodschap niet overkwam 'I noticed that the message did not come across'

This spelling convention reflects that these word combinations are felt as lexical units, with often idiosyncratic meaning aspects. On the other hand, these separable complex verbs are not words in the morphological sense, as they cannot appear in second position in main clauses. In Section 3.4 I will come back to this issue.

Dutch orthography requires compounds to be written without an internal space. However, many users of Dutch occasionally do insert a space between the two parts of a compound. This may be partially due to the influence of English orthography in which many compounds are written with an internal space.

Another factor might be that from a phonological point of view compounds are similar to phrases in that each constituent word forms a phonological word of its own. For instance, the N+N compound *tandextractie* 'tooth-extraction' consists of the phonological words /tɑnd/ and /ɛkstrɑksi/. These two words form separate domains of syllabification. Hence, the first part *tand* is a syllable of its own. This implies that the underling final /d/ of *tand* is in syllable-final position, and not in the onset of a syllable with the vowel /ɛ/ as its nucleus. It is therefore subject to the constraint of Dutch that obstruents are voiceless in coda position (Auslautverhärtung), and thus *tand* is pronounced as [tɑnt], and the phonetic form of *tandextractie* is [tɑntɛkstrɑksi].

This phonological similarity between compound constituents and phrasal constituents, which both consist of more than one phonological word, may lead to uncertainty as to how spell compounds properly.

2.3 Phonological properties

Are there phonological properties that distinguish compounds from phrases? In the case of nominal compounds, main stress is in most cases on the first constituent, but there are exceptions, such as *boerenzóon* 'farmer's son'. In nominal phrases, on the other hand, main stress is on the head, except when contrastive stress is involved. That is, the location of stress is dependent on information structure. Thus, stress location may not always differentiate between nominal compounds and nominal phrases, but does so in pairs like *ópoefiets* (compound) versus *opoe's fíets* (phrase). A+A compounds and A+A phrases also vary in stress location, again dependent on information structure, that is, on what counts as new and what as old information. For instance, the A+A compound *donker+geel* can be pronounced as *donker+géel* or, with emphatic or contrastive stress, as *dónker+geel*. Hence, stress location does not provide an unambiguous clue to the formal status of A+A sequences.

In verbal compounds of the type N+V and A+V, main stress is on the N and A respectively. The same applies to the corresponding separable complex verbs. Therefore, stress location cannot be used to distinguish between these compound verbs and the corresponding separable complex verbs. In verbal compounds with prepositions or adverbs as first constituents, however, main stress is on the second constituent, whereas in the corresponding separable complex verbs it is located on the non-verbal part. Thus we get a contrast between, for instance *over+kómen* 'to happen to' (compound) versus *óver+komen* 'to come across' (particle verb). Hence, stress can differentiate here between compounds and phrasal predicates. Because of this stress difference, the unstressed first constituents of these complex

words may be considered prefixes, as Dutch native verbalizing prefixes such as *be-* do not carry the main stress of a complex verb either (cf. Section 3.4).

2.4 Morphological properties

Morphological properties can also be used to distinguish compounds from phrases. In present-day Dutch there is no regular case marking anymore. Hence, when morphemes such as *s*, *en*, or *e*, historically case or stem endings, appear in the middle of a word sequence, they are linking elements, as in:

(9) *koning+s+zoon* 'king's son'
her+en+huis lit. gentleman's house, 'mansion'
zonn+e+schijn[3] 'sun shine'

The presence of a linking element is a clear mark of compound status. The only apparent exceptions to this criterion are nouns used in the possessive construction (Booij 2010: 216–222). The N+N sequence *opoe-s fiets* 'grandma's bike', for example, is a phrase: the *-s* is not a linking element here, but a marker of the possessive construction. This word sequence exhibits the normal flexibility of phrases, witness a phrase like *opoe's zwarte fiets* 'grandma's black bike'. The stress pattern is also revealing, as in this word sequence the word *fiets* can carry main stress.

In the case of A+N sequences, the presence of the inflectional ending *-e* on the adjectives reveals the phrasal status of such sequences. In Dutch, prenominal adjectives have an ending *-e*, unless the noun phrase as a whole is singular indefinite, and the head noun is neuter. In the examples (10), the noun *boek* 'book' is neuter, and the word *vrouw* 'woman' has common gender:

(10) *een goed boek* 'a good book'
het goed-e boek 'the good book'
(*de*) *goed-e boeken* '(the)good books'

een mooi-e vrouw 'a beautiful woman'
de mooi-e vrouw 'the beautiful woman'
(*de*) *mooi-e vrouwen* '(the) beautiful women'

3 In *zonneschijn*, the final schwa of *zonne* 'sun' has disappeared in present-day Dutch, and *zon* is now the Dutch word for 'sun'.

The inflection of prenominal adjectives indicates that these adjectives are words by themselves; within compounds an adjectival modifier cannot be inflected (compare the compound *snel+trein* 'fast train, intercity train' with *snelle trein* 'fast train'). It is only the head of a compound that can carry inflectional markers.

There are two complications, however. The first one is that in some types of noun phrases the adjective does not carry an overt inflectional marker (Booij 2002a: 43ff.; Tummers 2005). This applies to adjectives ending in *-en* /ən/ (11a), where a sequence of two syllables with a schwa as vowel is avoided. It also holds for adjectives in A+N phrases that denote an individual (11b), the function of an individual (11c), or an institution (11d), where the presence of the inflectional marker *-e* is optional:

(11a) *het open* / **opene boek* 'the open book'
(11b) *een wijs* / *wijze man* 'a wise man'
(11c) *een toegepast* / *toegepaste taalkundige* 'an applied linguist'
(11d) *het gemeentelijk* / *gemeentelijke museum* 'the municipal museum'

In these cases, the absence of the inflectional ending *-e* should not be taken as an indication of compound status. The stress pattern is that of noun phrases, with main stress on the nominal head.

The second complication is that some A+N phrases with inflected adjectives have undergone univerbation, and are now considered as one word, as reflected in the orthography:

(12a) *jonge+mán* 'young man'
rode+kóol 'red cabbage'
(12b) *hóge+priester* 'high priest'
wítte+brood 'white bread'

The words in (12a) have final stress, like phrases, but the words in (12b) carry initial stress. The word status of these A+N sequences can be concluded from the way in which they form diminutives, in contrast to regular phrases:

(13) *een jongemannetje* 'a little boy' versus *een jong mannetje* 'a young little man'
een wittebroodje 'a small white sandwich' versus *een wit broodje* 'a white small loaf of bread'

Diminutives are neuter nouns, and hence they require a prenominal adjective without *-e* in indefinite singular phrases of which they form the head. The exam-

ples in (13) show that both uses of the same A+N sequence are sometimes possible. In their use as words, they function as names, whereas in their phrasal use they have a descriptive interpretation.

A+N phrases frequently occur as left constituents of nominal compounds, as in

(14) [*oude*+*mannen*]+*huis* 'old men's home'
[*hete*+*lucht*]+*ballon* 'hot air balloon'
[*zwarte*+*kousen*]+*kerk* lit. black stockings church, 'orthodox protestant church'

These sequences are words, and they are to be written without internal spaces: *oudemannenhuis*, *heteluchtballon*, *zwartekousenkerk*. The inflectional ending *-e* of the adjectives *oude*, *hete* and *zwarte* shows that here A+N phrases have been made parts of words. In the orthography, these compounds can be distinguished from phrases like *oude mannenhuis* 'old house for men' and *hete luchtballon* 'air balloon that is hot'. The presence of a linking element *s* after the phrasal constituent confirms the compound status, as in *oude-dag-s-voorziening* lit. old-day-s-provision, 'pension'.

In conclusion, there are a number of criteria for distinguishing between compounds and phrases. In a few cases two structural interpretations of two-word-sequences are possible, and in this case there is variation in the way language users deal with such word sequences.

3 Competition and complementarity in naming

In this section I discuss how compounds and phrases with a naming function complement each other, or are in competition. In Section 3.1 I discuss the competition between A+N and N+N compounds on the one hand, and A+N phrases on the other. Section 3.2 deals with N+A compounds and phrases that express a comparison. In Section 3.3 we have a look at the complementarity of N+V compounding and N+V phrases. Section 3.4 analyses the relation between particle verbs and compound verbs with a prepositional or adverbial first constituent, and Section 3.5 deals with the nominalization of particle verbs by means of compounding.

3.1 Nominal compounds and A+N phrases

As pointed out by Schlücker (2014), the main, though not the only, function of A+N and N+N compounds is that of classification. These words create names for subclasses of entities. The same classifying function can be performed by A+N phrases (Booij 2002b, 2009a, 2010: 183ff.). Compare first N+N compounds with A+N phrases:

(15)	*atoom+fysica*	*atom-aire fysica*
	'nuclear physics'	'nuclear physics'
	structuur+analyse	*structur-ele analyse*
	'structure analysis'	'structural analysis'
	konings+huis	*konink-lijk huis*
	'king-s house'	'royal house'
	muziek+scholing	*muzik-ale scholing*
	'music(al) training'	'music(al) training'
	wetenschaps+beleid	*wetenschapp-elijk beleid*
	'science policy'	'science policy'

In (15) we see that an N+N compound may correspond to an A+N phrase. Typically, in these phrases the adjective is a denominal adjective that belongs to the class of relational adjectives. This is a productive class of adjectives in Dutch, mainly, but not exclusively non-native in character. Both options are grammatical, and both types function as names. This may be expected for these A+N phrases since relational adjectives do not describe properties, but denote the relation between the head noun of the phrase and the base noun of the adjective. In principle both options are available, and which one is used is partially a matter of convention. For me as speaker of Dutch, *muzikale scholing* is the conventional name for this type of education, but *muziekscholing* is also found on the internet. The compound *koning-s-besluit* 'king-s-decision' is not used as an alternative for the A+N phrase *koninklijk besluit* 'royal decision', nor *koningsfamilie* 'king-family' besides *koninklijke familie* 'royal family', even though these N+N compounds are well-formed. The advantage of using the adjective *koninklijk* 'royal' instead of the compound constituent *koning* 'king' is that it may also be used for denoting queens.

This kind of competition between words and phrases is similar to the competition between words that is known as 'blocking'. Blocking is the phenomenon that the formation of a complex word is blocked by the existence of another (simplex or complex) word with the same meaning. The formation of the deverbal noun *lieg-er* 'liar' in Dutch, for instance, is blocked by the existing complex word

leugenaar 'liar'. This does not mean that *lieger* is ill formed, but that it does not belong to the language convention (the norm) of the Dutch-speaking community. The fact that we find this type of competition between words and phrases as well confirms that both types of lexical units must be stored in the lexicon, and that the use of one of the relevant (morphological or syntactic) constructions for the formation of a new expression can be blocked by a stored instantiation of a competing construction. This implies that there cannot be a strict separation of morphology and syntax in the grammar of Dutch.

The second type of competition is that between A+N compounds and A+N phrases, a topic discussed in Hüning (2010), Hüning/Schlücker (2010), Schlücker (2014) and Schuster (2016). Here are some examples:

(16a)	A+N compound	classifying or descriptive A+N phrase
	rood+koraal 'red coral'	*rode koraal* 'red coral'
	rood+vos 'red fox'	*rode vos* 'red fox'
	**rood+wijn* 'red wine'	*rode wijn* 'red wine'
(16b)	A+N compound	descriptive A+N phrase
	hard+glas 'safety glass'	*hard glas* 'hard glass'
	hard+hout 'hardwood'	*hard hout* 'hard wood'
	rood+huid 'redskin, Indian'	*rode huid* 'red skin'

The compounds have initial stress on the first constituent, the phrases carry stress on the head noun, that is, final stress. The data in (16a) illustrate that both A+N compounds and A+N phrases are possible as names, and do not necessarily block each other. A compound such as *roodwijn*, however, is odd. In some cases the compounds differ in semantic interpretation from the phrasal correlates, as shown in (16b): the compounds are names, but the corresponding phrases are used as descriptions.

A+N phrases that function as names have a restricted syntax compared to other A+N phrases (Booij 2010: 178): they cannot be modified, or split by another word. For instance, we cannot say **heel gele koorts* 'very yellow fever', and a phrase like *gele and hevige koorts* 'yellow and high fever' is also odd. When we coin the phrase *heel rode wijn* 'very red wine', we coerce *rode wijn* into a description, denoting wine with a very red color. This lack of syntactic flexibility of phrases with a naming function makes them more similar to compounds than other kinds of phrases.

Dutch more often opts for A+N phrases as names for entities in comparison to A+N compounds than German (Booij 2002b; Hüning 2010). There are two structural factors that play a role in this difference. First, given the rich adjectival inflection of German, A+N phrases in German have quite a number of different

forms, whereas in Dutch there is only marginal variation in the shape of the adjective (usually ending in *-e*, occasionally in ø). Hence, in the case of German the compound option has the advantage of reducing the form variation of the adjective, as only its stem is used (Hüning 2010). For instance, the Dutch phrase *rode wijn* 'red wine' and the German compound *Rot-wein* both have a constant form for the adjective (*rode*/*rot*). This makes use of the phrasal alternative more feasible for Dutch. A second factor is that in Dutch A+N compounds the adjective has to be simplex (Schlücker 2014). This excludes the use of relational adjectives in A+N compounds. For instance, the compound *wetenscháppelijk+domein* 'scientific domain' is ill-formed, whereas this combination is fine as a phrase: *wetenschappelijk doméin*. This restriction also excludes the use of the various non-native relational adjectives in A+N compounds, a common pattern in German A+N compounds:

(17)

Dutch phrase	German compound
collectieve schuld	*Kollektiv+schuld* 'collective guilt'
nationale vlag	*National+flagge* 'national flag'
primaire literatuur	*Primär+literatur* 'primary literature'
sociale verzekering	*Sozial+versicherung* 'social security'
verbale aanval	*Verbal+attacke* 'verbal attack'

This does not mean that A+N compounds with non-native adjectives are completely excluded in Dutch, but they are relatively rare, and often considered as loan translations form German (Schlücker 2014: 234). This applies to compounds such as *nationaal-socialist* 'national-socialist', *normaal+verdeling* 'standard distribution', *speciaal+zaak* 'specialist shop', and *spectraal+analyse* 'spectral analysis'.

As to the choice between A+N compounds and A+N phrases, it has been argued for German that paradigmatic analogy plays an important role (Schlücker/Plag 2011; Rainer 2013; Schlücker, this volume). Schlücker/Plag (2011: 1546) argue that "the larger the compound family of an item, the more likely it is that participants choose the compound, and the larger the phrasal family of an item, the more likely it is that participants choose the phrase". This role of paradigmatic analogy in the choice between compounds and phrases has been confirmed for Dutch by Schuster (2016) on the basis of an investigation of Dutch dictionaries and corpora.

The role of paradigmatic analogy can be observed in the use of color adjectives. For example, Dutch color adjectives such as *geel* 'yellow', *rood* 'red', and *zwart* 'black' are used in A+N compounds that function as names for animals and for human beings (in some cases with a possessive interpretation):

(18) *geel+bek* lit. yellow+mouth, 'fledgling'
geel+gors 'yellow hammer (type of bird)'
geel+vink 'serin finch'

rood+forel 'red trout'
rood+baard lit. red+beard, 'person with read beard'
rood+staart lit. red+tail, 'redstart (bird with red tail)'

zwart+hemd lit. black+shirt, 'fascist'
zwart+kop 'black-cap (type of bird)'
zwart+rok lit. black+coat, 'person wearing a blackcoat'

On the other hand, we find these color adjectives in phrasal names such as *gele kaart* 'yellow card' and *rode kaart* 'red card', names for the cards used for indicating improper actions in a football match (a *kaart*-family). Likewise, there is a family of phrasal names with *zwart* 'black', as in *zwarte markt* 'black market, *zwart geld* 'black money', *zwarte doos* 'black box', and *zwarte kunst* 'black magic', a *zwart*-family with *zwart* being used with the meaning 'illegal, opaque'. These observations confirm that analogy to similar compounds or phrases plays an important role in the choice between compound and phrase.

3.2 N+A compounds and adjectival phrases

Dutch N+A compounds can be used as an alternative to phrases that express a comparison (Hoeksema 2012: 7):

(19)

compound	adjective phrase	gloss
dons+zacht	(*zo*) *zacht als dons*	'soft as down'
honds+trouw	(*zo*) *trouw als een hond*	'faithful as a dog'
ijs+koud	(*zo*) *koud als ijs*	'cold as ice'
kaars+recht	(*zo*) *recht als een kaars*	'straight as a candle'
sneeuw+wit	(*zo*) *wit als sneeuw*	'white as snow'

According to Hoeksema (2012) the choice of the compound structure over the phrasal alternative is determined by two advantages of the compound option: compactness and expressiveness. There is always a phrasal alternative for the compound, but not vice versa. For instance, the comparison *sterk als een paard* 'strong as a horse' cannot be expressed by the compound *paardesterk*. The phrasal alternative might, however, not carry exactly the same meaning: *ijzersterk* 'iron-strong' can be used in contexts where the phrasal expression is odd.

For instance, *een ijzersterk verhaal* 'a very strong story' cannot be properly paraphrased as *een verhaal sterk als ijzer* 'a story strong as iron' (ibid.). Similar observations have been made for German (Schlücker, this volume), and Italian (Masini, this volume). The same applies to compounds with *reuze* (an allomorph of *reus* 'giant'), as in *reuze-groot* 'giant-big, very big' where the phrase *zo groot als een reus* 'as big as a giant' may not be a proper paraphrase. In these compounds the nouns *ijzer* and *reuze* have acquired a more general meaning of intensification. These compounds are called elative compounds and express that the property denoted by the head is present to a high degree. This elative use is the source of the development of these nouns into intensifier affixoids. For instance, besides *bloed+rood* 'red as blood' we find compounds like *bloed+saai* lit. blood-boring, 'very boring' and *bloed+mooi* lit. bloed-beautiful, 'very beautiful', which cannot be paraphrased as *saai* / *mooi als bloed* 'boring / beautiful as blood'.

This difference between compounds and phrases can also be observed for another class of N+A compounds of the type *dood+ziek* lit. dead-ill, 'so ill that it may cause death'. Again, some of these nominal modifiers have acquired a more general meaning of intensification, and in such cases a phrasal paraphrase is not adequate:

(20) *dood+gewoon* 'very ordinary'
dood+simpel 'very simple'

This development of nominal (and other) modifiers into affixoids, that is, words with a more abstract meaning of intensification when embedded in compounds, is discussed in detail in Booij/Hüning (2014) and Hüning/Booij (2014).

3.3 N+V compounds and phrases

Unlike nominal and adjectival compounding, the formation of verbal compounds is not a productive process in Dutch. This does not mean that there are no verbal compounds whatsoever. The main source of such compounds is backformation from nominal compounds with the form $[[N][V\text{-}er]_N]_N$ or $[[N][V\text{-}ing]_N]_N$. Examples are:

(21)	*beeld+houwen*'	< *beeld+houw+er*
	lit. to image-hew, 'to sculpture'	'sculptor'
	honger+staken	< *honger+stak+ing*
	lit. to hunger-strike, 'to go on hungerstrike'	'hungerstrike'

vaat+wassen	< *vaat+wass+er*
'to dish-wash'	'dish washer'
tekst+verwerken	< *tekst+verwerk+ing*
'to text-process'	'text processing'

A second type of verbal compounds are verbs like *klapper+tanden* lit. chatter-tooth.INF, 'to have chattering teeth' and *kwispel+staarten* lit. wag-tail.INF, 'to wag one's tail'. They have the structure $[VN]_V$, and are exceptional in that they are left-headed. There are also a few V+V compounds like *hoeste+proesten* lit. to cough-sneeze, 'to cough and sneeze', but again, this is not a productive process of word formation (Booij 2002a: 164f.).

The productive alternative for N+V compounds are phrasal word sequences that consist of a bare noun and a verb. An example is the N+V sequence *piano+spelen* 'to play the piano'. This word sequence can be used as a verb phrase, but the noun can also be quasi-incorporated into the verb:

(22) ... *dat Julian* {*piano kan spelen* / *kan pianospelen*}
... that Julian {piano can play / can play piano}
'... that Julian can play the piano'

Verb phrases with a bare noun are often used as names for denoting a certain kind of activity. For instance, *piano spelen* is a specific type of musical activity. The word *piano* does not denote a specific referent here. This may be contrasted with a verb phrase like *de piano bespelen* 'to play on the piano', where, by using a definite noun phrase, the identifiability of a specific referent of *piano* is presupposed. When count nouns are used as bare nouns, without the normally expected determiner, this evokes an interpretation as name instead of description of the verbal phrase in which that bare noun is used. Note that in a compound like *pianospeler* 'piano player' the word *piano* likewise has no referential power.

In the second variant in (22), the noun and the verb form a syntactically closer unit than in the first variant, and are adjacent. This unit can be qualified as a case of quasi-noun incorporation. Noun incorporation is the process in which a noun is incorporated into a verb, and thus creates a verbal compound. However, in Dutch the incorporation process does not lead to compounds in the morphological sense. This is shown by the fact that the N+V sequence cannot appear in the position for finite verbs (the second position) in main clauses, unlike a real verbal compound like *beeldhouwen* 'to sculpture':

(23) *Julian* {**pianospeelt graag* / *speelt graag piano*}
'Julian likes playing the piano'

Amber beeldhouwt graag
'Amber likes sculpturing'

This is why Dahl (2004) calls this process quasi-incorporation: there is incorporation and formation of lexical units, but these lexical units are not words. Quasi-noun incorporation in Dutch is discussed in detail in Booij (2010: Chapter 4), and the account below is mainly based on this chapter.

The strong bond between N and V in the incorporated variant can also be seen in two syntactic constructions, the verb raising construction and the progressive construction. In the verb raising construction the verb of the main clause forms a unit with the verb of the embedded clause. The incorporated noun can appear in between the two verbs (24a), whereas this is impossible for a full noun phrase (24b). The first option in (24a) is that with quasi-incorporation, and Dutch orthography requires the quasi-incorporated word combination to be spelled as one word, without an internal space:

(24a) ... *dat Barbara* {*wil pianospelen* / *piano wil spelen*}
... that Barbara {wants pianoplay / piano wants play}
'... that Barbara wants to play the piano'
(24b) ... *dat Barbara* {**wil de piano bespelen* / *de piano wil bespelen*}
... that Barbara {wants the piano play / the piano wants play}
'... that Barbara wants to play on the piano'

The second construction that functions as a litmus test for quasi-noun incorporation is the progressive construction of the form *aan het V-infinitive*:

(25) *Matthias is aan het lezen*
Matthias is at the read.INF
'Matthias is reading'

Matthias is {*aan het pianospelen* / *piano aan het spelen*}
Matthias is {at the piano-play.INF / piano at the play.INF}
'Matthias is playing the piano'

Matthias is {*de piano aan het bespelen* / **aan het de piano bespelen*}
Matthias is {the piano at the PREF.play.INF / at the piano PREF.play.INF}
'Matthias is playing on the piano'

Verbs with an incorporated noun can function as a unit in the progressive construction, and thus appear after *aan het*. This applies to the N+V sequence *piano+spelen*. On the other hand, the prefixed verb *bespelen* 'to play on' is an

obligatorily transitive verb that does not allow for noun incorporation. Like verbal phrases with bare nouns, the quasi-incorporation structure is used to express that the action referred to is a conventional action. In other words, it creates names for types of action. Whatever is considered as a conventional action by the language user can be expressed in this form. For instance, *auto+wassen* 'to wash cars' is a conventional action, whereas buying a car is not conceived as a conventional action, and therefore there is no verb phrase *auto kopen*, or quasi-compound *autokopen* (instead, the proper phrase for naming this action is *een auto kopen*, with an indefinite determiner). Hence the difference in syntactic behavior between *auto+wassen* en *auto+kopen*:

(26) ... *dat Peter gaat* {*auto+wassen* / **auto+kopen*}
... that Peter goes {car+wash.NF / car+buy.INF}
'... that Peter is going to {car+wash / *car+buy}'

Conventional actions can also be expressed with verbs + plural nouns. For instance, *aardappels schillen* lit. potatoes-peel, 'to peel potatoes' can be conceived as a conventional action, and hence we can say:

(27) *Geert is aan het aardappels schillen* 'Geert is peeling the potatoes'
... *dat Geert wil aardappels schillen* '... that Geert wants to peel potatoes'

However, when the noun is plural, the N+V sequence is not spelled as one word.

The use of the term 'quasi-incorporation' may suggest that these quasi-compounds always derive from a regular phrase, but this is not the case. There are many N+V sequences where the bare noun cannot be interpreted as an object-NP. This applies to, for instance, the following cases (Booij 2010: 112):

(28) *buik+spreken* lit. to stomach speak, 'ventriloquizing'
koord+dansen lit. to rope dance, 'walking a tightrope'
mast+klimmen lit. to pole climb, 'climbing the greasy pole'
steen+grillen lit. to stone grill, 'stone-grilling'
stijl+dansen lit. to style dance, 'ballroom-dancing'
vinger+verven lit. to finger paint, 'finger-painting'
zak+lopen lit. to bag walk, 'running a sack-race'
zee+zeilen lit. to see sail, 'ocean-sailing'

These quasi-compounds are referred to as immobile verbs in the linguistic literature (cf. Vikner 2005), because they cannot appear in second position, as illus-

trated here for *zee-zeilen* (29a). At the same time, they cannot be split (29b), but are fine if they are not split (29c, d):

(29a) **Mijn vader zee+zeilt vaak*
My father sea+sails often
'My father often sails at sea'
(29b) **Mijn vader zeilt vaak zee*
My father sails often see
'My father often sails at sea'
(29c) *Mijn vader is vaak aan het zee+zeilen*
My father is often at the sea+sail.INF
'My father often sails at sea'
(29d) *... dat mijn vader vaak zee+zeilt*
... that my father often sea+sails
'... that my father often sails at sea'

The conclusion drawn from these facts in Booij (2010: Chapter 4) is that there are N+V combinations that are neither regular compounds nor regular syntactic phrases. Instead, they are quasi-compounds without a corresponding verbal phrase: a word sequence such as *zee zeilen* cannot be used as a well-formed phrase.

For a proper account of the distribution of quasi-compounds, their structure should be different from that of phrases and that of morphological compounds. They may be considered syntactic compounds. In a syntactic verbal compound a bare N^0 is adjoined to a V^0, and together they form a V^0:

(30) $[[zee]_{N0}\ [zeil]_{V0}]_{V0}$

Their syntactic compound status prohibits them from being split in main clauses (29a). At the same time they cannot appear in second position in main clauses as this position allows only for a single verb (29b). When the bare noun functions as an object, as in *pianospelen*, the quasi-compound corresponds with a verbal phrase with a bare noun that can be split. Hence, the two possible word orders in sentences like (22). Thus, the grammar of Dutch provides three different structures for N+V combinations that function as names:

(31)

morphological compound	$[[honger]_N[staak]_V]_V$	$[[vaat]_N\ [was]_V]_V$
syntactic compound	$[[piano]_{N0}\ [speel]_{V0}]_{V0}$	$[[zee]_{N0}\ [zeil]_{V0}]_{V0}$
verb phrase	$[[[piano]_{N0}]_{NP}\ [speel]_{V0}]_{VP}$	

Since quasi-compounds cannot be used as finite verbal forms in main clauses, the usual strategy is to use the progressive *aan het V-infinitive*-construction as an alternative, as illustrated by the sentences in (25).

This type of quasi-compound structure is also possible for A+V combinations:

(32) *dood+vriezen* lit. dead+freeze, 'freeze to death'
goed+keuren lit. good+judge, 'to approve'
schoon+maken lit. clean+make, 'to clean'
vreemd+gaan lit. strange+go, 'to sleep around'
vrij+geven lit. free+give, 'to release'
wit+wassen lit. white+wash, 'money-laundering'
zoek+maken lit. missing+make, 'to mislay'

These A+V combinations are not words in the morphological sense, and are therefore split in main clauses, just like the N+V combinations. They exhibit the same word order variation as that shown in (22):

(33) ... *dat de directeur het voorstel* {*wilde goedkeuren* / *goed wilde keuren*}
... that the director the proposal {wanted good-judge / good wanted judge}
'... that the director wanted to approve the proposal'
... *dat Ton het boek* {*heeft zoekgemaakt* / *zoek heeft gemaakt*}
... that Ton the book {has missing-made / missing has made}
'... that Ton has mislaid the book'

In other words, what we see here are A+V combinations, often idiosyncratic in meaning, that are structurally interpreted either as verbal phrases with a bare adjective as complement, or as quasi-compounds.

Both types of compounds have past participles in which the participial prefix *ge-* appears before the verbal stem, which confirms their phrasal status:

(34) *Jan heeft piano+gespeeld* 'Jan has played the piano'
Wij hebben dit voorstel goed+gekeurd 'We have approved this proposal'

The adjectives of the quasi-compounds cannot be modified, that is, they cannot head an adjectival phrase. When we add a modifier, this leads to an ungrammatical result, or another, more literal interpretation. For instance, the verb phrase *heel vreemd gaan* lit. very strange go, 'to go very strange', with the degree adverb *heel*, cannot be interpreted as 'sleep around intensively'.

In conclusion, the lack of productivity of verbal compounding in Dutch is compensated by the availability of (i) verbal phrases with a bare noun or adjective as complement such as *piano spelen* and *goed keuren*, and (ii) by quasi-compounds with a verbal head and a nominal or adjectival adjunct (spelled without an internal space) such as *pianospelen*, *goedkeuren*, and *zeezeilen*. They function as names for conventional, nameworthy activities. The class of quasi-compounds is larger than that of the verbal phrases with bare complements, because in quasi-compounds the noun need not be licensed syntactically by the verb. For instance, in *zeezeilen*, the noun *zee* does not function as an object-NP, and hence its occurrence is not licensed by syntax. Nevertheless, it can combine with a verb into a syntactic compound.

3.4 Prefixed verbs and particle verbs

Dutch has a number of complex verbs which might be considered compounds because they consist of a preposition or an adverb followed by a verbal stem:

(35a) *aan+bidden* lit. at+pray, 'worship'
achter+halen lit. behind+fetch, 'recover'
voor+komen lit. for+come, 'prevent'
(35b) *door+zoeken* lit. through+search, 'search through'
om+geven lit. around+give, 'surround'
onder+schatten under+estimate, 'underestimate'
over+spoelen lit. over+wash, 'wash over'
(35c) *mis+lukken* lit. wrong+succeed, 'fail'
weer+houden lit. back+hold, 'restrain'
vol+brengen lit. full+bring, 'to finish'

The types of verb with *aan-*, *achter-* and *voor-* exemplified in (35a) are unproductive, just like those with the adverbs *mis-* and *weer-* and the adjective *vol-* shown in (35c). The types exemplified in (35b) with *door-*, *om-*, *onder-*, and *over-*, however, are productive. In reference grammars of Dutch they are usually considered prefixed words, because unlike what is normally the case for Dutch compounds, the main stress in these words is located on the second constituent (instead of the first constituent). Thus, from the point of view of stress location, they pattern with prefixed verbs such as *be-hálen* 'to acquire' and *ver-zóeken* 'to request'. Moreover, the meaning contribution of these morphemes in complex verbs may differ from that of the corresponding morphemes when used as words by them-

selves. In other words, these words have grammaticalized into prefix-like morphemes. Prefixes like *be-* and *ver-* also originate from words that are parts of compounds, but their phonological form has been reduced as well, with a reduced vowel /ə/. Hence, in present-day Dutch there are no identical lexical counterparts for these prefixes.

The number of productive processes of verbalizing prefixation in Dutch is quite restricted, and therefore, there is a huge range of meanings for the expression of which phrasal verbal predicates with a corresponding make-up can be used. This is the class of particle verbs, with the particles corresponding to prepositions like *binnen* 'inside', postpositions like *mee* 'with', and adverbs like *neer* 'down'. The number of types is quite big, and I list here only a few for the purpose of illustration. Complete lists can be found in De Haas/Trommelen (1993), and on Taalportaal (www.taalportaal.org):

(36) *binnen+komen* lit. inside come, 'enter'
mee+vallen lit. with fall, 'turn out better than expected'
op+bellen lit. up phone, 'to phone up'
rond+lopen lit. around walk, 'walk around'
neer+vallen lit. down fall, 'to fall down'
weg+lopen lit. away walk, 'walk away'
voorop+lopen lit. in front walk, 'walk in front'

Particle verbs are lexical units, but phrasal in nature, just like verbal predicates such as *piano+spelen* and *schoon+maken* discussed in Section 3.3. They are split in main clauses, and can function as verbal phrases. At the same time, they can also be used as quasi-compounds, that is, behave like a tight syntactic unit in verb raising constructions. In this latter use, they are spelled as one word. These two syntactic options are illustrated by the following sentences:

(37) ... *dat Hans zijn moeder* {*op wilde bellen* / *wilde opbellen*}
... that Hans his mother {up wanted phone / wanted up-phone}
'... that Hans wanted to call his mother'

Morphologically, particle verbs also behave as phrases, since the prefix *ge-* of the past participle appears in between the particle and the verb: *op-ge-beld*, not **ge-op-beld*. When we nominalize a particle verb by means of the prefix *ge-*, it also appears before the verbal stem, as in *op-ge-bel* 'calling'.

The proper grammatical analysis of Dutch particle verbs is discussed in detail in Booij (2010: Chapter 5), and in Los et al. (2012). The gist of this analysis is that each class of particle verbs has to be represented in the grammar of Dutch as a

constructional idiom. Constructional schemas are schemas that specify the systematic correspondence between form and meaning of a construction. A constructional idiom is a constructional schema in which one or more slots are lexically fixed. Each type of particle verb will be represented by a constructional idiom with that particle specified. The meaning of the particle sometimes corresponds with that word used in isolation, but in other cases it has acquired a specific meaning. For instance, the particle *door* 'through' has acquired, among others, the aspectual meaning of 'to continue with', as in *door+fietsen* 'to continue cycling' and *door+eten* 'to continue eating', unlike the preposition *door* 'through'. Hence, I assume the following constructional idioms for *door+V*, one without, and one with quasi-incorporation. In the first case we have a phrasal verbal predicate, labeled as V', in the second case a syntactic compound:

(38)

form	$[[door]_{Prt}\ V_i]_{V'}$	≈	$[[door]_{Prt}\ V^0{}_i]_{V0}$
meaning	Continue SEM_i		Continue SEM_i

where SEM_i stands for the meaning of the verb V_i, and the symbol ≈ indicates the paradigmatic relationship between the two constructional schemas.

For a number of morphemes we saw that they are used in Dutch either as prefix or as particle. This applies in particular to *door*, *om*, *onder*, and *over*, which can be used productively as prefixes. In these cases there is no competition between prefixed verbs and particle verbs, but complementarity, since they differ in meaning. These morphemes in their prefixal use create transitive verbs that denote an action that completely affects the object in a specific manner, as illustrated in (39) (examples from Los et al. 2012: 184):

(39) *het huis door+zoeken*
the house through-search
'to search (through) the house'

het kasteel om+geven
the castle around-give
'to surround the castle'

het gebouw onder+kelderen
the building under-cellar
'to make a cellar under the building'

het land over+spoelen
the land over-wash
'to wash over the land'

There are a few minimal pairs for prefixed verbs / particle verbs with semantic differences, for example:

(40)	*door+zóeken*	*door+zoeken*
	lit. through-search, 'to search'	lit. through-search, 'to continue searching'
	voor+kómen	*vóor+komen*
	lit. for-come, 'to prevent'	lit. fore-come, 'to occur'

In sum, prefixed verbs and particle verbs coexist, the number of prefixed verb types is restricted, and the high number of particle verb types provides an extensive range of names for activities and events.

3.5 Nominalizations of particle verbs

Phrasal and morphological expressions exhibit an interesting type of cooperation in the nominalization of particle verbs. The crucial observation is that particle verbs often select an unproductive type of nominalization, and in that case they select the same unproductive nominalization type as the corresponding base verb (Booij 2015). In the default case, verbs are nominalized by means of the suffix *-ing*, or by using the infinitive form. A number of verbs, however, have an unproductive type of nominalization. For instance, the nominalization of *komen* 'to come' is *komst*, and the particle verb *aan+komen* 'to arrive' has the parallel nominalization *aan+komst* 'arrival'. In order to account for this parallelism, we should analyze *aankomst* as the compound $[[aan]_{Part}\ [kom\text{-}st]_N]_N$. Because *komst* is listed as derived word, it can combine with a particle into a compound. This implies that we are confronted with an asymmetry between meaning and form, since the nominalizing suffix *-st* has semantic scope over the particle verb *aankom* (the stem of *aankomen* 'to arrive') as a whole. This systematic choice of an unproductive type of nominalization by particle verbs is shown in (41) (data from Booij 2015):

(41)	verbal stem	nominalization
(41a)	no formal change (conversion)	
	val 'fall'	*val* 'fall'
	aan+val 'attack'	*aanval* 'attack'
	in+val 'raid'	*inval* 'raid'
(41b)	with vowel change	
	grijp 'seize'	*greep* 'grip'
	in+grijp 'interfere'	*ingreep* 'interference'
	mis+grijp 'miss one's hold'	*misgreep* 'blunder'

(41c) stem change and/or suffixation

gaan 'go'	*gang* 'going'
af+gaan 'fail'	*afgang* 'failure'
door+gaan 'continue'	*doorgang* 'taking place'
neer+gaan 'go down'	*neergang* 'going down'
op+gaan 'rise'	*opgang* 'rise'
in+gaan 'enter'	*ingang* 'entrance'
geef 'give'	*gave / gifte* 'gift'
aan+geef 'report'	*aangifte* 'report'
op+geef 'state'	*opgave* 'statement'
uit+geef 'spend'	*uitgave* 'expense'
kom 'come'	*kom-st* 'arrival'
aan+kom 'arrive'	*aankom-st* 'arrival'
op+kom 'rise'	*opkom-st* 'rise'

This observation concerning the selection of a particular unproductive type of nominalization for the particle verb is accounted for straightforwardly by an analysis in which nominalizations of particle verbs are compounds that consist of a particle plus the nominalized form of the simplex verb. Hence, the form part of the general construction schema for these particle verb nominalizations is:

(42) $[\text{Particle } [[x]_V\ z]_N]_N$

where $[[x]_V\ z]_N$ stands for the nominalized form of the simplex verb. The variable x stands for (an allomorph of) the verbal stem, and the variable z stands for a suffix or zero. All instantiations of unproductive types of nominalization have of course to be listed. Hence, listed nouns like *gang* and *komst* will be available for combining with a particle into a compound. Thus, it is predicted that the nominalized form of a particle verb corresponds to that of the nominalized form of the corresponding simplex verb.

The structure for compounds of the form (42) has to be available anyway in the grammar of Dutch, as there are a number of compounds of this form without a corresponding particle verb. This applies to, for instance, the following nouns:

(43)

compound word	lacking particle verb
af+dronk 'after-taste'	*afdrinken*
bij+slag 'bonus'	*bijslaan*
toe+gang 'access'	*toegaan*

The meaning of particle compounds has to be specified as being the nominalization of the corresponding particle verb, if available, which often has an idiosyncratic meaning. This is expressed by the following set of paradigmatically related constructional schemas:

(44) form $[\text{Particle}_i\ [[x]_{Vj}\ z]_N]_N \approx [\text{Particle}_i\ V_j]_{V'k}$
meaning Event of SEM_k SEM_k

Recall that the symbol ≈ denotes a paradigmatic relationship. The formal and semantic correspondences between the two schemas are specified by means of co-indexation. Such a schema of schemas is called a second order schema. For instance, given the particle verb *aankomen* with the meaning 'to arrive', second order schema (44) states that the compound noun *aankomst* is interpreted as the event of arriving.

This case shows that there might be an asymmetry between form and meaning in morphological constructions. The meaning of the particle compound is a compositional function of the meaning of the particle verb, even though the particle verb is not a formal subconstituent of the corresponding compound. Instead, there is a paradigmatic relationship between the particle compound schema and the schema for particle verbs. This kind of asymmetry can be accounted for by relating schemas paradigmatically in second order schemas (Booij/Masini 2015). Schema (44) is a second order schema, as it relates the constructional schema for particle compounds to the constructional schema for particle verbs.

This implies that the grammar of Dutch requires access to the meaning of phrasal lexical expressions in order to account for the meaning of particle compounds. This is another type of complementarity between compounds and phrasal lexical items, and shows again that we need a grammar in which morphological and phrasal lexical units can interact.

4 The construction of numeral expressions

Compounding and phrasal expressions are used in tandem in the construction of complex numeral expressions in Dutch (Booij 2010: Chapter 8). The use of compound structure can be observed in cardinal numbers like the following:

(45) *drie+honderd* 'three-hundred'
vijf+duizend 'five-thousand'

In the compounds *driehonderd* and *vijfduizend* there is a relation of multiplication between the first and the second constituent, the first constituent is the multiplier.[4] These numerals are spelled as one word.

Phrasal structure is used in the form of coordination by means of the conjunction *en* 'and', as in:

(46) *drie+en+zestig* 'three and sixty, 63' spelling: *drieënzestig*
honderd+(en)+drie 'hundred and three, 103' spelling: *honderd(en)drie*

In (46) we see the use of syntactic coordination by means of *en*. This corresponds with the semantic effect of addition. This phrasal pattern is subject to a specific restriction, however, that does not apply to syntactic coordination in general: there is a fixed order in which the two numbers have to appear, the lower digit before the higher digit in numbers < 100, the higher digit before the lower one in numbers > 100. You cannot say *zestig-en drie* '63' or *drie-en-honderd* '103'. Moreover, the conjunction *en* is optional in numbers > 100, an optionality that does not apply to regular coordination. In other words, phrasal coordination is used here for the expression of addition, but is subject to specific restrictions. Additional construction-specific properties for this use of coordination are that the conjunction *en* /ɛn/ can be pronounced either as [ɛn] or as [ən] in numbers < 100, and can be optionally omitted in numbers > 100.

Compounding and phrasal coordination are used together in the formation of complex numerals: the numeral compounds are building blocks of the coordination construction, as in:

(47) *acht+honderd(en)drie+en+twintig* 'eight hundred three and twenty, 823'

with the structure:

(48) $[[[acht]_{Num}\ [honderd]_{Num}]_{Num}\ ([en]_{Conj})\ [[drie]_{Num}\ [en]_{Conj}\ [twintig]_{Num}]_{NumP}]_{NumP}$

where Num = Numeral, and NumP = Numeral Phrase.

4 The word sequence *zes miljoen* 'six million' looks similar to these compounds, but is considered a phrase, as reflected by its spelling with an internal space. This means that *miljoen* is interpreted as a measure noun, similar to nouns like *gulden* 'guilder' and *uur* 'hour' which also appear in their singular form after a cardinal > 1: *drie gulden*, *drie uur*. However, this interpretation is not chosen for words with *honderd* and *duizend*. *Honderd*, *duizend*, en *miljoen* can all function as nouns, and may appear in plural form: *honderden*, *duizenden*, *miljoenen*.

The orthography of numerals reflects their hybrid nature. The compounds and the coordinated numerals are spelled as one word, except that there is a space after *duizend*. Moreover, the words *miljoen* and *miljard* are always spelled as separate words. Thus we get spellings like *achthonderd* (800), *drieëntwintig* (23), *achthonderdendrieëntwintig* (823), *tweeduizend drieënveertig* (2,043), and *vijf miljoen achthonderdduizend driehonderdentwintig* (5,800,320).

These numeral phrases seem to feed word formation in the construction of ordinals, as in:

(49) *acht+honderd(en)drie+en+twintig-ste* '823th'

The spelling of this ordinal is *achthonderd(en)drieëntwintigste*. The ordinal suffix *-ste* is attached to the last word of this complex expression, but its semantic scope includes the part *achthonderd* as well. Hence, we see another asymmetry here between the formal structure and the semantic interpretation of complex expressions.

5 Construction Grammar and Construction Morphology

The data discussed in Sections 3 and 4 provide strong evidence for a view of the organization of the grammar in which there is no strict separation between morphology and syntax. This is one of the core hypotheses of constructionist approaches to morphology and syntax. Here are the main points:

(i) Morphological and syntactic constructions may compete; both can be used for creating names, and hence, there are blocking effects between morphological and phrasal constructs.

(ii) Phrasal constructions may be subject to specific restrictions when used as names. For instance, in A+N phrasal names, the adjective cannot be separated from the head noun, nor be modified. In a constructionist approach we can account for the properties of such phrasal names by phrasal constructional schemas which derive from general syntactic schemas, but with specific formal and semantic properties specified. The same applies to the description of specific forms of coordination in the construction of complex numeral expressions.

(iii) Morphological processes may be unproductive, or unavailable for the expression of certain types of names. In Dutch, phrasal structures fill

those gaps, hence we may speak of periphrastic word formation. This is the case for separable complex verbs of various types: N+V, A+V and particle verbs. There is a clear complementarity between morphological and syntactic ways of creating names.

(iv) The interpretation of complex words may depend on the meaning of paradigmatically related phrasal lexical constructions. This is the case for nominalizations of particle verbs. Paradigmatic relationships between constructional schemas, morphological or phrasal, can be expressed by second order schemas.

These kinds of finding form underpinnings of the model of Construction Morphology proposed in Booij (2010), and further articulated in a number of publications on Dutch referred to in this article. In Construction Grammar (Hoffmann/Trousdale (eds.) 2013) and Construction Morphology, the grammar is seen as a multidimensional web of syntactic and morphological constructions of various degrees of abstractness. Constructional schemas form a hierarchy: more abstract schemas dominate more concrete ones, and constructions are instantiated by fully lexically specified constructions, which may be listed in the lexicon. For example, there are, in increasing order of concreteness, a general schema for Dutch right-headed compounds, a subschema for N+A compounds, a constructional idiom $[[dood]_N A]_A$ 'very A', and listed instantiations of this constructional idiom such as *doodziek* 'very ill' and *doodnormaal* 'very normal'. Syntactic constructions are also specified in terms of schemas. Phrasal names of the type A+N are specified by a subschema of the general schema for Noun Phrases, with certain restrictions imposed, such as linear adjacency of A and N and bareness of the adjective. Similarly, the grammar of Dutch contains a general syntactic schema for syntactic coordination, which dominates specific subschemas for numeral expressions in which the properties mentioned in Section 5 are specified. Thus, the idea of periphrastic word formation finds its natural expression in Construction Grammar.

Since in Construction Grammar both morphological schemas and syntactic schemas, and their lexicalized instantiations are listed, there is potentially a competition between morphological and syntactic expression of the same meaning. This predicts the observed blocking effects.

Paradigmatic relations between schemas and between concrete constructions are expressed by means of co-indexation. They give expression to the existence of word families and phrase families. The presence of a network of paradigmatic relations in the grammar provides a natural interpretation for the observation that paradigmatic analogy co-determines the choice between compound and phrase when coining a name.

The claim that morphology and syntax cannot be separated in grammar does not mean that there is no formal distinction between morphological and phrasal constructions. This formal distinction is necessary for a proper account of the syntactic behavior of the various types of names. At the same time, since compound schemas and phrasal schemas are not split in different components of the grammar, they can interact: phrasal constituents may form parts of compounds and vice versa, and compounds may function as nominalizations of particle verbs which themselves are phrasal expressions. These observations led to the conclusion that second order schemas (paradigmatic relations between constructional schemas) form part of the grammar.

Since morphology often derives historically from syntax, it should not come as a surprise that there are transitional cases such as quasi-compounds, verbs with incorporated particles, and cardinal numerals of the type *drieëntwintig* '23' where the conjunction *en* can also be interpreted as a linking element. These phenomena underscore Hermann Paul's remarks on the blurred boundary between syntax and word formation quoted in the introduction of this article. As we saw above, a Construction Grammar approach can do justice to these transitional cases.

References

Booij, Geert (1985): Coordination reduction in complex words: a case for prosodic phonology. In: Van der Hulst, Harry/Smith, Norval (eds.): Advances in non-linear phonology. Dordrecht: Foris. 143–160.

Booij, Geert (2002a): The morphology of Dutch. Oxford: Oxford University Press.

Booij, Geert (2002b): Constructional idioms, morphology, and the Dutch lexicon. In: Journal of Germanic Linguistics 14. 301–329.

Booij, Geert (2002c): Separable complex verbs in Dutch: a case of periphrastic word formation. In: Dehé, Nicole et al. (eds.): Verb-particle explorations. Berlin: De Gruyter. 21–42.

Booij, Geert (2009a): Phrasal names: A constructionist analysis. In: Word Structure 2. 219–240.

Booij, Geert (2009b): Lexical integrity as a morphological universal, a constructionist view. In: Scalise, Sergio/Magni, Elisabetha/Bisetto, Antonietta (eds.): Universals of Language Today. Dordrecht: Springer. 83–100.

Booij, Geert (2010): Construction Morphology. Oxford: Oxford University Press.

Booij, Geert (2015): The nominalization of Dutch particle verbs: schema unification and second order schemas. In: Nederlandse Taalkunde 20. 285–314.

Booij, Geert/Hüning, Matthias (2014): Affixoids and constructional idioms. In: Boogaart, Ronny/Colleman, Timothy/Rutten, Gijsbert (eds.): Extending the scope of Construction Grammar. Berlin: De Gruyter. 77–105.

Booij, Geert/Masini, Francesca (2015): The role of second order schemas in word formation. In: Bauer, Laurie/Körtvélyessy, Lívia/Štekauer, Pavol (eds.): Semantics of complex words. Cham i. a.: Springer. 47–66.
Dahl, Östen (2004): The growth and maintenance of linguistic complexity. Amsterdam/Philadelphia: Benjamins.
De Haas, Wim/Trommelen, Mieke (1993): Morfologisch handboek van het Nederlands. Den Haag: SDU Uitgeverij.
Hoeksema, Jack (2012): Elative compounds in Dutch: properties and developments. In: Oebel, Guido (ed.): Intensivierungskonzepte bei Adjektiven und Adverben im Sprachenvergleich/Crosslinguistic comparison of intensified adjectives and adverbs. Hamburg: Verlag Dr. Kovač. 97–142.
Hoffman, Thomas/Trousdale, Graeme (eds.) (2013): The Oxford Handbook of Construction Grammar. Oxford: Oxford University Press.
Hüning, Matthias (2010): Adjective + noun constructions between syntax and word formation in Dutch and German. In: Onysko, Alexander/Michel, Sascha (eds.): Cognitive perspectives on word formation. Berlin/New York: De Gruyter. 195–215.
Hüning, Matthias/Booij, Geert (2014): From compounding to derivation. The emergence of derivational affixes through 'constructionalization'. In: Folia Linguistica 48. 579–604.
Hüning, Matthias/Schlücker, Barbara (2010): Konvergenz und Divergenz in der Wortbildung. Komposition im Niederländischen und im Deutschen. In: Dammel, Antje/Kürschner, Sebastian/Nübling, Damaris (eds.): Konvergenz und Divergenz in der Wortbildung. (= Germanistische Linguistik 206–209). Hildesheim i.a.: De Gruyter. 783–825.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.): Word formation. An international handbook of the languages of Europe. Vol. 1. (= Handbooks of Linguistics and Communication Science (HSK) 40.1). Berlin/Boston: De Gruyter. 450–467.
Koefoed, Geert (1993): Benoemen. Een beschouwing over de faculté de langage. Amsterdam: Meertens-Instituut.
Levelt, Willem J.M./Meyer, Antje (2000): Word for word: Multiple lexical access in speech production. In: European Journal of Cognitive Psychology 12, 4. 433–452.
Los, Bettelou/Blom, Corrien/Booij, Geert/Elenbaas, Marion/van Kemenade, Ans (2012): Morphosyntactic change. A comparative study of particles and prefixes. Cambridge, UK: Cambridge University Press.
Nooteboom, Sieb (2011): Self-monitoring for speech errors in novel phrases and phrasal lexical items. In: Yearbook of Phraseology 2. 1–16.
Paul, Hermann (1898): Prinzipien der Sprachgeschichte. Halle/Saale: Max Niemeyer [^{1}1880].
Rainer, Franz (2013): Can relational adjectives really express any relation? An onomasiological perspective. In: Skase Journal of Theoretical Linguistics 10. 12–40.
Schlücker, Barbara (2014): Grammatik im Lexikon. Adjektiv-Nomen-Verbindungen im Deutschen und Niederländischen. Berlin: De Gruyter.
Schlücker, Barbara/Plag, Ingo (2011): Compound or phrase? Analogy in naming. In: Lingua 121. 1539–1551.
Schuster, Saskia (2016): Variation und Wandel. Zur Konkurrenz morphologischer und syntaktischer A+N-Verbindungen im Deutschen und Niederländischen seit 1700. Berlin: De Gruyter. Internet: www.degruyter.com/view/product/456743 (last access: 4.5.2018).

Schutz, Rik/Permentier, Ludo (2016): Met zoveel woorden. Gids voor trefzeker taalgebruik. Amsterdam/Leuven: Amsterdam University Press/Davidsfonds Uitgeverij.
Sprenger, Simone A. (2003): Fixed expressions and the production of idioms. Ph.D. dissertation. University of Nijmegen.
Sprenger, Simone A./Levelt, Willem J.M./Kempen, Gerard (2006): Lexical access during the production of idiomatic phrases. In: Journal of Memory and Language 54. 161–184.
Tummers, José (2005): Het naakt(e) adjectief. Kwantitatief-empirisch onderzoek naar de adjectivische buigingsalternantie bij neutra. Leuven: Katholieke Universiteit Leuven [Ph.D. dissertation].
Vikner, Sten (2005): Immobile complex verbs in Germanic. In: Journal of Comparative Germanic Linguistics 8. 83–105.

Kristel Van Goethem/Dany Amiot

Compounds and multi-word expressions in French

1 Introduction

French compounds differ from Germanic compounds in two important aspects. First, while Germanic compounding complies with the Right-hand Head Rule (e.g. English *postage stamp*, German *Briefmarke*, Dutch *postzegel*), French, like other Romance languages (see the chapters by Masini (Italian) and Fernández-Domínguez (Spanish) in this volume), has a general tendency towards left-hand headed compounding (e.g. *timbre-poste* lit. stamp-post). Second, whereas languages such as Dutch and German establish a clear demarcation between compounds and lexicalized phrases on the basis of formal criteria (spelling, prosody, linking elements, loss of adjectival inflection in [A N] compounds), French compounds are not easily distinguishable from syntactic expressions, and true compounds in Germanic languages often correspond to syntactic multi-word units in French (e.g. English *admission ticket* vs. French *billet d'entrée* (lit. ticket of entrance)) (Zwanenburg 1992: 222; see also the chapters by Booij (Dutch), Schlücker (German) and Bauer (English) in this volume).

Contrary to Germanic languages, French has no distinctive word stress, only phrase stress. Moreover, whereas Germanic compounds may present linking elements (e.g. Dutch *zonnebril*, German *Sonnenbrille* 'sunglasses'), these do not occur in French. Furthermore, the spelling of French multi-word units is characterized by many inconsistencies and irregularities: many combinations can be spelled with or without a hyphen (e.g. *bébé*(-)*éprouvette* 'test-tube baby' (lit. baby(-)test tube), *porte*(-)*monnaie* 'coin purse' (lit. carry(-)money)) or even as one word (e.g. *portefeuille* 'wallet, billfold' (lit. carrysheet) (Lehmann/Martin-Berthet 2008). Spelling of complex lexical units as one word occurs (e.g. *vinaigre* 'vinegar' (lit. wineacid)), but it is far from being the rule (cf. French *vin rouge* vs. German *Rotwein*), and the French spelling rules are systematically updated by orthographic reforms.[1] Finally, many French compound-like expressions have

1 The orthographic reform of 1990 proposed, for instance, to hyphenate complex numerals greater or lower than one hundred (e.g. *vingt-trois* 'twenty-three', *cent-cinquante-huit* 'one hundred and fifty-eight'), whereas this was only the case for numerals lower than one hundred before. The French Academy also suggested writing as one word a list of complex lexical units

https://doi.org/10.1515/9783110632446-005

internal inflection markers (e.g. *beaux-arts* 'fine arts'), while these are generally attributed to syntactic formations.

As a result, none of the formal criteria typically applicable in Germanic languages[2] allow for a straightforward differentiation between compounds and (lexicalized) multi-word phrases in French, and, accordingly, the term 'compound' is not always used in a consistent way in the literature on French morphology. As a matter of fact, 'compounding' is often used to refer to various types of complex lexical units regardless of the formation process (morphological or syntactic) (for an overview, see, for example, Van Goethem 2009 and Villoing 2012).

Van Goethem (2009) illustrates this in the domain of [A N] units. The Dutch compound *zuurkool* 'sauerkraut' (lit. sour-cabbage) can be distinguished from the lexicalized phrase *zure regen* 'acid rain' and the non-lexicalized syntactic phrase *zure kers* 'sour cherry' on the basis of its spelling (written as one word), its stress pattern (prominent stress on *zuur* in *zúurkool* while *zúre kérs* has double stress and *zure régen* has prominent stress on the noun *regen*, cf. De Caluwe 1990: 17) and the lack of inflection of the adjectival component *zuur* in the compound (cf. Booij 2002: 314). In French, however, these criteria do not apply and Van Goethem (2009) concludes that, leaving aside some exceptions that do not conform to regular modern French syntax (e.g. *rouge-gorge* 'robin' (lit. red-throat) and *grand-mère* 'grandmother', cf. Van Goethem 2009: 246f.), French [N A] and [A N] units are phrases and not compounds, whatever their spelling may be: whether written as two separate words (e.g. *premier ministre* 'prime minister'), hyphenated (e.g. *cordon-bleu* 'master chef' (lit. cord-blue)) or even as one single word (e.g. *vinaigre* 'vinegar' (lit. wineacid)).

In this paper, we will turn the focus to [N_1 N_2] units, but before doing so we will present the different approaches to complex lexical units in French and show

previously written as separate words (with or without a hyphen), for example *chauvesouris* 'bat' (lit. bald-mouse), *millepattes* 'centipede' (lit. thousand-legs), *passepartout* 'pass key' (lit. pass-everywhere), *portemonnaie* 'coin purse' (lit. carry-money) and *véloski* 'skibob' (lit. bike-ski). (Internet: www.lalanguefrancaise.com/guide-complet-nouvelle-orthographe, last access: 18.4.2017).

2 In this respect, English may be considered to occupy an intermediary position: the traditional distinctive criterion applicable to English is the stress pattern, compounds being typically characterized by fore-stress (e.g. *black bírd* vs. *blackbird*, cf. Bauer 2004 and this volume), but even this criterion is not always straightforward and many mismatches can be observed: as shown by Bauer (2004), a lexicalized phrase such as *prímary school* has first-element stress (or compound stress), whereas *first-áid*, with the two components hyphenated and unified, has second element stress (or phrase stress). These inconsistencies also apply to [N N] formations: *péanut oil*, for instance, has fore-stress, whereas *olive oil* may have end stress (cf. Bauer 1998, this volume; Giegerich 2009a, 2009b).

how true morphological formations (i.e. compounds) can be distinguished from multi-word phrases (Section 2). At the end of this section, the possible benefits of a constructionist approach to the issue will be highlighted. Section 3 will concentrate on [N_1 N_2] lexical units, which turn out to be the most problematic case in French since it is not easy to determine whether this formation belongs to syntax or morphology. In Section 4, a specific subtype, that of subordinative [N_1 N_2] units, will be examined because the latter most severely challenge this morphology-syntax divide. Whereas Fradin (2009) considers these formations to be true compounds, we will show that this only holds for the classifying subtype, and not for the qualifying one. Section 5, finally, will be devoted to a constructionist account of qualifying subordinative [N_1 N_2] formations, followed by the conclusion in Section 6.

2 Complex lexical units in French: four approaches

The notion of compounding generally has a more extensive scope in French morphology than in the literature on Germanic languages. Van Goethem (2009) identifies three different approaches. The common view is 'non-restrictive' in the sense that it includes all kinds of complex lexical units, regardless of whether they are formed in morphology or syntax (2.1). According to the 'scalar' approach (2.2), compounds are considered the endpoint of a scale of 'lexicalization' (used here to refer to the process of becoming a lexical item). The 'restrictive' or 'lexicalist' approach (2.3) aims to establish a clear demarcation between compounds and multi-word phrases. In what follows, we will outline these three different approaches. In 2.4, finally, we will add a fourth perspective and briefly show how complex lexical units can be accounted for from a Construction Grammar perspective.

2.1 The non-restrictive approach

In their overview article of multi-word expressions, Hüning/Schlücker (2015: 454ff.) convincingly show that (syntactic) multi-word expressions and word-formation units (i.e. compounds) share a set of properties. Both are complex expressions with (potential) status as a lexical unit, and both expressions typically serve as linguistic signs for specific concepts (i.e. they have a 'naming function', cf. also Schlücker/Hüning 2009). Lastly, lexicalized phrases and compounds may have compositional or non-compositional semantics and may contain constituents with metaphorical semantics.

In French, formations such as [N *de* N] (e.g. *fil de fer* 'iron wire' (lit. wire of iron)), [N *à* N] (e.g. *verre à vin* 'wine glass' (lit. glass to wine)), [N *à* Det N] (e.g. *sauce à l'ail* 'garlic sauce' (lit. sauce to the garlic)), [A N] (e.g. *Moyen Âge* 'Middle Ages') and [N A] (e.g. *poids lourd* 'heavyweight' (lit. weight heavy)) (Fradin 2003: 199; Booij 2010: 172) are constructed by means of syntactic rules, as manifested through the presence of prepositions, determiners and adjectival inflection. Nevertheless, like compounds, they are productively used in name formation and it is therefore not surprising that the notion of compounding is often extended to all kinds of complex lexical units with a naming function, regardless of the formation rules. This approach can be illustrated by Mathieu-Colas's (1996) classification of French compounds, which includes, for instance, lexicalized [A N] and [N A] units such as *premier ministre* 'prime minister' and *table ronde* 'round table meeting' (lit. table round), even though these comply with the syntactic formation rules, including adjectival inflection.

2.2 The scalar approach

A second approach is to establish a scale of lexicalization ranging from free syntactic phrases over (semi-)lexicalized phrases to true compounding. Such a scale contains, by definition, a large transition zone in which it is not easy to decide whether we are dealing with syntactic phrases or with compounds.

This idea of a scale of lexicalization of complex units can be found in studies by Gross (1988, 1996), who argues that lexicalized phrases and compounds can be distinguished from free syntactic phrases by means of semantic and syntactic parameters of lexicalization ('figement'). Semantically, lexicalized phrases and compounds such as *fait divers* 'novelty, piece of news, news item' (lit. fact diverse) are typically characterized by 'non-compositionality', in contrast to free syntactic phrases such as *fait évident* 'obvious fact', which have compositional semantics. Syntactically, in lexicalized [A N] or [N A] expressions the adjective loses the possibility of 'actualization' (1) and of 'predication' (2) (cf. Gross 1996: 31–34).

(1)	*un fait maintenant évident* 'a now obvious fact'	vs.	**un fait maintenant divers* 'a now diverse fact'
(2)	*Nous avons constaté un fait qui est évident* 'we have observed a fact that is evident'	vs.	**Nous avons constaté un fait qui est divers* 'we have observed a fact that is diverse'

On the basis of these parameters[3], Gross (ibid.) distinguishes between different degrees of lexicalization. *Cordon solide* 'solid rope', *cordon électrique* 'power cord' (lit. cord electric) and *cordon*(-)*bleu* 'master chef' (lit. cord(-)blue) illustrate three different degrees of lexicalization: *cordon solide* is a free syntactic noun phrase ('groupe nominal libre'), *cordon électrique* is considered a semi-lexicalized noun phrase or compound ('un groupe nominal ou nom composé semi-figé') and *cordon*(-)*bleu* is called a lexicalized compound ('un nom composé figé').

However, as rightly observed by Corbin (1992: 36), Gross still uses the term 'compounds' ('mots composés') to refer to all lexicalized and semi-lexicalized combinations: both *cordon électrique* and *cordon-bleu* are called 'noms composés', whatever the differences may be in structure or degree of lexicalization. In other words, similar to the non-restrictive approach, the notion of compound is still applied to all structures with a naming function, including syntactic expressions.

2.3 The restrictive or lexicalist approach

In a modular approach to grammar, it has to be accepted that phrasal multi-word expressions and compounds, notwithstanding significant similarities, are different, the most crucial distinction being the fact that they are constructed according to the rules of different components of the language system (syntax vs. morphology).

A third theoretical tradition in French morphology, whether or not inspired by the 'lexicalist' approach in Generative Grammar (Di Sciullo/Williams 1987) and represented by Benveniste (1974), Corbin (1992, 1997), Zwanenburg (1992), Fradin (2003, 2009) and Villoing (2012), among others, follows this view and argues that a clear distinction should be made between compounds and lexicalized phrases. Although both strategies may have the same naming function, they obviously fit into different parts of grammar, compounds belonging to morphology and phrases to syntax.

These authors argue, for instance, that [N Prep N] combinations such as *pomme de terre* 'potato' (lit. apple of ground) and *sac à main* 'handbag' (lit. bag to hand), commonly considered compounds in French, should be analyzed as lexicalized syntactic phrases since they respect the general principles of word order and syntax in French.

3 Cf. also ten Hacken's (1994) tests (such as insertion, substitution, anaphora from one constituent of the sequence).

The most extreme position can be found in Di Sciullo/Williams (1987), who claim that French does not have any compounds at all:

> It now appears that French (and no doubt Spanish) lacks compounding altogether. Once we have subtracted fixed syntactic phrases (idioms) such as *timbres-poste* and phrases reanalyzed as words (syntactic words) such as *essui-glace* <sic>, there are no candidates left. (ibid.: 83)

Corbin (1992, 1997) is less restrictive and preserves the term 'compound' to refer to lexical units of the type [N_1 N_2] (e.g. *timbre-poste* 'postage stamp') and [V N] (e.g. *essuie-glace* 'windscreen wiper') because they are formed according to lexical composition rules, specific to the lexicon and different from syntactic rules. Corbin (1997) uses the notion of 'polylexematic units' ('unités polylexématiques') as a general term for covering both compounds and lexicalized phrases. However, both naming strategies are distinguished on the basis of the 'division of labor principle' between morphology and syntax. According to this principle, also labeled the 'Lexical Integrity Hypothesis' (LIH hereafter), syntax has no access to morphological operators or infralexical units and, conversely, morphology has no access to syntactic operators:[4]

> Les règles syntaxiques n'ont accès ni aux opérateurs morphologiques ni à des unités infralexicales. Les règles morphologiques n'ont pas accès aux opérateurs syntaxiques. (ibid.: 83)

On the one hand, this implies that affixed polylexematic units such as *fil-de-fér**iste*** 'high wire walker' (lit. wire-of-iron-ist) belong to morphology, since syntax cannot attach affixes. On the other hand, polylexematic units containing a syntactic operator, a preposition as in *verre **à** vin* 'wine glass' (lit. glass to wine) or a determiner as in *hors-**la**-loi* 'outlaw' (lit. outside-the-law), necessarily belong to syntax.[5] In other words, polylexematic units are exclusively formed either by syntax or by morphology, and the idea of a scale is thus rejected:

---

4 Corbin's analysis is in line with the strong lexicalist hypothesis: 'The syntax neither manipulates nor has access to the internal structure of words' (Anderson 1992: 84). On this topic, see, among many others, Lieber (1992), Plag (2003) and, for an overview, Lieber/Scalise (2007).

5 There seems to be a contradiction in Corbin's analysis, which considers *fil-de-fériste* as a morphological unit despite the presence of the preposition *de* 'of'. However, Corbin (1997: 83) argues that the morphological insertion of the suffix *-iste* is subsequent to the insertion of the preposition *de* and that only the final step of the formation should be taken into account: the word is a morphological construct (application of the suffix *-iste*) on the basis of a syntactically constructed stem, *fil de fer*, which can be considered a lexical unit.

> En vertu du partage des tâches entre les modules d'une grammaire, les séquences engendrables syntaxiquement ne le sont pas morphologiquement et réciproquement. (ibid.: 84)[6]

On the same grounds, Fradin (2009: 418) excludes expressions such as *sans-papiers* 'person without identity papers, illegal immigrant' (lit. without papers) and *pied-à-terre* 'pied-à-terre, holiday cottage' (lit. foot-on-ground) from true compounding because they correspond to phrases that can be generated by syntax (cf. *Il s'est retrouvé sans papiers* 'he ended up without (identity) papers' and *Le cavalier mit pied à terre* 'the horseman dismounted' (lit. put foot on ground)). He relabels Corbin's proposal as 'Principle A':

> Principle A: Compounds may not be built by syntax (they are morphological constructs) (ibid.: 417)

Whereas in Corbin's (1997) view, only [N N] and [V N] configurations can be considered true compounds, Fradin (2009) concludes that not two but four productive compounding patterns can be retained in French: [V N] (e.g. *brise-glace* 'ice-breaker' (lit. break-ice)), [A A] (e.g. *sino-coréen* 'Sino-Korean'), [N N] coordinates (e.g. *auteur-compositeur* 'author-composer') and [N N] subordinates (e.g. *poisson-chat* 'catfish' (lit. fish-cat)). Villoing (2012: 36) adds to this a particular subclass of [A N] compounds with a color adjective as head (e.g. *bleu-ciel* 'sky blue' (lit. blue-sky)). She argues that all these formations should be considered true compounds because they all display syntactic anomalies:

- VN compounds: the absence of a determiner between the verb and the noun, and a diverse range of semantic relations between the verb and the noun (*ouvre-boîte* 'can opener' (lit. open-can)),
- coordinated NN (*horloger-bijoutier* 'jeweler-watchmaker' (lit. watchmaker-jeweler)) and AA (*aigre-doux* 'sweet and sour' (lit. sour-sweet)) compounds: the absence of a coordinating conjunction between the constituents,
- all other NN compounds (*poisson-chat* 'catfish' (lit. fish-cat), *pause-café* 'coffee break' (lit. break-coffee)): hyponymic interpretation,
- AN compounds (*bleu-ciel* 'sky blue' (lit. blue-sky)): the presence of an adjectival rather than a nominal head.

  (paraphrased from Villoing 2012: 36)

6 Our translation: 'By virtue of the division of tasks between the modules of a grammar, sequences that are possibly generated by syntax are not generated by morphology and vice versa'.

Villoing (2012: 30) specifies that French native compounding[7] 'is prototypically formed of two lexemes of the current lexicon of French, without any linking element; the internal order of constituents is XY, where X is the governing element'. Furthermore, the composing lexemes belong, by definition, to the major word classes (noun, verb, adjective), and are uninflected. This implies that 'no constituent is marked by inflection: no modality, tense, person or aspect marking on the verb in VN compounds, no number on the N, and no gender or number on adjectives, disregarding cases of agreement' (ibid.: 31f.).[8] Examples are *poisson-chat* 'catfish' (lit. fish-cat), *wagon-fumeur* 'smoking car' (lit. car-smoker), *ouvre-boîte* 'can opener' (lit. open-can) and *vert-pomme* 'apple green' (lit. green-apple).[9]

This view implies that many other multi-word units that are often considered compounds do in fact belong to syntax and, therefore, need to be analyzed as lexicalized phrases. According to Villoing (ibid.: 35f.), the following French multi-word units should not be analyzed as compounds:

- Complex units composed of non-lexemes, such as complex prepositions and complex conjunctions: e.g. *par-dessus* 'from above', *de sorte que* 'such that'[10]
- Lexicalized syntactic constructions, namely NPs (3), PPs (4) and VPs (5) that behave like lexical units:

7 Villoing (2012) distinguishes native compounds from neoclassical compounds, which have different properties: the latter are 'prototypically composed of two bases of Greek or Latin origin that are not syntactically autonomous in French, connected by a linking element; the internal order of constituents is YX, where X is the governing element' (Villoing 2012: 30) (e.g. *lud**o**-thèque* 'game library', *hom**i**-cide* 'manslaughter', *cyn**o**-céphale* 'dog head').

8 However, Villoing (2012: 34) rightly observes that some compounds actually display inflected forms of the lexeme: for instance, many [V N] compounds include a plural N, orthographically and/or phonologically marked (e.g. *presse-fruits* 'fruit press' (lit. press-fruits), *protège-yeux* 'eye protector' (lit. protect-eyes)). Villoing argues that this plural inflection is not the result of syntactic marking, but of inherent and semantically motivated inflection.

9 This approach, in line with Corbin (1992), Villoing (2009), Bonami/Boyé (2003, 2014) and Fradin (2009), among others, implies that the V in French $[V\ N]_N$ compounds (e.g. *ouvre-boîte* 'can opener') is not an inflected form of the verb (imperative or present indicative), but a stem of the lexeme.

10 Although Zwanenburg (1992) starts from the same syntax-morphology divide principle, his analysis leads to completely different results: he concludes that real compounding in French is precisely restricted to nouns, adjectives and verbs with a modifying preposition or adverb (e.g. *sous-chef* 'deputy' (lit. under-boss), *arrière-pays* 'hinterland' (lit. behind-land), *maltraiter* 'maltreat'). Paradoxically, this implies that French compounding would be right-headed, similar to Germanic compounding.

(3) *brosse à dents* 'toothbrush' (lit. brush at teeth)
coffre-fort 'safe' (lit. box strong)
case départ 'start, square one' (lit. box departure)

(4) *sans-papiers* 'illegal immigrant' (lit. without-papers)

(5) *boit-sans-soif* 'drunk' (lit. drinks-without-thirst)

– Lexicalized phrasal expressions that behave like lexical units: for instance, *rendez-vous* 'appointment, date' (lit. go-you), *qu'en-dira-t-on* 'gossip' (lit. what about it-will say-one).

Villoing (ibid.: 36) admits, nevertheless, that the boundary between compounds and syntactic units is most problematic in the case of [N_1 N_2] sequences. This can also be derived from her examples: *horloger-bijoutier* 'jeweler-watchmaker' is considered a compound, whereas *case départ* 'square one' is analyzed as a lexicalized syntactic construction. It does indeed appear that French [N_1 N_2] sequences can be constructed by both morphology and syntax and that a subcategorization of [N_1 N_2] formations is needed. We will therefore focus on this particular formation type in Sections 3 and 4.

2.4 A constructionist perspective to complex lexical units

It can be concluded from the preceding overview that the term 'compounding' is not used consistently in the French linguistic tradition and often covers much more than, strictly speaking, morphological complex lexical units. Hüning/Schlücker (2015) point out the commonalities and differences found between compounds as word-formation units and syntactically formed multi-word expressions. In spite of the differences, both patterns may serve the same purpose and even enter into competition to do so. As for French, many examples of competition can be found between [N N] and [N Prep N] formations: *village*(-)*vacances* coexists with *village de vacances* 'holiday village, holiday resort' (lit. village (of) holidays) and the same holds for *point*(-)*rencontre* and *point de rencontre* 'meeting point' (lit. point (of) meeting) and *impression* (*par*) *laser* 'laser printing' (lit. printing (by) laser) (cf. also Section 3.1). These facts indicate that in French, too, the boundary between compounds and syntactic multi-word expressions is fuzzy and the data are suggestive of a lexicon-syntax continuum.

This non-modular view of language is precisely a basic assumption of Construction Grammar (cf. Goldberg 1995, 2006; Croft 2001; Booij 2010; Hoffmann/

Trousdale (eds.) 2013, a.o.). Crucial to this model is the concept of 'constructions': these are conventional pairings of form (referring to syntactic, morphological and phonological properties) and meaning (including semantic, pragmatic and discourse-functional properties) and are considered the fundamental units of the linguistic system. All levels of grammatical description involve such form-meaning pairings – not only words as in the Saussurean tradition – and constructions vary in size, degree of schematicity and complexity (cf. Goldberg 2009), the minimal linguistic construction being the word in Booij's (2010) model of Construction Morphology. Furthermore, constructions, both syntactic and morphological, are linked to each other by (vertical) inheritance relations and also by (horizontal) connectivity links (Norde 2014; Norde/Morris 2018). As a consequence, language can be considered a complex network of constructions. Substantive constructions (e.g. *petit mais vaillant* 'small but tough', *position clé* 'key position') are instances of semi-schematic constructions (e.g. [Adj_1 *mais* Adj_2], [N_1 *clé*]), which – in turn – inherit properties from more general schematic constructions (e.g. [Adj_1 CONJ Adj_2], [N_1 N_2]). Moreover, constructions may also inherit properties from multiple-parent constructions via so-called 'multiple inheritance' (cf. Trousdale 2013; Trousdale/Norde 2013).[11]

It is not surprising that many recent studies in the field of multi-word expressions are in the constructionist vein. In this approach, it can be assumed that both compounds and phrasal structures with a naming function can act as conventionalized form-meaning pairings or 'constructions', and we should accept the existence of what Booij (2010: 190) calls 'lexical phrasal constructions': these are syntactic formations that should be stored as lexical units in the mental lexicon, such as *fil de fer* 'iron wire' (lit. wire of iron) and *moulin à vent* 'windmill' (lit. mill at wind). These formations demonstrate that there is no strict boundary between the lexicon and syntax, or, as Booij (ibid.: 191) puts it, 'syntax permeates the lexicon because syntactic units can be lexical'.

Compounds and phrasal structures are not only closely linked in the constructional network; they may also compete or interact with each other. The process of 'multiple inheritance' may even produce hybrid constructions that inherit properties from parent constructions belonging to different domains, such as morphology and syntax. We believe that these insights from Construction Gram-

11 The idea of 'multiple inheritance' could be seen as the synchronic representation of the complexity of language change. Diachronic developments do not always follow linear pathways from one source construction to another target construction; a complex interplay between different sources and processes is often at stake (cf. De Smet/Ghesquière/Van de Velde's (eds.) 2013 volume *On multiple source constructions in language change*).

mar are useful to account for problematic cases that cannot be univocally classified as morphological or syntactic constructs, such as French [N_1 N_2] subordinatives. In Sections 3 and 4 we will therefore focus on these particular cases and in Section 5 we will propose an analysis in line with the constructionist insights.

3 French [N_1 N_2] sequences: compounds or phrases?

In Section 2.3, we observed that both Fradin (2009: 428f.) and Villoing (2012: 36) admit that the boundary between morphological and syntactic units in French is most difficult to apply in the case of [N_1 N_2] formations. We will therefore now concentrate on Fradin's arguments to retain only [N_1 N_2] coordinates and subordinates as true French compounds, at the expense of other types of [N_1 N_2] sequences, namely so-called 'two-slot nominal constructs' and identificational [N_1 N_2] constructs (3.1). In Section 3.2, we will focus on subordinate [N_1 N_2] formations and show that their status is more problematic than acknowledged by Fradin (2009).

3.1 Fradin's (2009) typology of [N_1 N_2] sequences

Fradin (2009) distinguishes between four types of [N_1 N_2] sequences: coordinates, subordinates, two-slot nominal constructs and identificational constructs; the first two are assigned to morphology and the others to syntax.

First, two types of [N_1 N_2] ***coordinates*** can be distinguished: in (6) each N has a distinct referent and the compound's denotatum is the sum of these referents; the compounds in (7), however, denote a unique referent combining properties of both N_1 and N_2 (ibid.: 429f.):

(6) *Bosnie-Herzégovine* 'Bosnia-Herzegovina'
physique-chimie 'physics-chemistry (as a teaching discipline)'

(7) *chanteur-compositeur* 'singer-composer'
hôtel-restaurant 'hotel-restaurant'

As also argued by Villoing (2012: 36), the absence of a coordinating conjunction between the constituents excludes these sequences being generated by syntax, and they should therefore be considered true compounds.

Unlike coordinate compounds, ***subordinate compounds*** only denote the referent expressed by N_1 (i.e. the head noun), while N_2 (i.e. the modifier) refers to

one of its salient properties. According to Fradin (2009: 430f.), this property may concern a physical dimension (shape, length, weight) (8), an intrinsic capacity (slowness, quickness, strength, duration) (9) or a function (10), and is metaphor-based.

(8) *requin-marteau* 'hammerhead shark' (lit. shark-hammer)
homme-grenouille 'frogman' (lit. man-frog)

(9) *justice escargot* 'slow justice' (lit. justice snail)
guerre éclair 'blitzkrieg' (lit. war lightning)
attaquant-bulldozer 'offensive forward' (lit. attacker-bulldozer)
discours fleuve 'lengthy discourse' (lit. discourse river)

(10) *camion-citerne* 'tanker truck' (lit. truck-tanker)
voiture-balai 'broom wagon' (lit. car-broom)
livre-phare 'leading book' (lit. book-lighthouse)

Even though Fradin recognizes that the morphological status of these compounds is open to debate (cf. Section 4), he claims that the regular interpretative patterns found in these subordinate compounds are similar to those of some derived lexemes, such as French adjectives derived with the suffix *-able* (Fradin 2003). In the same way as productive suffixes, the N_2 of subordinate [N_1 N_2] formations can be combined with a broad range of stems and forms a productive constructional pattern with a regular interpretation. This similarity with derivation is taken as an argument in favor of their morphological status.

Whereas coordinate and subordinate [N_1 N_2] sequences follow a constrained pattern and have a regular semantic relationship between the constituents, this is not the case with ***two-slot nominal constructs*** (Fradin 2009: 432f.) and identificational [N_1 N_2] sequences. The examples in (11) all denote the referent expressed by N_1, but they completely differ from subordinate compounds because N_2 does not refer to an intrinsic and salient property of N_1. Moreover, the sequence usually corresponds to a syntactic phrase in which N_2 forms part of a prepositional phrase (12), which suggests a syntactic origin.

(11) *impression laser* 'laser printing' (lit. printing laser)
espace fumeurs 'smoking area' (lit. space smokers)
accès pompiers 'firemen entrance' (lit. entrance firemen)

(12) *impression par laser* (lit. printing by laser)
espace pour (*les*) *fumeurs* (lit. space for (the) smokers)
accès pour (*les*) *pompiers* (lit. entrance for (the) firemen)

Fradin (2009: 433f.) likewise argues for ***identificational [N_1 N_2] sequences*** (cf. also Noailly 1990):

(13) *la catégorie adjectif* 'the adjective category'
l'institution Opéra 'the Opera institution'

N_2 identifies N_1 ('N_2 is an N_1') and from this point of view, these sequences are equivalent to syntactic (appositional) [N_1 N_2] constructs in which N_2 is a proper noun and N_1 expresses a socially recognized category (e.g. *le président Mandela* 'President Mandela', *la région Bourgogne* 'the region of Burgundy').

3.2 Discussion: morphological and syntactic approaches to [N_1 N_2] subordinatives

We agree with Fradin that [N_1 N_2] coordinates are true compounds and cannot be the result of syntactic formation. We also subscribe to his view on two-slot nominal and identificational [N_1 N_2] constructs: both sequences can be shown to correspond to syntactic phrases. However, subordinate [N_1 N_2] formations are more problematic than acknowledged by Fradin (2009) and it can be demonstrated that the examples mentioned for this class are not all of the same kind. At first glance, it can, for instance, be observed that some of them permit degree modification of N_2 while others do not (*discours vraiment fleuve* 'really lengthy discourse' (lit. discourse really river) vs. **requin vraiment marteau* 'really hammerhead shark' (lit. shark really hammer)), and some but not all N_2s form productive series (e.g. *discours-fleuve* 'lengthy discourse' (lit. discourse-river), *roman-fleuve* 'novel cycle', *film-fleuve* 'lengthy movie', *débat-fleuve* 'lengthy debate', etc.), while no series formation is possible for [N-*marteau*], for instance. We will discuss these differences more extensively in Section 4.

As already mentioned, these formations have been the subject of some debate. Amiot/Van Goethem (2012: 350ff.) and Van Goethem (2012: 77–81) provide an overview of the different accounts, which range from purely syntactic analyses (cf. Noailly 1990 and Goes 1999) to strictly morphological accounts, like the one by Fradin (2009).

With regard to the syntactic approaches, a distinction can be made between analyses where the second component of the phrase is still considered a noun in spite of some adjectival properties (cf. Noailly 1990, who labels N_2 as 'substantif épithète' and Arnaud/Renner 2014, who detect adjective-like syntactic behavior to some extent), and others like Lehmann/Martin-Berthet (2008: 206), who claim

that N_2 is converted into an adjective if it complies with a set of criteria typical of adjectives (such as degree modification and predicative use).

With regard to the morphological approaches, we can contrast Fradin's classification of French compounding with the general typology of compounds by Scalise/Bisetto (2009) (applied to French by Villoing 2012), according to whom these 'problematic' compounds are not subordinatives but belong to the ATAP (attributives-appositives) class, and more particularly to the subclass of appositives:

> Attributive compounds can actually be defined as formations whose head is modified by a non-head expressing a 'property' of the head, be it an adjective or a verb: actually, the role of the non-head categorial element should be that of expressing a 'quality' of the head constituent. Appositives, to the contrary, are compounds in which the non-head element expresses a property of the head constituent by means of a noun, an apposition, acting as an attribute. (Scalise/Bisetto 2009: 51)

As these definitions show, attributives (e.g. *high school*) and appositives (e.g. *snailmail*, *swordfish*) belong to the same ATAP class because they have similar functions. The metaphorical value of the modifier is argued to be an important distinctive criterion between $[N_1\ N_2]$ subordinatives (e.g. *mushroom soup*), on the one hand, and $[N_1\ N_2]$ appositives (e.g. *mushroom cloud*), on the other:

> In appositives that, together with attributives, make up the ATAP class, the noun plays an attributive role and is often to be interpreted metaphorically. Metaphoricity is the factor that enables us to make a distinction between, e.g. *mushroom soup* (a subordinate *ground* compound) and *mushroom cloud*, where *mushroom* is not interpreted in its literal sense but is rather construed as a 'representation of the mushroom entity' (...) whose relevant feature in the compound under observation is shape. (ibid.: 52)

In the next section, we will take a closer look at this specific type of formation and will argue that we need to distinguish between two different subclasses: classifying and qualifying $[N_1\ N_2]$ subordinatives, of which only the former undoubtedly belong to morphology.

4 Classifying vs. qualifying $[N_1\ N_2]$ subordinatives

In this section we will argue that two types of $[N_1\ N_2]$ subordinatives should be distinguished: classifying (e.g. *requin-marteau* 'hammerhead shark' (lit. shark-hammer)) and qualifying (e.g. *guerre éclair* 'blitzkrieg' (lit. war lightning)). The difference can essentially be found in the different role of N_2 with respect to

N_1.[12] Despite their similarities (in all these subordinate compounds, N_2 denotes a salient, metaphor-based property of N_1), N_2 has a classifying role in some [N_1 N_2] formations (e.g. *requin-marteau*) but a qualifying role in others (e.g. *guerre-éclair*). We will present the distinguishing properties of both types of [N_1 N_2] subordinatives in 4.1 and 4.2, respectively.

4.1 Classifying [N_1 N_2] subordinatives

Classifying [N_1 N_2] subordinatives are characterized by a number of particular semantic and syntactic properties:

(i) Semantically, they behave like designations ('names'): they refer to stable concepts (Kleiber 1984), but their reference is established in a motivated way: N_1, the semantic head, is the hyperonym and N_2, which does not have a referential meaning, refers to a salient property of N_1 that allows the [N_1 N_2] sequence to be distinguished from other N_1s. Hence, N_2 expresses a classifying property of N_1.[13] This is why, at least when they denote biological species, classifying [N_1 N_2] sequences are often the vernacular denominations corresponding to scientific taxonomies: for instance, *serpent-tigre* corresponds to *Notechis Scutatus*, *pin-parasol* to *Pina Pinea* and *oiseau-lyre* to *Menura Superba*, etc. (cf. Ureña/Faber 2010 for English compounding). When [N_1 N_2] is not a vernacular denomination corresponding to a scientific taxonomy, it can at least integrate a hierarchical folk categorization (Wierzbicka 1996): for example, a *fauteuil-crapaud* 'squat armchair' (lit. armchair-toad) is a kind of armchair (*fauteuil*), in the same way as a club chair or a rocking chair. And, in turn, an armchair is a piece of furniture, etc. This signals the relationship of inclusion [X *is a* Y], typical of the hierarchy between a hyponym and its hyperonym.

(ii) N_1 often denotes a biological species, especially animals (14a), vegetables (14b) or sometimes human beings (14c). More exceptionally, compounds denoting artefacts can also be found (14d):

12 This category merges what Arnaud (2003: 13) calls the 'composés équatifs-analogiques' ('equative analogical compounds') and the 'composés méronymiques-analogiques' ('meronymic analogical compounds'), i.e. *poisson-chat* 'catfish' vs. *poisson-scie* 'sawfish', respectively.

13 To a certain extent, such sequences correspond to the 'generic-specific compounds' in Arnaud (2003), but the author classifies them as 'equative/analogical compounds', because of the metaphorical use of N_2.

(14a) *poisson-scie* 'sawfish' (lit. fish-saw)
oiseau-lyre 'lyrebird' (lit. bird-lyre)
serpent-tigre 'tiger snake' (lit. snake tigre)
(14b) *saule têtard* 'silver willow' (lit. willow tadpole)
pin-parasol 'umbrella pine' (lit. pine-umbrella)
tomate-cerise 'cherry tomato' (lit. tomato-cherry)
(14c) *homme-grenouille* 'frogman' (lit. man-frog)
femme-objet 'woman as object' (lit. woman-object)
enfant-roi 'spoilt child' (lit. child-king)
(14d) *voiture-bélier* 'ram-raid' (lit. car-ram)
fauteuil-crapaud 'squat armchair' (lit. armchair-toad)
noeud-papillon 'bow tie' (lit. bow-butterfly)

(iii) In these cases, and as opposed to coordinate compounds, the two nouns denote concrete entities that do not belong to the same semantic class and the metaphor that underpins the relation between N_1 and N_2 is often based on physical resemblance: the nose of a *poisson-scie* is shaped like a saw (*scie*) and a *saule têtard* has roughly the shape of a tadpole (*têtard*): a big head like the upper part (the foliage) of the willow, and a short tiny bottom part (like the trunk). In our examples, the only sequences that do not instantiate this relation are *enfant-roi*, *femme-objet* and *voiture-bélier*, in which the metaphor is based on behavioral resemblance. For example, an *enfant-roi* is a child (*enfant*) who is treated like a king (*roi*) and who often becomes a 'domestic tyrant'.

(iv) Syntactically, all the linguistic tests usually used to measure the lexical integrity of a sequence (cf. Sections 2.2 and 2.3) show that these classifying $[N_1\ N_2]$ formations are words, insofar as they do not accept any of these manipulations, unlike the qualifying $[N_1\ N_2]$ subordinatives that we will study in Section 4.2.

(v) The last property to be mentioned is the fact that, unlike the qualifying $[N_1\ N_2]$ formations, these classifying subordinatives do not give rise to productive series.

We can conclude from this survey that the subordinate $[N_1\ N_2]$ formations like those exemplified under (14) are binominal words and true compounds in which N_2 metaphorically denotes a classifying property of N_1.

4.2 Qualifying $[N_1\ N_2]$ subordinatives

Qualifying $[N_1\ N_2]$ subordinatives can be distinguished from the classifying subtype on the following grounds:

(i) All kinds of nouns may instantiate N_1: nouns denoting artefacts (15a), social roles (15b), time or slots of time (15c), events (15d), and even abstract nouns (15e):

(15a) *livre-phare* 'landmark book' (lit. book-lighthouse)
établissement-pilote 'pilot institution' (lit. institution-pilot)
film-culte 'cult movie' (lit. movie-cult)
(15b) *acteur-clé* 'key actor' (lit. actor-key)
attaquant-bulldozer 'offensive forward' (lit. attacker-bulldozer)
(15c) *moment-charnière* 'pivotal moment' (lit. moment-hinge)
date-limite 'deadline' (lit. date-limit)
(15d) *discours-fleuve* 'lengthy discourse' (lit. discourse-river)
guerre-éclair 'blitzkrieg' (lit. war-lightning)
(15e) *justice-escargot* 'slow justice' (lit. justice-snail)

(ii) According to Fradin (2009), N_2s often refer to a metaphoric intrinsic property of N_1 (cf. Section 3.1): slowness (e.g. *justice-escargot*), quickness (e.g. *guerre-éclair*), strength (e.g. *attaquant-bulldozer*) or duration (e.g. *discours-fleuve*). To a certain extent, they often express intensity, as in *livre-phare*, *acteur-clé*, *moment-charnière*: a *livre-phare*, for example, is a *very* famous book that attracts a lot of attention. However N_2 does not have a categorization function (a *livre-phare* is not a kind of book, an *acteur-clé* is not a kind of actor, etc.): the [N_1 N_2] sequences exemplified under (15) are not designations that could be included in a hyperonymy/hyponymy hierarchy. Instead, N_2 has a qualifying role and, moreover, it can often be substituted with a qualifying adjective: an *acteur-clé* is a very important actor (in a given context), a *justice-escargot* is very slow justice, and so on.

(iii) It is precisely the qualifying role of N_2 that could, in our view, explain the specific behavior of these [N_1 N_2] formations, and particularly their lack of lexical integrity (cf. 2.3):

(a) Both N_1 and N_2 can be instantiated by a complex (i.e. multi-word) sequence. The examples under (16) represent formations with a 'complex N_1':

(16a) *Wilo Salmson France représente un acteur économique* ***clé*** *de la région*. (www)[14]
'Wilo Salmon France represents a key economic actor in the region'

14 All examples followed by (www) were taken from the web via Google searches in May 2017.

(16b) *Wall Street 2 adopte la forme d'une saga familiale* ***fleuve*** (www)
'Wall Street 2 takes the form of a very long (lit. river) family saga'
(16c) *L'affiche du film d'animation* ***culte*** *Akira a eu droit à de nombreuses parodies* (www)
'The poster of the cult animated movie Akira spawned many parodies'
(16d) *d'un coup de poing* ***éclair***, *elle dévie le ballon* (www)
'with a lightning punch (lit. punch-of-fist), she deflects the ball'

In these examples, the N_1s resemble phrases: they result from the association of a noun and an adjective (16a–b) or of a noun and a prepositional phrase (16c–d).

The N_2 slot can also be filled by a complex item, but this is more exceptional:

(17) *La compagnie de gendarmerie* [...] *a mobilisé des effectifs lors de l'opération* ***coup de poing*** *menée vendredi* (www)
'The police [...] mobilized officers on Friday for the lightning [lit. punch-of-fist] raid'

Interestingly, a lexicalized multi-word expression such as *coup de poing* can fill, in its literal meaning ('punch'), the N_1 slot or, in its metaphorical meaning ('lightning'), the N_2 slot.

It should be noticed that, since the complex sequences that may fill the N_1 or N_2 slots are lexicalized phrases, this is less problematic for the LIH than if they were free, compositional phrases (cf. Booij's (2010) use of 'lexical phrasal constructions' in 2.4).

(b) Most N_2s can be modified by an adverb of degree, as shown in (18):

(18a) *on avait le sentiment d'assister à un moment* ***vraiment*** *charnière* (www)
'we had the feeling of witnessing a truly pivotal (lit. hinge) moment'
(18b) *un conseil* ***vraiment*** *éclair* (www)
'a really whirlwind (lit. lightning) council meeting'
(18c) *la multiplicité des voix de ce roman* ***vraiment*** *fleuve* (www)
'the multiplicity of voices in this really lengthy (lit. river) novel'

This second property is more challenging for the LIH: the lexical integrity of the $[N_1 N_2]$ sequences is undoubtedly called into question by the insertion of an adverb of degree between the two Ns. This is why some authors put forward a weakened

version of the hypothesis, including Ackema/Neeleman (2004), Booij (2005) and Lieber/Scalise (2007).[15]

Our previously conducted corpus research (Amiot/Van Goethem 2012; Van Goethem 2012, 2015) indicate that the most frequently inserted adverb is *vraiment* 'really, truly', as in (18), but other degree adverbs can be found too: *absolument* 'absolutely' (19), *réellement* 'really' (20), *extrêmement* 'extremely' (21) and even, but more rarely, *très* 'very' (22):

(19) *Les années 1970 constituent en effet une période* ***absolument*** *charnière dans la vie des communautés* [...] (www)
'The 1970s constituted an absolutely pivotal (lit. hinge) period in the life of communities [...]'

(20) *Nous reviendrons sur ce point* ***réellement*** *clé pour la suite de la réflexion* (www)
'We will return to this point, which is really key (lit. this really key point) for the continuation of the discussion'

(21) [...] *une version raccourcie d'un texte* ***extrêmement*** *fleuve qu'il a publié quelques années plus tôt* (www)
'[...] an abridged version of an extremely lengthy (lit. river) text that he published a couple of years before'

(22) *le match a été une orgie offensive avec un score* ***très*** *fleuve* (42–24 en faveur des Parisiens) (www)
'the match was an offensive orgy with a very crushing (lit. river) score (42–24 in favor of the Parisians)'

The presence of such adverbs conflicts not only with the lexical integrity of the [N_1 N_2] sequence, but also with the nominal status of N_2: usually an adverb of degree modifies a gradable adjective, not a noun. However, in the context of the qualifying [N_1 N_2] sequences, N_2 seems to switch to adjectival status.

Syntactically, evidence for this adjectival status is not only provided by the possibility of modification by an adverb, but, like a qualifying adjective, N_2 can also be inserted into a comparative construction:

15 Cf. also the 'Italian *trasporto latte*-type constructions' (Lieber/Scalise 2007), in which both components can be modified by an adjective, e.g. *produzione scarpe* 'shoe production' → *produzione* (*accurata*) *scarpe* (*estive*) '(accurate) production of (summer) shoes'.

(23a) *pour moi c'est* [*la préadolescence*] *une période bien* ***plus charnière que*** *l'adolescence* (www)
'For me it [pre-adolescence] is a much more pivotal (lit. hinge) period than adolescence'
(23b) *La proximité de commerces est* ***moins clé que*** *pour une résidence senior* (www)
'The proximity of shops is less key than for a senior housing complex'

Semantically, in all the examples under (19–23), N_2 could be paraphrased by an evaluative adjective, for example:

(24) [...] *une période absolument* ***charnière*** / ***cruciale***
'an absolutely pivotal (lit. hinge)/crucial period'
[...] *ce point réellement* ***clé*** / ***important***
'this really key/important point'
[...] *un texte extrêmement* ***fleuve*** / ***long***
'an extremely lengthy (lit. river)/long text'

This demonstrates the qualifying value of N_2 vis-à-vis N_1. We will return to this in Section 5, but it is worth noting for the time being that this behavior distinguishes the qualifying subordinative [N_1 N_2] from the classifying subordinative (Section 4.1).

(c) Some N_2s can be used predicatively. Predicative use is the most prototypical use of qualifying adjectives. In some cases, 'N_2' can fill the slot of an adjective in a predicative construction (25) with or without degree marking:

(25a) *La <u>période</u>* ***est charnière*** *également sur le plan économique* (www)
'The period is also pivotal in economic terms'
(25b) *Leur <u>rôle</u>* ***est*** *ainsi plus clé que jamais* (www)
'Their role is thus more key than ever'
(25c) *c'est déjà arrivé quand l'<u>interview</u>* ***est*** *vraiment* ***fleuve*** (www)
'It has already happened when the interview is really lengthy'

In this use, the [N_1 N_2] construction (*période charnière* in (25a), *rôle clé* in (25b) and *interview fleuve* in (25c)) is broken up, and N_2 acquires autonomous adjectival behavior. This separation of compound-like sequences has been labeled 'debonding' by Norde (2009) (cf. also Amiot/Van Goethem 2012; Van Goethem 2012; Norde/Van Goethem 2014; Van Goethem/De Smet 2014; and Van Goethem 2015).

5 A constructionist analysis of qualifying [N_1 N_2] subordinatives

As can be concluded from the preceding section, besides coordinate [N_1 N_2] sequences, only classifying [N_1 N_2] subordinatives should be regarded as true compounds in French, whereas the qualifying [N_1 N_2] formations display hybrid behavior in the sense that they may, to a greater or lesser extent, undergo syntactic operations. We will now demonstrate how the idea of 'multiple inheritance' (cf. Section 2.4) can be fruitfully applied to account for these hybrid qualifying [N_1 N_2] subordinative constructions.

Two phases can be distinguished in the emergence of qualifying subordinatives (cf. Amiot/Van Goethem 2012 and Van Goethem 2015 on [N_1 *clé*] subordinatives).

The first step is the emergence of a productive constructional idiom – via so-called 'constructionalization' (Traugott/Trousdale 2013; Hüning/Booij 2014) – in which N_2 develops a specific (metaphoric) qualifying meaning when combined with an N_1 in a compound(-like) sequence (e.g. *question-clé* 'key question', *moment charnière* 'pivotal moment', *réunion marathon* 'marathon meeting', *cas limite* 'borderline case', etc.). This qualifying meaning may be seen as the result of 'coercion' (cf. Audring/Booij 2016) in which the metaphoric meaning sometimes already available for the noun outside the compound-like pattern (e.g. *la clé du succès* 'the key of success') is selected ('coercion by selection') and/or in which N_2 develops adjective-like (semantic and formal) properties within the [N_1 N_2] pattern ('coercion by override'). This semi-schematic construction, applied to the example of [N *charnière*] formations, can be represented as follows:

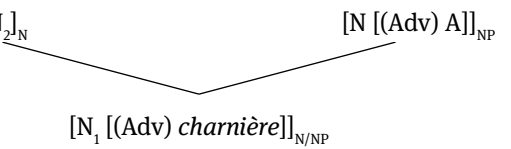
(26) $[[X]_{Ni}\ [\textit{charnière}]_{N}]_{Nj} \leftrightarrow [\text{pivotal, crucial SEM}_i]_j$

However, the constructionalization of N_2 goes beyond this morphological stage, since it may occur in innovative syntactic constructions with the same semantics (cf. Section 4.2). As already suggested by Amiot/Van Goethem (2012) and Van Goethem (2015), the adjective-like uses of N_2 can be seen as the result of an interaction between the closely related morphological $[N_1\ N_2]_N$ and syntactic $[N\ A]_{NP}$ constructions.[16] The fact that N_2s such as *charnière*, *clé*, *fleuve*, *limite* and so on

16 The schematic representations are a bit simplified since, as we have seen in 4.2, N_1 and N_2 can include a multi-word sequence, and the A can be instantiated by a phrase in the case of degree modification (e.g. *une période vraiment charnière*).

developed a qualifying meaning in the former construction, typical of adjectives, may have favored this constructional ambiguity. In constructional terms, this interaction can be translated as an instance of 'multiple inheritance'. Schematically, this multiple inheritance can be represented as in (27):

(27)



The $[N_1\ [(Adv)\ \textit{charnière}]]_{N/NP}$ sequence inherits its properties from two distinct parent constructions, the morphological qualifying compound $[N_1\ N_2]_N$ pattern (e.g. *moment-charnière* 'pivotal moment') and the syntactic $[N\ [(Adv)\ A]]_{NP}$ pattern (e.g. *un moment* (*vraiment*) *crucial* 'a (really) crucial moment'). As a consequence, and as shown in Section 4.2, it is a hybrid between a morphological and a syntactic construction and N_2 can, in some cases, gradually develop more adjective-like syntactic uses, such as the predicative use.

This approach indicates that French $[N_1\ N_2]$ subordinatives, and especially the subclass of formations with a qualifying N_2, are in reality closely related to [N A] or [A N] formations. As we have seen in Section 3.2, Scalise/Bisetto (2009) merge $[N_1\ N_2]$ appositives and [N A]/[A N] attributives within the class of ATAP compounds because the modifier in both cases expresses a qualifying property of the head noun. We can therefore conclude that their classification for these types of formations is highly insightful. However, what is still missing in this approach is the fact that this ATAP class contains not only pure (morphological) compounds, but also hybrid constructs with both morphological and syntactic properties.

6 Conclusion

Compared with Germanic languages, it turns out to be very difficult to delineate French compounds from syntactic multi-word units. In the first part of this contribution, we outlined three different approaches dealing with compounding in the French tradition: non-restrictive, scalar and restrictive (lexicalist). Although we believe morphological formations should be distinguished from syntactic formations, it is insightful to highlight their shared potential for expressing the same denominative functions. We therefore added a fourth approach: we believe a constructionist, non-modular approach to the language system provides a more

appropriate account. From this perspective, both compounds and phrasal structures with a naming function can act as conventionalized form-meaning pairings or 'constructions' and we should accept the existence of what Booij (2010: 190) calls 'lexical phrasal constructions', namely phrasal constructions that are stored in the (mental) lexicon.

Another advantage of this constructionist approach is that it can deal with structurally ambiguous formations, such as $[N_1\ N_2]$ structures with a qualifying N_2. As shown throughout this paper, these sequences are particularly difficult to deal with in a modular approach because, on the one hand, they formally and semantically resemble $[N_1\ N_2]$ (subordinative) compounds, but, on the other hand, they allow syntactic operations to a greater or lesser extent. In a conception of language as a constructionist network, these hybrid formations can be fruitfully accounted for by the mechanism of 'multiple inheritance'. Following this process, we have argued that the hybrid properties of French qualifying $[N_1\ N_2]$ sequences result from the inheritance of properties from both a morphological and a syntactic parent construction.

References

Ackema, Peter/Neeleman, Ad (2004): Beyond Morphology. Oxford: Oxford University Press.

Amiot, Dany/Van Goethem, Kristel (2012): A constructional account of French *-clé* 'key' and Dutch *sleutel-* 'key' as in *mot-clé*/*sleutelwoord* 'key word'. In: Morphology 22. 347–364.

Anderson, Steven (1992): A-Morphous Morphology. Cambridge, UK: Cambridge University Press.

Arnaud, Pierre J. (2003): Les Composés *timbre-poste*. Lyon: Presses Universitaires Lyon.

Arnaud, Pierre/Renner, Vincent (2014): English and French $[NN]_N$ lexical units: A categorial, morphological and semantic comparison. In: Word Structure 7. 1–28.

Audring, Jenny/Booij, Geert (2016): Cooperation and coercion. In: Linguistics 54. 617–637.

Bauer, Laurie (1998): When is a sequence of two nouns a compound in English? In: English Language and Linguistics 2. 65–86.

Bauer, Laurie (2004): Adjectives, compounds and words. In: Nordic Journal of English Studies 3. 7–22.

Bonami, Olivier/Boyé, Gilles (2003): Supplétion et classes flexionnelles. In: Langages 152. 102–126.

Bonami, Olivier/Boyé, Gilles (2014): De formes en thèmes. In: Villoing, Florence/David, Sophie/Leroy, Sarah (eds.): Foisonnements morphologiques. Études en hommage à Françoise Kerleroux. Presses Universitaires de Paris Ouest. 17–45.

Benveniste, Emile (1974): Problèmes de linguistique générale. Vol. 2. Paris: Gallimard.

Booij, Geert (2002): Constructional idioms, morphology, and the Dutch lexicon. In: Journal of Germanic Linguistics 14. 301–329.

Booij, Geert (2005): Compounding and derivation. Evidence for construction morphology. In: Dressler, Wolfgang U. et al. (eds.): Morphology and its Demarcations. Selected Papers

from the 11th Morphology Meeting, Vienna, February 2004. Amsterdam: Benjamins. 109–132.
Booij, Geert (2010): Construction Morphology. Oxford: Oxford University Press.
Corbin, Danielle (1992): Hypothèses sur les frontières de la composition nominale. In: Cahiers de Grammaire 17. 27–55.
Corbin, Danielle (1997): Locutions, composés, unités polylexématiques: lexicalisation et mode de construction. In: Martins-Baltar, Michel (ed.): La locution entre langue et usages. Paris: ENS. Editions Fontenay/Saint-Cloud. 53–101.
Croft, William (2001): Radical Construction Grammar: Syntactic theory in typological perspective. Oxford: Oxford University Press.
De Caluwe, Johan (1990): Complementariteit tussen morfologische en syntactische benoemingsprocédés. In: De Caluwe, Johan (ed.): Betekenis en productiviteit. Gentse bijdragen tot de studie van de Nederlandse woordvorming. (= Studia Germanica Gandensia 19). Gent: Seminarie voor Duitse taalkunde, Rijksuniversiteit Gent. 9–25.
De Smet, Hendrik/Ghesquière, Lobke/Van de Velde, Freek (eds.) (2013): On multiple source constructions in language change. Special issue of Studies in Language 37, 3. Amsterdam i.a.: Benjamins.
Di Sciullo, Anna Maria/Williams, Edwin (1987): On the definition of word. Cambridge, MA: MIT Press.
Fradin, Bernard (2003): Nouvelles approches en morphologie. Paris: Presses Universitaires de France.
Fradin, Bernard (2009): IE, Romance: French. In: Lieber/Štekauer (eds.). 417–435.
Giegerich, Heinz (2009a): Compounding and lexicalism. In: Lieber/Štekauer (eds.). 178–200.
Giegerich, Heinz (2009b): The English compound stress myth. In: Word Structure 2. 1–17.
Goes, Jan (1999): L'adjectif. Entre nom et verbe. Paris/Bruxelles: Duculot.
Goldberg, Adele (1995): Constructions. A Construction Grammar approach to argument structure. Chicago: University of Chicago Press.
Goldberg, Adele (2006): Constructions at Work. The Nature of Generalization in Language. Oxford: Oxford University Press.
Goldberg, Adele (2009): The nature of generalization in language. In: Cognitive Linguistics 20. 93–127.
Gross, Gaston (1988): Degré de figement des noms composés. In: Langages 90. 57–72.
Gross, Gaston (1996): Les expressions figées en français: noms composés et autres locutions. Paris: Ophrys.
Hacken, Pius ten (1994): Defining Morphology. Hildesheim i.a.: Olms.
Hoffmann, Thomas/Trousdale, Graeme (eds.) (2013): The Oxford Handbook of Construction Grammar. Oxford: Oxford University Press.
Hüning, Matthias/Booij, Geert (2014): From compounding to derivation. The emergence of derivational affixes through "constructionalization". In: Folia linguistica 48, 2. 579–604.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.): Word-formation. An international handbook of the languages of Europe. Vol. 1. (= Handbooks of Linguistics and Communication Science (HSK) 40.1). Berlin/Boston: De Gruyter. 450–467.
Kleiber, Georges (1984): Dénomination et relations dénominatives. In: Langages 76. 77–94.

Lehmann, Alise/Martin-Berthet, Françoise (2008): Introduction à la lexicologie. Sémantique et morphologie. 3rd ed. Paris: Armand Colin.
Lieber, Rochelle (1992): Deconstructing morphology: Word formation in syntactic theory. University of Chicago Press.
Lieber, Rochelle/Scalise, Sergio (2007): The Lexical Integrity Hypothesis in a New Theoretical Universe. In: Booij, Geert et al. (eds.): *On-line Proceedings of the Fifth Mediterranean Morphology Meeting* (*MMM5*), Fréjus 15–18 September 2005. University of Bologna. 1–24. Internet: www.lingue.unibo.it (last access: 11.9.2018).
Lieber, Rochelle/Štekauer, Pavol (eds.) (2009): The Oxford handbook of compounding. Oxford: Oxford University Press.
Mathieu-Colas, Michel (1996): Essai de typologie des noms composés français. In: Cahiers de lexicologie 69. 71–125.
Noailly, Michèle (1990): Le substantif épithète. Paris: Presses Universitaires de France.
Norde, Muriel (2009): Degrammaticalization. Oxford: Oxford University Press.
Norde, Muriel (2014): On parents and peers in constructional networks. Paper presented at *Coglingdays* 2014, Ghent: Ghent University.
Norde, Muriel/Morris, Caroline (2018): Derivation without category change: A network-based analysis of diminutive prefixoids in Dutch. In: Van Goethem, Kristel et al. (eds.): Category change from a constructional perspective. Amsterdam: Benjamins. 47–90.
Norde, Muriel/Van Goethem, Kristel (2014): Bleaching, productivity and debonding of prefixoids. A corpus-based analysis of 'giant' in German and Swedish. In: Lingvisticae Investigationes 37, 2. 256–274.
Plag, Ingo (2003): Word-formation in English. Cambridge, UK: Cambridge University Press.
Scalise, Sergio/Bisetto, Antonietta (2009): The classification of compounds. In: Lieber/Štekauer (eds.). 34–53.
Schlücker, Barbara/Hüning, Matthias (eds). (2009): Words and Phrases – Nominal expressions of naming and description. Special Issue of Word Structure 2, 2. Edinburgh: Edinburgh University Press.
Traugott, Elizabeth Closs/Trousdale, Graeme (2013): Constructionalization and Constructional Changes. Oxford: Oxford University Press.
Trousdale, Graeme (2013): Multiple inheritance and constructional change. In: Studies in Language 37, 3. 491–514.
Trousdale Graeme/Norde, Muriel (2013): Degrammaticalization and constructionalization: two case studies. In: Language Sciences 36. 32–46.
Ureña, José Manuel/Faber, Pamela (2010): Reviewing imagery in resemblance and non-resemblance metaphors. In: Cognitive Linguistics 21, 1. 123–149.
Villoing, Florence (2009): Les mots composés VN. In: Fradin, Bernard/Kerleroux, Françoise/Plénat, Marc (eds.): Aperçus de morphologie du français. Saint-Denis: Presses Universitaires de Vincennes. 75–197.
Villoing, Florence (2012): French compounds. In: Probus 24. 29–60.
Van Goethem, Kristel (2009): Choosing between A+N compounds and lexicalized A+N phrases: The position of French in comparison to Germanic languages. In: Word Structure 2, 2. 241–253.
Van Goethem, Kristel (2012): Le statut des séquences 'N+N à N2 productif': le cas de N-clé. In: Lingvisticae Investigationes 35, 1. 76–93.
Van Goethem, Kristel (2015): Cette mesure est-elle vraiment clé? A constructional approach to categorial gradience. In: Journal of French Language Studies 25, 1. 115–142.

Van Goethem, Kristel/De Smet, Hendrik (2014): How nouns turn into adjectives. The emergence of new adjectives in French, English and Dutch through debonding processes. In: Languages in Contrast 14, 2. 251–277.
Wierzbicka, Anna (1996): Semantics. Primes and universals. Oxford: Oxford University Press.
Zwanenburg, Wiecher (1992): Compounding in French. In: Rivista di linguistica 4. 221–240.

Francesca Masini

Compounds and multi-word expressions in Italian

1 When two (or more) words come together

It is often observed that compounds, being complex words formed by two (or more) words, are the morphological constructions closest to syntactic constructions, and that this is the reason why drawing a line between compounds and phrases is often difficult. Other complex lexical units challenge – possibly even more – the distinction between syntax, morphology and the lexicon: these are generally known as *multi-word expressions* (henceforth MWEs). MWEs are larger than morphological words and are nonetheless stored into our lexicon. The very existence of such MWEs poses a number of theoretical questions regarding (i) the organization of the lexicon, and (ii) the relationship between MWEs and compounds.

The first question has been addressed, among others, by Jackendoff (1995, 1997), who proposes to extend the lexicon to "multiword constructions" (1997: 153), including so-called "constructional idioms" (Jackendoff 1990: 221; cf. also Booij 2002a), since these phenomena are too pervasive to be regarded as a peripheral part of the grammar. This enlarged view of the lexicon is viable under such approaches as the Parallel Architecture (Jackendoff 2010), Construction Morphology (Booij 2010) and Construction Grammar in general (Hoffmann/Trousdale (eds.) 2013).

If we accept MWEs as part of our lexicon, we may want to address the second question, which is exactly what the present volume does. More specifically, we may ask:

a) Is there a way to distinguish between MWEs and compounds? On the basis of which criteria? Are there criteria that would hold crosslinguistically?
b) What kind of role do MWEs and compounds play in the construction of the lexicon? Is there competition between them?

These questions emerge quite naturally, given that both MWEs and compounds are, in a way, complex (multiword) lexical units. Yet, relatively little attention has been devoted to these specific issues, mainly because compounds and MWEs are topics that traditionally belong to different linguistic fields: morphology on the one hand, lexicology and phraseology on the other. In this paper I will address

https://doi.org/10.1515/9783110632446-006

the matter by discussing data from Italian, with a view to contributing some answers to questions in a) and b) above.

First, I briefly describe the state of the art as far as Italian compounds and MWEs are concerned (Section 2). In Section 3 I address demarcation issues concerning compounding and MWEs. Section 4, instead, explores possible areas of competition between compounds and MWEs.

2 Italian: a brief overview

In this section I offer a (necessarily brief and sketchy) overview of Italian compounds and MWEs, which will serve as background knowledge for subsequent sections.

2.1 Italian compounds

Research on Italian compounding has by now a long-stading tradition (cf., among many others, Scalise 1992; Bisetto/Scalise 1999; Bisetto 2004; Ricca 2010; Masini/Scalise 2012; Radimský 2015). Whereas, as widely known, compounding in Italian and Romance languages is not as productive as in Germanic languages, compounds are well-documented in these varieties. In what follows, I illustrate some basic facts about Italian compounds, taking into account the morphological type of the input elements, the lexical categories involved, and the relation among the constituents.

Typically, Italian compounds are made of full (sometimes inflected) words (1a), although we may also find stems (like verbal stems in VN compounds, cf. *cava-* in (1b)), as well as neoclassical formatives or semiwords (cf. (1c), where LV stands for 'linking vowel').

(1a) *pesce-cane*
fish-dog
'shark'

(1b) *cava-tappi*
extract-corks
'corkscrew'

(1c) *crimin-o-logo*
crime-LV-logist
'criminologist'

Compounding in Italian productively feeds mostly the word classes of nouns (2) and adjectives (3), not verbs. As for input elements, productive patterns creating nouns and adjectives involve mostly nouns, adjectives and verbs, secondarily prepositions, as showed in (2)–(3) (where the head is underlined, when present).[1]

(2) Productive compound nouns

(2a) NA *carro armato*
cart armed
'tank'

(2b) NN *agenzia viaggi*
agency travels
'travel agency'

(2c) VN *asciuga-mani*
dry-hands
'towel'

(2d) PN *dopo-guerra*
after-war
'post war period'

(3) Productive compound adjectives

(3a) AN *giallo oro*
yellow gold
'golden yellow'

(3b) AA *marxista-leninista*
Marxist-Leninist
'Marxist-Leninist'

(3c) VN *salva-spazio*
save-space
'space-saving'

As far as the classification of compounds is concerned, Italian displays all six classes identified by Scalise/Bisetto (2009), as summarized in Table 1 (taken from Masini/Scalise 2012: 77).

1 These observations are taken from Masini/Scalise (2012). Patterns with semiwords are not included.

Table 1: Classes of Italian compounds

	Subordinate	Attributive	Coordinate
Endocentric	*capo-stazione* (chief-station) 'stationmaster'	*cassa-forte* (case/box-strong) 'safe'	*poeta pittore* (poet painter) 'poet painter'
	transporto latte (transportation milk) 'milk transportation'	*viaggio lampo* (journey lightening) 'very fast journey'	*divano-letto* (sofa bed) 'sofa bed'
Exocentric	*porta-lettere* (carry-letters) 'mailman'	*viso pallido* (face pale) 'facepale'	*Emilia Romagna* (Emilia Romagna) 'Emilia Romagna'
	sotto-scala (under-stairway) 'closet under the stairway'	*piedi piatti* (feet flat) 'cop'	*dormi-veglia* (sleep-wake) 'drowsiness'

It is worth noting that NN compounds encode the highest number of relations among the constituents, since they may be attributive (ATT), coordinate (CRD) and subordinate (SUB):

(4) Classes of NN compounds

(4a) ATT *pesce spada*
fish sword
'sword fish'

(4b) CRD *divano letto*
sofa bed
'sofa bed'

(4c) CRD *nord-est*
North-East
'North-East'

(4d) SUB *vendita latte*
sale milk
'milk shop'

(4e) SUB *agenzia viaggi*
agency travels
'travel agency'

In attributive NN compounds (4a), the non-head expresses a property of the head noun (often via some metaphorical mechanism), despite not being an adjective. Coordinate (CRD) NN compounds may have two semantic heads (see (4b), where

divano letto is both a sofa and a bed, hence a hyponym of both its input elements), or no internal head at all, like in *nord-est* (4c). Subordinate (SUB) NN compounds also comprise two subtypes, depending on the nature of the head noun, that may be deverbal (like *vendita* in (4d)) or not (like *agenzia* in (4e)).

Finally, one should note that Italian displays at least three productive patterns of exocentric compounds: coordinate NN compounds (cf. (4c)), PN compounds (cf. (2d)) and VN compounds, giving rise both to nouns (2c) and adjectives (3c). The latter is one of the most productive types of compounds in contemporary Italian (cf. Ricca 2010). Hence, exocentricity is well-attested in Italian compounding.

2.2 Italian MWEs and phrasal lexemes

Multi-word expression is widely used as an umbrella term to refer to a large set of linguistic objects (cf. Baldwin/Kim 2010 and Hüning/Schlücker 2015 for an overview), including verbal idioms (5a) and other kinds of idiomatic expressions (e.g. (5b)), sayings (5c), lexicalized sentences (5d), formulae (5e), complex nominals (5f), irreversible binomials (5g), verb-particle constructions (5h) and other complex predicates such as light verb constructions (5i).

(5a) *alzare* *il* *gomito*
raise the elbow
'to drink too much'

(5b) *fuori* *di* *testa*
out of head
'out of one's mind'

(5c) *mai* *dire* *mai*
never say never
'never say never'

(5d) *fai-da-te*
do-from-you
'do-it-yourself'

(5e) *stai* *scherzando?*
stay.2.SG joking
'Are you kidding me?'

(5f) *armi* *di* *distruzione* *di* *massa*
weapons of destruction of mass
'weapons of mass destruction'

(5g) *vivo* *e* *vegeto*
alive and thriving
'alive and well'

(5h) *mettere* *sotto*
put down
'to run over (with a vehicle)'

(5i) *dare* *luogo* (*a*)
give place (to)
'to give rise (to)'

Most of these expressions have been investigated separately from word formation, and within other scholarly traditions. Idioms and collocations, for instance, are typically the realm of phraseology (cf., e.g., Cowie (ed.) 1998) and corpus linguistics (cf., e.g., Moon 1998), but also psycholinguistics (cf., e.g., Cacciari/Tabossi (eds.) 1993) and syntax (cf., among others, Everaert et al. (eds.) 1995).

Morphologists, on the other hand, have always devoted little attention to these multiword phenomena. A notable exception regards complex predicates (cf., e.g., Butt 1995, Ackerman/Webelhuth 1997) – in particular verb-particle constructions in Germanic (cf., e.g., Dehé et al. (eds.) 2002) but also Romance (cf. Iacobini/Masini 2007; cf. also below) languages.

Recently, morphologists have started devoting more attention to this area, especially within the framework of Construction Morphology (Booij 2010; henceforth CxM). This is little surprising – as also observed by Hüning/Schlücker (2015) – given that CxM is linked to Construction Grammar (Hoffmann/Trousdale (eds.) 2013; henceforth CxG), a model whose foundations lie on studies on idiomatic structures, from Fillmore/Kay/O'Connor (1988) onwards.

In CxM, both words and word formation patterns are seen as 'constructions', i.e. conventionalized form-meaning pairings: morphological constructions may differ in size, complexity and schematicity, and are organized into a hierarchical lexicon. Besides, units that are larger than a morphological word but nonetheless conventionalized and stored into our lexicon are also regarded as constructions, as complex signs. Indeed, CxM has originated from work on phenomena in-between morphology and syntax, in particular separable complex verbs in Dutch, which have been treated as a case of 'periphrastic word formation' by Booij (2002b).

In other words, within CxG and CxM, MWEs are seen as part of our lexicon, as anticipated in Section 1. Some MWEs have the same distribution of sentences (sayings) or full VPs (idiomatic expressions); formulaic expressions may also

serve as full utterances (but note that formulae may be constituted also by one single word). Some other MWEs, in particular those that have been called *phrasal lexemes* or *lexical phrases* (Booij 2009a, 2010; Masini 2009, 2012) are closer than other MWEs to morphological words (especially compounds), hence I will mainly focus on these.

Phrasal lexemes are those MWEs that are closest to words in terms of both distribution and function, i.e., they have a word-like distribution (so sentence-level MWEs would *not* be phrasal lexemes) and they have the same concept-naming function of words, thus contributing to lexical enrichment (cf. Masini 2012). They correspond to various patterns and can in principle belong to all lexical categories, at least in Italian, e.g.: nouns (6a), adjectives (6b), verbs (6c), adverbs (6d), prepositions (6e), conjunctions (6f), interjections (6g), pronouns (6h).

(6a) *parte* *del* *discorso*
part of.the speech
'part of speech'

(6b) *felice* *e* *contento*
happy and glad
'happily ever after'

(6c) *stare* *su*
stay up
'to get up'

(6d) *volente* *o* *nolente*
willing or not-willing
'willing or not'

(6e) *di* *fronte* *a*
of front at
'in front of'

(6f) *fino* *a* *che*
until at that
'as long as'

(6g) *porca* *miseria!*
bloody misery
'for God's sake!'

(6h) *se* *stesso*
oneself same
'oneself'

These items are not *words* in the proper sense, since they have a phrase-like structure; some of them may even be separable under certain conditions.[2] At the same time, however, they present a unitary, often conventionalized semantics, and display a higher degree of internal cohesion than free phrases.

As an example, let us take phrasal lexemes that belong to the noun category, i.e., *phrasal nouns*.[3] Italian presents a variety of patterns that fill this class (cf. Masini 2012), including:

(7a)	NPN	*casa*	*di*	*riposo*
		home	of	rest
		'nursing home, hospice'		
(7b)	NPArtN	*parte*	*del*	*discorso*
		part	of.the	speech
		'part of speech'		
(7c)	NA	*anno*	*accademico*	
		year	academic	
		'academic year'		
(7d)	AN	*prima*	*serata*	
		first	evening	
		'prime time'		
(7e)	NConjN	*coltello*	*e*	*forchetta*[4]
		knife	and	fork
		'cutlery'		

Phrasal nouns of the NP(Art)N type, for instance, look like normal noun phrases (formed by a noun plus a prepositional phrase), but are more cohesive than free phrases: indeed, they generally resist various operations (with some variation)

2 This is especially true of verbal expressions: *stare su* (6c), for instance, may be interrupted by a light adverb, e.g. *stai subito su*! (lit. stay immediately up) 'get up immediately!'. On this topic cf. Voghera (2004), who claims that the (different) degree of cohesiveness displayed by these expressions partially depends on the lexical category they belong to, with prepositional and conjuctional phrasal lexemes being more cohesive than adverbial and adjectival ones, the latter being more cohesive than nominal ones, whereas verbal expressions are the least cohesive of all.

3 These items have been named in many different ways in the literature, including, e.g., "phrasal compounds" / "prepositional compounds" (Delfitto/Melloni 2009; Rio-Torto/Ribeiro 2009), and "improper compounds" (Rainer/Varela 1992). The distinction between phrasal nouns and compound nouns is not always trivial, as we will see in Section 3.

4 Coordinate phrasal nouns can also be formed by two verbs, e.g. *va e vieni* (lit. go and come) 'coming and going / toing and froing' (cf. Masini/Thornton 2008).

including interruption (8a), insertion of determiners (8b) or paradigmatic substitution (8c) (cf. Masini 2009 for more details).

(8) *casa di riposo* (lit. home of rest) 'retirement home, hospice'
(8a) **casa* ***rinomata*** *di* *riposo*
home renown of rest
Intended reading: 'renown retirement home'
(8b) **casa* ***del*** *riposo*
home of.the rest
Intended reading: 'retirement home'
(8c) ****abitazione*** *di* *riposo*
dwelling of rest
Intended reading: 'retirement home'

What is crucial about these items is that they are not just univerbations or lexicalized phrases that emerge diachronically. Some certainly are, but a number of them are actually neologisms productively created by speakers to name new concepts. Sometimes, they are calques from other languages. Take for instance the three following examples, from the ONLI database:[5]

(9a) *cibo* *di* *strada*
food of street
'street food'
(9b) *popolo* *della* *rete*
people of.the Internet
'people who use the Internet'
(9c) *città* *digitale*
town digital
'a town endowed with digital technology that inhabitants can use to access public information and services'

Cibo di strada (9a) is a calque from English *street food* which is however rendered in Italian with a NPN phrasal noun rather than a NN compound, which possibly points to the higher availability of the former type. *Popolo della rete* and *città digitale* are new coinages that have been introduced into the Italian language by exploiting the NPArtN and NA patterns, respectively. All examples in (9) are there-

5 Osservatorio Neologico della Lingua Italiana: www.iliesi.cnr.it/ONLI (last access: 11.6.2018).

fore conventionalized phrasal nouns with a naming function. Although wide-ranging quantitative data are still unavailable, it is reasonable to think that phrasal nouns constitute a significant part of neologisms in contemporary Italian. In this respect, it is useful to remind that Émile Benveniste claimed, already in 1966, that NPN is the true, productive compounding pattern in French (called by the author "synapsie", e.g. *clair de lune* lit. light of moon 'moonlight', *moulin à vent* lit. mill at wind 'windmill').

In addition to nouns, Italian has a variety of phrasal means to form complex predicates. This is important in view of the fact that Italian lacks verbal compounding altogether; therefore, multiword verb formation may be seen as a way to compensate this part of the Italian lexicon. There are two patterns that are especially prominent in this domain: verb-particle constructions and light verb constructions. As is well-known by now, Italian, despite being a Romance language, also has particle verbs (e.g., Masini 2005, Iacobini/Masini 2007, Iacobini 2015), although the phenomenon is not as pervasive as in English (see (10)). Also light verb constructions are quite widespread (11): they are formed by a light, generic verb plus a predicative noun (cf., e.g., Jezek 2004).

(10a) *andare* *su*
go up
'to go up(wards) / to ascend'

(10b) *mettere* *sotto*
put down
'to run over (with a vehicle)'

(10c) *guardare* *avanti*
look forward
'to look forward / to look to the future'

(10d) *buttare* *via*
throw away
'to throw away / to waste'

(11a) *mettere* *fretta*
put hurry
'to hurry (causative)'

(11b) *prendere* *freddo*
take cold
'to get cold'

(11c) *avere* *paura*
have fear
'to be afraid'

(11d) *dare* *vita* (*a*)
give life (to)
'to create'

Since phrasal lexemes (and other MWEs) can be seen as constructions within CxM – exactly like simple and complex words, as well as word formation schemas and subschemas – we expect them to interact in various ways with word-formation processes. Hüning/Schlücker (2015) claim that "MWEs and compounds are largely a complementary means for creating lexical units". In Section 4, I offer some data and reflections about the relationship between these two strategies in terms of competition. Before that, however, it is necessary to discuss some demarcation issues.

3 Demarcation issues

Starting from the idea that we have two sets of complex lexical constructions that are used to form stable (stored), complex denotations in the world's languages, namely compounds and MWEs, we may ask if they can actually be distinguished, and on which ground. In addition, we may want to ask whether their demarcation is clear-cut or not in every language, and if the criteria to be used are valid crosslinguistically. The expectation is that crosslinguistic validity is hardly achievable, since the demarcation between compounds and MWEs ultimately has to do with the demarcation between morphology and syntax, between words and phrases, which is a well-known, unsolved question, especially in a typological perspective (cf., e.g., Haspelmath 2011).

Compounds as purely morphological objects have been defined by Guevara/Scalise (2009: 108) as complex words formed by two (or more) words whose general structure is captured by the formula in (12). Let us take this operational definition as a starting point for our discussion.

(12) $[X\ \Re\ Y]_Z$
where X, Y and Z represent major lexical categories, and $\Re$ represents an implicit relationship between the constituents (a relationship not spelled out by any lexical item)

First, it is interesting to note that, according to (12), compound words should belong to major lexical categories only, i.e. to open classes that can be synchronically enriched with new members. As we have seen in Section 2, phrasal lexemes in Italian can also belong to minor lexical categories. Hence, this is one possible difference between compounds and MWEs in Italian (not necessarily valid in every language). However, MWEs belonging to minor lexical categories (prepositions, conjunctions, etc.) are basically the result of a diachronic process of lexicalization or univerbation, whereas at least some of the MWEs belonging to major lexical categories seem to result from a synchronic process of lexical creation. Therefore, synchronically speaking, both compounds and MWEs feed the same (open) classes (as is natural to expect).

Second, not all major lexical categories are equally fed by compounding and MWEs: languages differ in this respect. In Italian, compounds are mostly nouns (which is also the primary input category) and secondarily adjectives, whereas compound verbs and adverbs are basically absent. MWEs, on the other hand, feed also verbs and adverbs.

Third, the restriction to lexical categories implies that higher level structures (e.g. sentences) are excluded from compounding (obviously so), whereas we know that some MWEs may coincide with full sentences and utterances (e.g. sayings and formulaic expressions) or full VPs (cf. especially verbal idioms, e.g. like *mettere le mani avanti* lit. put the hands forward 'to prevent an unpleasant situation').

So, overall, we can conclude that in Italian (and possibly other languages) compounds and MWEs have a partially different distribution: whereas compounds function as word-level elements, MWEs may also correspond to full phrases and even sentences. Of course, there is a subset of MWEs – named here *phrasal lexemes* – that are closer to compounds in that, as anticipated in Section 2.2, they: i) have the same concept-naming function of compounds and words in general; ii) have the same distribution of a word (e.g. *carta di credito* 'credit card', which functions syntactically like a noun, with which it may be substituted: *pagare con la carta di credito* 'to pay with credit card' vs. *pagare con i contanti* 'to pay with cash').

Then, how can we distinguish compounds and phrasal lexemes in Italian?

Let us concentrate on phrasal nouns, since noun is the preferred output category for Italian compounding (but also crosslinguistically: cf. Guevara/Scalise 2009). Before focusing on Italian, I briefly discuss some examples from various languages that are meant to illustrate some of the criteria proposed in the literature to distinguish between compound nouns and phrasal nouns.

In Dutch, AN compounds and AN phrasal lexemes can be formally distinguished since the latter display agreement inflection on the adjective (see the suffix *-e* in (13a–b)),[6] whereas the former do not (13c–d) (cf. Booij 2009a).

(13a) *donker-**e** kamer* (Dutch)
'dark room'
(13b) *mager-**e** yoghurt*
'fat-free yoghurt'
(13c) *fijn-stof*
'fine-grained dust'
(13d) *vroeg-geboorte*
'premature birth'

German works very similarly (cf. Schlücker/Hüning 2009): like in Dutch, in German AN compounds, the adjective is not inflected, bears the main stress and is generally monomorphemic (14a), whereas in AN phrasal lexemes the adjective is inflected, does not bear the main stress, and can be complex (cf. (14b–c)).

(14a) *Rot-wein* (German)
'red wine'
(14b) *werdende Mutter*
'mother-to-be'
(14c) *werdender Vater*
'father-to-be'

In Russian, phrasal nouns display regular agreement (15a) or government (15b) among the constituents (which are independent words), whereas compounds do not, since the first member is typically a *root* (hence a bounded element) connected to the second constituent by a linking vowel (16) (cf. Masini/Benigni 2012).

(15a) *suchoe* *moloko* (Russian)
dry.NOM.SG.NEUT milk.NOM.SG.NEUT
'powdered milk'
(15b) *točka* *zrenija*
point.NOM.SG.F view.GEN.SG.NEUT
'point of view'

6 As Booij (2009a: 224) states, "[t]he pre-nominal adjective ends in the suffix *-e*, unless the NP is indefinite and the head noun is singular and neuter (in the latter case the ending is zero)".

(16) *sux-o-frukty* (Russian)
dry-LV-fruit.M.PL.NOM
'dry fruit'

Quite expectedly, the criteria vary from language to language, depending on language-specific properties. Some criteria may be shared by more than one language (e.g. Dutch and Russian share the loss of agreement inflection, although the phenomenon is more consistent in Russian), whereas others may not (e.g. Russian linking vowels can be used to distinguish compounds from phrases, but not all languages feature these items). Furthermore, some criteria are themselves questionable: it is not clear, for instance, whether "semantic transparency" would be a reliable criterion, as we will discuss below.

What about Italian? How can we say, for instance, that the expressions in (17) are phrasal nouns and not compounds?

(17a) *cart-a* *telefonic-a* $[NA]_N$
card-F.SG of_the_phone-F.SG
'phone card'

(17b) *terz-o* *mond-o* $[AN]_N$
third-M.SG world-M.SG
'Third World'

(17c) *casa* *di* *cura* $[NPN]_N$
house of treatment
'nursing home'

(17d) *botta* *e* *risposta* $[NConjN]_N$
blow and answer
'cut and thrust, verbal crossfire'

It seems to me that the following criteria might be used for Italian, taking the definition in (12) and lexical integrity as reference points:

(18a) **internal agreement** (absent in compounds, present in phrasal lexemes);
(18b) **explicit relational markers**, such as conjunctions and prepositions (absent in compounds, present in phrasal lexemes);
(18c) **minor lexical categories**, such as articles (absent in compounds, present in phrasal lexemes);
(18d) **bounded elements**, such as roots/stems or linking vowels (present in compounds, absent in phrasal lexemes).

Agreement in number and gender is present in (17a) and (17b), as shown by the glosses. The presence of explicit relational markers is displayed by both (17c) (preposition *di* 'of') and (17d) (conjunction *e* 'and'). The presence of minor lexical categories is shown by the examples in (19): in (19a) the two nouns are linked by a preposition with article (*della* 'of the' = *di* 'of' + *la* 'the.F.SG'), whereas in (19b) we have a lexicalized expression containing an article. Finally, bounded elements show up in compounds only (cf. Section 2.1).[7]

(19a)	*macchina*	*della*	*verità*
	machine	of.the	truth
	'lie detector'		
(19b)	*cessate*	*il*	*fuoco*
	cease.IMP.PL	the	fire
	'ceasefire'		

It is worth noting that the proposed criteria are formal, not semantic. Bisetto (2004) proposes a semantic criterion to distinguish between compounds and so-called *polirematiche* (an Italian standard term for *phrasal lexemes*): compounds would be the result of a productive process, thus tending to be hyponyms of their heads, whereas *polirematiche* would arise from lexicalization and thus typically display a non-compositional meaning. In our view, this semantic criterion is not really deciding: on the one hand, we may have compounds that are formed productively and can be readily interpreted by the hearer (cf. (20a), where *capostazione* is actually a type of *capo*) and compounds that are more lexicalized and whose semantics is not as transparent (20b); on the other hand, phrasal nouns may either be created on the basis of a productive and interpretable pattern (21a) or arise from lexicalization or idiomatization of a phrase (21b).

(20a)	*capo-stazione*	
	head-station	
	'stationmaster'	

7 A possible counterexample would be the phrasal lexemes with two coordinated verbs mentioned in footnote 4 (e.g. *va e vieni* lit. go and come 'coming and going'). As argued by Masini/Thornton (2008), the verbal forms used in these expressions are homophonous to the 2nd person singular imperative, exactly like the verbal forms occurring within VN compounds. If we analyze this verbal form as some sort of morphomic stem used in Italian morphology, we end up with a clash: use of a bounded form on the one hand, and presence of an explicit relational marker (*e* 'and') on the other.

(20b) *capo-cielo*
head-sky
'canopy erected over a high altar'

(21a) *mulino* *a* *vento*
mill at wind
'windmill'
(21b) *luna* *di* *miele*
moon of honey
'honeymoon'

The application of the criteria proposed above is not always straightforward and may produce unexpected results. Take for instance the agreement criterion. This is pretty efficient in Dutch and Russian, but less so in Italian, since agreement takes place in virtually all combinations of a noun and an adjective. This means that even an expression like *croce-rossa* (cross.F.SG-red.F.SG) 'Red Cross', which is traditionally regarded as a compound in the literature (like many others, e.g., *cassaforte* 'safe' in Table 1), should instead be considered as a phrasal lexeme by this criterion, exactly like *carta telefonica* and *terzo mondo* in (17a–b).

Along these lines, one may argue that also "internal inherent inflection" (i.e. inflectional marking occuring inside the word, not triggered by agreement, such as number for nouns) should be considered as a criterion to be added to the list in (18). Also in this case, we would end up regarding many Italian items (traditionally analyzed as compounds) as phrasal lexemes, such as left-headed NN compounds of the *capostazione* type (20a), in which the plural marker applies to the left (head) constituent: *capo-stazione* (lit. head-station) 'stationmaster' turns into *capi-stazione* (lit. heads-station) 'stationmasters', and not **capo-stazioni* (head-stations), with plural marker on the right (as we would expect from a "true word"). However, we can also observe that, despite internal inflection, *capostazione* is still (at least partly) compound-like due to the absence of any relational element (see criterion (18b)) between *capo* and *stazione* (cf. the corresponding phrasal expression *capo della stazione* lit. head of.the station). Therefore, the compound-phrasal noun demarcation may be a matter of degree rather than clear-cut (cf. also footnote 7): for instance, compounds that display no internal agreement and no internal (inherent) inflection (e.g. *dopoguerra* 'post war period' or *asciugamani* 'towel', cf. (2), Section 2.1) are more compound-like than *capostazione* (which is split by inflection in the plural). In other words, the concepts of compound and phrasal lexeme may be seen as prototypes, or radial categories, that can be defined on the basis of a complex interaction of properties, rather than on a set of necessary and sufficient features.

All in all, based on the observations above, one may note that the demarcation between noun compounds and phrasal nouns in Italian largely relies on criteria that typically distinguish words from phrases, with the complication that phrasal nouns are not free, full-fledged phrases.[8] As is well-known, word(hood) is far from being a simple concept with crosslinguistic validity (Haspelmath 2011). However, CxM assumes that "cohesiveness is the defining criterion for canonical wordhood" (Booij 2009b: 97). And cohesiveness obviously manifests itself in different ways in different languages, depending on the morphological and syntactic properties of the language in question. So, the exact criteria to be used should be identified on a language-specific basis, but the same general principle applies.

Given these premises, we might expect to have languages in which the *formal* differences between compounds and phrasal lexemes are evident and easily detectable (e.g. Russian, i.e. a language where compounds are mostly *root*-compounds), languages in which these are vague or even non-existent (e.g. English, where it is very difficult to state whether conventionalized AN combinations such as *black board* are compounds or phrases, cf. Giegerich 2005, 2009), and languages, such as Italian, that are in-between, since they offer at least some evidence in favor of maintaining such a division.

In conclusion, we may regard the demarcation between compounds and phrasal lexemes as an element of variation among the languages of the world that possibly correlates with their morphological type: with the limited data gathered so far, we may hypothesize that this demarcation is clearer in highly inflectional languages displaying root compounding, whereas in isolating languages the boundary is definitely more blurred, if not absent.

4 Competition issues

Competition in morphology and the lexicon is generally viewed as a relation holding between different word-level strategies that compete to realize the same grammatical or lexico-conceptual meaning. However, recent work has claimed that morphological words also compete with MWEs (Booij 2010; Hüning/Schlücker 2015; Masini 2016, to appear). The relationship between morphological words and MWEs, however, is still underinvestigated and calls for further research.

8 The following general properties keep phrasal lexemes apart from true, free phrases: greater internal cohesion, paradigmatic fixedness (i.e., they resist lexical substitution) and conventionalized (though not necessarily idiomatic) meaning (cf. Section 2.2, example (8)).

In this section I show that competition between compounds and MWEs may result in the blocking of specific lexical items, and that these blocking effects may operate in both directions.[9] More specifically, I briefly illustrate three case-studies regarding the competition between compounds and phrasal lexemes in the nominal domain, namely: i) NP(Art)N phrasal nouns (e.g. *macchina della verità* 'lie detector') in comparison with NN compounds (e.g. *capostazione* 'stationmaster') (Section 4.1); ii) the simile construction with color adjectives (e.g. *rosso come il fuoco* lit. red as the fire 'red as fire') in comparison with the corresponding compound pattern (e.g. *rosso fuoco* lit. red fire 'fire-like red') (Section 4.2); iii) irreversible binomials (e.g. *sano e salvo* 'safe and sound') as compared with coordinate compounds of the *sordomuto* 'deaf-mute' type (Section 4.3).

4.1 Complex nominals: NP(Art)N phrasal nouns vs. NN compounds

NN compounding is attested in Romance languages, including Italian (cf., e.g., Masini/Scalise 2012, Radimský 2015). At the same time, we have NP(Art)N phrasal nouns (cf. Section 2), which are another productive way to form complex nominals in Romance languages (especially – but not exclusively – in special languages), as already noted by Benveniste (1966) for French (cf. also Voghera 2004 and Masini 2009 for Italian; Bernal 2012 for Catalan; Rio-Torto/Ribeiro 2012 for Portuguese). See some examples below.

(22a)	*giacca*	*a*	*vento*	(Italian)
	jacket	at	wind	
	'windbreaker'			
(22b)	*moulin*	*à*	*vent*	(French)
	mill	at	wind	
	'windmill'			
(22c)	*mal*	*de*	*cap*	(Catalan)
	pain	of	head	
	'headache'			
(22d)	*cadeira*	*de*	*rodas*	(Portuguese)
	chair	of	wheels	
	'wheelchair'			

9 For a broader picture of the competition between MWEs and all kinds of morphological words, including simple words and derived words, cf. Masini (2016, to appear).

Given that NN compounds and NP(Art)N phrasal nouns coexist in Italian, and that both are used to coin new complex nominals, competition between these two patterns is likely to emerge. As a case-study, let us consider Italian NN compounds where *capo* 'head, boss' is the head (leftmost) constituent. This pattern of compounding is pretty productive in Italian and is associated with the meaning 'head/boss of N'.[10]

(23a) *capo-stazione*
head-station
'stationmaster'
(23b) *capo-classe*
head-class
'class president'
(23c) *capo-gruppo*
head-group
'group leader'
(23d) *capo-famiglia*
head-family
'head of the family'

Other possible *capo*+N compounds could be:

(24a) °*capo-stato*
head-state
(24b) °*capo-governo*
head-government
(24c) °*capo-polizia*
head-police

However, these perfectly well-formed items are not actually produced (the ° sign marks well-formed but non-existent expressions). The reason for this is that they are *blocked* by already existing NPN phrasal nouns featuring the same constituent words, namely:

10 Note that not all *capo*+N compounds have this semantics: some mean 'chief N', such as *capo-redattore* (lit. head-editor) 'editor in chief'.

(25a) *capo* *dello* *stato*
head of.the state
'head of state'
(25b) *capo* *del* *governo*
head of.the government
'Prime Minister'
(25c) *capo* *della* *polizia*
head of.the police
'chief of police'

The reverse may also occur: for instance, the expression *capo della classe* (26) is perfectly grammatical and interpretable as 'class president'; however, it is not used with this specific intended reading, because the same meaning is already conveyed by the established compound *capoclasse* (cf. (23b)).

(26) °*capo* *della* *classe*
head of.the class
'class president'

This type of competition in Italian can be compared to a similar case in Dutch and German. In these languages, AN combinations could be realized either as phrasal nouns (cf. (27a), (28a)) or as compounds (cf. (27b), (28b)) (cf. Booij 2009a, 2010 for Dutch; Hüning/Schlücker 2015 for German; cf. also Section 3). If we try to create the corresponding combination (cf. (27a'–b'), (28a'–b')), we get a possible but non-existent or non-conventionalized expression (in the intended reading).

(27a) *grüne Welle* (lit. green wave) 'progressive signal system' (German)
(27a') °*Grünwelle*
(27b) *Dunkelkammer* 'darkroom'
(27b') °*dunkle Kammer*

(28a) *wilde gans* (lit. wild goose) 'brant' (Dutch)
(28a') °*wildgans*
(28b) *sneltrein* (lit. fast-train) 'express train'
(28b') °*snelle trein*

These data may be interpreted either as cases of *lexical blocking* (Rainer 2016) (*token blocking* in Rainer 1988) or rather as an effect of a more general tension between two competing patterns, namely, in the Italian case, between NN com-

pounding on the one hand and NP(Art)N phrasal lexemes on the other. Both views are viable in a constructionist view of morphology and the lexicon, where constructions are arranged into an inheritance hierarchy where abstract schemas generalize over more specific constructions. Hence, which type of blocking is actually at work is an empirical question.

4.2 Complex color expressions: simile constructions vs. compounds

In many languages we find simile constructions with an intensifying meaning headed by an adjective of the type exemplified in (29) for English (cf. Kay 2013) and (30) for German (cf. Hüning/Schlücker 2015, Schlücker this volume). Most are conventionalized and qualify as MWEs.

(29) [A *as* NP] (English)
(29a) *dead as a doornail* 'quite dead'
(29b) *light as a feather* 'extremely light'
(29c) *flat as a pancake* 'completely flat'

(30) [(*so*) A *wie* NP] (German)
(30a) (*so*) *weiß wie Schnee* '(as) white as snow'
(30b) (*so*) *flink wie ein Wiesel* '(as) quick as a flash'
(30c) (*so*) *schlank wie eine Gerte* '(as) slender as a whip'

A very similar pattern is found in Italian, where *come* corresponds to *as* and *wie*:

(31) [A *come* NP]

(31a)	*vecchio*	*come*	*il*	*mondo*
	old	as	the	world
	'very old'			
(31b)	*bello*	*come*	*il*	*sole*
	beautiful	as	the	sun
	'very beautiful'			
(31c)	*liscio*	*come*	*l'*	*olio*
	smooth	as	the	oil
	'very smooth'			

If we search the [A *come* NP] pattern in a large corpus[11] and rank the results for frequency, what we find is that many of the top ranked occurrences contain a color term (32), most notably *nero* 'black', *bianco* 'white' and *rosso* 'red' (but also other colors, e.g. *azzurro* 'light-blue', *giallo* 'yellow', *blu* 'blue', *verde* 'green'). The simile construction with color terms apparently retains the intensification meaning associated with the general [A *come* NP] construction.

(32) [A_{COLOR} *come* NP]

(32a)	*nero*	*come*	*la*	*pece*
	black	as	the	pitch
	'pitch black'			
(32b)	*bianco*	*come*	*la*	*neve*
	white	as	the	snow
	'snow-white'			
(32c)	*rosso*	*come*	*il*	*sangue*
	red	as	the	blood
	'blood red'			

Interestingly, some of the AN pairs occurring within the simile construction are also found as compounds in German, as noted by Hüning/Schlücker (2015):

(33a)	*weiß*	*wie*	*Schnee*		~	*schneeweiß*	(German)
	white	as	snow			'snow-white'	
(33b)	*flink*	*wie*	*ein*	*Wiesel*	~	*wieselflink*	
	nimble	as	a	weasel		'quick as a flash'	
(33c)	*schlank*	*wie*	*eine*	*Gerte*	~	*gertenschlank*	
	slender	as	a	whip		'(as) slender as a whip'	

The same holds for Italian, but only for a subset of expressions, namely those containing a color adjective (34). Similar doublets are not found in Italian with other kinds of adjectives (cf. (35), corresponding to (31)).

11 The data for this analysis are taken from the *Italian Web 2010* (or *itTenTen10*) corpus, a web corpus of approx. 2,5 billion words searched through the SketchEngine (www.sketchengine.co.uk, last access: March 2017).

(34)	[A_{COLOR} *come* NP] MWE				~	[A_{COLOR} N] compound
(34a)	*bianco*	*come*	*il*	*latte*	~	*bianco latte*
	white	as	the	milk		
	'milk-white'					
(34b)	*nero*	*come*	*il*	*carbone*	~	*nero carbone*
	black	as	the	coal		
	'coal-black'					
(34c)	*azzurro*	*come*	*il*	*cielo*	~	*azzurro cielo*
	light_blue	as	the	sky		
	'sky-blue'					
(35a)	*vecchio*	*come*	*il*	*mondo*	~	**vecchio mondo*
	old	as	the	world		
	'very old'					
(35b)	*bello*	*come*	*il*	*sole*	~	**bello sole*
	beautiful	as	the	sun		
	'very beautiful'					
(35c)	*liscio*	*come*	*l'*	*olio*	~	**liscio olio*
	smooth	as	the	oil		
	'very smooth'					

Compounds of the [A_{COLOR} N] type are relatively common in Italian (cf. D'Achille/Grossmann 2010, 2013). The color A is the head of the compound (and is generally invariable), whereas the N serves as a modifier: more precisely, it denotes a referent that typically exemplifies the shade of the color in question. The expression *giallo canarino* (lit. yellow canary), for instance, denotes a kind of yellow that is typically exemplified by canary birds.

Therefore, we have a domain, that of complex color adjectives, where there seem to be two competing strategies that form expressions with similar content: [A_{COLOR} *come* NP] multiword simile constructions and [A_{COLOR} N] compounds. How much do they actually overlap?

In order to answer this question, I generated frequency lists of both the [A_{COLOR} N] and the [A *come* NP] pattern for five color terms (*nero* 'black', *bianco* 'white', *rosso* 'red', *azzurro* 'light-blue', *verde* 'green'), using the itTenTen10 corpus (cf. footnote 11), and then I compared the top results of the (manually revised) lists, in order to see if the two constructions occur with the same nouns. It turned out that the two constructions share quite a lot of nouns, thus producing a considerable number of doublets. As an exemplification, see the 15 top ranked hits for *rosso* 'red' in Table 2, where the grey cells highlight the nouns that both constructions occur with.

Table 2: Comparing [*rosso* N] and [*rosso come* NP]: top ranked results from the itTenTen10 corpus

[*rosso* N] 'N-red'	Ns
rosso fuoco	(**fire**)
rosso rubino	(ruby)
rosso sangue	(**blood**)
rosso porpora	(purple)
rosso ciliegia	(**cherry**)
rosso mattone	(brick)
rosso corallo	(**coral**)
rosso tramonto	(**sunset**)
rosso fiamma	(flame)
rosso fragola	(strawberry)
rosso pomodoro	(**tomato**)
rosso ruggine	(rust)
rosso vino	(wine)
rosso passione	(**passion**)
rosso papavero	(**poppy**)

[*rosso come* NP] 'red as NP'	Ns
rosso come il sangue	(**blood**)
rosso come il fuoco	(**fire**)
rosso come un peperone	(pepper)
rosso come un pomodoro	(**tomato**)
rosso come un gambero	(shrimp)
rosso come la passione	(**passion**)
rosso come un papavero	(**poppy**)
rosso come un tacchino	(turkey)
rosso come il cuore	(heart)
rosso come una ciliegia	(**cherry**)
rosso come un peperoncino	(hot pepper)
rosso come la terra	(earth)
rosso come la brace	(embers)
rosso come il tramonto	(**sunset**)
rosso come il corallo	(**coral**)

A similar picture emerged for other colors. For instance: *nero* 'black' frequently occurs with *pece* 'pitch', *notte* 'night', *carbone* 'coal', *inchiostro* 'ink' and *petrolio* 'oil' in both constructions (vs. e.g. *morte* 'death', which selects only the simile construction: *nero come la morte* lit. black like the death 'intense black'); *bianco* 'white' frequently occurs with *latte* 'milk', *avorio* 'ivory', *marmo* 'marble', *neve* 'snow', *carta* 'paper' and *cadavere* 'corpse' in both constructions (vs. *cencio* 'rag' and *crema* 'cream', which occur only in one construction: *bianco come un cencio* lit. white as a rag 'very pale', *bianco crema* lit. white cream 'cream-like white'). Therefore, these two constructions share quite a lot of environment and actually seem to compete with each other.

At this point, one may inquire whether they are really equivalent. Take for instance the pairs in (36)–(39), where the (a) examples are taken from the itTenTen10 corpus and the (b) examples contain the corresponding (either MWE or compound) expression.

(36a) *Una villa* ***bianco neve*** *si stagliava su un pendio scosceso*
'A snow-white villa stood out on a steep slope'
(36b) *Una villa* ***bianca come la neve*** *si stagliava su un pendio scosceso*

(37a) *Per arrivarci bisogna guadare a piedi un fiume* [...] ***rosso come la ruggine***
'To get there you have to cross on foot a rust-like red river'
(37b) *Per arrivarci bisogna guadare a piedi un fiume* [...] ***rosso ruggine***

(38a) *Le occhiaie* ***nero pece*** *mi ricordano della nottata appena trascorsa*
'The pitch-black bags under my eyes remind me of the night that has just passed'
(38b) *Le occhiaie* ***nere come la pece*** *mi ricordano della nottata appena trascorsa*

(39a) *I suoi occhi*, ***azzurri come il ghiaccio***, *mandavano lampi gelidi*
'His eyes, blue as the ice, were sending icy flashes'
(39b) *I suoi occhi*, ***azzurro ghiaccio***, *mandavano lampi gelidi*

In these pairs, the two expressions seem quite interchangeable. However, a closer analysis of a number of examples showed that interchangeability is possible in specific contexts that meet certain semantic properties, to which I now turn.

I mentioned above that compounds of the [A_{COLOR} N] type denote, quite neutrally, a kind of color that is typically exemplified by N, whereas the simile construction with color terms, besides denoting a type of color, shares the intensification meaning with the general [A *come* NP] construction. This intensifying effect is especially prominent when N refers to an object that is associated with an intense shade, or with the focal shade of the color in question (40a). The intensification effect diminishes when N identifies a referent that *is not* associated with such an intense or "prototypical" shade (40b). At the same time, when the compound features an N that identifies a referent that *is* associated with such an intense or "prototypical" shade of the color at hand (41a), some slight intensification emerges, otherwise absent in this construction (41b).[12]

(40a) *rosso come il sangue* (lit. red as the blood) 'blood-red'
⇒ true/intense red

12 Incidentally, the association with an entity (N) that is regarded as a prototypical example of the property conveyed by A might actually be at the basis of the intensification meaning conveyed by the more general [A *come* NP] construction.

(40b) *rosso come la ruggine* (lit. red as the rust) 'rust-like red'
≠ true/intense red

(41a) *bianco neve* (lit. white snow) 'snow-like white / snow-white'
⇒ true/pure white
(41b) *bianco avorio* (lit. white ivory) 'creamy-white'
≠ true/pure white

The two patterns are more likely to be interchangeable when they tend to "converge", i.e. when the intensification value is low in the [A_{COLOR} *come* NP] pattern (cf. (37), (39)) and when some intensification emerges in the [A_{COLOR} N] pattern (cf. (36), (38)), depending on the kind of N used. This said, it must be added that even in these specific situations, the two constructions are not totally equal semantically, because the simile construction always has a higher degree of expressiveness, probably inherited by the more general simile construction of which it is an instance. Compounds, on the other hand, are more objective and neutral. In those contexts where they are interchangeable, the two expressions may thus be seen as propositional synonyms (Cruse 2004), i.e. as denotationally equivalent but different in expressive meaning.

Besides semantics, there are a number of formal properties, partially derived from their phrasal vs. morphological status, that differentiate the two constructions. First of all, in the [A_{COLOR} *come* NP] pattern the color adjective is variable (see e.g. (39a), where *azzurri* agrees in number and gender with *occhi*: plural, masculine), whereas in the compound pattern it is primarily invariable:[13]

(42a)	*una*	*maglia*	*verd**e***	*prato*
	a	sweater.SG	green.SG	lawn
	'a lawn-like green sweater'			
(42b)	*due*	*magli**e***	*verd**e***	*prato*
	two	sweater.PL	green.SG	lawn
	'two lawn-like green sweaters'			
(42c)	?**due*	*magli**e***	*verd**i***	*prato*
	two	sweater.PL	green.PL	lawn

Second, in the [A_{COLOR} *come* NP] pattern, the color adjective can only be an adjective, whereas the compound may also be used as a noun:

13 Although D'Achille/Grossmann (2013) observed some variation in corpora.

(43a) *Il **rosso fuoco** non ti si addice*
'Fire-like red doesn't befit you'
(43b) **Il **rosso come il fuoco** non ti si addice*
'Red as fire doesn't befit you'

Third, although the two constructions share a lot of nouns, not all nouns are equally likely to occur in both constructions. For instance, the combination of *azzurro* 'light blue' and *polvere* 'dust' seems to occur within the compound pattern only (44a), whereas the combination of *nero* 'black' and *buio* 'dark' seems to work only within the simile construction (44b).

(44a)	*azzurro*	*polvere*		
	light_blue	dust		
	'dust-like light blue'			
	°*azzurro*	*come*	*la*	*polvere*
	light_blue	as	the	dust
(44b)	*nero*	*come*	*il*	*buio*
	black	as	the	dark
	'intense black'			
	°*nero*	*buio*		
	black dark			

In some cases, the attempt to apply a given A-N combination occurring in one construction to the other construction results in an unacceptable string. This typically happens when N is an abstract noun (45a–a'), when a metonymy is at work (cf. (45b–b'), where the entity referred to is not a cardinal, but the cardinal's cassock), and when the association with N has a purely intensifying effect, like in (45c–c'), where there is no obvious relationship between rags and whiteness.

(45a)	*giallo*	*tradimento*		
	yellow	betrayal		
	'typical yellow (color associated with betrayal)'			
(45a')	**giallo*	*come*	*il/un*	*tradimento*
	yellow	as	a/the	betrayal
(45b)	*rosso*	*cardinale*		
	red	cardinal		
	'cardinal red'			
(45b')	**rosso*	*come*	*un*	*cardinale*
	red	as	a	cardinal

(45c) *bianco* *come* *un* *cencio*
white as a rag
'very pale'
(45c') **bianco* *cencio*
white rag

In conclusion, what emerges from this overall picture is that the two constructions are not really equivalent in terms of both meaning and form. In some specific instances the two versions – compound and multiword – are pretty close and possibly competing with one another (although the multiword version is generally more expressive), but even in these cases they have partially different structural properties. Besides, they do not share the whole array of possible A-N pairs. In other words, the two constructions seem to do their best not to overlap too much, and to differentiate from each other.

4.3 Coordination in the lexicon: irreversible binomials vs. compounds

The last case-study I am going to briefly discuss concerns morphological and multiword coordinating constructions. As exemplified below, Italian displays both coordinate compounds (46) and so-called "irreversible binomials" (cf., among many others, Malkiel 1959; Lambrecht 1984; Masini 2006 for Italian) (47):

(46a) *sordo-muto*
deaf-mute
'deaf-mute'
(46b) *studente-lavoratore*
student-worker
'student-worker'
(46c) *agro-dolce*
sour-sweet
'sweet and sour, bittersweet'
(46d) *ceco-slovacco*
Czech-Slovak
'Czechoslovak'

(47a) *sano* *e* *salvo*
healthy and safe
'safe and sound'

(47b) *vivo* *e* *vegeto*
alive and thriving
'alive and well'
(47c) *anima* *e* *corpo*
soul and body
'body and soul'

Along the lines of the pattern for coordinate compounds, we might theoretically form compounds like those in (48): however, these expressions are not actually created by speakers because the corresponding binomials already exist (47a–b).

(48a) °*sanosalvo*
healthy-safe
(48b) °*vivo-vegeto*
alive-thriving

The reverse situation may also occur: for instance, the existence of an established coordinate compound like *sordomuto* (46a) blocks the formation, or lexicalization, of the corresponding irreversible binomial (49), which would be technically well-formed.

(49) °*sordo* *e* *muto*
deaf and mute

To which extent are these two patterns – coordinate compounds and irreversible binomials – actually equivalent? Let us take a step back.

Arcodia/Grandi/Wälchli (2010: 178) propose a macro-distinction between: i) "hyperonymic coordinate compounds" (what Wälchli 2005 calls "co-compounds"), which express superordinate-level concepts, i.e. their referent is in a superordinate relationship to the meaning of the parts (cf. (50a)); ii) "hyponymic coordinate compounds", which express subordinate-level concepts, i.e. their referent is in a subordinate relationship to the meaning of the parts (cf. (50b)). They also claim that, whereas the latter are common in Standard Average European (SAE) languages, including of course Italian (cf. also Grandi 2011), the former are more typically found in East and South East Asia.

(50a) *dāo-qiāng* (Mandarin)
sword-spear
'weapons'

(50b) *lanza-espada* (Spanish)
spear+sword
'a spear with a blade, i.e. a spear which is a sword at the same time'

In addition, Wälchli (2005) shows that co-compounds in the world's languages may be classified into different semantic types according to the relationship between the whole and the constituents. Most of the (non-compositional) meanings identified by Wälchli (2005: 138) for co-compounds crosslinguistically are not found in Italian coordinate compounds (which are typically of the "hyponymic" type), like for instance the generalizing meaning (51), the collective meaning (52), or the approximate meaning (53).

(51) Generalizing (= the output universally quantifies over the input)
t'ese-toso (Mordvin)
here-there
'everywhere'

(52) Collective (= the output is a hypernym of the input items)
sĕt-śu (Chuvash)
milk-butter
'dairy products'

(53) Approximate (= the output is an approximation w.r.t. the input)
ob peb (White Hmong)
two three
'some'

However, Masini (2006, 2012) shows that most of these functions are actually found in Italian, but they are conveyed by irreversible binomials (cf. (54a), (55a), (56a)). Most likely the same holds for other SAE languages: see for instance the English examples in (54b), (55b) and (56b).

(54) Generalizing
(54a) *giorno e notte* (Italian)
day and night
'day and night, always'
(54b) *high and low* (meaning 'everywhere') (English)

(55) Collective
(55a) *coltello* *e* *forchetta* (Italian)
knife and fork
'cutlery'
(55b) *bra and panties* (meaning 'lingerie') (English)

(56) Approximate
(56a) *poco* *o* *niente* (Italian)
little or nothing
'very little, almost nothing'
(56b) *two or three* (meaning 'some') (English)

Therefore, despite their structural resemblance, the actual competition between the two coordinating strategies under examination is quite limited, since the two patterns are similar but not equivalent: although they might compete in some specific cases (cf. the semantic similarity between *sordomuto* 'deaf-mute', being the sum of deaf and mute, and *vivo e vegeto* 'alive and well', being the sum of alive and thriving), overall the two patterns are specialized for different functions, compensating, so to speak, for one another.

5 Towards a unified treatment of complex lexical items

In this paper I dealt with complex lexical items in Italian, namely proper compounds and MWEs. Specifically, I focused on so-called phrasal lexemes, which are closer to compounds in distribution and function than other (e.g. sentence-level) MWEs. Whereas compounds mostly feed nouns and adjectives in Italian, phrasal lexemes – beside creating expressions belonging to nouns/adjectives – may also feed other major word classes, most notably verbs and adverbs, thus apparently compensating the limits of compounding in these specific areas. What should be stressed, once again, is that phrasal lexemes are not just the product of diachronic lexicalization: some instances certainly are, but some others are the result of synchronic lexical creation that relies on stored naming patterns (i.e., constructions).

The demarcation between compounds and phrasal lexemes turned out to be a non-trivial issue. I proposed four tentative criteria for the Italian language, i.e. presence/absence of: internal agreement, explicit relational markers, minor lexical categories, bounded elements. However, this set of criteria has no pretense of

crosslinguistic validity: in fact, each language will display a specific set of properties that help distinguishing between these two kinds of constructions (when this is actually possible). Ultimately, these criteria trace back to the traditional distinction between morphology and syntax, which is however not clear-cut within a constructionist view of the grammar.

I also contributed some data and observations on the competition that – quite expectedly – emerges between compounds and phrasal lexemes, given their shared function. I showed that this competition may lead to bidirectional blocking: compounds may block the establishment of a phrasal lexeme in the lexicon, and an established phrasal lexeme may block the creation of a new compound. From the data examined so far, it seems that these two competing patterns tend to differentiate, by specializing for different functions (cf. especially Sections 4.2 and 4.3). This goes into the direction advocated for by Aronoff (2016, to appear) in recent work, where competition leads to either extinction of one of the competitors, or to differentiation in terms of form, meaning or distribution, as a result of a "struggle for existence" between linguistic expressions.

In conclusion, the discussion of demarcation and competition issues carried out in this chapter suggests a view of the mental lexicon where both compounds and phrasal lexemes are stored, on a par with each other: they share the same function and distribution, they may compensate for each other at the most abstract level, and they definitely compete with each other for the expression of lexico-conceptual meanings.

References

Ackerman, Farrell/Webelhuth, Gert (1997): The composition of (dis)continuous predicates: Lexical or syntactic? In: Acta Linguistic Hungarica 44, 3/4. 317–340.

Arcodia, Giorgio Francesco/Grandi, Nicola/Wälchli, Bernhard (2010): Coordination in compounding. In: Scalise, Sergio/Vogel, Irene (eds.). 177–197.

Aronoff, Mark (2016): Competition and the lexicon. In: Elia, Annibale/Iacobini, Claudio/Voghera, Miriam (eds.): Livelli di analisi e fenomeni di interfaccia. Roma: Bulzoni. 39–52.

Aronoff, Mark (to appear): Competitors and alternants in linguistic morphology. In: Rainer, Franz/Gardani, Francesco/Dressler, Wolfgang U./Luschützky, Hans Christian (eds.).

Baldwin, Timothy/Kim, Su Nam (2010): Multiword expressions. In: Indurkhya, Nitin/Damerau, Fred J. (eds.): Handbook of Natural Language Processing. Boca Raton: CRC Press. 267–292.

Benveniste, Émile (1966): Différentes formes de la composition nominale en français. In: Bulletin de la Société de Linguistique de Paris 61, 1. 82–95.

Bernal, Elisenda (2012): Catalan compounds. In: Probus 24, 1. 5–27.

Bisetto, Antonietta (2004): Composizione con elementi italiani. In: Grossmann, Maria/Rainer, Franz (eds.). 33–51.
Bisetto, Antonietta/Scalise, Sergio (1999): Compounding. Morphology and/or syntax? In: Mereu, Lunella (ed.): Boundaries of Morphology and Syntax. Amsterdam/Philadelphia: Benjamins. 31–48.
Booij, Geert (2002a): Constructional idioms, morphology and the Dutch lexicon. In: Journal of Germanic Linguistics 14, 4. 301–329.
Booij, Geert (2002b): Separable complex verbs in Dutch: A case of periphrastic word formation. In: Dehé, Nicole et al. (eds.): Verb-Particle Explorations. Berlin/New York: De Gruyter. 21–42.
Booij, Geert (2009a): Phrasal names: A constructionist analysis. In: Word Structure 2, 2. 219–240.
Booij, Geert (2009b): Lexical integrity as a formal universal: A constructionist view. In: Scalise, Sergio/Magni, Elisabetta/Bisetto, Antonietta (eds.). 83–100.
Booij, Geert (2010): Construction Morphology. Oxford: Oxford University Press.
Butt, Miriam (1995): The structure of complex predicates in Urdu. Stanford: CSLI.
Cacciari, Cristina/Tabossi, Patrizia (eds.) (1993): Idioms: Processing, structure, and interpretation. Hillsdale: Psychology Press.
Cowie, Anthony (ed.) (1998): Phraseology: Theory, analysis, and applications. Oxford: Oxford University Press.
Cruse, Alan (2004): Meaning in language. Oxford: Oxford University Press.
D'Achille, Paolo/Grossmann, Maria (2010): I composti aggettivo + aggettivo in italiano. In: Iliescu, Maria/Siller-Runggaldier, Heidi M./Danler, Paul (eds.): Actes du XXVe Congrès International de Linguistique et de Philologie Romanes (3–8 sept. 2007, Innsbruck), VII. Berlin/New York: De Gruyter. 405–414.
D'Achille, Paolo/Grossmann, Maria (2013): I composti <colorati> in italiano tra passato e presente. In: Casanova Herrero, Emili/Calvo Rigual, Cesáreo (eds.): Actas del XXVI Congreso Internacional de Lingüística i Filología Románicas (Valencia, 6–11 de septiembre 2010). Berlin/New York: De Gruyter. 523–537.
Dehé, Nicole et al. (eds.) (2002): Verb-particle explorations. Berlin: De Gruyter.
Delfitto, Denis/Melloni, Chiara (2009): Compounds don't come easy. In: Lingue e Linguaggio VIII(1). 75–104.
Everaert, Martin et al. (eds.) (1995): Idioms: Structural and psychological perspectives. Hillsdale.
Fillmore, Charles/Kay, Paul/O'Connor, Mary Catherine (1988): Regularity and idiomaticity in grammatical constructions: The case of *let alone*. Language 64, 3. 501–538.
Giegerich, Heinz J. (2005): Associative adjectives in English and the lexicon-syntax interface. In: Journal of Linguistics 41, 3. 571–591.
Giegerich, Heinz J. (2009): The English compound stress myth. In: Word Structure 2, 1. 1–17.
Grandi, Nicola (2011): La coordinazione tra morfologia e sintassi. Tendenze tipologiche ed areali. In: Massariello Merzagora, Giovanna/Dal Maso, Serena (eds.): I luoghi della traduzione/Le interfacce. Roma: Bulzoni. 881–895.
Grossmann, Maria/Rainer, Franz (eds.) (2004): La formazione delle parole in italiano. Tübingen: Niemeyer.
Guevara, Emiliano/Scalise, Sergio (2009): Searching for universals in compounding. In: Scalise, Sergio/Magni, Elisabetta/Bisetto, Antonietta (eds.). 101–128.

Haspelmath, Martin (2011): The indeterminacy of word segmentation and the nature of morphology and syntax. In: Folia Linguistica 45, 1. 31–80.
Hoffmann, Thomas/Trousdale, Graeme (eds.) (2013): The Oxford Handbook of Construction Grammar. Oxford: Oxford Handbooks.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.). 450–467.
Iacobini, Claudio (2015): Particle-verbs in Romance. In: Müller, Peter O. et al. (eds.). 627–659.
Iacobini, Claudio/Masini, Francesca (2007): The emergence of verb-particle constructions in Italian. In: Morphology 16, 2: 155–188.
Jackendoff, Ray (1990): Semantic structures. Cambridge, MA: MIT Press.
Jackendoff, Ray (1995): The boundaries of the lexicon. In: Everaert, Martin et al. (eds.). 133–165.
Jackendoff, Ray (1997): The architecture of the language faculty. Cambridge, MA: MIT Press.
Jackendoff, Ray (2010): Meaning and the lexicon: the Parallel Architecture 1975–2010. Oxford: Oxford University Press.
Jezek, Elisabetta (2004): Types et degrés de verbes supports en italien. In: Linguisticae investigationes 27, 2: 185–201.
Kay, Paul (2013): The limits of (Construction) Grammar. In: Hoffmann, Thomas/Trousdale, Graeme (eds.). 32–48.
Lambrecht, Knud (1984): Formulaicity, frame semantics, and pragmatics in German binomial expressions. In: Language 60, 4. 753–796.
Malkiel, Yakov (1959): Studies in irreversible binomials. In: Lingua 8. 113–160.
Masini, Francesca (2005): Multi-word expressions between syntax and the lexicon: the case of Italian verb-particle constructions. In: SKY Journal of Linguistics 18. 145–173.
Masini, Francesca (2006): Binomial constructions: Inheritance, specification and subregularities. In: Lingue e Linguaggio 5, 2. 207–232.
Masini, Francesca (2009): Phrasal lexemes, compounds and phrases: A constructionist perspective. In: Word Structure 2, 2. 254–271.
Masini, Francesca (2012): Parole sintagmatiche in italiano. Roma: Caissa Italia.
Masini, Francesca (2016): Morphological words and multiword expressions: Competition or cooperation? Paper given at the 17th International Morphology Meeting (IMM17), Vienna, 18–21 February 2016.
Masini, Francesca (to appear): Competition between morphological words and multiword expressions. In: Rainer, Franz/Gardani, Francesco/Dressler, Wolfgang U./Luschützky, Hans Christian (eds.).
Masini, Francesca/Benigni, Valentina (2012): Phrasal lexemes and shortening strategies in Russian: The case for constructions. In: Morphology 22, 3. 417–451.
Masini, Francesca/Scalise, Sergio (2012): Italian compounds. In: Probus 24, 1. 61–91.
Masini, Francesca/Thornton, Anna M. (2008): Italian VeV lexical constructions. In: On-line Proceedings of the 6th Mediterranean Morphology Meeting (MMM6). University of Patras. 146–186.
Moon, Rosamund (1998): Fixed expressions and idioms in English: A corpus-based approach. New York: Clarendon Press.
Müller, Peter O. et al. (eds.) (2015): Word-formation. An international handbook of the languages of Europe. Vol. 1. (= Handbooks of Linguistics and Communication Science (HSK) 40.1). Berlin/Boston: De Gruyter.
Radimský, Jan (2015): Noun+Noun compounds in Italian: A corpus-based study. České Budějovice: University of South Bohemia in České Budějovice.

Rainer, Franz (1988): Towards a theory of blocking: The case of Italian and German quality nouns. In: Booij, Geert/van Marle, Jaap (eds.): Yearbook of Morphology 1988. Dordrecht: Springer. 155–185.
Rainer, Franz (2016): Blocking. In: Aronoff, Mark (ed.): The Oxford Research Encyclopedia of Linguistics. Internet: DOI: 10.1093/acrefore/9780199384655.013.33.
Rainer, Franz/Varela, Soledad (1992): Compounding in Spanish. In: Rivista di Linguistica 4, 1. 117–142.
Rainer, Franz/Gardani, Francesco/Dressler, Wolfgang U./Luschützky, Hans Christian (eds.) (to appear): Competition in inflection and word formation. Cham: Springer.
Ricca, Davide (2010): Corpus data and theoretical implications with special reference to Italian V-N compounds. In: Scalise, Sergio/Vogel, Irene (eds.). 237–254.
Rio-Torto, Graça/Ribeiro, Sílvia (2009): Compounds in Portuguese. In: Lingue e Linguaggio VIII(2). 271–291.
Rio-Torto, Graça/Ribeiro, Sílvia (2012): Portuguese compounds. In: Probus 24, 1. 119–145.
Scalise, Sergio (1992): Compounding in Italian. In: Italian Journal of Linguistics/Rivista di Linguistica 4, 1. 175–199
Scalise, Sergio/Bisetto, Antonietta (2009): The classification of compounds. In: Lieber, Rochelle/Štekauer, Pavol (eds.): The Oxford handbook of compounding. Oxford: Oxford University Press. 34–53.
Scalise, Sergio/Magni, Elisabetta/Bisetto, Antonietta (eds.) (2009): Universals of language today. Berlin: Springer.
Scalise, Sergio/Vogel, Irene (eds.) (2010): Cross-disciplinary issues in compounding. Amsterdam: Benjamins.
Schlücker, Barbara/Hüning, Matthias (2009): Compounds and phrases. A functional comparison between German A+N compounds and corresponding phrases. In: Italian Journal of Linguistics 21, 1. 209–234.
Voghera, Miriam (2004): Polirematiche. In: Grossmann, Maria/Rainer, Franz (eds.). 56–69.
Wälchli, Bernhard (2005): Co-compounds and natural coordination. Oxford: Oxford University Press.

Jesús Fernández-Domínguez

Compounds and multi-word expressions in Spanish

1 Introduction

Compounds have been customarily defined as lexical units that consist of two lexemes. They are morphological entities.[1] Phrases, for their part, may be made up by one unit, but often comprise two elements when they carry internal modification. They are syntactic entities. Provisionally satisfactory though these descriptions are, linguists are often in trouble when having to decide on which basis a two-word structure is morphological or syntactic, as a formation may be argued to be a compound on some grounds but at the same time display phrasal features. Certainly, in any category, it is quite common for some members not to meet all the prototypical features of the group; this is in fact statistically likely. Compounding is no exception as can be seen regarding the many exceptions to initial stress placement (e.g. '*snowball* vs. *rubber* '*ball*), or to spelling (*market place* vs. *market-place* vs. *marketplace*). A dilemma arises, however, if peripheral membership is not an exception but the norm (Bauer 1998: 65).

The Spanish word-formation system offers an array of means for the creation of neologisms, typically classified within the categories of derivation, compounding and minor processes. Derivation is by far the most fruitful resource and includes prefixation (*pintar* lit. paint 'to paint' > *repintar* lit. re.paint 'to repaint'), suffixation (*admirar* lit. admire 'to admire' > *admirable* lit. admir.able 'admirable') and infixation (*cantar* lit. sing 'to sing' > *cant.urre.ar* lit. sing.SUF 'to hum'). A number of other processes may be distinguished, for example, parasynthesis (*largo* lit. long 'long' > *alargar* lit. a.long.ar 'to lengthen'), back-formation (*comprar* lit. purchase 'to purchase' > *compra* lit. purchase 'purchase'), blending (*documental* 'documentary' + *drama* 'drama' = *docudrama* 'docudrama'), acronymy (**P**equeña **Y M**ediana **E**mpresa 'small and medium-sized business' > *PYME*), and clipping (*colegio* 'school' > *cole*). The affinities between compounds and phrases have also been noted for Spanish, with the fundamental difference that compounds have a morphological origin, serve a naming function and are often spe-

1 I would like to thank Francesca Masini, Barry Pennock, Vincent Renner, Barbara Schlücker and an anonymous reviewer for sensible advice on the form and content of this article.

https://doi.org/10.1515/9783110632446-007

cialized in meaning, while phrases are syntactic, tend to be semantically compositional and have a more descriptive nature. This view, heavily influenced by the Lexicalist Hypothesis, oversimplifies the picture and is not completely faithful to reality (Bustos Gisbert 1986: 69–72; Booij 2009: 220; Pafel 2017). The fact is that Spanish compounding constitutes a rather fuzzy category comprising a range of different formations, from genuine compounds to phrasal units. This categorial heterogeneity is probably caused by the rigid rules of Spanish compounding and by the fact that both compounds and *multi-word expressions* (MWEs) may qualify as lexical units, a fact which has eventually blurred the limits of both types. MWEs are constructions which comprise various constituents but nevertheless display meaning non-compositionality and referential stability. Because of their multifaceted nature, MWEs have attracted the attention of different fields of linguistics, although they are most frequently dealt with by phraseology, a discipline whose limits with other language areas have not been established hitherto (Gries 2008; Colson 2016). To date, phraseology has been more oriented towards practical (often corpus-based) applications than towards a theoretical delimitation of its boundaries. The label MWE, borrowed from computational linguistics, is an alternative to traditional terms like *idiom* or *phraseological unit* (cf. Hüning/Schlücker 2015: 450).

This article describes Spanish compounds and MWEs and aims to provide an up-to-date view of their nature and limits. It is arranged as follows: after this introduction, Section 2 offers a theoretical overview of compounding and neighbouring formations. Section 3 examines the demarcation between compounds and MWEs for Spanish nouns, verbs and adjectives, and the consequences of this relationship for the language system are discussed in Section 4. Section 5 contains the conclusions of the study.

2 The characterization of compounds and MWEs

One definition states that a compound is a complex lexeme made up by two or more lexemes in a relationship of dependency (subordinate compounds) or non-dependency (coordinate compounds). In subordinate compounds there is a relationship of dependency where the non-head modifies the head, e.g. *sun* modifying *light* in (1a) 'light that is produced by the sun', while in coordinate compounds both constituents stand at the same level, e.g. *girl* and *friend* in (1b) 'a girl that is also a friend' (Rainer/Varela 1992: 125–130; Bisetto/Scalise 2005: 326f.).

(1a) *sunlight*
(1b) *girlfriend*

Two broad comparable constructions have been distinguished in Spanish. The first is *lexical compounds*, where the relationship between constituents has special phonological, combinatorial and semantic properties, as in *coliflor* lit. cabbage.i.flower 'cauliflower', where *-i-* is a linking vowel. Compounds like *coliflor* are the product of native (now unproductive) word-formation rules and are generally scarce in Romance languages. If, however, the constituents are more loosely conjoined, we are faced with a second type of units whose internal structure is identical to that of syntactic objects, as in *fin de semana* lit. end of week 'weekend'. Because these formations emerge from syntax, their morpho-phonological integrity is looser than that of lexical compounds, and this has resulted in a variety of often unclear labels, e.g. *syntagmatic*, *phrasal* or *phraseological compound*. Some authors have plainly rejected an analysis of multi-word constructions as compounds due precisely to their phrasal nature (Rainer/Varela 1992). We henceforth employ the term MWE for such formations due to its "pre-theoretical" (Masini 2005: 145) nature.

The most widely debated argument is the exact position of MWEs in the language system, as compounds are morphological in nature, while idioms, collocations or proverbs in principle belong to phraseology. Admittedly, Spanish grammars have in general paid little attention to phraseological units, although attempts have been made at systematizing their description (Montoro del Arco 2008). In the case of nominal MWEs, one problem is that they are often examined together with syntactic objects, like verbal expressions or idioms. This makes it difficult to isolate and depict nominal MWEs because the levels of morphology, syntax and phraseology are mixed up. The limits between Spanish noun phrases and noun compounds lie close due to a number of coincidental aspects:

a) Both types resort to previously existing elements for their creation.
b) Both types are in general semantically left-headed.
c) Both types can perform a naming function.
d) Some compounds, and many MWEs, display idiomaticity.
e) Some compounds, due to the previous fact, may display varying levels of semantic and linear fixity.

In the face of these similarities, one unavoidable step when studying compounds and MWEs is the description of the morphology-syntax interface (Gaeta/Ricca 2009; Masini 2009; Buenafuentes 2010, 2014; Pafel 2017). The most frequently adduced factors concern several language areas:

(i) Referential uniqueness. While the conceptual unity of compounds is generally agreed upon, MWEs may also represent a semantic entity coherent

with extralinguistic reality (Gaeta/Ricca 2009: 36f.). The constituents of phrases like those in (2) have retained their basic semantics but at the same time their referent is a particular reality that differs from that of the lexical base, here *huelga* 'strike' and *vale* 'coupon'.

(2a)	*huelga patronal*	lit. strike employer	'lockout'
(2b)	*vale descuento*	lit. coupon discount	'discount coupon'

(ii) Idiomaticity. Often, the combination of lexemes adds a semantic dimension to the new construction which may not be deducible from its constituents. This semantic alteration is acute in exocentric compounds, i.e. those whose semantic head is not contained in one of the compound constituents. Thus, a unit like (3a) refers to an agent, but nothing in its structure prevents a possible instrumental meaning (e.g. a machine which picks up spare balls). Renner (2006: 23) speaks of *retrospective transparency*: the global sense of a compound X.Y stems from X and Y but it is not entirely predictable from their co-occurrence. Idiomaticity is especially relevant in metonymic/metaphorical extensions of meaning, both with human (3b) and non-human (3c) referents.

(3a)	*recoge.pelotas*	lit. pick up.balls	'ball boy'
(3b)	*agua.fiestas*	lit. spoil.parties	'spoilsport'
(3c)	*ojo de buey*	lit. eye of ox	'porthole'

(iii) Atomicity. Once created, compound constituents show an invariable arrangement that renders them unreadable by syntax. Atomicity is an indication of the lexical status of a construction because no element can be inserted between the compound constituents (4). For the same reason, the non-head cannot be anaphorically designated, the case of *esmalte* 'polish' and *quitaesmalte* 'polish remover' in (5):

(4a)	*hora punta*	lit. hour peak	'peak hour'
(4b)	**hora muy punta*	lit. hour very peak	

(5) **Usó el quita*$_1$*esmalte*$_2$*, pero no lo*$_2$ *pudo borrar*
Use-3SG-PAST the remove$_1$polish$_2$, but NEG D.OBJ-M-SG$_2$ could erase
'She used the polish remover, but couldn't erase the polish'

(iv) Semantic fixity. The head of a compound cannot be replaced by a semantically close lexeme. This has been described as an indispensable feature of MWEs, but it occurs in compounds too. The non-existence of (6b) reveals the lexical and semantic cohesion of *guerra fría*. This is no impediment, as pointed out by one of the reviewers, for the existence of (6c) as proof of occasional rule-breaking lexical creativity.

(6a)	*guerra fría*	lit. war cold	'cold war'
(6b)	**pelea fría*	lit. fight cold	
(6c)	*contienda fría islámica*	lit. dispute cold Islamic	'cold Islamic dispute'

(v) Linear fixity. Inflection is a common test for compoundhood because it is assumed that it should not occur within a lexeme, be it a derivative or a compound. This is the case with orthographic compounds, where plurality is applied peripherally (7), but it is different with MWEs, which typically display internal inflection (8). In such cases we find no external inflectional mark, except for the right-hand member in formations like *cajas de ahorros*, where *ahorros* 'savings' is pluralized also when the MWE is singular. This implies that inflection is only valid as a criterion for orthographic compounds (which are nevertheless unproblematic because they are clearly morphological in nature). In my view, inflection is precisely the differentiating factor between compounds and MWE, since an inflected first constituent is sufficient proof of an element's syntactic character (compounds forbid internal inflection; cf., however, Bauer 2017: 19ff.).

(7a)	*punta.pié*	lit. tiptoe.foot	'kick'
(7b)	*punta.piés*	lit. tiptoe.feet	'kicks'

(8a)	*caja de ahorros*	lit. bank of savings	'savings bank'
(8b)	*cajas de ahorros*	lit. banks of savings	'savings banks'

(vi) Frequency. The constituents of a multi-word formation should regularly co-occur to acquire lexical status. The syntax-derived example in (9a), because of its semantic coherence and repeated usage, is a lexical unit, while (9b) is not because, despite its semantic coherence, its constituents do not habitually co-occur.

(9a)	*libro de cocina*	lit. book of cuisine	'cookbook'
(9b)	*gorra de metal*	lit. cap of metal	'metal cap'

Most of the above criteria are either syntactic (atomicity, fixity, locus of inflection) or semantic (naming unity, idiomaticity), although productivity is a characteristic of morphology. Besides these, stress has proved crucial for the differentiation between phrases and compounds in other languages (e.g. German and Dutch), but it is not decisive in Spanish, as compounds may display single but also double stress (Rao 2015: 90f.). Bustos Gisbert (1986) reviews the phonetics of Spanish compounds and concludes that stress assignment is caused by the interaction of factors like the number of syllables, the semantic relationship between the constituents or the compound's headedness. Rao (2015) provides interesting experimental findings on the influence of orthography over prosodic interpretation, or the apparently minimal effect of the semantic relation between constituents on stress assignment.

The above generalizations represent general tendencies regarding prototypical compounds and prototypical syntactic entities but, crucially, most of these features may be displayed by both compounds and phrases and cannot individually provide conclusive evidence with respect to the compound-phrase divide.

3 Spanish compounds and MWEs: between morphology, syntax and phraseology

Formations of a nominal, verbal and adjectival type are taken into consideration in this section, particularly those made up of nouns, adjectives and verbs. There is a consensus in the literature that Spanish compounding is largely endocentric and that, while adjective and verb compounds are right-headed, noun compounds are left-headed, with the exception of specific right-headed types. The following subsections explore, in turn, nominal (Section 3.1), adjectival (Section 3.2), and verbal (Section 3.3) compounds and MWEs.

3.1 Nouns

Spanish noun compounds most often consist of two members whose grammatical categories may be the same, e.g. noun+noun (N+N), or different, e.g. noun+adjective (N+A) or verb+noun (V+N). A preposition may also be involved as a link between the two main constituents in the productive MWE type noun+preposition+noun (N+p+N), cf. (9a). A three-lexeme structure is less common but possible, as in *limpiaparabrisas* or *portacuentakilómetros*, which does not alter the

binary structure of the compound: *limpia* ('clean') + *parabrisas* ('windshield'), *porta* ('carry') + *cuentakilómetros* ('odometer').

(10a) *limpia.parabrisas* 'windshieldwiper'
lit. clean.windshield
(10b) *porta.cuentakilómetros* 'odometer-holder'
lit. carry.odometer

Spanish nominal compounding is characterized by left-headedness (11), although right-headed constructions are possible as well, cf. (12). In both cases the head transfers its syntactic and semantic features to the compound.

(11a) *hoja.lata* 'tinplate'
lit. blade.tin
(11b) *pez espada* 'swordfish'
lit. fish sword

(12a) *tele.novela* 'soap opera'
lit. TV.novel
(12b) *zarza.mora* 'blackberry'
lit. bramble.berry

Given the apparent detachment between morphology and phraseology, it comes as no surprise that the above morphological categories may overlap to some extent with those proposed by phraseologists for structurally parallel constructions. One of these is *collocations*, i.e. word combinations whose members display a high co-occurrence rate and are semi-idiomatic, but where the rules of phrase grammar are normally observed (Ruiz Gurillo 2002). The differences between compounds and collocations are gradual and not unambiguous, since both types display shared features (e.g. frequency of co-occurrence, lack of stress unity, being formed by several words) but also points of divergence, for instance having a naming function and being paradigmatically related to other units, which are typical of compounds and impossible in collocations. The following formations have been regarded as compounds in some views and as collocations in other, with the form N+A (13a), N+N (13b) or N+p+N (13c) (see Table 1 in Section 4):

(13a) *león marino* 'sea lion'
lit. lion marine
(13b) *paquete bomba* 'mail bomb'
lit. parcel bomb

(13c) *ciclo de conferencias* 'conference series'
lit. cicle of conferences

Within the range of existing nominal structures, several types stand out in Spanish: orthographic constructions (Section 3.1.1), where several different word-classes are found as input, together with the syntagmatic types N+N (Section 3.1.2), N+p+N (Section 3.1.3) and N+A (Section 3.1.4).

3.1.1 Orthographic nominal constructions

Spelling may be indicative of a unit's lexical status. That is the case of constructions which unequivocally qualify as compounds and as such are spelt as one word. The following units, for example, are made up of a preposition and a noun:

(14a) *sin.vergüenza* 'scoundrel'
lit. without.shame
(14b) *sobre.peso* 'overweight'
lit. over.weight

These compounds are characterized by morphological indivisibility and single stress, and many of them are highly lexicalized. The most productive type of Spanish orthographic compounding is the pan-Romance V+N (Kornfeld 2009: 438f.; Moyna 2011), in which the verb is the predicate and the noun its direct object, and where the resulting unit may be an agent (15a), an instrument (15b), or more marginally an activity (15c). V+N nouns are exocentric but fully transparent (even if not always predictable; see (3) and related discussion) and are always inseparable. The invariable plural form of the second constituent is caused by its semantic notion of habitual/repeated activity (15), although it is kept in singular in uncountable nouns (16). Therefore, the following units can be unambiguously declared genuine compounds; further proof of their morphological nature is provided by their solid spelling.

(15a) *aparca.coches* 'valet'
lit. park.cars
(15b) *para.rrayos* 'lightning conductor'
lit. stop.lightning.PL
(15c) *cumple.años* 'birthday'
lit. reach.years

(16) *guarda.rropa* 'cloakroom'
lit. keep.clothing

In contrast to such units, there are compounds whose form fluctuates between a single-word and a two-word spelling. These are phrasal structures with an increasingly tighter compound status, which forms a continuum from phrases to completely settled compounds and intermediate hybrid formations. It is therefore possible to come across *guardia civil* and *guardiacivil*, or *retrato-robot* and *retrato robot* (cf. Van Goethem 2009 for a scalar proposal on French A+N units):

(17a) *guardia civil* 'Civil Guard'
lit. guard civil
(17b) *retrato-robot* 'photofit portrait'
lit. portrait-robot

A more extreme although less frequent situation is that in which both compound constituents are pluralized, be the compound as a whole in plural or not. Constructions like (18) have been analyzed as exocentric compounds with morphological mismatches in their number and gender (RAE 2010: 193, 199; Scalise/Fábregas 2010: 122; Buenafuentes 2014: 4). These formations are dealt with in detail in the following sections.

(18) *relaciones públicas* 'public relations'
lit. relations public.PL

(19) *María es la relaciones públicas de la empresa*
'María is the public relations of the company'

3.1.2 Noun + Noun

N+N formations are one of the most frequently studied phenomena within the morphology-syntax divide, with a variety of labels revealing their ambiguous condition (*binominals*, *coordinate compounds*, *dvandvas*) in a good number of languages (Bauer 1998; Bisetto/Scalise 2005; Booij 2009). The constituents of N+N constructions concatenate with no tying formal mark, although a hyphen occasionally signals lexical status.[2] Because Spanish is, with the exception of the

2 A subtype of binominal compounds incorporates a linking vowel, as in *aj.i.aceite* lit. garlic.i.oil 'sauce made of garlic and olive oil'. This type is not further discussed due to its unproduc-

formations in Section 3.1.1, reluctant to accept novel nominal compounds, N+N units are significant from a lexical perspective. Synchronically, this is a productive type, and one for which constraints are not easily found. This alleged fertility leads to a wide range of possible syntactic and semantic options, which are classified as appositional (20a), specifying (20b) and classifying (20c) in the phraseological literature (Ruiz Gurillo 2002). Under this view, such formations are collocations whose first constituent is the base (*merienda*, *efecto*) and the second is the collocate (*cena*, *invernadero*):

(20a)	*merienda cena* lit. afternoon-snack supper	'late afternoon-snack / early supper'
(20b)	*efecto invernadero* lit. effect greenhouse	'greenhouse effect'
(20c)	*perro policía* lit. dog police	'police dog'

For Booij (2009: 223), the naming function shared by compounds and phrases complicates their demarcation especially in languages with left-headed compounding, since certain formations can be seen as compounds but also as phrases followed by an apposition. The fact that plural inflection tends to appear on the first constituent only (*merienda**s** cena*, *efecto**s** invernadero*) substantiates access from syntax to these formations, and Booij's (2009) interpretation is hence a phrasal one. This argument is not refuted by the use of these units in more colloquial registers, where inflection is attested for both members (*perro**s** policía**s***, *muble**s**-bar**es***). The degree of fixity is also significant here. Val Álvaro (1999: 4782) puts forward that an inflexible layout is typical of coordinate compounds because, given the paratactic relationship of their constituents, it is an optimal means to make the first member more salient. This happens, for example, when the first member precedes the second one chronologically (*merienda cena*) or when it is cognitively more relevant (*perro policía*). Similarly, sets of N+N formations may display a shared second (21) or first constituent (22):[3]

tive status. The same applies to minor types like V+V reduplicatives (*pilla.pilla* lit. catch.catch 'tag', a playground game) or V+V formations (*duerme.vela* lit. sleep.stay up 'slumber'), which are not illustrative of current trends (Val Álvaro 1999: 4804–4807).

3 An analogous series is comprised by *visita relámpago* lit. visit lightning 'lightning visit', *guerra relámpago* lit. war lightning 'blitzkrieg', *viaje relámpago* lit. trip lightning 'lightning trip', etc. These have been rejected as nominal MWEs because appositive nouns like *clave* or *relámpago* are not restricted in their co-occurrence, and they can be accompanied by almost any noun. That

(21a) *cuestión clave* 'key matter'
lit. matter key
(21b) *decisión clave* 'key decision'
lit. decision key
(21c) *hombre clave* 'key man'
lit. man key

(22a) *hombre anuncio* 'sandwich-board man'
lit. man ad
(22b) *hombre rana* 'frogman'
lit. man frog
(22c) *hombre araña* 'spiderman'
lit. man spider

One variant is represented by the examples in (23), which have been analyzed either as morphological or as syntactic structures. These constructions have also been considered collocations on the basis that nouns concatenate with no preposition whatsoever (Ruiz Gurillo 2002), but their referential uniqueness makes such reading unadvisable. Then again, an interpretation in terms of compounding is hindered essentially by plural inflection, which materializes internally (*foto**s** tamaño carnet*, *crema**s** tipo pomada*). This, together with the possibility of recovering an elided preposition *de* 'of' after the left-most member (24), seems sufficient evidence for a syntactic nature, in line with the units in (20).

(23a) *foto tamaño carnet* 'ID size photo'
lit. photo size ID-card
(23b) *crema tipo pomada* 'ointment-like cream'
lit. cream type ointment

(24a) *foto **de** tamaño carnet* 'ID size photo'
lit. photo of size ID-card
(24b) *crema **de** tipo pomada* 'ointment-like cream'
lit. cream of type ointment

such constructions can be inflected for number in standard registers (*viajes relámpagos*, *guerras relámpagos*) points to semantic specialization and suggests that they are more akin to standard modifying phrases (cf. Val Álvaro 1999: 4785; Montoro del Arco 2008: 133f.).

3.1.3 Noun + preposition + Noun

N+p+Ns are a fertile kind of construction that links a noun to a simple (25a) or deverbal noun (25b) by way of a preposition. N+p+Ns are head-initial formations whose right-hand constituent is subordinated and displays adjective-like behavior:[4]

(25a) *banco de datos* 'databank'
lit. bank of data
(25b) *máquina de escribir* 'typewriter'
lit. machine of writing

The first hurdle in the description of N+p+N units is that they are derived from a syntactic pattern whereby a nominal head is postmodifed by a prepositional phrase, but at the same time they perform a naming function that is typical of compounding. Several criteria have been put forward to test the compoundhood of N+p+N constructions. One is whether an equivalent lexeme exists in a different language (26), or whether a synonymous structure has been attested in Spanish through a different word-formation process, as in (27), although these do not seem entirely reliable criteria. Both features hint at the lexical status of N+p+Ns but do not evidence a morphological origin which, together with the syntactic provenance of these formations, has led to their rejection as compounds (cf. Rainer/Varela 1992). *Telaraña*, for example, developed out of lexicalization from *tela de araña*, a process that has nothing to do with morphology and can be more accurately described as *univerbation* than as compounding (Gaeta/Ricca 2009: 44f.).

(26a) *dolor de cabeza* 'headache'
lit. ache of head
(26b) *máquina de afeitar* 'shaver'
lit. machine of shaving

(27a)	*abridor de latas*	*abrelatas*	'can opener'
	lit. opener of cans	lit. to open.cans	
(27b)	*tela de araña*	*telaraña*	'spider web'
	lit. fabric of spider	lit. fabric.spider	

4 The label *syntagmatic compound* has been widely employed but it is also regarded as inaccurate on the grounds that the phraseologization of these constructions converts them into lexical units, not compounds, as they do not originate in morphology.

It was discussed above that whether or not constituents can be modified is a good indication of the morphological status of a construction. The following examples show how, for two N+p+N formations (cf. (28a) and (29a)), postmodification is permitted (cf. (28b) and (29b)), while internal separability is ungrammatical (cf. (28c) and (29c)):

(28a) *toque de queda* 'curfew'
lit. call of remain
(28b) *toque de queda* ***reglamentario*** 'obligatory curfew'
lit. call of remain obligatory
(28c) **toque* ***reglamentario*** *de queda*
lit. call obligatory of remain

(29a) *botas de montar* 'riding boots'
lit. boots of riding
(29b) *botas de montar* ***hechas a mano*** 'handmade riding boots'
lit. boots of riding handmade
(29c) **botas* ***hechas a mano*** *de montar*
lit. boots handmade of riding

Inflection in N+p+N units is customarily placed on the head, although the right-hand member may display permanent plural if it refers to a plural notion. Even in the latter case, the plural marker of the whole compound appears on the head (*agencia**s** de viajes*, *tren**es** de mercancías*, *cuento**s** de hadas*):

(30a) *agencia de viajes* 'travel agency'
lit. agency of travels
(30b) *tren de mercancías* 'freight train'
lit. train of merchandise.PL
(30c) *cuento de hadas* 'fairy tale'
lit. tale of fairies

Semantically speaking, N+p+N constructions are versatile, and several semantic relations may be found between N1 and N2: ORIGIN (31a), CONTENT (31b), MANNER (31c), MATERIAL (31d), LOCATION (31e) or PURPOSE (31f) (Lang 1992: 122). As it often happens in descriptions of compound semantics, categories show fuzziness and areas of overlap are evident, e.g. between ORIGIN and LOCATION, or between CONTENT and MATERIAL. This potentiality of meanings touches upon the indeterminacy of the preposition *de* 'of', by far the most frequent one but the least explicit semantically, which has favored the use of other prepositions as a dis-

ambiguation strategy (32). The existence of these constructions, however, does not prevent the coinage of equivalent ones with *de*, and hence the existence of doublets (*camisa de cuadros*, *esmalte de uñas*, etc.; cf. Piunno 2016: 16–19).

(31a) *almeja de río* 'marsh clam'
lit. clam of river
(31b) *gota de rocío* 'dew drop'
lit. drop of dew
(31c) *sierra de mano* 'hand saw'
lit. saw of hand
(31d) *diente de oro* 'gold tooth'
lit. tooth of gold
(31e) *cielo de la boca* 'roof of the mouth'
lit. sky of the mouth
(31f) *bestia de carga* 'beast of burden'
lit. beast of burden

(32a) *camisa a cuadros* 'checked shirt'
lit. shirt with squares
(32b) *televisión por satélite* 'satellite TV'
lit. TV through satellite
(32c) *café con leche* 'white coffee'
lit. coffee with milk
(32d) *fabricación en serie* 'mass production'
lit. production in series
(32e) *hockey sobre patines* 'roller hockey'
lit. hockey on skates
(32f) *esmalte para uñas* 'nail polish'
lit. polish for nails

In phraseological studies, the most frequently studied N+p+N constructions are those where the left-hand noun refers to a set or portion of what is designated by the right-hand noun, that is, partitive formations. The first noun is often semantically selected by the second (33), although this is not a requirement (34). A certain degree of variability is possible, as in (34) (however, *rebanada de pan* 'slice of bread' vs. **rebanada de chocolate* 'slice of chocolate'), but idiomaticity is non-existent, which reveals the regular semantic contribution of the constituents. Such N+p+N units must consequently be analyzed as collocations.

(33a) *banco de peces* 'shoal'
lit. shoal of fish
(33b) *ramo de flores* 'bouquet'
lit. bouquet of flowers

(34a) *pizca de sal* 'pinch of salt'
lit. pinch of salt
(34b) *pizca de pan* 'piece of bread'
lit. pinch of bread
(34c) *pizca de tabaco* 'pinch of tobacco'
lit. pinch of tobacco

Nominal phrases constitute a different subtype, with full fixity and idiomaticity. These are infrequent constructions with a significant degree of lexicalization and metaphorical meanings, which bears witness to their phraseological status (exocentricity is impossible in syntactic formations). Some of such metaphorical units, e.g. (35b), may perform a limited range of syntactic roles at the clause level, usually direct object or subject complement, and never subject. This goes against an analysis of these formations as compounds because their use seems to be limited to comparative constructions, as in (36):

(35a) *caballo de batalla* 'important issue'
lit. horse of battle
(35b) *la carabina de Ambrosio* 'useless object, person or situation'
lit. the carbine of Ambrosio

(36) *Lo que usted propone es la carabina de Ambrosio*
'What you are suggesting is completely useless'
(Davies 2002–)

In the vast majority of such constructions no article is found between the preposition and the right-hand member, although there are exceptions, for example when an article denotes a well-known entity:

(37a) *abogado de**l** diablo* 'devil's advocate'
lit. advocate of **the** devil
(37b) *pipa de **la** paz* 'peace pipe'
lit. pipe of **the** peace

3.1.4 Noun + Adjective / Adjective + Noun

Spanish features abundant A+N and N+A nouns. Because of native left-headedness, the former are less numerous even if, due to their spelling, they stand out more clearly within the field of compounding than the latter. Moyna attributes a syntactic origin to these formations, which explains why "[...] they are the hardest to distinguish from non-compounded phrases" (2011: 181; cf. Gaeta/Ricca 2009: 51ff. for Italian). Most A+N units are endocentric and display a relationship of modification between their constituents (38a), although heads can become opaque over time and acquire metaphorical readings (38b). It is possible to find exocentric formations too, as in example (38c), which is not 'a kind of table' but 'a kind of meeting'.

(38a) *media.noche* 'midnight'
lit. half.night
(38b) *alta.voz* 'loudspeaker'
lit. high.voice
(38c) *mesa redonda* 'round table'
lit. round table

Spanish N+A compounds (39) are difficult to distinguish from N+A phrases (40) if one only looks at their meaning, since in both kinds the head can be a concrete noun (39a), a noun denoting physical state (39b), or an abstract noun (39c). As in other Romance languages, orthography is not reliable by itself, especially in formations with a separate spelling, since it does not necessarily reflect stress assignment (cf. Van Goethem 2009).

(39a) *agua bendita* 'holy water'
lit. water holy
(39b) *dolor crónico* 'chronic pain'
lit. pain chronic
(39c) *poder adquisitivo* 'purchasing power'
lit. power purchasing

(40a) *agua limpia* 'clean water'
lit. water clean
(40b) *dolor ficticio* 'imaginary pain'
lit. pain fictitious
(40c) *poder efímero* 'ephemeral power'
lit. ephemeral power

The adjectives in N+A constructions can be described as mainly relational (*budista* 'Buddhist', *carnívoro* 'carnivorous', *sindical* 'unionist') and qualitative (*grande* 'big', *ancho* 'wide', *rojo* 'red') (Koike 2001: 119f.). These adjectives tend to be polysemous and highly frequent (the former perhaps as a consequence of the latter), while the noun is semantically autonomous and determines the meaning of the adjective. N+A formations exhibit the semantic coherence that is characteristic of lexical units and, despite not being orthographically a single word, equivalents in other languages exist too (41). Combinations parallel to N+A constructions can be found also in N+p+N units, where a prepositional phrase replaces the adjective (42):

(41a) *escalera mecánica*
lit. staircase mechanical
huelga patronal
lit. strike employer

(41b) *escalator*

lockout

(42a) *cita médica*
lit. appointment medical
crisis petrolera
lit. crisis oil

(42b) *cita del médico*
lit. appointment of doctor
crisis del petróleo
lit. crisis of the oil

The customary syntactic tests of compoundhood may be applied in the distinction of N+A compounds and phrases: attributive use of the adjective (43), premodification of the adjective (44), swapping positions between adjective and noun (45), internal interruptibility (46), and replacement of the modifier by a synonym (47). These features reveal whether a given formation is more similar to a phrase, as in the examples labelled (a) below, or to a compound, as in those labelled (b):

(43a) *mesa espaciosa* — *la mesa es espaciosa*
'spacious table' — lit. the table is spacious
(43b) *ingeniero electrónico* — *?el ingeniero es electrónico*
'electronic engineer' — lit. the engineer is electronic

(44a) *charla animada* — *charla muy animada*
'lively chat' — lit. the chat is lively
(44b) *registro civil* — *?registro muy civil*
'civil registry' — lit. very civil registry

(45a) *objeción principal* — *principal objection*
'main objection' — lit. objection main

(45b)	*poder especial*	**especial poder*
	'special power'	lit. power special
(46a)	*público joven*	*público masculino joven*
	lit. audience young	lit. young male audience
(46b)	*oso hormiguero*	**oso grande hormiguero*
	lit. bear ant	lit. bear big ant 'anteater'
(47a)	*amor eterno*	*amor imperecedero*
	lit. love eternal	lit. love everlasting
(47b)	*caja fuerte*	**caja robusta*
	lit. box strong	lit. box robust

Interestingly, the adjectives in the above collocations, (43a)–(47a), have an intensifying role, while those in compounds, (43b)–(47b), share a classifying or determinative function. The label *lexical collocation* has been employed for cases where the semantic contribution of the adjective depends to a great extent on that of the noun, such as *fiesta nacional* 'national holiday' or *campaña electoral* 'election campaign', which would otherwise be categorized as compounds. Regardless of the term, this partly explains why compounds are less flexible than collocations in their structure, which in turn causes a wider variability in collocations and is a good argument for the listing of compounds in the lexicon. At the semantic level, N+A collocations are compositional but compounds show some degree of non-compositionality. Ruiz Gurillo (2002: 334) discusses *agua bendita* 'holy water', whose meaning is achieved by summing up the semantics of the two nouns plus additional features from the mental lexicon. In principle, the higher the degrees of compositionality and motivation, the closer a unit stands to compounding; the less isomorphic and motivated it is, the closer it stands to collocations. Schlücker/Hüning (2009; also Bauer 2017: 12f.), in contrast, point out that semantic specialization or compositionality are not definitive criteria.

One problem for the compoundhood of N+A formations is that, notwithstanding sporadic hesitation, plural inflection is normally applied to both constituents, and this is characteristic of phrases. As with juxtaposed nouns (Section 3.1.2), this violates the Lexical Integrity Principle, a behavior expected from the noun-adjective relationship in Spanish. This is compelling proof of the syntactic origin of these units:

(48a)	*bomba**s** lacrimógena**s***	'tear gas canisters'
	lit. bombs tear-producing.PL	

(48b) *llave.**s** inglesa.**s*** 'monkey wrenches'
lit. keys English.PL

In contrast to N+p+Ns, N+A units may undergo derivation, in which case the whole construction serves as lexical base, as in *agua bendita* 'holy water' and *cuenta corriente* 'current account', from which *-era* and *-ista* generate an instrument (49a) and an agent (49b). This test proves the semantic unity of such constructions and ratifies their lexical nature, although it tells us nothing about their morphological status. In addition, the test is of limited application from a morphological viewpoint because operating derivation on Spanish N+A constructions most frequently leads to ungrammatical formations (Bustos Gisbert 1986: 139).

(49a) *aguabendit**era*** 'home stoup'
lit. water.holy.er
(49b) *cuentacorrent**ista*** 'current account holder'
lit. account.current.ist

N+A units show heterogeneous behaviors, and disparities exist regarding their endocentricity/exocentricity, ability to undergo derivation or locus of inflection. In particular, some authors have argued for a level intermediate between N+A compounds and phrases. This would involve sets of MWEs that are compositional but at the same time share one of their constituents, e.g. *negro* 'black' in (50). Here, *negro* contributes a regular figurative sense throughout different examples, while the other member of the construction adds a literal meaning. These certainly behave as mixed nominal phrases insofar as they have a fixed component and an idiomatic one (Ginebra 2002: 148–151).

(50a) *dinero negro* 'dirty money'
lit. money black
(50b) *lista negra* 'blacklist'
lit. list black
(50c) *mercado negro* 'black market'
lit. market black

3.2 Adjectives

Complex adjectival constructions are less challenging than nominal ones thanks to their spelling, which may be closed or hyphenated but always reveals their lexical nature. The formal makeup of these constructions is adjective+adjective

(A+A) or N+A. The A+A examples in (51) are lexicalized and represent a synchronically unproductive type, while those in (52) are profuse and stand unambiguously within morphology. The former type is limited to adjectives expressing colors and judgement, while the latter displays a much wider semantic scope and is frequently recursive.

(51a) *agri.dulce* 'bittersweet'
lit. bitter.sweet
(51b) *verde.azul* 'green-blue'
lit. green.blue

(52a) *político-laboral* 'related to politics and labour'
lit. political labour
(52b) *nacional-cultural-social* 'national-cultural-social'
lit. national-cultural-social

Two main kinds of N+A constructions exist: one where the noun refers to salient body parts (53), an exocentric and usually non-compositional type, and a small group where the noun is the name of a language and the head is a participle meaning 'to speak', cf. (54). Both are analyzable as compounds as they receive external inflection (e.g. *pelirrojo**s*** lit. hair.red.PL, *vascoparlante**s*** lit. Basque. speaking.PL) and forbid internal modification (**castellano*.***muy.****hablante* lit. Spanish.very.speaking).

(53a) *pelirrojo* 'red-haired'
lit. hair.red
(53b) *paticorto* 'short-legged'
lit. leg.short

(54a) *castellanohablante* 'Spanish-speaking'
lit. Spanish.speaking
(54b) *vascoparlante* 'Basque-speaking'
lit. Basque.speaking

Some adjective compounds denote a color which is derived from the colors expressed by their constituents (55), while others denote nationalities (56). Plural marking varies, since most formations are peripherally inflected to the right (57a), but some remain uninflected (57b); gender is always expressed in the right-most member, both in the singular and the plural forms (58). The phraseological nature of these constructions can be observed in the restricted selection of their compo-

nents (**azul.i.blanco* lit. blue.i.white), which makes compoundhood relevant only diachronically.

(55a) *blanqu.i.azul* 'blue and white'
lit. white.i.blue
(55b) *roj.i.blanco* 'red and white'
lit. red.i.white

(56a) *hispano-francés* 'Hispanic-French'
lit. Hispanic-French
(56b) *anglo-eslovaco* 'Anglo-Slovak'
lit. Anglo-Slovak

(57a) *vaca.**s** blanqu.i.marron.**es*** 'white and brown cows'
lit. cows white.i.brown.PL
(57b) *camisa.**s** azul marino* 'navy blue shirts'
lit. shirts blue navy

(58a) *aficionad.**as** verd.i.negr.**as*** 'green and black fans'
lit. fans.FEM green.i.black.FEM-PL
(58b) *cumbre.**s** ruso-judí.**as*** 'Russian-Jewish summit'
lit. summits Russian.Jewish.FEM-PL

Adjective compounds therefore resemble phrases, but can be unproblematically analyzed as compounds, as plural and gender inflection is applied externally. For the same reason, only the constituents in phrases can be independently modified (59b). This is ungrammatical in compounds (59a).

(59a) **pat.i.**muy**.corto*
lit. leg.i.very.short
(59b) *muy ancho de espaldas* 'having a wide back' (person)
lit. very wide of back

3.3 Verbs

Genuine verbal compounding is so marginal in Spanish that it is altogether omitted from some works (Lang 1992), while others portray it as "virtually absent" (Klingebiel 1989: 1). Unlike noun and adjective compounds, this type cannot be formed by concatenating two lexemes from the word-class of the compound, here

verbs. The most representative type of verbal compounding is N+V, described sometimes as back-formation from adjectives (Val Álvaro 1999) and sometimes as noun incorporation that comes from Latin (Moyna 2011), cf. (60). The rule's current productivity is null, with the exception of a few recent formations derived by back-formation, e.g. (61a) from *boquiabierto* 'open-mouthed', or (61b) from *publicontratación* 'crowdsourcing'.

(60a) *maniobrar* 'to maneuver'
lit. hand.to act
(60b) *pelechar* 'to grow new fur'
lit. fur.to grow

(61a) *boquiabrir* 'to open one's mouth'
lit. mouth.to open
(61b) *publicontratar* 'to crowdsource'
lit. public.to hire

These formations aside, the verbal procedure that most closely resembles compounding is that of light verb constructions (62), where a semantically void verb is accompanied by a noun to create a conceptual unit. These constructions are compositional and have a corresponding synthetic lexical verb which often expresses the same meaning. Even if they cannot be called morphological objects, these are not regular verb phrases and resemble compounds because of their highly regular and frequent occurrence (cf. Val Álvaro 1999: 4830–4834).

(62a) *Pedro hizo mención de Luis* *Pedro mencionó a Luis*
'Pedro made mention of Luis' 'Pedro mentioned Luis'
(62b) *Pedro dio aviso del fuego* *Pedro avisó del fuego*
'Pedro gave notice of the fire' 'Pedro warned about the fire'

Despite their verbal nature, formations like (63) stand apart from regular verbs and from light verb constructions due to the fact that it is impossible to replace their constituents by synonyms (64), to internally modify the noun (65), or to apply sentence transformation on their structure (66) (cf. Val Álvaro 1999: 4831).

(63a) *tomar el pelo* 'to pull somebody's leg'
lit. to take the hair
(63b) *estirar la pata* 'to kick the bucket'
lit. to stretch the leg

(64a) **coger el pelo* lit. to catch the hair
(64b) **extender la pata* lit. to extend the leg

(65a) **tomar el pelo bonito* lit. to take the hair beautiful
(65b) **estirar la pata izquierda* lit. to stretch the leg left

(66a) **El pelo le fue tomado por Luis a Pedro*
'Pedro's leg was pulled by Luis'
(66b) **¿Qué ha estirado Pedro?*
'What has Pedro pulled?'

One peculiarity of verbal MWEs is their lack of predisposition towards orthographic fusion, which would lead to noun incorporations such as **pelotomar* (lit. hair.to take) or **pataestirar* (lit. leg.to stretch). One likely explanation is the possibility to bring the verbal complement into theme position, thus suggesting that the components in these structures are not morphological and retain at least some syntactic independence. This is observable in *brillar por su ausencia* 'to be conspicuous by its absence' and *hilar fino* 'to split hairs', and makes it simpler to set boundaries between verb compounds and verbal MWEs.

(67a) *Por su ausencia no brilla*
'By its absence it is not conspicuous'
(67b) *Por muy fino que hiles no lo conseguirás*
'Many hairs though you split, you will not achieve it'

Verbal collocations must be taken into account as well (cf. (68)). Here, the head is a verb that is complemented by a noun (68a), a preposition plus a noun (68b) or an adverb (68c). These exhibit different degrees of idiomaticity and fixity, and must be regarded as syntactic.

(68a) *estallar una revolución/rebelión/protesta*
lit. to break out a revolution/rebellion/protest
(68b) *gozar de popularidad/fama/renombre/tirón*
lit. to enjoy of popularity/fame/renown/momentum
(68c) *dormir plácidamente*
lit. to sleep placidly

As happens in Italian (Iacobini 2009), it may be the case that Spanish verbal MWEs are proportionally more widely employed than MWEs of other word classes because of the low productivity of verbal compounding, although this has yet to

be substantiated. Even though verbal MWEs do occur, it seems safe to assert that the native procedures for phrasal or multi-word verbs are not powerful if compared to Germanic languages or even Romance languages like Catalan or Italian (Guevara 2012; Bisetto 2015).

4 Reconciling compounds and MWEs

The previous sections have evidenced the heterogeneous and unequal performance of Spanish compounds and MWEs for the categories noun, adjective and verb. This section reconsiders these views and describes their competitive vs. cooperative relationship.

Scholars have ascribed a range of attributes and behaviors to MWEs and compounds. This has brought about a catalogue of discriminating measures designed to allocate a structure to morphology, phraseology or syntax. One thorough approach is Ruiz Gurillo (2002), where features are reviewed at the phonological, syntactic, lexico-semantic and pragmatic levels. Table 1 outlines the most prominent characteristics and indicates if they are possible (+), impossible (–) or optional (±) in synchronic compounds, phrases and collocations.[5]

Table 1: Cross-categorial features (from Ruiz Gurillo 2002)

Features	Compound	Phrase	Collocation
Multi-word nature	+	+	+
Naming ability	+	+	–
Consolidated formation	±	+	±
Frequent co-occurrence	+	+	+
Paradigm membership	+	–	–
Lack of stress unity	+	+	+
Fixed lexical components	+	+	+

5 These features have been discussed at different points in the present article and appear here with specialized Spanish terminology. Paradigm membership, for example, refers to the fact that, if a construction is coined via a synchronic syntactic procedure, it will be placed together with the previous constructions created by that rule. The body of constructions built through the same structure would therefore constitute its consolidation as a paradigm. Similarly, isomorphism is a variable of a unit's idiomaticity, since it indicates to what extent a unit can be broken down in meaningful subcomponents.

Features	Compound	Phrase	Collocation
Variability of lexical components			
Plural inflection	+	+	+
Insertion of modifiers	–	–	±
Isomorphism	+	–	+
Meaning compositionality	+	–	+
Metaphors and tropes	–	+	±
Idiomaticity	–	+	±
Lexical selection	–	–	+

Table 1 makes manifest an uneven distribution pattern of features, with the result that some are possible in all three constructions (e.g. the ability to be made up of multiple words), others are largely optional (e.g. making up a consolidated formation), and others are impossible (e.g. insertion of modifiers), although exceptions have been noted for most of the categories. Taken together, this causes a cross-categorial overlap which leads to descriptive vagueness and fuzzy borders. Depending on the degree of concurrence of these features, we will be faced with a more or less prototypical morphological, phraseological or syntactic unit. The combination of these characteristics also demarcates two features often associated with phraseology: fixity and idiomaticity. In principle, the more fixed and idiomatic a unit is, the more it can be considered as unambiguously phraseological, even if less prototypical constructions may be phraseological too (Gries 2008: 5f.). There are hence archetypal compounds and archetypal phraseologisms, depending on their overall reaction to the above criteria. In view of their border properties, Gaeta/Ricca (2009) accommodate compounds and phrases into a quadripartite system that distinguishes the feature of being listed in the lexicon from that of being the output of morphology. For these authors, lexicalization and compoundhood are independent notions and each may be present or absent in a particular construction. This materializes in a four-level typology (69) which embraces prototypical compounds (69a), prototypical phrases (69d), and two intermediate positions (69b) and (69c):

(69a) [+ morphological], [+ lexical]
(69b) [+ morphological], [– lexical]
(69c) [– morphological], [+ lexical]
(69d) [– morphological], [– lexical]

The rationale is that, just like there are "[...] lexical units that are not compounds, but syntactic units, we should also find compounds (morphological units) which are not lexical units" (Gaeta/Ricca 2009: 40). An example of (69a) is *compraventa* 'buying and selling', and one of (69d) is *gorra de metal* 'metal cap'. Type (69c) involves syntactic elements that have a conceptual referent, e.g. *dolor de cabeza* 'headache', while (69b) is *a priori* an unexpected kind: compounds that are not lexically listed. This is possible for extremely productive morphological processes, whose output is large, and not all of which is lexicalized. In Spanish, it is the case of V+N compounding, as in *espantacucarachas* 'cockroach scarer' (cf. Section 3.1.1).

This leads us to the competitive vs. cooperative behavior of compounds and MWEs. The fact that many phrasal constructions (e.g. *guerra fría* 'cold war', *café con leche* 'white coffee') have a denominative role and are accompanied by a definition in lexicographic studies is proof of their naming ability, which in turn sets them up as potential competitors for word-formation (Booij 2009: 220). This is evident for example in doublets formed by one morphological and one phraseological construction, as in (42): *cita médica* 'medical appointment' vs. *cita del médico* 'appointment of the doctor'. Occasionally, one of the units becomes established and blocks the other, e.g. **guerra de*(*l*) *frío* lit. war of the cold (vs. *guerra fría* 'cold war'), although coexistence is not rare. The exact nature of this interaction depends on language-specific factors (Hüning/Schlücker 2015 on German; Masini this volume on Italian), not extensively discussed in the Spanish literature.

The consequence deriving from this behavior is what one would expect: genuine compounding is not a frequent lexical resource in Spanish, and this causes the interference of MWEs as a naming device. In the case of nouns, Section 3.1 discusses orthographic constructions which unequivocally qualify as compounds and the three configurations N+N, N+p+N and N+A. In the case of adjectives (Section 3.2), broad agreement exists on their morphological origin, which is why adjectival MWEs (70) are not generally required to fulfil a naming function.

(70) *estar hecho polvo* 'to be exhausted'
lit. to be made dust

Finally, the role of verbal compounding in Spanish (Section 3.3) is so negligible that most constructions are derived from phrasal processes. It seems that Spanish resorts to MWEs differently for each word-class: adjectival compounding is practically self-sufficient and requires almost no additional support, verbal compounding stands at the opposite extreme, so phraseology is often activated for verbal MWEs, and nominal compounding stands midway. Unsurprisingly, the

differentiation between morphological and syntactic objects is the most problematic in those areas where compounding and MWEs interact closely, i.e. N+A nouns, N+N nouns and N+p+N nouns.

Bearing this in mind, the relationship between compounding and phraseological processes must be characterized as partly competitive and partly cooperative. There is competition when two processes are synchronically productive and struggle to coin naming units, so speakers may resort to both of them, at which time doublets arise. On the other hand, the cooperation between compounding and phrase-formation becomes manifest when the latter produces units for morphologically unavailable compound types, thus guaranteeing that concepts can be named. When both processes are available, compounding seems to be hierarchically superior (which is in keeping with the basic naming function of word-formation). This can be noticed in adjectival formations, where compounding is prevalent and MWEs are far less common despite being synchronically available. In contrast, in verbal compounding, where compounding is either unproductive or lexicalized, phraseological formations abound. This versatility of MWE formation in Romance languages has been explained by its fruitful use of prepositions, which facilitates the creation of p+N strings that "[...] may function as derivational suffixes where proper suffixes may not be admitted or may not exist" (Piunno 2016: 31). This view accounts for formations like *tren de mercancías* 'freight train' or *cuento de hadas* 'fairy tale' (30), where the prepositional modifiers (*de mercancías* 'of freight', *de hadas* 'of fairies') counterweigh the non-existence of adjectival derivations from *mercancía* and *hada*.

Bauer (1998: 83ff.) opines differently on the connection between MWEs and compounds. In discussing English N+N constructions, he wonders if we are faced not with two different prototypical categories plus midway cases, but with just one broader category whose members display contrasting features. This would certainly explain the oft-cited overlap of morphological and syntactic entities in various languages (Ruiz Gurillo 2002; Gaeta/Ricca 2009). Bauer invokes the Avoid Synonymy Principle (Kiparsky 1983), which accounts for the fact that the existence of a denominative unit (be it a compound or an MWE) prevents the use of its competitor, and he wonders about the nature of this single category: morphological or syntactic. At present we lack strong evidence for a definitive distinction between two types of N+N constructions, although that does not necessarily validate the existence of a single category. The main obstacle, if so, is which framework may embrace these formations, since their hybrid nature is irreconcilable with a modular view of grammar. As in other works dealing with MWEs (Masini 2005, 2009, this volume; Booij 2009, this volume; Piunno 2016), Construction Morphology (Booij 2010) is here deemed a suitable candidate since constructions are versatile form-meaning pairings whose complexity ranges from simple words

to complex idioms. As has been shown, the data available for Spanish is not favorable for a two-category distinction, and so the possibility of a single all-inclusive class is particularly welcome in this case. Turning to constructions of course implies allowing MWEs into the mental lexicon, meaning that MWEs and compounds co-exist, overlap somewhat in their forms and functions and are hence competitors for the naming act. This position is consistent with the depiction of the Spanish system presented above, and offers a middle-ground solution to the apparently irreconcilable nature of these two sets of units.

5 Conclusions

This article has offered a concise overview of MWEs and compounds in Contemporary Spanish. It has dealt with constructions that can be viewed as compounds, phrases or collocations depending on an analysis based on a combination of syntactic, phraseological and morphological features. A non-discrete demarcation of such units is the clearest outcome of the tests available, with several features shared by compounds and idiomatic expressions. These tests make it impossible to empirically separate morphological from phraseological formations due to idiosyncrasies and exceptions caused by semantic and functional similarities. The above arguments and examples indeed make a case for a gradient structure of MWEs, of which compounds and phrases are extreme positions.

Some Spanish compounds and MWEs stand in cooperative rivalry. This association is apparently inversely proportional to their respective lexical output, such that the more productive compounding is for a given category, the less productive MWE formation will be. This ensures that a linguistic resource for concept naming will always be available. In this sense, observation of the data makes it safe to assert that Spanish compounding is productive mainly for nouns and adjectives, and that MWE formation is exploited for other categories. It must be borne in mind, however, that the environment of Spanish morpho-syntax is different from that of English, from which most current linguistic frameworks and theories of word-formation have emerged. The contrast between the Spanish and English systems is evident for example in the allegedly poor output of Spanish compounding or very high productivity of the exocentric V+N pattern, measures which will by need seem unsatisfactory if English is taken as the benchmark. It may then be the case that a strict application of Germanic models on Romance phenomena will most likely project an imperfect picture. The present situation calls for an approach which considers MWEs in other languages but does not impose external models to native patterns (e.g. Booij 2009; Gaeta/Ricca 2009; Masini 2009).

In elucidating the status of MWEs, the need for agreement among linguistic disciplines is urgent, a task that has been neglected so far. For decades, research into morphology has made little headway in the analysis of phrase-like compounds, and phraseologists have unsuccessfully struggled in explaining various levels of multi-word formations. Joint efforts may thrive in precisely locating MWEs in the language system, not through separate investigations, but by looking at the common goals of morphology and phraseology: "a proper theory of the relation between morphological and syntactic naming constructions is called for" (Booij 2009: 220). Let us remember that phraseology is a young field whose conceptual foundations seem to be under development. Gries puts it as follows (2008: 22; also Colson 2016):

> Many phraseologists [...] have focused on rather descriptive work on phraseology (or, more narrowly, idioms) and have often not been concerned with integrating their accounts of phraseologisms in particular and other patterns more generally into a larger theory of the linguistic system.

Hopefully, this dearth of theoretical descriptions will eventually be overcome and develop into a robust treatment of MWEs which will allow us to explain borderline cases like the above. Morphology and phraseology are undoubtedly on track to achieving a comprehensive account of multi-word lexical phenomena, but a concerted effort is needed to reach this end; until then, a definitive description of MWEs will be on hold.

References

Bauer, Laurie (1998): When is a sequence of two nouns a compound in English? In: English Language and Linguistics 2, 1. 65–86.

Bauer, Laurie (2017): Compounds and compounding. (= Cambridge Studies in Linguistics 155). Cambridge, UK: Cambridge University Press.

Bisetto, Antonietta (2015): Do Romance languages have phrasal compounds? A look at Italian. In: Language Typology and Universals (STUF) 68, 3. 395–419.

Bisetto, Antonietta/Scalise, Sergio (2005): The classification of compounds. In: Lingue e Linguaggio 4, 2. 319–332.

Booij, Geert (2009): Phrasal names: A constructionist analysis. In: Word Structure 2, 2. 219–240.

Booij, Geert (2010): Construction Morphology. Oxford: Oxford University Press.

Buenafuentes de la Mata, Cristina (2010): La composición sintagmática en español. San Millán de la Cogolla: Cilengua.

Buenafuentes de la Mata, Cristina (2014): Compounding and variational morphology: The analysis of inflection in Spanish compounds. In: Borealis: An International Journal of Hispanic Linguistics 3, 1. 1–21.

Bustos Gisbert, Eugenio (1986): La composición nominal en español. Salamanca: Universidad de Salamanca.
Colson, Jean-Pierre (2016): Editorial: Phraseology at the intersection of grammar, culture and statistics. In: Yearbook of Phraseology 7. 1–2.
Davies, Mark (2002–): El Corpus del Español. Internet: www.corpusdelespanol.org (last access: 11.6.2018).
Gaeta, Livio/Ricca, Davide (2009): Composita solvantur: Compounds as lexical units or morphological objects? In: Rivista di Linguistica 21, 1. 35–70.
Ginebra, Jordi (2002): Las unidades del tipo dinero negro y dormir como un tronco: ¿Naturaleza léxica o gramatical? In: Veiga, Alexandre/González Pereira, Miguel/Gómez, Montserrat Souto (eds.): Léxico y gramática. (= Linguas e Lingüística 3). Lugo: Tris Tram. 147–154.
Gries, Stefan Th. (2008): Phraseology and linguistic theory. A brief survey. In: Granger, Sylviane/Meunier, Fanny (eds.): Phraseology: An interdisciplinary perspective. Amsterdam u.a.: Benjamins. 3–25.
Guevara, Emiliano R. (2012): Spanish compounds. In: Probus 24, 1. 175–195.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.): Word-formation. An international handbook of the languages of Europe. Vol. 1. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK) 40.1). Berlin/Boston: De Gruyter. 450–467.
Iacobini, Claudio (2009): Phrasal verbs between syntax and the lexicon. In: Rivista di Linguistica 21, 1. 97–117.
Kiparsky, Paul (1983): Word-formation and the lexicon. In: Ingemann, Frances J. (ed.): Proceedings from the 1982 Mid-America Linguistics Conference. Lawrence, KS: University of Kansas. 3–32.
Klingebiel, Kathryn (1989): Noun + verb compounding in Western Romance. (= University of California Publications in Linguistics 113). Berkeley/Los Angeles, CA: University of California Press.
Koike, Kazumi (2001): Colocaciones léxicas en el español actual: Estudio formal y léxico-semántico. Alcalá de Henares: University of Alcalá de Henares.
Kornfeld, Laura Malena (2009): IE, Romance: Spanish. In: Lieber, Rochelle/Štekauer, Pavol (eds.): The Oxford handbook of compounding. Oxford: Oxford University Press. 436–452.
Lang, Mervyn F. (1992): Formación de palabras en español: Morfología derivativa productiva en el léxico moderno. Madrid: Cátedra.
Masini, Francesca (2005): Multi-word expressions between syntax and the lexicon: The case of Italian verb-particle constructions. In: SKY Journal of Linguistics 18. 145–173.
Masini, Francesca (2009): Phrasal lexemes, compounds and phrases: A constructionist perspective. In: Word Structure 2, 2. 254–271.
Montoro del Arco, Esteban T. (2008): Relaciones entre morfología y fraseología: Las formaciones nominales pluriverbales. In: Montoro del Arco, Esteban T./Pérez, Ramón Almela (eds.): Neologismo y morfología. Murcia: Universidad de Murcia. 121–146.
Moyna, María Irene (2011): Compound words in Spanish. Theory and history. Amsterdam u.a.: Benjamins.
Pafel, Jürgen (2017): Phrasal compounds and the morphology-syntax relation. In Trips, Carola/Kornfilt, Jaklin (eds.): Further investigations into the nature of phrasal compounding. (= Morphological Investigations 1). Berlin: Language Science Press. 233–259.
Piunno, Valentina (2016): Multiword modifiers in some Romance languages. Semantic formats and syntactic templates. In: Yearbook of Phraseology 7. 3–34.

RAE: Real Academia Española y Asociación de Academias de la Lengua Española (2010): Nueva gramática de la lengua española. Manual. Madrid: Espasa.

Rainer, Franz/Varela, Soledad (1992): Compounding in Spanish. In: Rivista di Linguistica 4, 1. 117–142.

Rao, Rajiv (2015): On the phonological status of Spanish compound words. In: Word Structure 8, 1. 84–118.

Renner, Vincent (2006): Les composés coordinatifs en anglais contemporain. Unpublished PhD dissertation. Université Lumière Lyon 2. Internet: https://tel.archives-ouvertes.fr/tel-00565046 (last access: 10.12.2016).

Ruiz Gurillo, Leonor (2002): Compuestos, colocaciones, locuciones: Intento de delimitación. In: Veiga, Alexandre/González Pereira, Miguel/Gómez, Montserrat Souto (eds.): Léxico y gramática. Lugo: Tris Tram. 327–339.

Scalise, Sergio/Fábregas, Antonio (2010): The head in compounding. In: Scalise, Sergio/Vogel, Irene (eds.): Cross-disciplinary issues in compounding. Amsterdam u.a.: Benjamins. 109–125.

Schlücker, Barbara/Hüning, Matthias (2009): Compounds and phrases. A functional comparison between German A+N compounds and corresponding phrases. In: Rivista di Linguistica 21, 1. 209–234.

Val Álvaro, José Francisco (1999): La composición. In: Demonte, Violeta/Bosque, Ignacio (eds.): Gramática descriptiva de la lengua española. Vol. 3: Entre la oración y el discurso. Morfología. Madrid: Espasa-Calpe. 4757–4842.

Van Goethem, Kristel (2009): Choosing between A+N compounds and lexicalized A+N phrases: The position of French in comparison to Germanic languages. In: Word Structure 2, 2. 241–253.

Maria Koliopoulou

Compounds and multi-word expressions in Greek

1 Introduction

Complex lexical units include compounds as well as multi-word expressions displaying mixed morphosyntactic properties.[1] These mixed properties are determined by language-specific characteristics. Moreover, a diversity of properties is observed among the different types of multi-word expressions; in some cases even within the same type of structure. Therefore, their status is rather unclear, as is also revealed by the strong name variation among scholars (Hüning/Schlücker 2015: 450f.), even within the same language. The different naming suggestions cannot be considered as one-to-one equivalents or synonyms. The selection of one of them is also determined by the theoretical approach adopted. Specifically, the selection or the creation of a new label depends on the type of grammatical model as well as on the role of the lexicon to the formation of new lexical units.

Multi-word expressions in Greek have caught the attention of linguists in the twentieth century. This type of lexical unit has been used more often in the form of loan translations from English and French (Anastassiadis-Symeonidis 1986, 1994). Since then it has been rather prominent in many terminological domains as well as in media language. Moreover, it constitutes a commonly selected formation type of lexical units for the naming of new concepts or the translation of borrowed terms gaining ground over the formation of typical compounds.

The phenomenon of terminological variation regarding multi-word expressions is also apparent in the literature of Greek. Different names that have been suggested among scholars are for instance *lexical phrases* (Anastassiadis-Symeonidis 1986; Ralli 1991), *multi-word compounds* (Ralli 1992; Anastassiadis-Symeonidis 1996; Christofidou 1997; Ralli/Stavrou 1998) and *loose multi-word compounds* (Ralli 2005, 2007; Koliopoulou 2006, 2008, 2009). Ralli (2013a, 2013b; cf. also Bağriaçik/Ralli 2015) adopts in her later studies the term *phrasal compounds*, inspired by Booij's (2009, 2010: 169–192) term *phrasal names*, in order to differen-

1 I wish to thank the editor of this volume, Anna Anastassiadis-Symeonidis, Pius ten Hacken as well as the two anonymous reviewers for their constructive comments and criticism. Needless to say, remaining mistakes and opinions expressed are of my own responsibility.

https://doi.org/10.1515/9783110632446-008

tiate specific types of complex lexical units from typical one-word compounds which are morphological objects. However, the use of the term *phrasal compound* to refer to this type of structure can be misleading, since it is also used to denote another kind of structure, namely compounds with a phrasal element at the non-head position, like *chicken and egg situation* in English. Such structures are not possible in Greek (cf. Section 2.1).

In this study, I adopt the term *multi-word expression* as a term that is general and theory-neutral – also suggested by Hüning/Schlücker (2015: 451) – to refer to different types of complex lexical units in Greek sharing morphological and syntactic features in various proportions. The aim of this study is to analyze their complicated properties and compare them to typical compounds without letting theoretical considerations override the data. After having analyzed in detail the different types of multi-word expressions in Greek, I will come back to more theoretical considerations regarding their interrelation with other comparable lexical units as well as their locus of realization in grammar.

Specifically, this study is structured as follows: Section 2 gives an overview of various complex lexical units found in Greek. Typical compounds, multi-word expressions as well as phrase-like structures are analyzed in detail and compared to each other. Section 3 discusses the interrelation between the various types arguing that they coexist in the lexicon as complementary resources of nominal naming units. However, coexistence in the lexicon does not exclude competition among types. Section 4 deals with the question of how complex lexical units can be accounted for in the lexicon and in grammar. Finally, Section 5 summarizes the conclusions.

2 Typical compounds vs. other complex lexical units

Compounding can be considered as the output of a morphological operation situated closer to syntax than any other morphological formation (Scalise 1992: 4). As a result of this closeness, it is sometimes rather difficult to differentiate compounds from phrases[2], and even more from intermediate structures displaying a

2 Many studies have been carried out on the distinction between compounds and phrases based on selected criteria mostly concerning the formal properties of a compound contrary to those of a syntactic phrase (e.g. Borer 1988; Scalise 1992; ten Hacken 1994; Bisetto/Scalise 1999; Bauer 2001; Olsen 2001; Donalies 2004; Gaeta/Ricca 2009; Schlücker/Hüning 2009). Despite the detec-

mix of properties of different types of structures. However, the demarcation between typical one-word compounds and intermediate structures is relatively clear with regard to the Greek data. The difficulty consists in the demarcation between the different types of intermediate structures, the analysis of the degree of structural connection with typical compounds or with regular syntactic phrases, as well as in the decision on whether these structures belong to morphology or to syntax.

2.1 Typical compounds

Compounding is one of the most productive morphological processes in Greek. One-word compound formations mostly built up from stems are found in both spoken and written language in various types of texts. The spontaneous creation of compounds that in some cases succeed to be established and to enter the speakers' mental lexicon is not rare.

Compounds in Greek involve all major lexical categories, namely nouns (1), adjectives (2) and verbs (3). Determinative compounds are right-headed, as shown by the examples below.[3]

(1)	κεφαλόσκαλο	←	κεφάλ(ι)$_{N}$	-ο-	σκαλ(ί)$_{N}$
	kefaloskalo		*kefal*(*i*)	*-o-*	*skal*(*i*)
	upper/wider step		head	LE[4]	step
(2)	εθιμοτυπικός	←	έθιμ(ο)$_{N}$	-ο-	τυπικός$_{A}$
	ethimotipikos		*ethim*(*o*)	*-o-*	*tipikos*
	formal/traditional		custom	LE	typical
(3)	κρυφοκοιτάζω	←	κρυφ(ά)$_{Adv}$	-ο-	κοιτάζω$_{V}$
	krifokitazo		*krif*(*a*)	*-o-*	*kitazo*
	watch secretly		secretly	LE	watch/look

tion of specific criteria, there is no agreement on a clear-cut distinction between compounds and phrases, at least not cross-linguistically.

3 Greek examples in this chapter are given in Greek as well as transliterated in the Latin script, before being translated into English.

4 The abbreviation stands for "linking element".

Nominal compounds consisting of two nouns are the most productive ones (1), as in a number of other languages, for instance in German. Verbal compounding is very productive in Greek in comparison to other European languages, either in the form of determinative structures, as in (3), or in the form of coordinative structures (e.g. *πηγαινοέρχομαι* 'come and go'). In German, for instance, the limited number of verbal compounds is the result of a backformation process from nominal compounds (Becker 1992: 20f.; Günther 1997: 6).

With regard to their structural properties, compounds in Greek usually consisting of stems form one phonological word written as one graphemic unit (cf. (1)–(3)). This phonological word has one main stress assigned either on the antepenultimate syllable of the entire compound formation (1), or on the regular stress position of the right-hand constituent (2, 3). Stress assignment is determined by two specific phonological rules applicable to all compound formations (Nespor/Ralli 1994: 201, 1996: 357). The form of these rules will not concern us here.

Moreover, compounds in Greek constitute one morphological unit, to which syntactic operations do not have access. In the following, I contrast the properties displayed by a compound formation (4) with those of a syntactic phrase (5), both consisting of an adjective and a noun, so that the analysis is comparable. The first indication of the word atomicity displayed by compounds is related to the fact that word internal inflection is not allowed (4b), contrary to syntactic phrases, whose components are inflected.

(4)	$[A\ N]_{compound}$				
(4a)	τρελόπαιδο$_{Neu.Nom.Sg}$	←	τρελ(ό)$_{Neu.Nom.Sg}$	-ο-	παιδ(ί)$_{Neu.Nom.Sg}$
	trelopedo		*trel*(*o*)	*-o-*	*ped*(*i*)
	crazy boy		crazy	LE	child
(4b)	*τρελ-ά-παιδ-α$_{Neu.Nom.Pl}$	←	τρελ(ά)$_{Neu.Nom.Pl}$	παιδ(ιά)$_{Neu.Nom.Pl}$	
	trel-a-ped-a		*trel*(*a*)	*ped*(*ia*)	
	crazy boys		crazy	children	

(5)	$[A\ N]_{phrase}$	
(5a)	τρελ-ό$_{Neu.Nom.Sg}$	παιδ-ί$_{Neu.Nom.Sg}$
	trel(*o*)	*ped-i*
	crazy	child/boy
(5b)	τρελ-ά$_{Neu.Nom.Pl}$	παιδ-ιά$_{Neu.Nom.Pl}$
	trel-a	*ped-ia*
	crazy	children/boys

Apart from this first distinctive characteristic, I will apply a number of diagnostic tests to both types of structure in order to verify the lexical integrity of compound structures. Some of the typical diagnostic tests found in the literature on compounding in Greek (cf. Ralli 2013a: 21, 24; Bağriaçik/Ralli 2015: 328f.; ten Hacken/Koliopoulou 2016: 130ff.) are the following: a) independent modification of the non-head (6), b) coordination of the components (7), c) reversing the word order (8).

(6a) *πολυ-τρελ-ό-παιδο
poli-trel-o-pedo
very crazy boy
(6b) πολύ τρελό παιδί
poli trelo pedi
very crazy boy

(7a) *τρελ-ο-και-χαζ-ό-παιδο
trel-o-ke-chaz-o-pedo
crazy and stupid boy
(7b) τρελό και χαζό παιδί
trelo ke chazo pedi
crazy and stupid boy

(8a) *παιδ-ό-τρελο
ped-o-trelo
boy crazy
(8b) παιδί τρελό
pedi trelo
boy crazy

With regard to the last test, according to which the order of the constituents of syntactic phrases can be reversed (8b), it should be mentioned that this possibility increases the emphasis on the syntactic phrase. Specifically, the property designated by the adjective is highlighted by this stylistic variation (ten Hacken/Koliopoulou 2016: 131f.). On the contrary, the word order of compound components is fixed (8a). Even in compounds consisting of components with the same lexical category, like noun-noun compounds, the change of the order of the two components is – at least in Standard Modern Greek – ungrammatical (cf. (1) *κεφαλόσκαλο*/**σκαλοκέφαλο* 'upper/wider step').

A further distinctive characteristic is related to the type of constituents participating in syntactic phrases or compounds. Phrases consist of words while compounds in Greek usually consist of stems. However, the possibility of a word constituent in one of the two positions or even in both positions of a compound formation cannot be excluded. Since both types of free lexemes can occupy any constituent position, four structural patterns result from all possible combinations (cf. Ralli 2005: 237f., 2013a: 16; Koliopoulou 2013: 24f.).

(9a)	[stem-stem]				
	καραβόπανο	←	καράβ(ι)[5]	-ο-	παν(ί)
	karav-o-pano		*karav*(*i*)	*-o-*	*pan*(*i*)
	sailcloth		ship	LE	cloth
(9b)	[stem-word]				
	θαλασσοταραχή	←	θάλασσ(α)	-ο-	ταραχή
	thalassotarachi		*thalassa*(*a*)	*-o-*	*tarachi*
	sea disturbance		sea	LE	disturbance
(9c)	[word-stem]				
	επτάψυχος	←	επτά	ψυχ(ή)	
	eptapsichos		*epta*	*psich*(*i*)	
	having seven lives		seven	soul	
(9d)	[word-word]				
	ξαναμιλάω	←	ξανά	μιλάω	
	ksanamilao		*ksana*	*milao*	
	talk again		again	talk	

The most productive pattern is that of stem-stem formations (9a), since stem-constituents are preferred in Greek compounds.

The preference for a specific type of constituent in the compound formations constitutes an important parameter determining various structural characteristics of compounds (cf. Koliopoulou 2013, 2014a), among others the possibility of the appearance of a linking element. Specifically, in case the first constituent is a stem (9a), (9b), the two constituents are linked to each other with the element *-o-*.[6] Its appearance is obligatory and rather systematic, motivated by the fact that compound constituents are usually stems.

5 A stem constituent in a Greek compound can also be indicated by the fact that the truncated inflectional ending is given in parentheses.

6 Other possible forms of the linking element are *-ι-* and *-α-* appearing in rare cases (cf. Ralli 2013a: 50–53).

The impact of the "word- vs. stem-based parameter" in compounding becomes obvious if we compare Greek with German compounds.[7] German compounds are mostly built out of words, without excluding cases of a stem constituent in the first position of the compound, as in *Stimmabgabe* ('voting', *Stimm*(*e*) 'vote/voice', *Abgabe* 'delivery'). They are also characterized by the appearance of a linking element as for instance *Arbeit-s-ablauf* ('workflow').

However, the linking element in German compounds displays very different properties compared to the Greek linking element *-o-* (Koliopoulou 2014b). Therefore, the appearance of a linking element is not systematic, its form is variable, while compounds without linking element are very productive (e.g. *Stimm-Æ-abgabe* 'voting').

The preference of a particular language to build compounds out of words or stems affects further characteristics of the compounding process, for instance the possibility of recursion. Specifically, compounds in German tend to be expanded through recursion either in the non-head or the head position, as shown in (10a) and (10b) respectively (cf. Bauer 2009: 350; Neef 2009: 386; Koliopoulou 2017: 123). By contrast, recursion in Greek compounds, as illustrated in (11), is a rather rare phenomenon (cf. Koliopoulou 2013: 29f.; Mukai 2013: 43).

(10a)	[[*Stadt*][*fahrplan*]]	←	*Stadt*	*Fahrplan*
	city timetable		city	timetable
(10b)	[[*Altstadt*][*plan*]]	←	*Altstadt*	*Plan*
	old town map		old town	plan
(11a)	[[ζαμπον]-ο-[τυρόπιτα]]	←	ζαμπόν	τυρόπιτα
	zampon-o-tiropita		*zampon*	*tiropita*
	ham-cheese pie		ham	cheese pie
(11b)	[[ποδοσφαιρ]]-ό-[φιλος]]	←	ποδόσφαιρο	φίλος
	podosfer-o-filos		*podosfero*	*filos*
	football fun		football	friend

The difference in the degree of recursion between German and Greek compounds is related to the type of constituents. Specifically, I argue that stem constituents exhibit more restrictions than word constituents, whose more independent character allows the connection with further compound members, either in the head or in the non-head position.

7 For an extensive comparison between Greek and German compounds cf. Koliopoulou (2013, 2014a, 2014c, 2015).

2.2 Multi-word expressions

Another type of complex lexical unit composed of free morphemes are multi-word expressions. These peculiar structures, found also in Greek, have been already studied by many scholars (cf. literature mentioned in Section 1 as well as Anastassiadis-Symeonidis 1986: 138–143, 203ff.; Koliopoulou 2012: 862) in comparison to one-word compounds, to syntactic phrases and even to each other, since the various types display different characteristics. Specifically, contrary to typical compounds constituting one morphological word to which syntax has no access, multi-word expressions in Greek are structures with some morphological properties (cf. Section 2.2.1) without though preventing syntax from having access to their internal structure. They can be considered as intermediate structures, since they behave similarly to one-word compounds, but they also bear features typical for syntactic phrases. Their mixed properties vary not only within the different types of intermediate structures, but in some cases even among the various examples of the same type (cf. (26)–(27)).

Specifically, multi-word expressions in Greek are nominal structures[8] composed either of an inflected adjective and a noun or of two nouns. They look like syntactic phrases since their components are independent phonological words, contrary to one-word compounds constituting a single phonological word, regardless of the type of the compound constituents. Moreover, multi-word expressions consist of two inflected words. Compounds, by contrast, are usually formed out of stems linked by the element *-o-*. Compound formations are inflected at the right edge of the structure.

To be more specific, there are four types of multi-word structures[9]:

a) [A N] expressions composed of an inflected adjective and a noun (12),
b) [N N_{GEN}] expressions consisting of two nouns, the second being in the genitive case (13),
c) [N $N_{Attr.}$] expressions consisting of two nouns in attributive relation (14),
d) [N $N_{App.}$] expressions composed by two nouns in appositive relation (15).

8 There are only nominal multi-word expressions in Greek, which should not be confused with other types of phrasal expressions, like *την κάνω* ($tin_{FEM.ACC.SG}$ $kano_{1P.SG}$, her make, 'I am going'), namely fossilized expressions with a very idiomatic meaning (cf. Ralli 2013a: 252).

9 Most examples are taken from Anastassiadis-Symeonidis (1986), the first linguist that mentioned and analyzed thoroughly these structures in Greek.

(12)	[A N]			
(12a)	ψυχρός πόλεμος	←	ψυχρός$_{\text{Masc/Nom/Sg}}$	πόλεμος$_{\text{Masc/Nom/Sg}}$
	psichros polemos		*psichros*	*polemos*
	cold war		cold	war
(12b)	τρίτος κόσμος	←	τρίτος$_{\text{Masc/Nom/Sg}}$	κόσμος$_{\text{Masc/Nom/Sg}}$
	tritos kosmos		*tritos*	*kosmos*
	third world		third	world
(12c)	μαύρη αγορά	←	μαύρη$_{\text{Fem/Nom/Sg}}$	αγορά$_{\text{Fem/Nom/Sg}}$
	mavri agora		*mavri*	*agora*
	black market		black	market
(12d)	μεγάλη οθόνη	←	μεγάλη$_{\text{Fem/Nom/Sg}}$	οθόνη$_{\text{Fem/Nom/Sg}}$
	megali othoni		*megali*	*othoni*
	cinema		big	screen
(13)	[N N$_{\text{GEN}}$]			
(13a)	αγορά εργασίας	←	αγορά$_{\text{Nom.Sg}}$	εργασίας$_{\text{Gen.Sg}}$
	agora ergasias		*agora*	*ergasias*
	job market		market	job
(13b)	τάγματα ασφαλείας	←	τάγματα$_{\text{Nom.Pl}}$	ασφαλείας$_{\text{Gen.Sg}}$
	tagmata asfalias		*tagmata*	*asfalias*
	security battalions		battalions	safety
(13c)	κρέμα ημέρας	←	κρέμα$_{\text{Nom.Sg}}$	ημέρας$_{\text{Gen.Sg}}$
	krema imeras		*krema*	*imeras*
	day cream		cream	day
(13d)	άρση βαρών	←	άρση$_{\text{Nom.Sg}}$	βαρών$_{\text{Gen.Pl}}$
	arsi varon		*arsi*	*varon*
	weightlifting		lift	weight
(14)	[N N$_{\text{Attr.}}$]			
(14a)	λέξη κλειδί	←	λέξη$_{\text{Nom}}$	κλειδί$_{\text{Nom}}$
	leksi klidi		*leksi*	*klidi*
	key word		word	key
(14b)	νόμος πλαίσιο	←	νόμος$_{\text{Nom}}$	πλαίσιο$_{\text{Nom}}$
	nomos plesio		*nomos*	*plesio*
	frame law		law	frame
(14c)	φόρος φωτιά	←	φόρος$_{\text{Nom}}$	φωτιά$_{\text{Nom}}$
	foros fotia		*foros*	*fotia*
	very high tax		tax	fire
(14d)	γυναίκα αράχνη	←	γυναίκα$_{\text{Nom}}$	αράχνη$_{\text{Nom}}$
	gineka arachni		*gineka*	*arachni*
	greedy, dishonest woman		woman	spider

(15) [N $N_{App.}$]
(15a) μεταφραστής διερμηνέας
metafrastis diermineas
translator-interpreter
(15b) σκηνοθέτης παραγωγός
skinothetis paragogos
director-producer
(15c) ηθοποιός τραγουδιστής
ihtopios tragudistis
actor-singer
(15d) δικηγόρος πολιτικός
dikigoros politikos
layer-politician

Multi-word expressions constitute naming units, many of them displaying an idiomatic meaning. The degree of semantic opacity is in some cases comparable to that of typical compounds. Consider, for instance, the example *μαύρη αγορά* ('black market', (12c)), denoting a very specific type of market, or the example *άρση βαρών* ('weightlifting', (13d)) denoting an athletic discipline. However, as stated e.g. by Gaeta/Ricca (2009: 36), the semantic criterion is unreliable and can even be misleading for the demarcation between morphological and syntactic structures (cf. Section 2.2.1). Therefore, the present analysis is mainly based on formal criteria.

With regard to headedness, all four types display the same order as comparable adjective-noun (16a) and noun-noun syntactic phrases (16b).

(16a)	μικρό καλάθι	←	μικρό$_{Neu/Nom/Sg}$	καλάθι$_{Neu/Nom/Sg}$
	mikro kalathi		*mikro*	*kalathi*
	small basket		small	basket
(16b)	πόρτα σπιτιού	←	πόρτα$_{Fem/Nom/Sg}$	σπιτιού$_{Neu/Gen/Sg}$
	porta spitiu		*porta*	*spitiu*
	house door		door	house

Particularly, the nominal right-hand constituent is the head in [A N] formations. In [N N_{GEN}] and [N $N_{Attr.}$] expressions the left-hand constituent bears the head properties. Interestingly, [A N] expressions share the same order also with adjective-noun one-word compounds displaying the head position at the right-hand constituent (e.g. *τρελόπαιδο* 'crazy boy', cf. (4)). Nominal expressions of the types [N N_{GEN}] and [N $N_{Attr.}$] display the reverse order in comparison to noun-noun compounds which are right-headed (e.g. *κεφαλόσκαλο* 'upper/wider step' (1),

καραβόπανο 'sailcloth' (9a)). Therefore, with regard to headedness, [A N] expressions share more characteristics with typical one-word compounds than [N N_{GEN}] and [N $N_{Attr.}$] expressions.

A further property that some multi-word expressions share with typical compounds is that they can be input to a derivation process, specifically to suffixation (cf. Koliopoulou 2006: 49, 2009: 62, 2012: 863; Ralli 2007: 232f., 2013a: 247f., 266; ten Hacken/Koliopoulou 2016: 132f.). Specifically, one-word compounds, regardless of the type and the lexical category of constituents they consist of, can become bases for derivational suffixation, cf. (17). The most common derivational suffix added to a complex base is the adjectival suffix -ικ(ός), as shown in the examples below.

(17)	χαρτοπαικτικός	←	χαρτ$_{N}$-ο-παίκτ(-ης)$_{N}$	[-ικος]$_{ADJ}$
	chartopektikos		*chart-o-pekt-is*	*-ikos*
	card play$_{ADJ}$		card-LE-player	
	καλογερικός	←	καλ$_{A}$-ό-γερ(-ος)$_{N}$	[-ικος]$_{ADJ}$
	kalogerikos		*kal-o-ger-os*	*-ikos*
	monk$_{ADJ}$		good-LE-old man	

[A N] expressions, like those given under (12), stripped off both inflectional endings and turned into one complex stem can also receive a derivational suffix, as shown in the examples given under (18). However, [A N] expressions are not the only structures that display this possibility. As Anastassiadis-Symeonidis (1986: 140) mentions, some [N N_{GEN}] expressions (13) can also be input to a suffixation process, as shown in (19). The suffixes that take part in this derivational process are the adjectival suffix *-ικ(ός)* and the nominal suffixes *-ίτ(ης)* and *-ίστ(ας)*.

(18a)	ψυχροπολεμικός	←	ψυχρ(ός)	πόλεμ(ος)	[-ικός]$_{ADJ}$
	psichropolemikos		*psichr*(*os*)	*polem*(*os*)	*-ikos*
	cold war		cold	war	
(18b)	τριτοκοσμικός	←	τρίτ(ος)	κόσμ(ος)	[-ικός]$_{ADJ}$
	tritokosmikos		*trit*(*os*)	*kosm*(*os*)	*-ikos*
	third world		third	world	
(18c)	μαυραγορίτης	←	μαύρ(η)	αγορ(ά)	[-ίτης]$_{NOUN}$
	mavragoritis		*mavr*(*i*)	*agor*(*a*)	*-itis*
	black marketer		black	market	

(19a)	ταγματασφαλίτης	←	τάγματ(α)	ασφαλεί(ας)	[-ίτης]$_{NOUN}$
	tagmatasfalitis		*tagmata*	*asfalias*	*-itis*
	security battalion member		battalions	safety	

(19b)	αρσιβαρίστας	←	άρσ(η)	βαρ(ών)	[-ίστας]$_{NOUN}$
	arsivaristas		*ars*(*i*)	*var*(*on*)	*-istas*
	weightlifter		lift	weight	

The possibility to become input to derivation is not applicable to all [A N] or [N N$_{GEN}$] expressions. *Μεγάλη οθόνη* ('cinema', (12d)) or *αγορά εργασίας* ('job market', (13a)), for instance, cannot be input to any derivation process. Although there are no certain criteria determining which structures can participate to further derivation processes, it can be argued that these structures share more morphological features with typical compounds.

[N N$_{App.}$] structures are different from the other types of multi-word expressions with regard to headedness. Particularly, the two components share the same formal and semantic properties and thus the head properties as well. Since the two components display the same lexical category, it is possible to reverse their order, as shown in (20).

(20a)	μεταφραστής διερμηνέας	/	διερμηνέας μεταφραστής
	metafrastis diermineas		*diermineas metafrastis*
	translator-interpreter	/	interpreter-translator
(20b)	σκηνοθέτης παραγωγός	/	παραγωγός σκηνοθέτης
	skinothetis paragogos		*paragogos skinothetis*
	director-producer	/	producer-director

Coordinative compounds in Greek that are possible in all major lexical categories (21) do not usually display this possibility except for very few [A A] compounds, such as (21b) (cf. Ralli 2007: 99; Koliopoulou 2013: 301).

(21a)	αλατοπίπερο	←	αλάτι$_{N}$	πιπέρι$_{N}$
	alatopipero		*alati*	*piperi*
	salt and pepper		salt	pepper
(21b)	μαυρόασπρος/ασπρόμαυρος	←	μαύρος$_{A}$	άσπρος$_{A}$
	mavroaspros/*aspromavros*		*mavros*	*aspros*
	black and white		black	white
(21c)	πηγαινοέρχομαι	←	πηγαίνω$_{V}$	έρχομαι$_{V}$
	pigenoerchome		*pigeno*	*erchome*
	come and go		go	come

Despite the fact that the order of the [N N$_{App.}$] components is more easily reversible, this possibility affects in some degree the meaning of the structure (Anastassiadis-Symeonidis 1986: 191f.; Ralli 2013a: 256). Specifically, the first member

bears a more prominent semantic role than the second one. Therefore, the meaning of the expression changes slightly in case the order of the constituents is reversed.

Moreover, coordinative compounds and [N $N_{App.}$] structures are not directly comparable, although some scholars treat them in this way (cf. Olsen 2001; Bisetto/Scalise 2005). In many studies it has been argued that [N $N_{App.}$] expressions display different characteristics in comparison to coordinative compounds (cf. Wälchli 2005: 7; Bauer 2008: 4; Gaeta/Ricca 2009: 50; Manolessou/Tsolakidis 2009: 30). In the case of Greek, there is a clear demarcation between the two types of formation since coordinative compounds constitute one phonological and morphological word (21). In contrast, [N $N_{App.}$] expressions consist of two phonologically and morphologically independent words. Moreover, coordinative compounds are not characterized by an appositional relation between the components. The most common type of semantic relation found in Greek coordinative compounds is the additive one (Ralli 2007: 80f., 98, 2013a: 163; Koliopoulou 2013: 297ff.).

Since multi-word expressions always consist of two inflected words, they do not display the morphological properties of one-word compounds (cf. Anastassiadis-Symeonidis 1986: 149, 174, 196). Particularly, the inflected components of [A N] expressions agree in gender, case and number, as shown in (22), like regular syntactic phrases.

(22a)	ψυχρός$_{Masc/Nom/Sg}$	πόλεμος$_{Masc/Nom/Sg}$	‘cold war’
	psichros	*polemos*	
(22b)	ψυχροί$_{Masc/Nom/Pl}$	πόλεμοι$_{Masc/Nom/Pl}$	
	psichri	*polemi*	
(22c)	ψυχρού$_{Masc/Gen/Sg}$	πολέμου$_{Masc/Gen/Sg}$	
	psichru	*polemu*	
(22d)	ψυχρών$_{Masc/Gen/Pl}$	πολέμων$_{Masc/Gen/Pl}$	
	psichron	*polemon*	

Similar characteristics of agreement in gender, case and number are also displayed by [N $N_{App.}$] expressions (cf. (15)), as illustrated below.

(23a)	μεταφραστής$_{Masc/Nom/Sg}$	διερμηνέας$_{Masc/Nom/Sg}$	‘translator-interpreter’
	metafrastis	*diermineas*	
(23b)	μεταφραστή$_{Masc/Gen/Sg}$	διερμηνέα$_{Masc/Gen/Sg}$	
	metafrasti	*dierminea*	
(23c)	μεταφραστές$_{Masc/Nom/Pl}$	διερμηνείς$_{Masc/Nom/Pl}$	
	metafrastes	*dierminis*	

(23d) μεταφραστών$_{\text{Masc/Gen/Pl}}$ διερμηνέων$_{\text{Masc/Gen/Pl}}$
metafraston *diermineon*

[N N$_{\text{GEN}}$] expressions show inflectional properties similar to syntactic phrases, like *πόρτα σπιτιού* ((16b), 'house door$_{\text{GEN}}$'). Specifically the first noun can be independently inflected, while the second one always appears in genitive case, triggered by the first noun, the head of the structure (cf. Koliopoulou 2012: 866), as shown below.

(24a) κρέμα$_{\text{Fem/Nom/Sg}}$ ημέρας$_{\text{Fem/Gen/Sg}}$ 'day creme'
krema *imeras*
(24b) κρέμες$_{\text{Fem/Nom/Pl}}$ ημέρας$_{\text{Fem/Gen/Sg}}$
kremes *imeras*
(24c) κρέμας$_{\text{Fem/Gen/Sg}}$ ημέρας$_{\text{Fem/Gen/Sg}}$
kremas *imeras*
(24d) κρεμών$_{\text{Fem/Gen/Pl}}$ ημέρας$_{\text{Fem/Gen/Sg}}$
kremon *imeras*

Moreover, the genitive case of the non-head is always singular regardless of the number value of the head, as presented in (24b) and (24d). The inflectional properties of the non-head are less variable than the inflectional properties of the non-head of equivalent regular phrases. Specifically, both constituents of a syntactic phrase can be variably inflected regarding the features of number, as illustrated in (25).

(25a) πόρτες $_{\text{Fem/Nom/Pl}}$ σπιτιού$_{\text{Neu/Gen/Sg}}$ 'house doors (of one house)'
portes *spitiu*
(25b) πόρτες $_{\text{Fem/Nom/Pl}}$ σπιτιών$_{\text{Neu/Gen/Pl}}$ 'house doors (of many houses)'
portes *spition*

[N N$_{\text{Attr.}}$] expressions display inflectional properties different from syntactic phrases. Despite the fact that the non-head displays a certain degree of inflectional autonomy, there are some restrictions with regard to the features of plural number and genitive case (cf. Koliopoulou 2009: 67, 2012: 866; Ralli 2013a: 254), as shown below.

(26a) λέξη$_{\text{Fem/Nom/Sg}}$ κλειδί$_{\text{Neu/Nom/Sg}}$ 'key word'
leksi *klidi*

(26b)	λέξεις$_{\text{Fem/Nom/Pl}}$	κλειδιά$_{\text{Neu/Nom/Pl}}$
	leksis	*klidia*
	?λέξεις$_{\text{Fem/Nom/Pl}}$	κλειδί$_{\text{Neu/Nom/Sg}}$
	leksis	*klidi*
(26c)	λέξης$_{\text{Fem/Gen/Sg}}$	κλειδί$_{\text{Neu/Nom/Sg}}$
	leksis	*klidi*
	*λέξης$_{\text{Fem/Gen/Sg}}$	κλειδιού$_{\text{Neu/Gen/Sg}}$
	leksis	*klidiu*
(26d)	λέξεων$_{\text{Fem/Gen/Pl}}$	κλειδί$_{\text{Neu/Nom/Sg}}$
	lekseon	*klidi*
	λέξεων$_{\text{Fem/Gen/Pl}}$	κλειδιά$_{\text{Neu/Nom/Pl}}$
	lekseon	*klidia*
	?λέξεων$_{\text{Fem/Gen/Pl}}$	κλειδιών$_{\text{Neu/Gen/Pl}}$
	lekseon	*klidion*

Interestingly, comparing two example of the same type of expression, *λέξη κλειδί* ('key word', (14a), (26)) and *νόμος πλαίσιο* ('frame law', (14b), (27)), it becomes obvious that not all examples have the same inflectional properties in comparable contexts (cf. Koliopoulou 2009: 67f., 2012: 866f., Ralli 2013a: 254f.). Specifically, the non-head of the expression *νόμος πλαίσιο* displays a higher degree of inflectional autonomy in comparison to the inflectional variation displayed by the non-head of the example *λέξη κλειδί* (cf. (26b)–(27b), (26d)–(27d)). Moreover, with regard to the features of plural and genitive case there are different grammaticality judgements among native speakers.

(27a)	νόμος$_{\text{Masc/Nom/Sg}}$	πλαίσιο$_{\text{Neu/Nom/Sg}}$	'frame law'
	nomos	*plesio*	
(27b)	?νόμοι$_{\text{Masc/Nom/Pl}}$	πλαίσια$_{\text{Neu/Nom/Pl}}$	
	nomi	*plesia*	
	νόμοι$_{\text{Masc/Nom/Pl}}$	πλαίσιο$_{\text{Neu/Nom/Pl}}$	
	nomi	*plesio*	
(27c)	?νόμου$_{\text{Masc/Gen/Sg}}$	πλαίσιο$_{\text{Neu/Nom/Sg}}$	
	nomu	*plesio*	
	νόμου$_{\text{Masc/Gen/Sg}}$	πλαισίου$_{\text{Neu/Gen/Sg}}$	
	nomu	*plesiu*	
(27d)	νόμων$_{\text{Masc/Gen/Pl}}$	πλαίσιο$_{\text{Neu/Nom/Sg}}$	
	nomon	*plesio*	
	*νόμων$_{\text{Masc/Gen/Pl}}$	πλαίσια$_{\text{Neu/Nom/Pl}}$	
	nomon	*plesia*	

*νόμων$_{\text{Masc/Gen/Pl}}$	πλαισίου$_{\text{Neu/Gen/Sg}}$
nomon	*plesiu*
*νόμων$_{\text{Masc/Gen/Pl}}$	πλαισίων$_{\text{Neu/Gen/Pl}}$
nomon	*plesion*

Regarding the variation in behavior of this type of multi-word expression, it has been argued that they are in a process of desyntacticization, passing from the status of syntactic phrases to that of intermediate structures, i.e. to the status of formations displaying morphosyntactic features (cf. Ralli 2007: 247ff., 2013a: 255; Koliopoulou 2012: 867). However, after more careful consideration, the only safe claim that can be made is that these expressions have not yet acquired a stable status and that their inflectional properties vary among the different instances of this type and among speakers. They are indeed in a transitional stage, although it is not clear if these expressions gradually gain more syntactic autonomy or if they tend to lose their syntactic status.

2.2.1 Syntactic fixedness

Despite the fact that multi-word expressions share basic properties with regular syntactic phrases, they share many properties with typical compounds as well. Specifically, all four types of multi-word expression in Greek display a certain degree of syntactic fixedness. Some expressions are more restricted than others with regard to the degree of access to syntactic operations, as illustrated by the result of applying a number of tests concerning their internal properties i.e. their degree of lexical integrity (Anderson 1992: 84). Their mixed morpho-syntactic properties have been studied in detail (Anastassiadis-Symeonidis 1986, 1994, 1996; Ralli 1991, 1992, 2005, 2007, 2013a, 2013b; Christofidou 1997; Ralli/Stavrou 1998; Koliopoulou 2006: 43–56, 2008, 2009, 2012; Bağriaçik/Ralli 2015; ten Hacken/Koliopoulou 2016). In most of these studies, the degree of lexical integrity of the multi-word expressions has been analyzed on the basis of diagnostic tests exploring how many properties they share with regular syntactic formations.

In the following, I use the tests applied to typical compounds (cf. (6)–(8)) in the previous section in order to determine the degree of syntactic fixedness displayed by the different types of multi-word expressions found in Greek (cf. (12)–(15)). Moreover, I use an additional test regarding the possibility of adjective-noun syntactic phrases to double the definite article for emphatic reasons, which is only applicable to [A N] expressions. I summarize the tests under (28):

(28a) independent modification of the non-head

(28b) coordination of the components
(28c) reversion of the word order
(28d) doubling of the definite article of [A N] structures

In (29), I apply the above tests contrastively to the [A N] expression *μεγάλη οθόνη* ('cinema', (12d)) as well as to the corresponding syntactic phrase *μεγάλη οθόνη* ('big screen'). The examples chosen for the contrastive analysis consist of the same constituents. However, the difference between them is clear since the [A N] expression denotes the cinema, whereas the meaning of the syntactic phrase is fully compositional, denoting a big screen.

(29)	[A N] expression		[A N] phrase
(29a)	*πολύ μεγάλη οθόνη *poli megali othoni* lit. very big screen	(29a′)	πολύ μεγάλη οθόνη *poli megali othoni*
(29b)	*μεγάλη και φωτεινή οθόνη *megali ke fotini othoni* lit. big and bright screen	(29b′)	μεγάλη και φωτεινή οθόνη *megali ke fotini othoni*
(29c)	… σε μια *οθόνη μεγάλη *… se mia othoni megali* lit. … in a screen big	(29c′)	… σε μια οθόνη μεγάλη *… se mia othoni megali*
(29d)	*η οθόνη η μεγάλη *i othoni i megali* lit. the screen the big	(29d′)	η οθόνη η μεγάλη *i othoni i megali*

It is obvious from the negative response of the [A N] expression to all diagnostic tests that the structure displays a certain degree of lexical autonomy, contrary to the corresponding syntactic phrase, which allows access of all syntactic operations to its structure.

In (30), I test the structural properties of the [N N_{GEN}] expressions by applying the tests (28a–c). Specifically, I take as an example the expression *αγορά εργασίας* ('job market', (13a)) contrastively to the syntactic phrase *αναζήτηση εργασίας* ('job search') which bears the same non-head (cf. Koliopoulou 2009: 63; Ralli 2013a: 248).

(30)	[N N_{GEN}] expression		[N N_{GEN}] phrase
(30a)	*αγορά μόνιμης εργασίας *agora monimis ergasias* lit. market permanent$_{Gen}$ job$_{Gen}$	(30a′)	αναζήτηση μόνιμης εργασίας *anazitisi monimis ergasias* search permanent$_{Gen}$ job$_{Gen}$

(30b) *αγορά εργασίας και απασχόλησης
agora ergasias ke apascholisis
lit. market job$_{Gen}$ and occupation$_{Gen}$

(30b′) αναζήτηση εργασίας και απασχόλησης
anazitisi ergasias ke apascholisis
search job$_{Gen}$ and occupation$_{Gen}$

(30c) *εργασίας αγορά
ergasias agora
lit. job$_{Gen}$ market

(30c′) εργασίας αναζήτηση
ergasias anazitisi
job$_{Gen}$ search

The negative response of the [N N$_{GEN}$] expression to the applied test reveals a degree of syntactic fixedness similar to that of the [A N] expressions.

The reaction of [N N$_{Attr.}$] expressions to the same tests is not different from that of the structures tested above, as illustrated in the following on the basis of the example *φόρος φωτιά* ('very high tax', (14c)).

(31a) *φόρος μεγάλη φωτιά
foros megali fotia
tax big fire

(31b) *φόρος φωτιά και καπνός
foros fotia ke kapnos
tax fire and smoke

(31c) *φωτιά φόρος
fotia foros
fire tax

With regard to the possibility of reversing the order of the constituents, most of the examples belonging to this type of expression have a negative response, proven by (31c) as well as by (32a′–c′). However, there are a few exceptions, e.g. (32d′), in which the inversion of the two constituents is allowed (Koliopoulou 2006: 52, 2009: 66), since in this way the property designated by the non-head can be highlighted.[10]

(32a) λέξη κλειδί
leksi klidi
lit. word key

(32a′) *κλειδί λέξη
klidi leksi
key word

10 By contrast, Anastassiadis-Symeonidis (1986: 197) mentions no exception regarding the possibility of reversing the order of the constituents.

(32b) νόμος πλαίσιο
nomos plesio
lit. law frame

(32b′) *πλαίσιο νόμος
plesio nomos
frame law

(32c) γυναίκα αράχνη
gineka arachni
lit. woman spider

(32c′) *αράχνη γυναίκα
arachni gineka
spider woman

(32d) εταιρία μαϊμού
eteria maimu
lit. company monkey
'fake company'

(32d′) μαϊμού εταιρία
maimu eteria
monkey company

The frequency of use or the degree of semantic compositionality (cf. Fellbaum 2011) are possible parameters that influence the varying degree of syntactic fixedness determining which structure may be characterized by a free word order.

The last type of multi-word expressions displays an appositional relation between the constituents. As already mentioned in the previous section, these expressions are double-headed. Therefore, the tests listed under (31) are almost inapplicable. Specifically, the application of the tests regarding the coordination of compounds (31b) as well as reversing of the order (31c) would not make much sense, since appositional structures are recursive and can be coordinated with further constituents attached to any of two members. Moreover, the order of the constituents is reversible (cf. (20)), since the structures are double-headed, despite the semantic restrictions.

I consider test (28a) in that I check the possibility of independent modification of one of the two constituents, although none of them is a non-head (33a). Moreover, I apply a further diagnostic test in order to investigate the degree of lexical integrity in their internal structure of the [N $N_{App.}$] expressions. Specifically, in (33b) I test the possibility of insertion of an uninflected adjective (*πρώην* 'former'), while in (33c) I test the possibility of insertion of a parenthetical element between the constituents.

(33a) *μεταφραστής ικανός διερμηνέας
metafrastis ikanos diermineas
lit. translator capable interpreter

(33b) ?μεταφραστής πρώην διερμηνέας
metafrastis proin diermineas
lit. translator former interpreter

(33c) ?ο μεταφραστής, όπως βλέπετε, διερμηνέας είναι ...
o metafrastis, *opos vlepete*, *diermineas ine* ...
lit. the translator, as you see, interpreter is ...

As illustrated above, the independent modification of one of the two members by a qualifying adjective is not possible, cf. (33a). However, this type of expression displays a limited degree of syntactic fixedness in comparison to the other types of multi-word expressions, since an element can intervene in their internal structure, as shown in (33b) and (33c).

2.2.2 Summary

In (34), I summarize the main points of the analysis of the four types of multi-word expression found in Greek regarding their degree of syntactic fixedness:

(34a) [A N] and [N N_{GEN}] expressions look like syntactic phrases and are inflected as such. However, both their inflectional properties as well as their behavior on the diagnostic tests show a certain degree of lexical integrity. Specifically, they share the most morphological characteristics with typical compounds compared to the other types of expressions. Moreover, they can be input to a suffixation process. Although both types of expressions are rather rigid with regard to their morphosyntactic features, not all instances may take part in a suffixation process.

(34b) [N $N_{Attr.}$] expressions display a rather unclear status. Not only their inflectional properties but also their response to the tests of syntactic fixedness varies among the different instances.

(34c) [N $N_{App.}$] expressions constitute a borderline case among multi-word expressions in Greek. Not only with regard to their inflectional properties but also with regard to their behavior in the diagnostic tests, they show the lowest degree of syntactic fixedness among all types of expressions considered in this study. However, they still show some signs of lexical autonomy, according to which their classification as multi-word expressions is justified.

2.3 Phrase-like structures

Although there is a clear distinction between typical compounds and multi-word expressions in Greek, the variety of structures sharing properties with compounds as well as with syntactic phrases creates a certain difficulty in differentiating them from each other and classifying them into distinctive types. Specifically, it has been argued that there are further types of [A N] and [N N_{GEN}] formations which can be classified neither as multi-word expressions nor as regular syntactic

phrases (Anastassiadis-Symeonidis 1986; Ralli/Stavrou 1998; Ralli 2005, 2007, 2013a: 257ff.; Koliopoulou 2006: 21ff., 36f., 2012: 863f.). In order to designate this extra type of intermediate structure, Koliopoulou (2012) uses the term "special noun phrases", while Ralli (2013a) prefers the term "constructs".

The argument that they differ from [A N] and [N N_{GEN}] multi-word expressions, although they display the same structure, is based on the observation that the two members display a special syntactic relation. Specifically, formations of the [A N] type consist of a relational (35a, b) or classifying adjective (35c, d). In [N N_{GEN}] the right-handed noun, i.e. the non-head, has the role of a head argument (36).

(35)	[A N]			
(35a)	θεατρική κριτική	←	θεατρική	κριτική
	theatriki kritiki		*theatriki*	*kritiki*
	theater review		theatrical	criticism/review
(35b)	βιομηχανική ζώνη	←	βιομηχανική	ζώνη
	viomichaniki zoni		*viomichaniki*	*zoni*
	industrial zone		industrial	zone
(35c)	πυρηνική δοκιμή	←	πυρηνική	δοκιμή
	piriniki dokimi		*piriniki*	*dokimi*
	nuclear testing		nuclear	testing
(35d)	ψηφιακό κύκλωμα	←	ψηφιακό	κύκλωμα
	psifiako kikloma		*psifiako*	*kikloma*
	digital circuit		digital	circuit

(36)	[N N_{GEN}]			
(36a)	επεξεργασία δεδομένων	←	επεξεργασία	δεδομένων$_{GEN}$
	epeksergasia dedomenon		*epeksergasia*	*dedomenon*
	data processing		processing	data
(36b)	εκπομπή αερίων	←	εκπομπή	αερίων$_{GEN}$
	ekpompi aerion		*ekpompi*	*aerion*
	gas emission		emission	gases

However, as it is obvious from the examples above, both types of structure display a certain degree of semantic opacity, like the corresponding types of multi-word expression.

According to their response to the diagnostic test of syntactic atomicity, both structures can be subjects to syntactic operations. Specifically, in (37) I consider the application of the tests (28a–d) to the [A N] structure *βιομηχανική ζώνη* (35b) and in (38) I apply the tests (28a–c) to the [N N_{GEN}] example *εκπομπή αερίων*, cf. (36b).

(37a) έντονα βιομηχανική ζώνη
entona viomichaniki zoni
lit. intensive industrial zone
(37b) βιομηχανική και μολυσμένη ζώνη
viomichaniki ke molismeni zoni
lit. industrial and polluted zone
(37c) ζώνη βιομηχανική
zoni viomichaniki
lit. zone industrial
(37d) η βιομηχανική η ζώνη
i viomichaniki i zoni
the industrial the zone

(38a) εκπομπή βλαβερών αερίων
ekpompi vlaveron aerion
lit. emission harmful gases
(38b) εκπομπή αερίων και θερμότητας
ekpompi aerion ke thermotitas
lit. emission gases and heat
(38c) ?αερίων εκπομπή
aerion ekpompi
lit. gases emission

It becomes clear from the above tests that syntactic operations have access to their internal structure, contrary to the [A N] and [N N_{GEN}] multi-word expressions which display a certain degree of lexical integrity, as shown in (29) and (30).

However, due to the argument structure displayed by these structures it can be argued that they are of a different nature from common syntactic phrases. Particularly, their structure resembles that of compounds consisting of a relational adjective and a noun (ten Hacken 1994: 89–98; Bisetto 2010: 65–85). Moreover, they constitute naming units which also supports the view that they are of a different nature than common syntactic phrases. Therefore, they constitute a further type of complex lexical units, which on the one hand differs from regular syntactic phrases, and on the other cannot be classified as belonging to the set of the multi-word expressions analyzed above. Moreover, they display more syntactic properties than the multi-word expressions. Thus, their demarcation from syntactic phrases is a rather difficult task, since it is only based on a few minor distinctive characteristics and not on their response with regard to the diagnostic tests of syntactic fixedness.

3 Complementation vs. competition

I have argued above for a distinction between three types of complex lexical units in Greek: one-word compounds, multi-word expressions and phrase-like structures. All three constitute nominal structures sharing a function, i.e. to name concepts, particularly complex concepts. Regarding their function, they are clearly different from syntactic phrases, which describe a concept but do not name it. Since they provide further means for naming concepts associated with various terminological areas, the set of naming devices in the nominal domain of the lexicon is extended through their existence. In this sense, the three types of complex lexical units constitute complementary resources of nominal naming units.

Complementation in the lexicon with regard to different naming strategies does not exclude competition among structures. Specifically, typical one-word compounds and multi-word lexical units do not exist in Greek side by side, although this scenario cannot be excluded for all languages. Take for instance lexical units in German (cf. ten Hacken/Koliopoulou 2016), like *grüner Tee* and *Grüntee* ('green tee') or *schwarzer Markt* and *Schwarzmarkt* ('black market'), coexisting synchronically. Their coexistence is explained by Hüning/Schlücker (2015: 459) on the grounds of stylistic variation and/or diachronic change arguing that the structure *schwarzer Markt*, for instance, has been gradually replaced by the compound *Schwarzmarkt*, which is synchronically more frequent than the equivalent phrase.

In Greek, the three types of complex lexical units compete with each other. However, there is no evidence supporting the existence of a blocking mechanism (cf. Rainer 2016), although the formation of typical compounds is much more productive and regular than the formation of multi-word structures. Moreover, I claim that the selection of a possible naming strategy depends on the characteristics of the concept. Specifically, a borrowed nominal concept or a complex concept meant for terminological use is a possible candidate for a type of multi-word lexical unit.

4 Complex lexical units in lexicon and grammar

In Greek, there is a clear demarcation between compounds and other complex lexical units, i.e. multi-word expressions and phrase-like structures. Among these, compounds are the only type of complex lexical units built in morphology. Taking into consideration the various types of multi-word formations and in some cases their variable features, the question arises how they can be accounted for.

They are neither morphological structures nor regular syntactic phrases; they are rather situated in between. Therefore, multi-word expressions in Greek have been often assigned to a continuum situated between the two components. In this sense, different grammatical models supporting the interaction between the two domains (Kiparsky 1982; Bybee 1985; Borer 1988) have been adopted by many scholars as the most sufficient way to deal with multi-word expressions and their variable features in Greek (cf. Ralli 1991, 1992, 2007: 245f.; Ralli/Stavrou 1998; Koliopoulou 2009: 69, 2012: 868).

In a similar context, Ralli (2013a: 261f., 266ff., 2013b: 183f., 194), based on Borer's (2009) analysis of comparable nominal constructs in Hebrew, argues that multi-word expressions in Greek are derived within the syntactic domain which interacts with morphology. Her argument is rather justified, since multi-word expressions and phrase-like structures in Greek look like syntactic phrases that consist of two phonologically and morphologically independent words. However, they are different from regular syntactic phrases, since their structure is not accessible to all syntactic operations. Moreover, they display a certain degree of lexical integrity coinciding in many cases with a non-compositional meaning, also displayed by typical compounds.

The fact that there is strong variation among the different types of multi-word structures with regard to their mixed morphosyntactic properties supports the view that there is no clear borderline between morphology and syntax and that the two domains are situated on a continuum.[11] Multi-word expressions in Greek which display a varying degree of structural visibility to syntactic operations occupy different positions on this continuum. [A N] and [N N_{GEN}] expressions in Greek are clearly nearer to the morphological domain, i.e. to typical compounds, than any other multi-word expression. The fact that some [A N] and [N N_{GEN}] formations can be input to a derivational process is a further argument in favor of the interaction between morphology and syntax, since structures generated in syntax are turned into one complex stem in order to undergo a morphological operation (cf. (18)–(19)). The other two types of nominal expressions are widespread on the continuum, specifically between the [A N] and [N N_{GEN}] formations and regular syntactic phrases. Phrase-like structures are situated near to the syntactic domain.

The various approaches that argue in favor of the existence of a continuum between the two grammatical components or the interaction among them are based on the assumption that the two grammatical domains are distinct. Although they may account for structures like multi-word expressions in Greek displaying

11 On the closeness of compounding to the syntactic domain cf. Koliopoulou (2014b).

mixed morphosyntactic properties, they do not throw any light on the grey zone between morphology and syntax. In this respect, the question arises whether the two grammatical domains are actually distinct and if not what kind of demarcation would allow us to differentiate between typical morphological structures, syntactic phrases and intermediate structures.

In order to distinguish compounds from phrases as well as from the in-between formations, Gaeta/Ricca (2009: 38f.) propose another type of demarcation. They argue in favor of a four-scaled classification based on two criteria: a) 'morphological', i.e. the output of a morphological operation and b) 'lexicalized', i.e. attributed to the lexicon taking into consideration not only idiosyncrasy but also token frequency and/or naming force. In this respect, typical compounds are characterized as [+ morphological] and [+ lexical], whereas syntactic phrases have a negative sign in both properties. Multi-word expressions – or phrase-like units in Gaeta/Ricca's terminology – are non-morphological but lexical units. In this view being a lexical unit is independent from being an output of a morphological operation.

On a similar basis, ten Hacken/Koliopoulou (2016: 134ff.), dealing with [A N] multi-word expressions in various languages, argue that the main criterion to demarcate [A N] intermediate structures from adjective-noun syntactic phrases is related to the function of these structures. Structures constituting a naming unit are lexical units, while descriptive phrases belong to the syntactic domain.

With regard to Greek, the different types of multi-word expressions and phrase-like structures, despite their varying morphosyntactic features, sometimes even within the same type, share the naming function (cf. Anastassiadis-Symeonidis 1986: 142f.). They are lexical units with a rule-based formation extending the naming device of the lexicon. This extended view of the lexicon is also supported by approaches such as the Parallel Architecture (cf. Jackendoff 2010) and Construction Morphology (cf. Booij 2010, this volume) on the basis of comparable multi-word, intermediate structures.

5 Conclusions

The demarcation between the various types of complex lexical units is primarily a language specific matter, although most of the criteria used to differentiate morphological from syntactic structures apply at an abstract level to all languages. It actually depends on the particular characteristics of wordhood and compoundhood, as displayed in each language. These two basic characteristics determine the morphological structures and the lexicon. The degree of resemblance between

typical morphological structures and other complex lexical units specifies the form of the lexicon in a particular language and the possibility of interaction between the grammatical domains.

Multi-word expressions and phrase-like structures in Greek are clearly distinct from typical compounds: their constituents are phonologically and morphologically independent words, a linking element is not required, they display head properties similar to syntactic phrases as well as internal inflection. In Greek, the degree of syntactic fixedness depends on the type of expression one deals with. Sometimes, there is variation of the syntactic characteristics even among the different examples of the same type of structure (cf. (26)–(27), (32)). Despite the fact that multi-word expressions and phrase-like structures in Greek cannot be assigned to morphology like typical compounds, all three types of complex lexical units share the same function, i.e. the naming function. They are generated by different lexical unit formation patterns which extend the naming strategies of the lexicon. The outcome of this formation process is lexical units stored in the speakers' mental lexicon.

Compounding in Greek is a very productive process and thus a main language naming device. However, new concepts have been introduced to the language in the last decades through the form of a multi-word expression or a phrase-like structure mostly found in specialized or newspaper texts. The appearance of such a lexical unit is an indication for native speakers of the terminological use of the concept. The emergence of various types of lexical units other than compounds shows a clear tendency to different types of naming units and indicates a silent process of language change regarding the naming of concepts, especially those borrowed from English.

References

Anastassiadis-Symeonidis, Anna (1986): Η Νεολογία στην Κοινή Νεοελληνική [Neology in standard Modern Greek]. Thessaloniki: Aristotle University of Thessaloniki.

Anastassiadis-Symeonidis, Anna (1994): Νεολογικός δανεισμός της Νεοελληνικής [Neological borrowing in modern Greek]. Thessaloniki: Self publishing.

Anastassiadis-Symeonidis, Anna (1996): Η νεοελληνική σύνθεση [Modern Greek compounding]. In: Katsimali, Georgia/Kavoukopoulos, Fotis (eds.): Ζητήματα νεοελληνικής γλώσσας: Διδακτική Προσέγγιση [Themes of the Modern Greek language: A didactic approach]. Rethymno: University of Crete. 97–120.

Anderson, Stephen R. (1992): A-morphous morphology. (= Cambridge Studies in Linguistics 62). Cambridge, UK: Cambridge University Press.

Bağrıaçık, Metin/Ralli, Angela (2015): Phrasal vs. morphological compounds: Insights from Modern Greek and Turkish. In: Language Typology and Universals (STUF) 68. 323–357.

Bauer, Laurie (2001): Compounding. In: Haspelmath, Martin et al. (eds.): Language typology and language universals. An international handbook. Vol. 1. Berlin/New York: De Gruyter. 695–707.
Bauer, Laurie (2008): Dvandva. In: Word Structure 1, 1. 1–20.
Bauer, Laurie (2009): Typology of compounds. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 343–356.
Becker, Thomas (1992): Compounding in German. In: Rivista di Linguistica 4, 1. 5–36.
Bisetto, Antonietta (2010): Relational adjectives crosslinguistically. In: Lingue e Linguaggio 9, 1. 65–85.
Bisetto, Antonietta/Scalise, Sergio (1999): Compounding: Morphology and/or syntax? In: Mereu, Lunella (ed.): Boundaries of morphology and syntax. (= Current Issues in Linguistic Theory 180). Amsterdam: Benjamins. 31–48.
Bisetto, Antonietta/Scalise, Sergio (2005): The classification of compounds. In: Lingue e Linguaggio 4, 2. 319–332.
Booij, Geert (2009): Compounding and construction morphology. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 201–216.
Booij, Geert (2010): Construction morphology. Oxford: Oxford University Press.
Borer, Hagit (1988): On the morphological parallelism between compounds and constructs. In: Booij, Geert/van Marle, Jaap (eds.): Yearbook of morphology. Dordrecht: Foris. 45–65.
Borer, Hagit (2009): Afro-Asiatic, Semitic: Hebrew. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 491–511.
Bybee, Joan (1985): Morphology: A study of the relation between meaning and form. (= Typological Studies in Language 9). Amsterdam: Benjamins.
Christofidou, Anastasia (1997): A textlinguistic approach to the phenomenon of multi-word compounds. In: Drachman, Gaberell et al. (eds.): Greek linguistics 95. Proceedings of the 2nd International Conference on Greek Linguistics. Graz: Neugebauer. 67–75.
Donalies, Elke (2004): Grammatik des Deutschen im europäischen Vergleich. Kombinatorische Begriffsbildung. Vol. 1: Substantivkomposition. Mannheim: Institut für Deutsche Sprache.
Fellbaum, Christiane (2011): Idioms and collocations. In: Maienborn, Claudia/von Heusinger, Klaus/Portner, Paul (eds.): Semantics. An international handbook of natural language meaning. Vol. 1. Berlin/New York: De Gruyter. 441–456.
Gaeta, Livio/Ricca, Davide (2009): Composita solvantur: Compounds as lexical units or morphological objects? In: Rivista di Linguistica 21, 1. 45–68.
Günther, Harmut (1997): Zur grammatischen Basis der Getrennt-/Zusammenschreibung im Deutschen. In: Dürscheid, Christa/Ramers, Karl-Heinz/Schwarz, Monika (eds.): Sprache im Fokus. Festschrift für Heinz Vater zum 65. Geburtstag. Tübingen: Niemeyer. 3–16.
Hacken, Pius ten (1994): Defining morphology. A principled approach to determining the boundaries of compounding, derivation and inflection. Hildesheim: Olms.
Hacken, Pius ten/Koliopoulou, Maria (2016): Adjectival non-heads and the limits of compounding. In: SKASE Journal of Theoretical Linguistics 13, 2. 122–139.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.): Word-formation. An international handbook of the languages of Europe. Vol. 1. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK) 40.1). Berlin/Boston: De Gruyter. 450–467.
Jackendoff, Ray (2010): Meaning and the lexicon: The parallel architecture 1975–2010. Oxford: Oxford University Press.
Kiparsky, Paul (1982): Lexical morphology and phonology. In: The Linguistic Society of Korea: Linguistics in the morning calm. Selected papers from SICOL-1981. Seoul: Hanshin. 3–91.

Koliopoulou, Maria (2006): Περιγραφή και ανάλυση των χαλαρών πολυλεκτικών συνθέτων της Νέας Ελληνικής [Description and analysis of the loose multi-word compounds of Modern Greek]. Patras: University of Patras. M.A. Thesis. Internet: http://nemertes.lis.upatras.gr/jspui/handle/10889/914 (last access: 14.6.2018).

Koliopoulou, Maria (2008): The loose multi-word compounds of Modern Greek under the prism of construction morphology. In: Lavidas, Nikolaos/Nouchoutidou, Elissavet/Sionti, Marietta (eds.): New perspectives in Greek linguistics. Newcastle upon Tyne: Cambridge Scholars Publishing. 213–224.

Koliopoulou, Maria (2009): Loose multi-word compounds and noun constructs. In: Patras Working Papers in Linguistics 1. Special issue: Morphology. 59–71.

Koliopoulou, Maria (2012): Μεταξύ συνθέτων και φράσεων [Between compounds and phrases]. In: Gavriilidou, Zoe et al. (eds.): Selected papers of the 10th International Conference on Greek Linguistics. Komotini: Democritus University of Thrace. 861–869.

Koliopoulou, Maria (2013): Θέματα σύνθεσης της Ελληνικής και της Γερμανικής: συγκριτική προσέγγιση. [Issues of Modern Greek and German compounding: A contrastive approach]. Patras: University of Patras. PhD Dissertation. Internet: http://nemertes.lis.upatras.gr/jspui/handle/10889/5962?locale=en (last access: 14.6.2018).

Koliopoulou, Maria (2014a): Issues of Modern Greek and German compounding: a contrastive approach. In: Journal of Greek Linguistics 14, 1. 117–125.

Koliopoulou, Maria (2014b): How close to syntax are compounds? Evidence from the linking element in German and Modern Greek compounds. In: Rivista di Linguistica 26, 2. 51–70.

Koliopoulou, Maria (2014c): Komposition im Deutschen und Neugriechischen: Eine kontrastive morphologische Analyse. In: Katsaounis, Nikolaos/Sidiropoulou, Renate M. (eds.): Sprachen und Kulturen in (Inter)Aktion. Vol. 2: Linguistik, Didaktik, Translationswissenschaft. (= Hellenogermanica 2). Frankfurt a.M.: Lang. 43–55.

Koliopoulou, Maria (2015): Possessive/bahuvrīhi compounds in German: An analysis based on comparable compounds in Modern Greek. In: Languages in Contrast 15, 1. 81–101.

Koliopoulou, Maria (2017): What can word formation offer to translation practice? A case study of German compounds and their English equivalents. In: Zybatow, Lew N. et al. (eds.): Übersetzen und Dolmetschen: Berufsbilder, Arbeitsfelder, Ausbildung. Ein- und Ausblick in ein sich wandelndes Berufsfeld der Zukunft. (= Forum Translationswissenschaft 21). Frankfurt a.M.: Lang. 117–136.

Lieber, Rochelle/Štekauer, Pavol (eds.) (2009): The Oxford handbook of compounding. Oxford: Oxford University Press.

Manolessou, Io/Tsolakidis, Symeon (2009): Greek coordinated compounds: Synchrony and diachrony. In: Patras Working Papers in Linguistics 1. 23–39.

Mukai, Makiko (2013): Recursive compounds and linking morpheme. In: International Journal of English Linguistics 3, 4. 36–49.

Neef, Martin (2009): IE, Germanic: German. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 386–399.

Nespor, Marina/Ralli, Angela (1994): Stress domains in Greek compounds: A case of morphology-phonology interaction. In: Philippaki-Warburton, Irene/Nicolaidis, Katerina/Sifianou, Maria (eds.): Themes in Greek linguistics I. Amsterdam: Benjamins. 201–208.

Nespor, Marina/Ralli, Angela (1996): Morphology-phonology interface: Phonological domains in Greek compounds. In: The Linguistic Review 13, 3/4. 357–382.

Olsen, Susan (2001): Copulative compounds: A closer look at the interface between syntax and morphology. In: Booij, Geert/van Marle, Jaap (eds.): Yearbook of morphology 2000. Dordrecht: Springer. 279–320.
Rainer, Franz (2016): Blocking. In: Aronoff, Mark (ed.): Oxford research encyclopedia of linguistics. 1–22. Internet: http://dx.doi.org/10.1093/acrefore/9780199384655.013.33 (last access: 14.6.2018).
Ralli, Angela (1991): Λεξική φράση: Αντικείμενο μορφολογικού ενδιαφέροντος [Lexical phrase: A morphological analysis]. In: Μελέτες για την Ελληνική Γλώσσα [Studies in Greek Linguistics 1990]. 205–221.
Ralli, Angela (1992): Compounds in Modern Greek. In: Rivista di Linguistica 4, 1. 143–174.
Ralli, Angela (2005): Μορφολογία [Morphology]. Athina: Patakis.
Ralli, Angela (2007): Η σύνθεση λέξεων: διαγλωσσική μορφολογική προσέγγιση [Compounding: A cross-linguistic morphological approach]. Athina: Patakis.
Ralli, Angela (2013a): Compounding in Modern Greek. Dordrecht: Springer.
Ralli, Angela (2013b): Compounding and its locus of realization: Evidence from Greek and Turkish. In: Word Structure 6, 2. 181–200.
Ralli, Angela/Stavrou, Melita (1998): Morphology-syntax interface: A-N compounds vs. A-N constructs in Modern Greek. In: Booij, Gert/van Marle, Jaap (eds.): Yearbook of morphology 1997. Dordrecht: Springer. 243–263.
Scalise, Sergio (1992): Compounding in Italian. Rivista di Linguistica 4, 1. 175–200.
Schlücker, Barbara/Hüning, Matthias (2009): Compounds and phrases: A functional comparison between German A + N compounds and corresponding phrases. In: Rivista di Linguistica 21, 1. 209–234.
Wälchli, Bernhard (2005): Co-compounds and natural coordination. Oxford: Oxford University Press.

Ingeborg Ohnheiser †

Compounds and multi-word expressions in Russian

Introduction

This chapter deals with the discussion of the relation between multi-word expressions, compounds, and derivations in the description of Russian and other Slavic languages. Referring to pertinent publications, the aim is to show how these descriptions have been influenced by particular theoretical conceptions (e.g. the onomasiological view adopted by Dokulil 1962) and the respective grammatical tradition (e.g. Russkaja grammatika 1980, generally known as "Grammatika-80": Švedova 1980). New approaches to the description of the relation between phrases and derivatives as well as between phrases and a special type of Russian compounds (the so-called stump compounds) from the viewpoint of Construction Grammar are presented with reference to works by Benigni/Masini (2010) and Masini/Benigni (2012). In view of recent linguistic developments, the competition between multi-word expressions and N+N compounds is discussed, which persists irrespective of the increasing productivity of this compound type in Russian.

The chapter does not provide a comprehensive overview of all naming processes in Russian, but rather focuses – also from the perspective of research history – on those types of nominal multi-word expressions and compounds (as well as one derivational type) that stand in a mutual relation of cooperation and/or competition. Particular attention will be paid to stylistic and pragmatic aspects.

The chapter is organized as follows: Section 1 gives a brief overview of the main findings of previous studies on complex lexical units in Russian and other Slavic languages. Section 2 presents compound and MWE patterns in Russian as well as their interrelation as discussed in Grammatika-80. Sections 3 and 4 discuss the co-existence and interaction between MWEs and various morphological patterns. The chapter ends with a conclusion in Section 5.

1 Some remarks on the current state of research

The interaction of various naming procedures in Slavic languages has been discussed from different angles.

https://doi.org/10.1515/9783110632446-009

1.1 "Condensation" of complex naming units

Isačenko (1958), for instance, paid special attention to the formal and semantic condensation of complex naming units, stating that "complex designations consisting of several lexical units have a clear tendency towards univerbation, i.e. to the compression of the semantic content into one word" (ibid.: 340; translated from Russian). This phenomenon manifests itself in different naming procedures:

a) Certain types of compounding (e.g., Slovak *svet-o-názor* [world.LV *o*-view][1] 'world view' < *svetový názor* [world.RA view] 'id.'[2,3])
b) Mergers (Czech *pravdě-podobný* [truth.DAT-similar] 'probable')
c) Ellipsis
   1) of the head (Russian *prjamaja* 'straight line' < *prjamaja linija* 'id.')
   2) of the modifier (Russian *plastinka* 'record' < *grammofonnaja plastinka* '(grammophone) record')
d) Affixal derivation (Russian *setčat-k-a* 'retina' < *setčataja oboločka* [net.A membrane] 'id.')
e) Binominals (appositional compounds), particularly in Russian, e.g., *ženščina-vrač* [woman-doctor] 'female doctor'
f) Different types of compounds with a clipped modifier ("stump compounds" in the terminology of Comrie/Stone 1978 and Comrie/Stone/Polinsky 1996) (Russian *zarplata* 'salary'< *zarabotnaja plata* [for.work.RA payment] 'id.'), but also of initialisms and acronyms (Russian *IMLI* < *Institut mirovoj literatury* [institute world.RA.GEN literature.GEN] 'Institute of World Literature' [of the Russian Academy of Sciences]). According to Isačenko, the dominance of this latter type in Russian is not accidental as it provides an important option to condensate MWEs with modifiers in the genitive case.
g) Formations of the type Russian *Glavryba* [*glav-* clipped stem of the adjective *glavnyj* 'main, principal' + *ryba* 'fish'] < *Glavnoe upravlenie rybnoj promyšlennosti* – the name of the Soviet central administration of the fishing industry. As has been pointed out by one of the reviewers, from a semantic point of view, *Glavryba* reflects a metonymic shift, because the modifier does not directly modify the noun, but a concept connected to the noun ("central

1 LV: linking vowel

2 RA: relational adjective

3 In Czech, the MWE still exists next to the compound (*světový názor* and *světonázor*) as an obvious calque of the German *Weltanschauung* 'id.'.

administration of the fishing industry").[4] In this respect formations like *Glavryba* differ from stump compounds like *glavvrač* < *glavnyj vrač* 'head physician'.

1.2 MWEs, compounds, and derivations from an onomasiological point of view

The relationship between MWEs, compounds, and derivations was dealt with in Czech linguistics in the description of word formation as part of naming procedures (Dokulil 1962). Thus for example Czech MWEs, compounds, and suffixed compounds (1a–e) are contrasted with suffixed derivatives (1a'–e') (with the same meaning) (ibid.: 31):

(1a) *malíř krajin* [painter landscape.PL.GEN] 'landscape painter'
(1b) *hráč na housle* [player on violin] 'violin player'
(1c) *žák první třídy* [pupil first.GEN grade.GEN] 'first-grader'
(1d) *kov-o-dělník* 'metalworker'
(1e) *dřev-o-rub-ec* 'woodcutter'

(1a') *krajin-ář*
(1b') *housl-ista*
(1c') *prvň-ák*
(1d') *kov-ák*
(1e') *dřev-ař*

Dokulil (ibid.) uses examples from terminology and technical language to show that the formation of multi-word designations is a very common naming procedure, as in (2):

(2a) A+N *vysoké napětí* 'high voltage'
(2b) $N+N_{GEN}$ *stupnice tvrdosti* [scale hardness.GEN] 'hardness scale'

In spite of this, such procedures appear "cumbersome" in an inflectional language, which is why single word names are preferred in everyday speech. This can also be seen from the so-called univerbations which – according to the tradi-

4 This formation type can also be found in more recent designations, e.g., *Glavlinza* [main lens], a leader brand for contact lenses. (www.glavlinza.ru, last access: 1.3.2017).

tional interpretation of the term in Slavic studies – means the transformation of MWEs into suffixed one-word designations. According to Dokulil, an important criterion of "univerbized" designations is the coexistence of a synonymous multi-word designation (generally of the structure/form RA+N), which should be a real, i.e. a fixed (established) naming unit, but not any free combination of words, e.g.:

(3) *čajová růže*[5] [tea.RA rose] 'tea rose' > *čajovka* tea.RA-stem-SUFF 'id.'[6]

The word *stař-ik* 'old man' should, however, be regarded as a deadjectival suffixed formation < *starý* 'old' and not as univerbation of *starý člověk* 'old man'. A significant extension of the concept of univerbation in Slavic studies has been proposed in a new monograph on Slovak (Ološtiak (ed.) 2015). In this study, the criterion of stability of the underlying MWEs is maintained. The results, however, are not restricted to suffix formations. Some examples are provided by Ološtiak (ed.) (ibid.: 308ff.):

(4a) MWEs and "traditional" suffixal univerbations with truncation of the stem and ellipsis of the head, e.g., Slovak *izolač-n-á páska* 'insulating tape' > *izolač-k-a* 'id.'

(4b) Combination of compounding and univerbation ("*kompozičná univerbizácia*"), e.g., Slovak *hráč prvej ligy* [player first.GEN division.GEN] > *prv-o-lig-ist-a* 'first division player'. In Russian grammars, the analog formation *pervoligist* < *pervaja liga* is described as suffixed compound

(4c) Clipping of the modifier of an MWE and formation of a compound, e.g. Slovak *alkoholový test* > *alkotest* 'alcotest' (however, this formation might also be a direct loan from English)

(4d) Phenomena like the following are also included:
Slovak *kompaktný fotoapparát* > *kompakt* 'compact camera'

(4e) The formation of acronyms from MWEs is also often regarded as univerbation:
Slovak *Mestská hromadná doprava* > *MHD*; coll. suffixed *MHD-čka* 'local public transport'[7]

5 In botanical nomenclature N+RA *růže čaj-ová* [rose tea-RA] 'tea rose' with inverted word order.

6 Another formally identical word can be based on the MWE *čajový salám* 'tea sausage (spread)'.

7 The vernacular suffixation of acronyms is more productive in Slovak than in Russian.

For a concise overview of different approaches to univerbation in the Slavic national philologies cf. Martincová (2015).

Kuchař (1963) took up the question of a possible systematic relation among different naming processes. His starting point was the following idea: if word formation is considered as name formation with morphological means, then there might also be similar processes on the syntactic level (i.e. syntactically complex forms) and on the semantic level (i.e. semantic shifts). The aim would then be to discover the common as well as the specific characteristics of the three naming procedures, as for example in Czech *hlup-ák* (stupid-SUFF) 'fool', *hloupý člověk* 'stupid, foolish person' and *osel* '[neutral donkey], dope, ass' with the same meaning.

For instance, in Czech causal relations are not realized as denominal verbs but as complex namings, e.g. *zemřít hladem* (die hunger.INSTR) 'die of hunger, starve', *zemřít žízní* 'die of thirst'. Purpose relations are realized as prepositional word combinations (cf. Czech *míchačka na beton* [mixer for concrete]) 'concrete mixer' in contrast to compounds such as Russian *betonmešalka* 'id.' and others.[8] If metaphor, metonymy, and synecdoche are viewed from an onomasiological perspective, similar types of onomasiological structures (according to Kuchař 1963) can be identified, which can be realized either by word formation, by syntactic means, e.g. conjunctions (Czech *jak*(*o*) 'like' – *jako had* 'like a snake') or alternatively by MWEs, expressing for instance similarity as in *hadí muž* [snake-RA man] 'snake man'. Metonymic shifts can be observed in many deverbal abstract nouns, describing not only the action but also the result. Part-whole relationships can be expressed in Russian by suffixal singulatives (*solom-inka* [straw-SUFF] 'blade of straw', *pesč-inka* [dust-SUFF] 'mote of dust'), contrasting with combinations $N+N_{GEN}$ in Czech (*stéblo slámy* [blade straw.GEN], *zrnko písku* [grain.DIM dust.GEN]).

2 Compounds and MWEs in Russian

This section provides a brief overview of compound and MWE patterns in Russian. We specifically focus on the question if and to what extent the relation between these two naming procedures is being paid attention to in Russian grammars.

8 For similar examples in Polish cf. Cetnarowska (this volume).

2.1 Compounds

We start with the classification of nominal compounds as provided by Grammatika-80 (Švedova 1980: 242ff.). This grammar distinguishes between two groups (cf. A and B in Table 1).[9]

Table 1: Patterns of Russian nominal compounds

A. Coordinate nominal compounds	
1.	N+LV+N
	Formations of this group are rather rare, e.g. *lesostep'* 'forest steppe'.
2.	N+N (with hyphenated spelling)
	In Grammatika-80 (Švedova 1980: 253) some formations of this type are still regarded as word combinations whose first component is no longer subject to declination, e.g. *divan krovat'* 'sofa bed'.[10]
B. Subordinate nominal compounds	
1.	N_{STEM}+LV+N[11]
	zvuk-o-režisser 'sound editor', *sen-o-uborka* 'hayharvest' (cf. Section 4.2.1)
2.	N+N
	Focussing on the absence of a linking vowel, Grammatika-80 forms a heterogeneous group of formations, including loans such as *džaz-orkestr* 'jazz orchestra'. On the activation of the N+N type cf. Section 4.2.2.

9 Attributive compounds are not considered as a group in their own right, cf. however Benigni/Masini (2009).

10 Some formations of this structure are not considered as compounds, but as appositive constructions and thus as syntactic phenomena, cf. *car'-ubijca* [tsar-murderer] 'a tsar who was a murderer' (in contrast to the determinative compound *careubijca* [tsar.LV.murderer] 'regicide'). Cf. also *inžener-fizik* [engineer-physicist] 'engineer and physicist'; *sudno-cholodil'nik* [ship-refrigerator] 'refrigerator ship'; more recent: *komp'juter-tabletka* 'tablet computer'.
The combination of two words is not considered appositive if they designate objects consisting of a larger number of elements or groups of persons and semantically resemble a single word. In Russian they are frequently used as a means of stylization, e.g., *čaški-bljudca* [cups-plates] 'dishes', *ruki-nogi* [hands-legs] 'limbs', *devočki-mal'čiki* [girls-boys] 'children' (cf. also Wälchli 2015, who uses the term "co-compound" for similar, but regular and stylistically neutral formations in various languages).

11 The lack of compounds with (de-)verbal modifiers is often compensated by phrases of the structure [A<V]+N (e.g., *stiral'naja mašina* 'washing machine'). Compare, however, some types of exocentric compounds, whose first constituent might be regarded as derived from an imperative, as in *sorvigolova* [bite-off-head] 'daredevil'.

B. Subordinate nominal compounds		
3.	Frequent first components (modifiers) + N	
a.	*samo-* 'self-'	*samoocenka* 'self-assessment'
b.	*vzaimo-* 'inter-, mutual'	*vzaimopomošč'* 'mutual aid'
c.	*lže-* (< *lož'* 'lie') 'pseudo-'	*lženauka* 'pseudoscience'
d.	*polu-*/*pol-* 'half-'	*polukrug* 'semicircle', *polčasa* 'half an hour'
e.	Some formations with hyphenated spelling, also known from folk literature, e.g., *čudo-* 'miracle, wonder' or *gore-* 'sorrow, misery'	*čudo-bogatyr'* '(epic) hero with magical stength', *gore-rukovoditel'* 'bad leader, manager'
4.	Clipped stems of nouns and/or adjectives (mostly internationalisms) as modifier + N	
	avto$_1$- (referring to *avtomobil'* 'car' and RA *avtomobil'nyj*; *avtobus* 'bus'/RA *avtobusnyj*)	*avtotransport* 'mototransport', *avtovokzal* 'bus terminal'
	avto$_2$- (referring to *avtomatičeskij* 'automatic')	*avtokormuška* 'automatic feeder'
	benzo- (referring to *benzin* 'petrol')	*benzozapravka* 'filling station' (next to the compound without clipped modifier *benzin-o-zapravka* 'id'.)
	others: *kosmo-* 'cosmic, referring to astronautics', *moto-* 'referring to motors; motorized', *ėnergo-* 'energy-; energetic', etc.	*kosmoplavanie* 'space flight', *motolodka* 'motorboat', *ėnergosnabženie* 'energy supply'
5.	Compounds with bound (mostly neoclassical) modifiers + N	
	avto$_3$- 'self-', *aėro-* 'air-', *video-*, *geo-*, *gidro-* 'hydro-', *nevro-* 'neuro-', *poli-* 'poly-', etc.	
6.	Compounds with bound (mostly neoclassical) heads	
	-graf '-grapher', *-fil* '-phile', *-fob* '-phobe', *-metr* '-meter', etc. and *-logija* '-logy', *-fobija* '-phobia', *-fil'stvo* '-philia', etc.	
7.	Suffixed compound nouns	
a.	[[N/A$_{STEM}$ +V$_{STEM}$]SUFF]	*zakonodatel'* [[law.LV.giv]er] 'legislator' *lesopil'nja* [[wood.LV.saw]SUFF] 'sawmill' *čaepitie* [[tea.LV.drink]SUFF] 'tea drinking' (N)
b.	[[A/Num$_{STEM}$ +N$_{STEM}$]SUFF]	*vtorogodnik* [[second.LV.year]SUFF] 'repeater' (a pupil who repeats a grade)

B. Subordinate nominal compounds		
8.	Compounds with "zero suffixes"[12] [N_{STEM} +LV+$STEM_{\varnothing}$] *-vod* '1. guide; 2. breeder, 3. grower', *-mer* 'meter', *-provod* 'conduit', etc.	*ėkskursovod* 'tourist guide', *pticevod* 'poultry farmer'; *vlagomer* 'hygrometer', *vodoprovod* 'water conduit'

2.2 Multi-word expressions

In Russian linguistics, the description of (non-idiomatic) subordinate word combinations of different structures, based on

a) agreement (*nov-aja kniga* [new.FEM book.FEM])
b) government (*čitať knigu* [read book.ACC]; *urok čtenija* [lesson reading.GEN] 'reading instruction'; *kniga dlja detej* [book for children.GEN 'children's book')
c) adjunction (*čitať vsluch* 'to read aloud')

has traditionally been regarded as a domain of syntax.

Following Vinogradov's maxims, coordinate word combinations are ignored in Grammatika-80 (Švedova 1980) while they are taken into consideration in other contributions (e.g. Belošapkova (ed.) 1989 and others).

Fixed subordinate multi-word expressions are described as an object of phraseological research. The distinction of three groups of phrasemes, depending on the degree of idiomatization, goes also back to Vinogradov (1946):

(5a) Phraseological fusions (Russian *frazeologičeskie sraščenija*) – demotivated opaque idioms, e.g.,
biť bakluši [split logs for the production of wooden household utensils] 'twiddle one's thumbs'

(5b) Phraseological unities (Russian *frazeologičeskie edinstva*) – partially metaphorically motivated, e.g.,
plyť po tečeniju 'go [Russian *swim*] with the flow'
New calques based on metaphorical motivation are also regarded as phrasemes, e.g.,

12 This type demonstrates once again the wide-spread distribution of compounds with a deverbal second component. Further, less productive formation types are not discussed here.

myl'naja opera [soap.RA opera] 'soap opera',
promyvanie mozgov [washing brain.GEN] 'brainwash'[13]

(5c) Phraseological word combinations (Russian *frazeologičeskie sočetanija*), e.g.,
skoropostižnaja smert' 'sudden death'. The adjective is exclusively combined with designations of death – Russian *smert'*, *končina*, but **skoropostižnyj ot"ezd* [sudden departure]

Phrasemes of the type (5a) and (5b) or their constituents can function as the basis of compounds or derivations, cf., e.g., *baklušničat'* as synonymous expression to *bit' bakluši* (5a) or the adjective *myl'noopernyi* 'similar to a soap opera' < *myl'naja opera* (5b). Numerous studies are devoted to the relations between phraseology and word formation, including a dictionary of Russian dephrasemic lexis (Alekseenko/Belousova/Litvinnikova (eds.) 2003).

Phrasemes with coordinate relations between the components are generally disregarded in the literature – as in the case of free word combinations (cf. however Benigni (2012: 5f.) on fixed coordinate phrases (*binomi coordinativi*) with a varying degree of idiomaticity, e.g. *mužčina i ženščina* 'man and woman', *sploš i rjadom* [pretty often and nearby] 'very often', *ni ryba ni mjaso* [neither fish nor flesh] 'neither fish nor fowl').

In continuation of Vinogradov's classification of phrasemes Šanskij ([1963] 1985) specifies a fourth group which proves to be of special importance for our topic:

(5d) Phraseological expressions (Russian *frazeologičeskie vyraženija*)

Just as the phrasemes of the other groups, they display the following characteristics: multi-word structure, reproducibility, fixedness (and thus belonging to the lexicon). They do not necessarily need to be idiomatic or metaphorical, however, cf., e.g., *medicinskaja sestra* [medical sister] 'nurse', *teplovaja ėnergija* [heat.RA energy] 'heat energy, thermal energy', *vysšee učebnoe zavedenie* [higher educational institution] 'institution of higher education', etc.

In recent Russian studies (cf. Droga 2010, for instance) such expressions are described as "complex designations" (Russian *sostavnye naimenovanija*), particularly those of the structure:

13 Cf. Mokijenko/Walter (2008: 105); the authors do however not adopt the traditional typology of phraseology.

(6a) A+N
paneľnyj dom [panel.RA house] 'panel house, prefabricated building'
(6b) $N+N_{GEN}$ (or – more rarely – other oblique cases)
sredstva massovoj informacii [media mass.RA.GEN information.GEN] 'mass media'
(6c) N+Prep+N
kniga dlja čtenija [book for reading] 'reader'

It should be mentioned here that such word combinations for a large part compensate for non-existent compound patterns in Russian, including the adaptation of compound loanwords (cf. Section 4). Complex designations of this kind are regarded as "phrasal nouns" by Masini/Benigni (2012: 422): Just like compounds they "generally cannot (a) be interrupted by lexical material, (b) undergo paradigmatic commutability, (c) be internally modified".

2.3 The interaction between different naming procedures in Russian academic grammars

According to Grammatika-80 (Švedova 1980), relations between certain word formation procedures (derivation, compounding) and MWEs only exist if an MWE forms the semantic basis of the word formation (from a formal point of view it is sufficient if only the stem of one constituent of the MWE is retained), e.g.:

(7a) MWE > suffixed one-word combinations, e.g., *večernjaja gazeta* [evening.RA newspaper] 'evening newspaper' > *večer-ka* 'id.'
(cf. Section 3.1)
(7b) MWE > compounds with clipped modifiers (very often internationalisms)
benzinovaja pila > *benzopila* 'power saw'
(7c) MWE > compounds with neoclassical constituents
ėkologičeskaja sistema 'ecological system' > *ėkosistema* 'id.'
(7d) Suffixed compounds (synthetic compounds)[14]
$[[N_{STEM}+V_{STEM}]$-SUFF]
Nouns: *kanatochod-ec* [[rope.LV.go]-SUFF] 'ropedancer; new: tightrope

14 "Complex words that contain at least three morphemes, with neither the combination of the first two nor of the last two existing as free words" (Neef 2015: 583); other studies use the term "parasynthetic compound".

walker' < *chodit' po kanatu* 'to walk on a wire' (in a certain way this also refers to *-vod*, *-mer* formations)
[[A+N]-SUFF]
nouns: *vodolyž-nik* [[water.LV.ski]-SUFF] 'water skier' < *vodnye lyži* 'waterski';
adjectives: *dal'nevostoč-nyj* [[far.LV.east]-SUFF] 'Far Eastern' < *Dal'nij vostok* 'Far East'; with alternation k > č; qualitative adjectives < free word-combinations, e.g., *dlinnonogij* [[long.LV.leg]-SUFF] 'long-legged' < *dlinnye nogi* 'long legs';

(7e) MWE > abbreviations
In Grammatika-80 formations of clipped components of MWEs are regarded as abbreviations,[15] e.g., *prodmag* < *prodovol'stvennyj magazin* [food.RA store] 'food store', etc. (cf. Section 3.2)

Synonymous word formations in the strict sense are listed systematically only for derivations in Grammatika-80 (e.g., *salat-nik*/*salat-nica* 'salad dish', *žad-ina*/*žad-juga* 'greedy person', *meri-l'nyj*/*meri-tel'nyj* 'measuring', *akcentovat'*/*akcentovirat'* 'accent, emphasize', *kratk-o*/*v-kratc-e* 'briefly in short'). Regarding adjectival compounds, reference is made to synonymous second components expressing similarity such as *-vidnyj*,*-obraznyj* (*šarovidnyj*, *šaroobraznyj* [globe/ball.LV. shaped] 'globular, round'). However, Grammatika-80 does not take into account parallel formations of MWEs and nominal compounds, cf. (8), as they are also found in Polish (cf. Cetnarowska, this volume, example (25)):

(8a) *vlag-o-mer* [wetness (in Russian non-derived) measure-∅] 'hygrometer'
(8b) *gigro-metr* 'id.'
(8c) *izmeritel'* (sometimes *meritel'*) *vlažnosti* [measurer wetness.GEN] (alongside rarer forms: *izmeritel' vlagi* 'id.')
Both in Russian and in Polish (cf. example (26) in the chapter on Polish) the genitive attribute can in turn be modified by another genitive, e.g., *izmeritel' vlažnosti vozducha* [meter humidity.GEN air.GEN) 'air humidity meter'

The next section discusses phenomena like those in (7a) and (7e) above in greater detail.

15 Masini/Benigni (2009) regard them as compounds.

3 Interaction between different naming procedures

3.1 MWEs and derivatives

Derivations like the above-mentioned *večerka* are regarded as "synonyms" of MWEs in Grammatika-80 (Švedova 1980: 167ff.).[16] (See, however, below for well-founded objections against this claim.) Formations with the suffix *-ka* (fem.) and its variants are the most frequent, as well as masculine suffixes such as *-nik*, *-jak* (masc.), as in (9):

(9a) *parusnoe sudno* [sail.RA boat] 'sailing boat' > *parus-nik* 'id.'
(9b) *tovarnyj poezd* [goods.RA train] 'freighttrain' > *tovarn-jak* 'id.'

According to Masini/Benigni (2012: 421), the MWEs the derivations are based on are "phrasal lexemes which have a naming function", e.g. *kreditnaja karta* [credit-RA card] 'credit card'. This means that the strategy at hand "consists in shortening a phrasal noun of the [ADJ N] type via ellipsis of the noun plus truncation of the adjective by means of a set of suffixes" (ibid.: 431).[17] They propose the following formal representation of the Russian [ADJ N] lexical construction (ibid.: 444, example (47)):

(10) FORM: $[[a]_{Adj^0x}\,[b]_{N^0y}]_{N'z}$
MEANING: < NAME for SEM_y with the property SEM_x (& SEM_w) $>_z$

For phrasal nouns such as Polish *telefon komórkowy* [phone cellular][18] 'mobile phone', Cetnarowska (this volume, example (31)) proposes the following representation:

(11) $[N^0_{\ i}\ A^0_{\ j}\ ]_k \leftrightarrow$ [NAME for SEM_i with some relation R to entity E of SEM_j $]_k$

16 Derivations that are not synonymous to MWEs are, for instance, *neotlož-ka* 'ambulance' < *neotložnaja pomošč'* [unpostponable aid] 'emergency service', *jader-ščik* 'nuclear physicist' < *jadernaja fizika* 'nuclear physics', *figure-ist* 'figure skater' < *figurnoe katanie* [figure-RA skating] 'figure skating'.

17 See above for the definition of the term "univerbation" in Slavic studies that does not explicitly mention the ellipsis of the head of the MWE.

18 The postposition of the RA typically applies to phrasal nouns in Polish, i.e. word combinations with a naming function.

As in other Slavic languages, this shortening process is very productive in Russian and typical of colloquial language. For this reason, such formations are rarely found in dictionaries (cf. ibid.: 434). This is, however, not quite true in the case of neologism dictionaries such as Uluchanov/Belentschikow (2007), which contains numerous derivations of phrasal nouns. The dictionary also contains (then new) MWEs that did not yet include single word formations. Thus, it can be used as a basis for determining registered single word neologisms. A new source is provided by the German-Russian dictionary of neologisms by Steffens/Nikitina (2014).

Most publications address the assignment of the formations to certain thematic areas. Traditionally, and constantly extended with neologisms (see our examples below), these comprise designations of:

(12a) medicines, cosmetics: (new) *kompaktka* < *kompaktnaja gruntovka* 'compact foundation'
(12b) pieces of clothing, etc.: *futzalki* < *futzaľnaja obuv'*/*futzaľnye tufli* 'shoes for indoor football'
(12c) means of transport and related items: *beskontaktka* < *beskontaktnaja mojka* 'touchless car wash'
(12d) public facilities: *mnogozalka* < *mnogozaľnyj* [multi.hall.RA], *kinoteatr* 'multiplex (cinema)', etc.
(12e) Numerous neologisms belong to professional and group jargon: in medicine: *preimplantacionka* < *pre-implantacionnaja genetičeskaja dignostika* 'preimplantation genetic diagnosis (PGD or PIGD)', or
(12f) in computational language, electronics, e.g., *sensorka* < *sensornyj ėkran* 'touch screen' and *sensornaja igra* 'sensor game'

The wide semantic range of the underlying relational adjective results in the occurrence of numerous homonyms, which are disambiguated in the context or the respective communicative situation; *ėlektronka*, for instance, can refer to 1. *ėlektronnaja kniga* 'e-book' or 2. *ėlektronnaja literatura* 'e-literature', but also to 3. *ėlektronnaja sigareta* 'electronic cigarette'.

Ološtiak (2015: 296) summarizes typical features, distinguishing Slovak MWEs and the results of "univerbation", as follows:

a) greater vs. lesser degree of formal explicitness and therefore
b) lack of ambiguity vs. greater degree of ambiguity,
c) lack of stylistic markedness vs. markedness,
d) official vs. unofficial character,
e) more pronounced association of MWEs with written language vs. underrepresentation of univerbation in written language.

Similarly, Masini/Benigni (2012: 441) state that Russian shortened lexemes with *-ka*, "despite having the same propositional meaning of corresponding full forms, have different pragmatic features". These features are implemented in the formal representation of the *-ka* construction in (13) (cf. ibid.: 445, example 48). The features of the *-ka* lexemes are compared to those described for diminutives with *-ka* which also display familiar/intimate characteristics. (Diminutives may, however, also imply negative or ironic traits, cf. Nagórko 2014: 784 on "quasi-diminutives").

(13) FORM: $[[c]_{N'z}\text{-}ka]_{N^0k}$ where $SYN_z = [[a]_{Adj^0x}\ [b]_{N^0y}]_{N'z}$ & $PHON_z$
= truncated ADJ
MEANING: < NAME for SEM_z & [+ familiar/intimate] (& [+ jargon J]) $>_k$

Thus, although the full forms (the phrasal nouns) and the shortened lexemes in *-ka* share the semantics, they differ with the respect to their pragmatic and textual properties and thus the formal difference between the constructions comes along with a difference in meaning. For this reason, they are not (fully) synonymous and they meet the non-synonymy constraint on constructions as proposed in Construction Grammar (cf. Masini/Benigni 2012: 446).

With respect to analogous forms in Polish Cetnarowska (this volume) states: "The interaction between phrasal lexemes and derivatives (or compounds proper), exemplified by univerbation, can be accounted for in Construction Morphology by means of second order schemas." The respective representation of "shortened phrasal nouns" in Polish can be found in (14) (cf. Cetnarowska, this volume, example (37)):

(14a) Polish *Szkoła budowlana* [school building.RA] 'secondary technical school of building' > *budowlan-ka*

(14b) $[N^0_i\ A^0_j\ ]_k \leftrightarrow$ [NAME for SEM_i with some relation R to entity E of $SEM_j\ ]_k > \approx <[\ A\ \text{-}ka]_{Nz} \leftrightarrow\ [SEM_k$ [+familiar]$]_z$

3.2 MWEs/phrases and stump compounds

Relations between MWEs and one-word designations do not only exist in the area of derivation but also with compounding. This includes formations which in Russian research are frequently described as *složnosokraščennye slova* ('stump compounds') as they represent a combination of compounding and shortening. The shortening process is not based on morphemes but on syllables, in contrast to compounds with clipped, mostly "neoclassical" modifiers, cf. Section 2.1. Stump

compounds have become productive since the end of the 19th century and the beginning of the 20th century and are frequently associated with Sovietisms, cf. (15):

(15a) *likbez* < *likvidacija bezgramotnosti* ($N+N_{GEN}$) 'liquidation of illiteracy' (in the 1920s),
komdiv < *komandir divizii* (Soviet military rank 1935–1940) 'divisional commander'
(15b) *kolchoz* < *kollektivnoe chozjajstvo* (RA+N) 'collective farm'.

Numerous formations have now become historical formations but lexical units based on these models can still be observed. New formations show a tendency to shorten the modifying components. Formations that contain stumps of both components are often proper nouns, e.g. names of Internet domains:

(16) *Dobro požalovať na oficiaľnyj sajt sportivnogo magazina* (sport.RA.GEN shop.GEN) *Sportmag*.[19]
'Welcome to the official site of the sports shop Sportmag'.

3.2.1 $N+N_{GEN}$ / $N+N_{INSTR}$ as underlying MWEs/phrases

Stump compounds consisting of two clipped elements are for instance found in the case of the semi-official namings/names of ministries (17a). The stump *min-* combined with the full form of the genitive is relatively rare (17b). The formation of the stump *obor* (from *oborony*) does obviously not comply with the preferred number of syllables (for the phonetic idiosyncracies of the first component of stump compounds cf. Billings 1998). Nevertheless, among the new formations of the type *min-*+ N_{GEN} there is a combination with the non-euphonic stump *obr* (however not in final position) (17c):

(17a) *Minkuľt* < *ministerstvo kuľtury* [ministry culture.GEN] 'ministry of culture', etc.
(17b) *Minoborony* < *Ministerstvo oborony* 'ministry of defence'
(17c) *Minobrnauki* < *Ministerstvo obrazovanija i nauki* [ministry education.GEN and science.GEN] 'Ministry of education and science'

19 This formation is viewed as an appellative in Acordia/Montermini (2013).

The genitive ending of the nominal modifier is also retained in some designations of deputies, e.g.:

(18) *zampredsedatelja* < *zamestitel' predsedatelja* 'deputy chairman' (alongside the older form *zampred*)

A comparatively small group comprises formations consisting of stumps of nominalized participles (19a, 19b) or a deverbal noun (19c) and the instrumental case of the object (according to the government of the bases – obsolete *zavedovat'* 'be in charge of', and *upravljat'* 'manage'):

(19a) *zavkafedroj* < *zavedujuščij kafedroj* 'head of the department'
(19b) *upravdelami* < *upravljajuščij delami* [manager affairs.INSTR] 'executive officer'[20]
(19c) *upravdelami* < *upravlenie delami* [administration affairs.INSTR] 'executive office (e.g., of the president, a governor)'

All formations with oblique case forms as second components cannot be inflected. In the adjectival derivation of the type (17b)–(19) which are generally informal, colloquial or ironically connotated, the case ending is clipped, e.g. *minoboron-skij gambit* 'the gambit of the Ministry of defense', *zamdekan-skij post* 'position of the vice-dean', or *zavkafedr-al'nyj kabinet* 'office of the head of the department'.

3.2.2 Adjectives (mostly relational adjectives) + nouns as underlying MWEs/phrases

Masini/Benigni (2012: 430) regard this formation type as another "shortening strategy associated with phrasal nouns", cf. (20):

(20) *fizkul'tura* < *fizičeskaja kul'tura* [physical culture] 'physical training, education'
zarplata < *zarabotnaja plata* [for.work.RA pay] 'salary'

20 See, however, the personal designation *upravdom* < *upravljajuščij domom* 'caretaker' where *dom* is the stump of the instrumental case *domom*. This formation can be inflected and is easier to use in colloquial language than the formations mentioned above.

The model is also productive in the formation of neologisms (see below). Compared to the corresponding MWEs/phrases, stump compounds may have the additional advantage of serving as bases for the derivation of relational adjectives, cf.:

(21) *sberegateľnyj bank* 'savings bank' > *sberbank* 'id.' >
sberbankovskij 'related to a savings bank'

Some stumps such as *kom-* 'communist' and *soc-* 'socialist' are mostly found in historical expressions of the Soviet era. Others are still productive, also as part of newly coined formations, such as *gos-* 'state.RA', e.g.:

(22) *goskorporacija* < *gosudarstvennaja korporacija* 'state corporation (a type of legal entity in Russia introduced in 1999)'

Others are new:

(23) *terakt* < *terrorističeskij akt* [terrorist(ic) action] 'terror(ist) attack'

In addition, certain stumps such as *polit-* < *političeskij* 'political' which are known from the notorious designation *politbjuro* 'politburo', can also be found in more recent forms, such as (24):

(24) *politkorrektnosť* 'political correctness', *politjumor* 'political humor'

These stump compounds which are common in politics, administration, press etc., contrast with formations that have become part of the general language. Most of these compounds are more frequent than the underlying phrases (number of hits of formations in the nominative according to Yandex[21]):

(25a) *roddom* (7 m.) < *rodiľnyj dom* (2 m.) [give birth/bear.A house]
'maternity clinic'; no corresponding stump compound (**rodklinika*) of the less frequent and more prestigious naming *rodiľnaja klinika* 'id.' is found, however.
(25b) *zapčasti* (212 m.) < *zapasnye časti* (15 m.) 'spare parts'

21 Yandex is the most frequently used Internet search engine in Russia.

Numerous new formations contain the stump *Ros-* < *rossijskij* 'related to Russia, Russian governmental institutions, enterprises with state participation etc.', e.g. *Rostelekom* 'Rostelecom'. *Ros-* is, however, predominantly found in proper names that are based only on parts of multi-word names. In the following example *Ros-* can be said to replace *Federal'nyj* 'federal':

(26) *Rospotrebnadzor* [Russ(ian) Consum(er) Supervision] = *Federal'naja služba po nadzoru v sfere zaščity prav potrebitelej i blagopolučija čeloveka*
'Federal Service for Surveillance on Consumer Rights Protection and Human wellbeing.'[22]

A similar formation principle is used for naming organizations or enterprises without an established multi-word designation to which the components might be related, cf. (27):

(27) *Rosėnergoatom* (also *RosĖnergoAtom*)
a corporation running nuclear power stations in Russia

As proper names such coinings provide more "convenient" constructions, even when complex multi-word terms exist in parallel.

3.2.3 Pragmatic and textual differences between phrasal nouns and corresponding shortened formations

The differences between stump compounds and suffix formations with *-ka* can be summarized as follows: Stumped compounds are generally used in the area of politics, administration and business. The underlying phrasal nouns, however, indicate a higher level of official status. A higher level of transparency is obtained with currently used stump compounds by not clipping the head. The clipped modifiers are less transparent than the respective word stem (which is retained in the deadjectival *-ka* formations), but they relate to a thematically more clearly restricted range of designation. Frequency specification of stump compounds in Russian newspapers from the year 2014 can be found in Milan Albertin (2013/14:

22 Compounds such as *Rostrud* [Russ(ian) labor], *Federal'naja služba po trudu i zanjatosti* 'Federal Labor and Employment Service' are reminiscent of the old type *Glavryba* (cf. the examples cited by Isačenko 1958 in Section 1.1).

76): state matters 35 %, military 15 %, occupations and functions 12 %, business 7 %, medicine 4 % and other 14 %.

Stump compounds are generally formed "top down", i.e. as planned designations. They are characterized by serial formation and – at least in present-day Russian – the clipped elements are only rarely homonymous.[23] Stump compounding as a semi-official type of word formation is sometimes also used in ironic occasionalisms, cf. *litnomenklatura* 'literary nomenklatura' (from the 1990s) (Uluchanov/Belentschikow 2007: 290) or the name of the Russian heavy metal band *Tjažmet* < *tjaželyj metall* 'heavy metal', consciously aiming at a contrast, as in the past and sometimes even today this stump compound is found in the meaning 'heavy metallurgy' as part of the official name of respective companies.

Derivations with *-ka* based on phrasal nouns are in general formed spontaneously and "from below", i.e. in oral communication. The preferred thematic areas are to be distinguished from those of stump compounds, cf. coll. *kožanka* 'leather jacket' < *kožanaja kurtka*, but not **kožkurtka* (in contrast to the common stump compound *kožizdelija* < *kožanye izdelija* 'leather ware, leather goods').

The following example may summarize the above said. A Russian passport can be referred to as follows:

a) in official use with a multi-word expression and a corresponding acronym: *obščegraždanskij zagraničnyj pasport* (OZP) [civil international passport],
b) in semi-official use with a stump compound: *zagranpasport*,
c) everyday use prefers derivatives like *zagranka* or *zagrannik*,
d) a further variant – the clipped stem *zagran* as noun – is found in social slang.

Masini/Benigni (2012: 447) regard the formation of such shortened lexemes also as a strategy of a highly inflectional language "to 'morphologize' lexical items that are larger than a word".

23 There are, however, older formations where the stump *kom* refers to *kommunističeskij* 'communist (A)', *komandir* 'commander' and *komitet* 'committee', for instance.

4 On the relation between MWEs of the type "relational adjective + noun" and compounds

4.1 RA+N combinations compensating a lack of nominal compound types

The preceding section has discussed the tendency of "morphologizing" word combinations. However, it is obvious that in Russian everyday speech there are also numerous relatively fixed designations of the type [RA<N]+N without shortened variants on *-ka*. These MWEs contrast with N+N compounds in English and German (leaving calques out of consideration), e.g.

a) *polevaja myš'* 'field mouse' (but see suffixal *polëvka* 'vole'), *vodjanaja ptica* 'water bird',
b) *utrennjaja smena* 'morning shift' (but see suffixal *utrennik* 'morning performance'), *nočnoj polet* 'nightflight',
c) *jabločnyj pirog* 'apple pie', *rapsovoe maslo*[24] 'rape oil' (see also parallel formations of the type N+Prep+N, e.g. with the preposition *s* 'with', *iz* 'of, from'),
d) *bannoe polotence* 'bath towel', *komp'juternye igry* 'computer games' (see also parallel formations of the type N+Prep+N, e.g. with the preposition *dlja* 'for').

The reservations that have been expressed about the listing of possible meaning relations between modifiers and non-deverbal heads of compounds (cf., e.g., Plag 2009: 150 with respect to English), may also hold for RA+N combinations.[25] However, it is obvious that there are certain typical relations, depending on the semantics of the modifier and the head of the MWE, i.e. local and temporal (a, b), purpose (c) or reference to the source or origin of what is referred to by the head (d).

Even if compounds can be formed, MWEs may be perceived as more canonical. This becomes evident from the persistence of RA+N combinations alongside older compound calques as well as from the different ways of adapting of new English N+N compounds and compound patterns.

---

24 Only occasionally the otherwise unknown/rare compound *rapsomaslo* [raps.LV.oil] is found in Internet forums.

25 Plag, for example, provides two interpretations of *marble museum* – 'a museum built with marble' and 'a museum in which marble objects are exihibited'. These can potentially also be found in the Russian *mramornyj muzej* (RA+N). Admittedly a search for *muzej* on the Internet typically renders associations with what is exhibited, cf. *Mramornyj muzej v Mramornom dvorce* 'The Marble museum in the Marble Palace (of Catherine the Great in Petersburg)'.

4.2 Compounds

4.2.1 "Classical" patterns (N+LV+N) and parallel patterns (RA+N, N+N_{GEN})

When dealing with Russian determinative compounds with N_{STEM} as modifier and non-derived head, it becomes obvious that their number is restricted. Compounds of the type N_{STEM} + [N<V] are much more productive. Although Grammatika-80 (Švedova 1980: 242) does not make such a distinction, examples such as *zvuko-režisser* [sound.LV.director] 'sound producer', *pticefabrika* 'poultry plant', *chlebo-zavod* [bread.LV.plant] 'bakery plant' (next to RA+N *chlebnyj zavod*), *gazoballon* 'gas bottle' (more frequently RA+N *gazovyj ballon*), *kino-teatr* [cinema-theatre] 'cinema' can be assigned to the first group, expressing primarily purpose relations. The second group includes compounds like *sen-o-uborka* 'hay harvest', *dač-e-vladelec* 'dacha-owner', *ovošč-e-chranilišče* 'vegetable store', reflecting the argument structure of the verb that underlies the head.

These differences become also apparent in the form of Russian equivalents of English compounds. Russian equivalents of English formations with deverbal heads are more frequently compounds of the form N_{STEM}+LV+N (or N+N_{GEN}) and less frequently RA+N patterns. Russian equivalents of other English compounds are, however, for the most part of the type RA+N, cf. Table 2:

Table 2: Compounds and multi-word expressions in Russian

English	**Russian**	
Compound N+N	Compound N+LV+N	Relational adjective + N or N+N_{GEN}
(1) *ship building*	*sudostroenie* (38 m.)[26]	*sudovoe stroenie* (267) *stroenie sudov* (3,000)
(2) *ship repair*	*sudoremont* (13 m.)	*sudovoj remont* (940) *remont sudov* (19 m.)
(3) *ship owner*	*sudovladelec* (5 m.)	*sudovoj vladelec* (sporadically) *vladelec sudna* (3 m.)
(4) *ship mechanic*	*sudomechanik* (132,000)	*sudovoj mechanik* (5 m.) *mechanik sudov* (2,000)

26 Here and subsequently: occurrences in the nominative/accusative in Yandex (January and November 2017).

English	Russian	
Compound N+N	Compound N+LV+N	Relational adjective + N or $N+N_{GEN}$
(5) *ship-broker*	*sudobroker* (14)	*sudovoj broker* (28 m.) *broker sudov* (108)
(6) *shipboard*	–	*sudovoj bort* (368) *bort sudna* (7 m.)
(7) *ship anchor chain*	–	*jakornaja cep' sudna* $[RA+N] + N_{GEN}$

Besides, we also have to consider that numerous English MWEs and compounds correspond to regular suffixal expressions in Russian, as in the case of a) denominal personal nouns: *parket-čik* 'parquet-layer' < *parket* 'parquet', *ryb-ak* 'fisherman' < *ryba* 'fish', *splet-nik* 'scandalmonger' < *spletni* 'rumors', *šachmat-ist* 'chess player' < *šachmaty* 'chess', and b) place nouns: *vinograd-nik* 'wine yard' < *vinograd* 'grapes; vine', *cvet-nik* 'flower garden' < *cvet*(*y*) 'flower(s)', *spaľnja* 'sleeping-room' < *spať* 'sleep', etc.

4.2.2 N+N compounds without linking vowel

Numerous older borrowed N+N compounds (without linking vowel) as a rule have RA+N equivalents, which sometimes are more frequent than the compound, cf.:

(28) *dizeľ-motor* (15,000) 'diesel engine' vs. *dizeľnyj motor* (13 m.) 'id.'; note, however, the use of the compound in names of business and the formation of a new common noun according to the structure N+N: *dizeľ-servis* "*Dizeľ-Motor*" 'diesel-service "Diesel-Engine"' (cf. Section 4.2.3), *vakuum-kamera* (45,000) 'vacuum chamber' vs. *vakuumnaja kamera* (25 m.) 'id.'

A similar relationship exists between some recent compounds (partial calques based on the English N+N model) and formations consisting of RA+N. In the case of

(29) *demping ceny* 'dumping prices' vs. *dempingovye ceny* [dumping.RA prices]

the reasons for the preference of RA+N are to be found in the enhanced syntactic availability or, more precisely, transparency. The ratio of the borrowed compound is considerably lower than the phrase in oblique cases, cf. Russian dative pl. *po*

demping cenam 'goods at dumping prices' (1,240) vs. *po dempingovym* [RA] *cenam* 'id.' (181,000), prepositive case pl. *o demping cenach* (not attested) 'about dumping prices' vs. *o dempingovych cenach* 'id.' (900). The genitive plural *demping cen* is obviously entirely avoided due to its homonymy with $N+N_{GEN}$ [dumping prices. GEN] 'price dumping'.

In addition to parallel formations of the patterns N+N (*marketing direktor* 'marketing director') and RA+N (*marketingovyj direktor*) alternative patterns of the form $N+N_{GEN}$ (*director martekinga*) and N+Prep+N (*direktor po marketingu* 'director of marketing') occur frequently, in particular with respect to professional titles and functional descriptions.

There are, however, numerous new N+N compounds, including compounds with abbreviated modifiers, that do not or only occasionally have RA+N "competitors":[27]

(30a) *biznes-vstreča* 'business meeting', *biznes-pravo* 'business law', *internet-opros* 'internet survey', *internet-magazin* 'internet shop'
(30b) *IT-specialist*, *IT-uslugi* 'IT services' (*IT* can be spelled in latin script, but it is more frequently rendered in Cyrillic.)

According to Benigni/Masini (2009: 179), a criterion for the productivity of N+N patterns in contemporary Russian is the fact that "not only loan words, but also native words occur in this pattern, especially in head position". N+N compounds are also the topic of an article by Kapatsinski/Vakareliyska (2013). According to the authors these new N+N compounds can be found in certain thematic areas, such as business (*cholding kompanija* 'holding company'), politics and media (*press-diskussija* 'press discussion'), music and entertainment (*lajting chudožnik* 'lighting artist'[28]), commerce, technology, computers and the Internet (see above), medicine and health, fashion and sexuality (ibid.: 71).

N+N compounds are also commonly used as names for businesses and events, e.g., *Nogti-Servis* 'Nail Service' (name of a manicure salon), etc. Kapatsinski/Vakareliyska (ibid.: 78) emphasize that this formation type "appears to have developed a distinct connotation: that is, it is not pragmatically synonymous

27 The preference for another compound pattern over MWEs with RA has been evident for some time in the formation of compounds with neo-classical modifiers, or, more generally, internationalisms, e.g., *tele-* (< *televizionnyj*) 'TV-, television-', cf. *telezritel'* (NOM.SG 50 m.) 'TV-viewer' vs. *televizionnyj zritel'* (NOM.SG 1,000).

28 Cf. also the direct, only grammatically adapted borrowing (here: NOM.PL) *lajting & šejding supervajzery* 'lighting and shading supervisors'.

with some other Russian constructions" (in accordance with the No Synonymy Principle as postulated in Construction Grammar). Whereas a possible stump compound like *gorzal* from *gorodskoj zal* [city.RA hall] 'civic hall' would have the connotation of a "Soviet holdover", the new N+N compound *Krokus Siti Choll* 'Crocus City Hall' (opened in 2009 near Moscow) "has a cosmopolitan, western association" (p. 81). In the case of other patterns that were already used in the Soviet era such as $N_{TOPONYM}$+N (e.g. *Tulaugol'* 'Tulacoal', name of a coal trust in the district of Tula), the pattern is retained but newly filled, e.g. *Tulabar*. Here, the "difference in connotations can be plausibly attributed to the interaction between the structure of the expression and the individual words that enter the structure, rather than to the structure per se" (Kapatsinski/Vakareliyska 2013: 81). By means of the new filling the pattern itself gains "a new prestige".

4.2.3 N+N compounds as proper names

As has been shown by some of the examples above, the idea that N+N compounds (without linking vowel) are on the increase is also suggested by their frequent occurrence in proper nouns, such as company names (e.g. *Ivent-Ėkspert* 'event expert' as the name of an agency for marketing solutions). Such names often adopt English patterns, which are also used for common nouns in English everyday speech. This is however not the case in Russian:

(31)	Proprial formations	Non-proprial formations (arranged by the frequency of occurrence of the respective formation type: $N+N_{GEN}$, RA+N, N+N with linking vowel *o* or *e*)
	Gazėksport 'Gas export'	*ėksport gaza*, *gazovyj ėksport*, *gazoėksport*
	Mebel'import	*import mebeli*, *mebel'nyj import*; a non proprial compound **mebeleimport* is not evidenced
	Mebel'Import[29]	
	Mebel' Import 'Furniture import'	

Similar observations apply to names of Internet domains, e.g. *Vodosport* (with linking vowel) 'water sports (equipment)' which has not (yet) been established as

29 It is striking that in many of these newly coined formations the second component is capitalized.

part of the general vocabulary, in contrast to the common appellative construction *vodnyj sport* (water.RA+N). As a common noun *vodosport* occurs only sporadically in Yandex,[30] as in the following example, possibly in analogy to other types of sports which are mentioned in the context, with international clipped initial components:

(32) *Nado vernuťsja v motosport, velosport. Vodosport vsegda byl v Murome populjaren.*[31]
'We have to return to motor sports, to cycle sports. Watersports were always popular in Murom.'

Vodopolo (in standard language RA+N *vodnoe pole*) 'water polo' and *vodolyži* (in standard language *vodnye lyži* 'water ski') are also found as common nouns in Internet texts, however with a linking vowel (!), i.e. not **Voda sport*. (In standard language the stem *vod-* 'water' is found only in compounds with a deverbal head, e.g. *vod-o-snabženie* 'water supply'.) It remains to be seen whether – under the influence of certain text types – the pattern N_{STEM}+LV+N will also occur with those implicit meaning relations that have only been used in RA+N combinations so far (cf. Section 4.1).

5 Conclusion

After a short overview of contributions of Slavic studies on the topic of the present volume this chapter explored some of the relations between non-idiomatic determinative MWEs/phrasal nouns and one-word designations in Russian, viz.:
a) MWEs and a (specific Slavic) type of condensed one-word designations,
b) MWEs/phrasal nouns und stump compounds,
c) MWEs/phrasal nouns and nominal compounds.

Particular emphasis was placed on functional-stylistic and pragmatic differences of referentially identical formations with different structures (cf. Section 3.2 on the relationship of MWEs/phrases and one-word designations, based on several shortening strategies). While suffixal derivations from MWEs/phrases can also be

30 There we also find *vodopolo* (in standard language RA+N *vodnoe pole*) 'water polo', *vodolyži* (in standard language *vodnye lyži*) 'water ski'.
31 https://kachevan.livejournal.com/tag/ %D1 %81 %D0 %BF %D0 %BE %D1 %80 %D1 %82 (last access: 30.4.2018).

found in other Slavic languages, stump compounds – as common names – are largely a specific characteristic of Russian.

With respect to nominal compounding, the chapter has focused on determinative compounds of the type N_{STEM} +LV+N-Type and the N+N type (without linking vowel) and parallel MWEs with the structure N+N_{GEN} or RA+N. (Relational adjectives are today still essential for the integration of numerous borrowed compounds as MWEs). In case of frequently occurring modifiers of N+N compounds a decrease of parallel RA+N formations can be observed. The Russian N+N type was already determined by borrowings in earlier times. In present-day language it is spreading due to the influence of English, since these compounds are no longer restricted to certain thematic areas. An increasing tendency to use the pattern also with non-borrowed words (particularly as head) can be observed. In addition, the spread of the N+N pattern is supported by the frequent use as proper names (cf. *Mebel'Import* 'furniture import') which is however still competing with appellative MWEs (*import mebeli* 'import of furniture').

Analyses of recent developments in the vocabulary of Russian often point to two opposing tendencies (cf. also Masini/Benigni 2012: 447): on the one hand, the increasing tendency towards analyticity (cf. the productive formation of MWEs/phrasal nouns), and, on the other, the persisting tendency towards synthesis (cf. the *-ka* formations derived from MWEs as well as the "condensation of complex nominals" in Slavic languages as mentioned in the introduction.)

The joint reflection on various naming procedures in consideration of their functional differences was determined especially by the onomasiologically orientated research of Slovak and Czech linguists and adopted in Russian research in the 1970s (cf. Serebrennikov (ed.) 1977a, 1977b). However, the interaction between different naming procedures has been considered in the Russian grammars only in the case of MWEs that form the basis of derivations or compounds and are being clipped or shortened. An appropriate theoretical framework for the common consideration of the various procedures is provided by Construction Grammar, and in particular Construction Morphology (Booij 2010), which is based on the fundamental assumption that there is no strict distinction between word formation and/or the lexicon on the one hand and syntax on the other hand. It seems that, for this reason, the simultaneous and in principle equal occurrence of morphological and syntactic naming procedures, as evidenced in this chapter for Russian, can be captured adequately by constructional frameworks. We therefore conclude by referring to Construction Grammar based analyses of compounds and MWEs in Russian and other Slavic languages in Benigni/Masini (2009), Masini/Benigni (2012), Cetnarowska (in this volume) as well as analyses on other languages in the volume at hand (amongst others Booij, this volume, Masini, this volume, Van Goethem and Amiot, this volume, and Schlücker, this volume).

References

Alekseenko, Michail A./Belousova, Taťjana P./Litvinnikova, Oľga I. (eds.) (2003): Slovar' otfrazeologičeskoj leksiki sovremennogo russkogo jazyka. Moskva: Azbukovnik.

Arcordia, Giogio Francesco/Montermini, Fabio (2013): Are reduced compounds compounds? Morphological and prosodic properties of reduced compounds in Russian and Mandarin Chinese. In: Renner, Vincent/Maniez, François/Arnaud, Pierre (eds.): Cross-disciplinary perspectives on lexical blending. Berlin/Boston: De Gruyter. 93–114.

Belošapkova, Vera A. (ed.) (1989): Sovremennyj russkij jazyk. 2nd ed. Moskva: Vysš. Škola.

Benigni, Valentina (2012): I binomi coordinativi in russo: un'analisi costruzionista. In: mediAzioni 13. Internet: www.mediazioni.sitlec.unibo.it/images/stories/PDF_folder/document-pdf/slavistica2012/01_benigni.pdf (last access: 1.9.2017).

Benigni, Valentina/Masini, Francesca (2009): Compounds in Russian. Lingue e Linguaggio VIII, 2. 171–193.

Benigni, Valentina/Masini, Francesca (2010): Nomi sintagmatici in russo. In: Studii Slavistici VII. 145–172.

Billings, Loren A. (1998): Morphology and Syntax. Delimiting stump compounds in Russian. In: Booij, Gert/Ralli, Angelij/Scalise, Sergio (eds.): Proceedings of the First Mediterranean Morphology Meeting. Patras: University of Patras. 99–110.

Booij, Geert E. (2010): Construction morphology. Oxford/New York: Oxford Univ. Press.

Comrie, Bernhard/Stone, Gerald (1978): The Russian language since the revolution. Oxford: Clarendon Press.

Comrie, Bernhard/Stone, Gerald/Polinsky, Maria (1996): The Russian language in the twentieth century. 2nd ed. Oxford: Clarendon Press.

Dokulil, Miloš (1962): Tvoření slov v češtině. Vol. 1: Teorie odvozování slov. Praha: Nakl. Československé Akad. Věd.

Droga, Marina A. (2010): Sostavnye naimenovanija v russkom jazyke. Belgorod: Belgorodskij Gosudarstvennyj Universitet. Internet: http://cheloveknauka.com/v/332425/d#?page=1 (last access: 11.9.2018).

Isačenko, Aleksandr V. (1958): K voprosu o strukturnoj tipologii slovarnogo sostava slavjanskich literaturnych jazykov. In: Slavia 27. 334–352.

Kapatsinski, Vsevolod/Vakareliyska, Cynthia M. (2013): [N[N]] compounds in Russian. A growing family of constructions. In: Constructions and Frames 5, 1. 69–87.

Kuchař, Jaroslav (1963): Základní rysy struktur pojmenování (Basic features of naming structures). In: Slovo a slovesnost 24, 2. 105–114. Internet: http://sas.ujc.cas.cz/archiv.php?art=1230 (last access: 11.9.2018).

Martincová, Olga (2015): Multi-word expressions and univerbation in Slavic. In: Müller, Peter O. et al. (eds.). 742–757.

Masini, Francesca/Benigni, Valentina (2012): Phrasal lexemes and shortening strategies in Russian. The case for constructions. In: Morphology 22. 417–451.

Milan Albertin, Isabella (2013/14): Analisi linguistica dei composti troncati in russo e del loro utilizzo nel linguaggio giornalistico. (MA thesis). Università Ca'Foscari Venezia. Internet: http://dspace.unive.it/bitstream/handle/10579/6112/987901-1193458.pdf?sequence=2 (last access: 1.11.2017).

Mokijenko, Walerij/Walter, Harry (2008): Leksičeskie i frazeologičeskie neologizmy: obščee i različnoe. In: Mokijenko, Walerij/Walter, Harry (eds.): Komparacja systemów i

funkcjonowania współczesnych języków słowiańskich. Vol. 3: Frazeologia. Opole: Wydawn. Uniw. Opolskiego. 101–108.
Müller, Peter O. et al. (eds.) (2015): Word-formation. An international handbook of the languages of Europe. Vol. 1. (= Handbooks of Linguistics and Communication Science (HSK) 40.1). Berlin/Boston: De Gruyter.
Nagórko, Alicja (2014): Diminutiva/Augmentativa. In: Kempgen, Sebastian et al. (eds.): Die slawischen Sprachen. Ein internationales Handbuch zu ihrer Struktur, ihrer Geschichte und ihrer Erforschung. Vol. 1. Berlin/New York: De Gruyter. 782–792.
Neef, Martin (2015): Synthetic compounds in German. In: Müller, Peter O. et al. (eds.). 582–593.
Ološtiak, Martin (ed.) (2015): Viacslovné pomenovania v slovenčine. Prešov: FF PU v Prešove.
Plag, Ingo (2009): Word-formation in English. 5th ed. Cambridge, UK: Cambridge University Press.
Šanskij, Nikolaj M. ([1963] 1985): Frazeologija sovremennogo russkogo literaturnogo jazyka. Moskva.
Serebrennikov, Boris A. (ed.) (1977a): Jazykovaja nominacija. Obščie voprosy. Moskva: Izd. Nauka.
Serebrennikov, Boris A. (ed.) (1977b): Jazykovaja nominacija. Vidy naimenovanija. Moskva: Izd. Nauka.
Steffens, Doris/Nikitina, Olga (2014): Deutsch-russisches Neologismenwörterbuch. Neuer Wortschatz im Deutschen 1991–2010. Mannheim: Inst. für Deutsche Sprache.
Švedova, Nataľja Ju. (ed.) (1980): Russkaja grammatika. Moskva: Izd. Nauka.
Uluchanov, Igor' S./Belentschikow, Renate (2007): Russko-nemeckij slovar' novych slov. Moskva: Izdat. Centr Azbukovnik.
Vinogradov, Viktor V. (1946): Osnovnye ponjatija russkoj frazeologii kak naučnoj discipliny. Leningrad.
Wälchli, Bernhard (2015): Co-compounds. In: Müller, Peter O. et al. (eds.). 707–727.

Bożena Cetnarowska

Compounds and multi-word expressions in Polish

1 Introductory: An overview of basic types of MWEs in Polish

The aim of this chapter is to discuss multi-word units in Polish, focusing on complex nominals (so-called juxtapositions), and to consider their interaction with compounds proper.[1]

Multi-word expressions (MWEs) are defined by Sprenger (2003: 4), Masini (2009: 245) and Hüning/Schlücker (2015: 450) as combinations of two or more words which are used as names for specific concepts. MWEs are intermediate between syntactic units and word-formation units. They show phrase-like syntactic complexity yet they resemble morphologically complex words (such as affixal derivatives and compounds) in exhibiting the naming function. Consequently, some scholars (e.g. Masini 2009; Booij 2010; Masini/Benigni 2012) refer to MWEs as "phrasal lexemes".

The layout of this chapter is as follows. A short overview of MWEs in Polish is given in the remainder of this section. Section 2 mentions basic types of Polish compounds proper and illustrates the occurrence of so-called "solid compounds". Section 3 offers a brief description of phrasal nouns (referred to as "juxtapositions" by Polish linguists). Section 4 discusses some criteria used in distinguishing between compounds proper, solid compounds and juxtapositions. The criteria in question involve prosodic pattern, orthographic form and inflectional properties of compounds. Section 5 examines syntactic fixedness and the internal complexity of juxtapositions. In Section 6 the issue of competition and complementariness between compounds proper and juxtapositions is explored. Section 7 demonstrates that a felicitous account of the interaction between morphological compounds and phrasal lexemes can be offered within the framework of Construction Morphology (as developed by Masini 2009; Booij 2010; Masini/Benigni 2012, among many others). A summary of conclusions is given in Section 8.

1 I would like to thank the editor of the volume and the anonymous reviewers for their useful comments on the previous version of this chapter.

https://doi.org/10.1515/9783110632446-010

Before presenting some examples of MWEs in Polish, we can add that instead of the term "multi-word unit" (Pol. *jednostka wielowyrazowa*), Polish linguists often use the term "phraseological unit" or "phraseme"[2] (Pol. *związek frazeologiczny*, *frazem*). According to the traditional classification[3] proposed by Stanisław Skorupka (e.g. Skorupka 1967), three types of phraseological units are distinguished on the basis of their formal structure: units which are nominal expressions (Pol. *wyrażenia*), such as *pies ogrodnika* (dog.NOM gardener.GEN) 'dog in the manger', verb-phrases (Pol. *zwroty*), e.g. *gryźć ziemię* (bite.INF earth.ACC) 'to bite the dust', and units which exhibit the structure of a sentence (Pol. *frazy*), e.g. *Do wesela się zagoi* (until wedding.GEN REFL heal.FUT.3SG) 'It will heal in no time'. Furthermore, phraseological units are divided into three types, depending on their degree of semantic non-compositionality and syntactic fixedness, into fixed idiomatic phraseological units (Pol. *związki stałe*), collocable phraseological units (Pol. *związki łączliwe*), and free syntactic combinations (Pol. *związki luźne*, lit. loose phraseological units). Fixed phraseological units, such as *biały kruk* (lit. white raven) 'rare specimen', resemble non-derived words in that their meaning does not follow from the meaning of individual components. In the case of collocable phraseological units, such as *dobry humor* 'good mood' and *pobudzić do działania* (wake.INF to action.GEN) 'to incite, to invigorate', their constituents retain literal meaning but show a preference to occur together. Loose phraseological units correspond to free syntactic strings, such as *młoda kobieta* 'young woman' or *zjeść jabłko* 'to eat (an/the) apple'.

Cross-linguistic typologies of phraseological units are discussed by, among others, Granger/Paquot (2008), Fellbaum (2011) and Hüning/Schlücker (2015: 45). I will follow the latter classification in a very brief presentation of types of multi-word expressions in Polish below.

Proverbs in Polish can be exemplified by such sentences as *Ręka rękę myje* (hand.NOM hand.ACC wash.PRES.3SG) 'You scratch my back and I'll scratch yours'. Commonplaces can be illustrated by truisms and tautologies based on everyday experience, e.g. *Żyje się raz* 'You only live once'. Quotations come from popular literary works, songs and films, e.g. *Kobieto*, *puchu marny* (woman.VOC fluff.VOC feeble.VOC) 'Woman, you wretched fluff'.

2 As is stated in the entry for "idiom" in Polański (ed.) (1999: 244), the term "phraseme" (Pol. *frazem*) in the narrow sense is employed to refer to multi-word expressions in which at least one item shows a literal meaning, e.g. *ślepa uliczka* 'blind alley', in contrast to idiomatic expressions whose meaning shows no relatedness to the meaning of particular constituents, e.g. *drzeć koty* (tear.INF cat.ACC.PL) 'to quarrel'.

3 For discussion of other classifications of phraseological units used in the Polish phraseological literature, cf. Lewicki (1976: 9–23), Żmigrodzki (2009: 100) and Szerszunowicz (2012).

Fossilised forms[4] include complex prepositions, such as *w związku z* (lit. in connection with) 'due to' and *naprzeciw* (lit. on opposite) 'opposite, across from'.

Routine formulas in Polish can be exemplified by such expressions as *na zdrowie* (lit. on health.ACC) 'Cheers!' and *do widzenia* (until seeing.GEN) 'good bye'.

Collocations are "prefabricated" semantically transparent combinations of words which show affinity, e.g. *zjełczałe masło* 'rancid butter' and *myć zęby* (wash teeth.ACC) 'to brush teeth'.

Among verbal idioms one can mention such phrases as *kopnąć w kalendarz* (kick.INF in calendar.ACC) 'to die'. Some verbal idioms (e.g. those given above) are based on metaphors. Metaphorical expressions include also prepositional phrases, adjectival phrases and noun phrases (or phrasal nouns), such as *pomiędzy młotem a kowadłem* (between hammer.INS and anvil.INS) 'between a rock and a hard place' and *pies ogrodnika* (dog.NOM gardener.GEN) 'dog in the manger'.

There are no phrasal verbs proper in Polish. However, the range of meanings exhibited by phrasal (or particle) verbs in Germanic languages corresponds largely to the meanings of prefixed verbs in Polish (and in other Slavonic languages). This is shown by the comparison of the prefixless verb *rzucić* 'to throw' and its prefixal derivatives, e.g. *narzucić* 'to throw (sth) on', *rozrzucić* 'to throw around', *wyrzucić* 'to throw away'.

Among fixed expressions in Polish, there occur combinations of nouns with verbs of general meaning,[5] such as *oddać* 'to give back', *zrobić* 'to do, to make', *wykonać* 'to perform', e.g. *oddać skok* 'to do a jump', *zrobić salto* 'to do a somersault', *wykonać przelew bankowy* 'to make a bank transfer'.

There are stereotyped comparisons among phraseological units in Polish, such as *silny jak byk* (strong as bull) 'as strong as an ox' and *pić jak szewc* (lit. drink like shoemaker) 'to drink like a fish'.

Binomial expressions can be illustrated by combinations of nouns, verbs, adjectives or adverbs linked by a conjunction, such as *mąż i żona* (lit. husband and wife) 'man and wife', *żyć i umierać* 'live and die'. They also include combina-

4 Solid compounds, such as *wniebowzięcie* 'assumption (of Virgin Mary)', can also be interpreted as frozen forms (cf. Section 2).

5 As pointed out to me by an anonymous reviewer, Buttler (1976) observes the expansion of analytic constructions in Polish. She (ibid.: 70) mentions the occurrence of verbo-nominal constructions, such as *ulec zepsuciu* (lit. undergo deterioration) 'deteriorate, go bad', and noun-adjective combinations such as *akcja szkoleniowa* (lit. action training.RA), which replace synonymous verbs or nouns, i.e. *zepsuć się* 'to deteriorate, go bad', and *szkolenie* 'training course'.

tions of nouns linked by a preposition, e.g. *ramię w ramię* (lit. shoulder in shoulder) 'shoulder to shoulder'.

Complex nominals, i.e. multi-word expressions with a naming function and with the internal structure of noun phrases, will be discussed in Section 3 (as juxtapositions).

First, however, in Section 2 some types of Polish compounds proper will be described.

2 Types of compounds proper and solid compounds in Polish

Polish composites are usually divided into three types (Grzegorczykowa/Puzynina 1984; Szymanek 2010; Nagórko 2016): compounds proper (which meet the criteria of morphological compounds, as shown in Section 4), solid compounds (Pol. *zrosty*), and juxtapositions (Pol. *zestawienia*).

Solid compounds originate from the coalescence (i.e. merging) of syntactic phrases (Długosz-Kurczabowa/Dubisz 1999: 60; Szymanek 2010: 224). They are written as one orthographic word, e.g. *Wielkanoc* 'Easter', which comes from *Wielka Noc* (lit. great night), *czcigodny* 'respectful', from *czci godny* (lit. respect-deserving), and *zmartwychwstały* 'resurrected', originating from the phrase *z martwych wstały* (lit. from dead arisen). According to Grzegorczykowa/Puzynina (1984: 396), solid compounds characteristically lack interfixes[6] or suffixes but they retain (compound-internal) inflectional elements.[7]

Compounds proper consist of two stems which are characteristically linked with a vocalic interfix (abbreviated here as LV, i.e. linking vowel), e.g. *drobn-o-ustrój* (small+LV+organism)[8] 'microorganism, microbe' and *słodk-o-gorzk-i* (lit. sweet+LV+bitter+NOM.SG) 'bittersweet'. In the case of compounds consisting of a verb stem followed by a nominal stem, the interfix is the vowel *-i-*/*-y-*, as in *gol-i-brod-a* (shave+LV+beard+NOM.SG) 'barber', and *mocz-y-mord-a* (soak+LV+trap+NOM.SG) 'sponge, drunkard'. When the left-hand constituent is the numeral

6 Consequently, Jadacka (2005: 121) regards other composites which lack a vocalic interfix as solid compounds, even if they do not originate from the "freezing" of syntactic phrases, e.g. *seksmasaż* 'sex massage', *biznespartner* 'business partner'.

7 Cf. Section 4 for more discussion of inflectional endings in solid compounds.

8 The compound nouns in question are normally written without hyphens. I use hyphens here to show the internal structure of the composites under discussion.

dw(*u*)- 'two', the interfix appears as the vowel *-u-*, e.g. *dw-u-znak* (two+LV+sign) 'digraph'. Some types of compounds proper, e.g. those with the numeral *trój-* 'three', or the element *pół-* 'half' contain no linking vowel, e.g. *trójskok* (three+jump) 'triple jump', *północ* (half+night) 'midnight, north'.

Compounds such as *drobnoustrój* 'microorganism' and *północ* 'midnight, north' can be compared to primary (root) compounds in English, in which two stems are combined without any intervention of derivational suffixes. The only formative that functions as the marker of composition is the vocalic interfix (if present).

On the other hand, in the case of compound nouns such as *król-o-bój-stw-o* (king+LV+kill+SUFF+NOM.SG) 'regicide', and *krwi-o-daw-c-a* (blood+LV+give+SUFF+NOM.SG) 'blood donor' both the linking vowel and the final derivational suffix act as co-formatives. Such Polish compounds, referred to as "interfixal-suffixal formations", are analogous to synthetic compounds in English, such as *proof-reading* or *truck-driver* (as observed by Szymanek 2010: 221). The right-hand verb stem with the nominalising suffix can either form an independently occurring word, e.g. *dawca* 'giver', or be unattested as a free form, e.g. **bójstwo* 'killing'.

There is yet another (formal) type of compounds proper, namely "interfixal-paradigmatic formations" (Grzegorczykowa/Puzynina 1984: 398; Szymanek 2010: 222), in which two elements act as co-formatives (signalling the operation of compounding): the linking vowel and the so-called paradigmatic formative (i.e. a change of the inflectional paradigm). The right-hand stems of the interfixal-paradigmatic compounds *paliw-o-mierz* (fuel+LV+measure+ø)[9] 'fuel indicator' and *dług-o-pis* (long+LV+write+ø) 'ballpen' are nominalised verb roots, which undergo conversion (i.e. paradigmatic derivation) into nouns. The resulting nominalised elements *-mierz* and *-pis* do not occur as nouns in isolation. Another type of interfixal-suffixal formations is exemplified by the compound noun *żmij-o-głów* (adder+LV+head+ø) 'snakehead fish', in which the right-hand stem does not show a category change but undergoes a shift of the paradigm (from feminine declension, as in *głow-a* (head+NOM.SG), to masculine declension).

If Polish compounds proper are divided into structural types (according to the cross-linguistic classification proposed by Scalise/Bisetto 2009), the compounds in (1) are recognised as subordinate compounds, in which one constituent is subordinated semantically and syntactically to the other so that a complement-head relation can be established between them. The left-hand constituent

9 The element ø represents here a paradigmatic formative (i.e. a zero morpheme), as in Szymanek (2010: 222) and Kolbusz-Buda (2014: 121).

in (1a–c) can be regarded as the object of the action of picking or indicating, and the result of the action of writing. In (1d) the left-hand constituent, i.e. the verb stem *wyrw-*, is syntactically superordinate to the following nominal stem *dąb*. The compound nouns in (1a) and (1b) are endocentric since they are hyponyms of their heads, e.g. *bajkopisarz* 'fabulist, writer of fables' is a kind of a writer. The compounds in (1c) and (1d) are regarded as exocentric by Grzegorczykowa/Puzynina (1984) and Szymanek (2010).[10]

(1a) *grzyb-o-bra-ni-e* (mushroom+LV+take+SUFF+NOM.SG) 'mushroom picking'
(1b) *bajk-o-pis-arz* (fable+LV+write+SUFF) 'fabulist, writer of fables'
(1c) *drog-o-wskaz* (road+LV+indicate+ø) 'signpost'
(1d) *wyrw-i-dąb* (pull_out+LV+oak) 'strong man, athlete'

In attributive compound nouns, such as those in (2), the modifying element expresses some property of the head noun. The compound in (2a) is endocentric, whereas those in (2b) and (2c) are exocentric.

(2a) *żyw-o-płot* (live+LV+fence) 'hedge'
(2b) *biał-o-głow-a* (white+LV+head+NOM.SG) '(obs.) woman'
(2c) *zielon-o-nóż-k-a* (green+LV+leg+DIM+NOM.SG) 'green-legged partridge'

Coordinate compounds in (3) consist of constituents whose status is equal. They can either be treated as endocentric formations which contain two heads, or as exocentric formations, in which the head is missing.[11]

(3a) *barman-o-kelner* (bartender+LV+waiter) 'waiter and bartender'
(3b) *gad-o-ptak* (reptile+LV+bird) 'archaeopteryx'
(3c) *spódnic-o-spodni-e* (skirt+LV+trouser+NOM.PL) 'skort, cullotes'

10 Grzegorczykowa/Puzynina (1984: 399) regard as exocentric formations those compound nouns which represent (mainly) the interfixal-paradigmatic type (e.g. *drog-o-wskaz* 'signpost') or the interfixal-suffixal type (*cudz-o-ziemi-ec* 'foreigner') and in which the right-hand (root+ø or root+SUFF) constituents do not occur as independent nouns, e.g. **wskaz* and **ziemiec*. The anonymous reviewer observes, however, that *drogowskaz* 'signpost' can be interpreted as an endocentric formation. Cf., among others, Grzegorczykowa/Puzynina (1984: 399–403) and Kolbusz-Buda (2014: 58–61, 133–162) for more discussion of the issue.

11 The endocentric/exocentric status of a coordinate compound depends to some extent on a particular semantic paraphrase (one of several available ones) which is employed (cf. Grzegorczykowa/Puzynina 1984: 399; Cetnarowska 2016).

Compound adjectives can be similarly divided into subordinate (e.g. (4a)), attributive (4b) and coordinate ones (4c).

(4a) *złot-o-daj-n-y* (gold+LV+give+SUFF+NOM.SG.M) 'gold-giving'
(4b) *zielon-o-ok-i* (green+LV+eye+NOM.SG.M) 'green-eyed'
(4c) *słodk-o-kwaś-n-y* (sweet+LV+acid+SUFF+NOM.SG.M) 'sweet and sour'

Compound verbs are rare in Polish. Nagórko (2016: 2838) suggests that many of them result from loan translation, e.g. *lekceważyć* 'to disrespect, to neglect' (from German *gering schätzen*[12]).

Długosz-Kurczabowa/Dubisz (1999: 50f.) point out that many compound nouns proper, solid compounds, and compound adjectives in Polish can be treated as calques. Some religious terms are translations of Latin compounds, e.g. *wszech-mogąc-y* (all+able+NOM.SG) 'almighty' (from Latin *omnipotens*). Polish compounds which are imitations of German compound lexemes include, among others, *list-o-nosz* (letter+LV+carry+ø) 'postman' (from *Briefträger*) and *ogni-o-trwał-y* (fire+LV+durable+NOM.SG) 'fireproof' (from *feuerfest*). The influence of Russian, on the other hand, can be observed in the case of such compounds as *brak-o-rób-stw-o* (dud+LV+do+SUFF+NOM.SG) 'wastage' (from *brakodielstvo*). Nevertheless, Długosz-Kurczabowa/Dubisz (ibid.: 75) argue for the recognition of compound formation in Polish as a native pattern (which can be traced back to Proto-Slavonic forms or the Old Polish period).

3 Juxtapositions ("phrasal nouns")

Juxtapositions show phrasal structure. The following syntactic types of juxtapositions, i.e. phrasal nouns, can be identified in Polish.

(5) N+N.GEN
(5a) *dom studenta* (house.NOM student.GEN.SG) 'dormitory, student hall of residence'
(5b) *mąż stanu* (man.NOM state.GEN.SG) 'statesman'

12 As is pointed out to me by the editor of the volume, the expression *gering schätzen* is not normally regarded as a compound in German.

(6) N+PP
(6a) *chustka do nosa* (kerchief.DIM.NOM for nose.GEN) 'handkerchief'
(6b) *dziurka od klucza* (hole.DIM.NOM from key.GEN) 'keyhole'

(7) N+A
(7a) *panna młoda* (maid young) 'bride'
(7b) *drukarka laserowa* (printer laser.ADJ) 'laser printer'
(7c) *krem odżywczy* (cream nourishing) 'nourishing cream'

(8) A+N
(8a) *biały kruk* (white raven) 'rare specimen'
(8b) *nocna zmiana* (night.ADJ shift) 'night shift'
(8c) *wieczne pióro* (eternal pen) 'fountain pen'

(9) N+N
(9a) *poeta-tłumacz* (poet translator) 'poet-translator'
(9b) *kobieta-guma* (woman rubber) 'female contortionist'
(9c) *wywiad-rzeka* (interview river) 'extended interview'

The constituents of juxtapositions exhibit the relation of government (as in N+N.GEN phrasal nouns) or agreement (as in N+A or A+N juxtapositions and in N+N juxtapositions). The adjective in N+A and A+N phrasal nouns is often a denominal one, i.e. a relational adjective such as *laserowy* (laser.RA) from the noun *laser* 'laser', and then the whole combination is a possible translation equivalent in Polish for a noun+noun compound in English or in other Germanic languages.[13] It needs to be added, though, that some N+A or A+N juxtapositions contain nonderived adjectives, e.g. *młoda* 'young' in *panna młoda* 'bride', or deverbal adjectives, e.g. *odżywczy* 'nourishing' from the verb *odżywiać* 'to nourish'.

When the tripartite structural typology of compounds proper is applied to juxtapositions, it can be noted that Polish juxtapositions behave similarly to those in Russian, discussed by Masini/Benigni (2012). N+N.GEN and N+PP phrasal nouns are often subordinate composites (as in 10), N+A and A+N combinations tend to be attributive (as in 11) while N+N combinations (in 12) are coordinate juxtapositions.

13 On the basis of translation equivalence between Germanic N+N compounds and Polish N+RA (or RA+N) units, ten Hacken (2013) argues that multi-word expressions in Polish consisting of nouns and relational adjectives should be treated as compounds.

(10a) *maszyna do szycia* (machine for sewing) 'sewing machine'
(10b) *dawca organów* (donor.NOM organ.GEN.PL) 'organ donor'

(11a) *stara panna* (old maid) 'old maid'
(11b) *panda wielka* (panda great) 'giant panda'

(12a) *torba-worek* (bag sack) 'large bag'
(12b) *kierowca-dostawca* (driver deliverer) 'delivery driver'

The relationship between the syntactic type and the structural classification of juxtapositions is not complete, though. N+N combinations (whose constituents show agreement) and N+N.GEN phrasal nouns in (13) require attributive interpretation.

(13a) *ryba-piła* (fish saw) 'sawfish'
(13b) *kobieta-guma* (woman rubber) 'female contortionist'
(13c) *człowiek honoru* (man.NOM honour.GEN) 'man of honour'

Damborský (1966) remarks that some N+N juxtapositions may have entered the Polish language as calques of French formations (e.g. *zegarek-bransoletka* 'watch-bracelet') or as calques of Russian complex lexemes (e.g. *miasto-bohater* 'hero city'). Nevertheless, he concludes that N+N juxtapositions represent mostly a native pattern of composite formation (as is also observed by Długosz-Kurczabowa/Dubisz 1999).

In the next section criteria which can be employed in distinguishing between compounds proper and juxtapositions will be presented.

4 Differences between compounds proper, solid compounds and juxtapositions

Polish compounds proper exhibit features expected of morphological compounds cross-linguistically (cf. Lieber/Štekauer 2009; Booij 2010). They are written as one orthographic word, though some compounds are hyphenated, e.g. *słodko-kwaśny* 'sweet and sour'.[14]

14 The hyphen is employed in the case of coordinate compound adjectives (e.g. *przemysłowo -rolniczy* 'industrial and agricultural') while attributive and subordinate compound adjectives

A compound proper constitutes one prosodic unit with respect to stress assignment. As is indicated here (for clarity) by the capitalization of the appropriate vowel, the main lexical stress falls on the penultimate syllable in compound nouns such as *długOpis* 'ballpen', and in compound adjectives, e.g. *ciemnoniebiEski* 'dark blue' (cf. Szymanek 2010: 225).[15]

Constituents of compounds proper in Polish form one morphological word, with the morphological head located on the right. The inflectional ending is attached to the right-hand stem, e.g. *-a* (NOM.SG) in (14a). In the case of exocentric compound nouns (as in 14b), the inflectional ending appears to attach to the whole compound stem, rather than to the right-hand stem, since the inflectional characteristics of those compound nouns often diverge from the inflectional properties of their right-hand constituents.[16]

(14a) *mebl-o-ścian-k-a*
furniture+LV+wall+DIM+NOM.SG
'wall unit'

(14b) *staw-o-nog-a*
joint+LV+foot+GEN.SG
'arthropod' (GEN.SG)

Solid compounds exhibit most of the properties of morphological compounds. They are written as one orthographic word and constitute one prosodic domain (with respect to stress assignment), as is shown by *WielkAnoc* 'Easter', as opposed to the free syntactic combination *wiElka nOc* 'great night'. The inflectional endings in solid compounds are usually attached only to the right-hand stems, e.g. *czcigodn-emu* (venerable.DAT.SG), and *duszpasterz-a* (priest.GEN.SG). The inflectional ending of the left-hand constituent (if present)[17] is 'frozen' inside the solid compound and it takes the function of the vocalic interfix, e.g. *-i* (GEN.SG) in *czcigodny* 'venerable'. In selected solid compound nouns both stems obligatorily

are written as single orthographic words (e.g. *roponośny* 'oil-bearing', *ciemnozielony* 'dark green').

15 In the case of polysyllabic compounds, apart from the main stress on the penultimate syllable, there may occur secondary stresses on the first constituent, e.g. *prAlkosuszArka* 'washer dryer', *ciEmnoniebiEski* 'dark blue'.

16 The compound noun *stawonóg* 'arthropod' is masculine, while its right-hand constituent *noga* 'foot' is feminine (cf. *nog-i* 'foot+GEN.SG').

17 There is no vocalic element linking the constituents *dusz* (soul.GEN.PL) and *pasterz* (shepherd.NOM.SG) since the marker of genitive plural in the first constituent is a morphological zero.

decline as independent morphological words,[18] in spite of constituting a single prosodic and orthographic unit, e.g. *Biał-y-stok* (white+NOM.SG+slope+NOM.SG) 'Białystok.NOM.SG' (a city in north-eastern Poland) and *Biał-ego-stok-u* (white+GEN.SG+slope+GEN.SG) 'Białystok.GEN.SG'.

Juxtapositions consist of constituents which are written as separate orthographic words, e.g. *maszyna do pisania* (machine for writing) 'typewriter', *kobieta pilot* (woman pilot) 'female pilot' and *prawa człowieka* (law.NOM.PL man.GEN.SG) 'human rights'. However, some attributive N+N compounds, e.g. *kobieta-guma* (woman rubber) 'female contortionist', and coordinate N+N compounds, e.g. *malarz-tapeciarz* 'painter-decorator', are hyphenated,[19] in which they resemble morphological compounds in other languages (cf. Lieber/Štekauer 2009) and coordinate adjectival compounds proper in Polish.

Each element of a juxtaposition takes its own inflectional endings. They can stand in either the relation of agreement (as in the case of N+A, A+N and N+N juxtapositions), or the relation of government (in the case of N+N.GEN or N+PP phrasal nouns). Constituents of juxtapositions also behave as independent units for the purpose of lexical stress assignment, as is shown by the stress pattern of *mAlarz-tapEciarz* 'painter-decorator', and *chUstka do nOsa* (lit. kerchief for nose) 'handkerchief'.

5 Syntactic fixedness

The Lexical Integrity Principle, postulated by Anderson (1992), does not allow rules of syntax to manipulate or have access to parts of words. Booij (2010: 177) points out that this principle can be split into two subparts (i.e. two subconstraints).

One subconstraint prohibits the operation of syntactic rules of case assignment and agreement on constituents of morphologically complex words. Inflectional endings do not occur inside affixal derivatives or inside compounds proper, cf. *czarn-o-biał-ego* (black+LV+white+GEN.SG) 'black-and-white.GEN.SG' and not **czarn-ego-biał-ego* (black+GEN.SG+white+GEN.SG). This subconstraint is vio-

18 There occur also solid compounds which allow alternative word-forms, e.g. *Wielk-a-noc* (great+NOM.SG/LV+night) 'Easter.NOM.SG', *Wielk-a-noc-y* (great+LV+night+GEN.SG) or *Wielki-ej-noc-y* (great+GEN.SG+night+GEN.SG) 'Easter.GEN.SG'.

19 According to current prescriptive recommendations, Polish coordinate compounds should be hyphenated while attributive compounds should not.

lated in the case of juxtapositions and some solid compounds, as was illustrated in the previous section.

The second subpart of the Lexical Integrity Principle predicts that words can be neither split by intervening constituents nor reordered. This subconstraint is met in the case of the majority of compounds proper and solid compounds in Polish. The left-hand modifiers of the compound nouns *dług-o-pis* (long+LV+write+ø) 'ballpen' and *grzyb-o-bra-ni-e* (mushroom+LV+take+SUFF+NOM.SG) 'mushroom picking' cannot be shifted to the right-hand position, as is shown by the ill-formedness of **pis-o-dług* and **brani-o-grzyb*. Moreover, those left-hand (modifier) stems cannot be modified themselves, as indicated by the unacceptability of **bardzo-dług-o-pis* (very+long+LV+write+ø) in the intended meaning 'ballpen which can write for a long time'. Constituents of coordinate compounds proper show some possibility of reordering, e.g. *czerwono-biały* 'red and white' and *biało-czerwony* 'white and red'.[20] However, one potential order of elements tends to be conventionalised, hence ?*suszark-o-pralk-a* (dryer+LV+washer+NOM.SG) and ?*robotnik-o-chłop* (worker+LV+peasant) sound decidedly odd when compared to the institutionalised forms *pralk-o-suszark-a* (washer+LV+dryer+NOM.SG) 'washer and dryer' and *chłop-o-robotnik* (peasant+LV+worker) 'a peasant farmer who also works in a factory'.

Juxtapositions resemble compounds proper in Polish in that their internal constituents cannot be modified (cf. Cetnarowska/Trugman 2012; Cetnarowska 2018).[21] If an adverbial modifier is inserted in front of the adjective in the N+A juxtaposition *foka szara* (seal grey) 'grey seal', the resulting string stops functioning as a naming unit and can be interpreted as a free syntactic combination, i.e. *foka bardzo szara* (seal very grey) 'seal whose fur is very grey'. Similarly, the addition of the demonstrative *tego* (this.GEN.SG) in front of the noun *człowieka* (man.GEN.SG) in the N+N.GEN phrasal noun *prawa człowieka* (law.NOM.PL man.GEN.SG) 'human rights' results in the reanalysis of the juxtaposition as a freely composed noun phrase, i.e. *prawa tego człowieka* (law.NOM.PL this.GEN.SG man.GEN.SG) 'this man's rights'. Some instances of phrasal nouns that contain internal pre- or post-modifiers (and complements) can be encountered, as shown in (15). It can be argued, though, that these are cases of complex phrasal nouns which contain

20 Nagórko (2016: 2837) remarks that there is a difference in meaning between *biało-czerwony* (white-red), which can be used to describe the flag of Poland, and *czerwono-biały* (red-white), which describes the colours of the flag of Monaco.

21 Consequently, adjectives and nouns are regarded as non-projecting categories (A^0 and N^0) in multi-word units in Polish by Cetnarowska (2018), as is suggested for MWEs in other languages by Booij (2010).

phrasal nouns as their subconstituents, e.g. *małe dziecko* 'small child' functions as a naming unit, hence it can become a part of another naming unit.

(15a)	*dom*	*dzieck-a*	
	house.NOM.SG	child+GEN.SG	
	'orphanage, children's home'		
(15b)	*dom*	*mał-ego*	*dzieck-a*
	house.NOM.SG	small+GEN.SG	child+GEN.SG
	'orphanage for small children'		
(15c)	*wod-a*	*mineral-n-a*	
	water+NOM.SG	mineral+RA+NOM.SG	
	'mineral water'		
(15d)	*gazowan-a*	*wod-a*	*mineral-n-a*
	aerated+NOM.SG	water+NOM.SG	mineral+RA+NOM.SG
	'sparkling mineral water'		

The issue of changes in the internal order of elements of juxtapositions is more complex. Constituents of coordinate N+N juxtapositions show a considerable degree of mobility,[22] e.g. *aktor-tancerz* (actor-dancer) and *tancerz-aktor* (dancer-actor), or *kobieta pilot* (woman pilot) and *pilot kobieta* (pilot woman).

N+N.GEN juxtapositions and N+PP juxtapositions resist internal reordering (except in poetry, artistic prose or journalese). Shifts in the order of their constituents result in the infelicity of the resulting phrasal noun, e.g. ??*honoru słowo* (honour.GEN.SG word.NOM.SG) vs. *słowo honoru* (word.NOM.SG honour.GEN.SG) 'word of honour', or ??*do szycia maszyna* (for sewing.GEN.SG machine.NOM.SG) vs. *maszyna do szycia* (machine.NOM.SG for sewing.GEN.SG) 'sewing machine'. Alternatively, such shifts may lead to the reinterpretation of the juxtaposition as a regular syntactic phrase, e.g. *małego dziecka dom* (small.GEN.SG child.GEN.SG house.NOM.SG) 'house of (a particular) small child'.

The mobility of constituents of A+N and N+A phrasal nouns depends on their semantic compositionality and the range of polysemy exhibited by a given adjective.

Cetnarowska/Pysz/Trugman (2011) and Cetnarowska/Trugman (2012) divide combinations of classifying adjectives and nouns (in any order) in Polish into

22 The internal word order is fixed in the case of some types of coordinate and quasi-coordinate juxtapositions, e.g. those that consist of a superordinate term followed by a hyponym, such as *lekarz ginekolog* (physician+gynecologist) 'gynecologist' or Kinship+Property coordinate juxtapositions, e.g. *syn prawnik* (son+lawyer) 'lawyer son'.

three groups: idiomatic A+N combinations, N+A 'tight units' and A+N/N+A combinations in which the classifying adjective is regarded as 'migrating'.

A+N juxtapositions which are regarded by Cetnarowska/Pysz/Trugman (2011) as lexicalised idiomatic phrases, such as *końsk*i *ogon* (horse.RA tail) 'ponytail', *lwia paszcza* (lion.RA jaw) 'snapdragon', and *boża krówka* (god.RA cow.DIM) 'ladybird', show syntactic fixedness. Their consitutents cannot be shifted, since the postposing of the adjective changes their meaning to non-idiomatic combinations, as shown in (16).

(16a)	*koń-sk-i*	*ogon*
	horse+RA+NOM.SG	tail.NOM.SG
	'ponytail'	
(16b)	*ogon*	*koń-sk-i*
	tail.NOM.SG	horse+RA+NOM.SG
	'tail of (a/the) horse'	

The elements of N+A 'tight units' are not (normally) reversible, either. Post-head classifying adjectives in tight units, such as *kurier dyplomatyczny* (courier diplomatic) 'diplomatic courier', *pancernik olbrzymi* (armadillo giant) 'giant armadillo' and *foka szara* (seal grey) 'grey seal', change their interpretation to those of qualifying adjectives, as indicated in (17) and (18).

(17a)	*kurier*	*dyplomat-yczn-y*
	courier.NOM.SG	diplomat+RA+NOM.SG
	'diplomatic courier'	
(17b)	*dyplomat-yczn-y*	*kurier*
	diplomat+RA+NOM.SG	courier.NOM.SG
	'tactful courier'	

(18a)	*pancernik*	*olbrzym-i*
	armadillo.NOM.SG	giant.A+NOM.SG
	'giant armadillo'	
(18b)	*olbrzym-i*	*pancernik*
	giant.A+NOM.SG	armadillo.NOM.SG
	'very large armadillo'	

'Migrating' classifying adjectives are felicitous in phrasal nouns both pre-nominally and post-nominally, without incurring any serious change in their interpretation (as in (19) and (20)). They can be analysed as intersective modifiers (as observed by Cetnarowska/Trugman 2012). The choice between placing a migrat-

ing classifying adjective in the pre- or post-head position is determined by a number of various syntactic and stylistic factors, one of them being the occurrence of additional classifying adjectives or genitive complements in a phrasal noun (cf. Szumska 2006; Cetnarowska/Pysz/Trugman 2011; Linde-Usiekniewicz 2013; Cetnarowska 2014 for more discussion).

(19a) *noc-n-y* *sklep*
night+RA+NOM.SG shop.NOM.SG
'night shop'

(19b) *sklep* *noc-n-y*
shop.NOM.SG night+RA+NOM.SG
'night shop'

(20a) *kurtk-a* *męsk-a*
jacket+NOM.SG male.NOM.SG
'men's jacket'

(20b) *męsk-a* *kurtk-a* *zim-ow-a*
male+NOM.SG jacket+NOM.SG winter+RA+NOM.SG
'men's winter jacket'

Syntactic flexibility in idioms can be regarded (cross-linguistically) as a consequence of their semantic transparency, as is argued by Nunberg/Sag/Wasow (1994). The behaviour of A+N and N+A phrasal nouns in Polish provides further evidence for such a conclusion, since idiomatic A+N juxtapositions are 'syntactically frozen'. Fellbaum (2011: 448) shows, however, on the basis of data from German and English, that even (more) opaque idioms may allow for morphological and syntactic variation, depending on their larger sentential context and on the presence of stylistic (or humorous) colouring. Some instances of the word-order modification in N+A 'tight units', to facilitate word play or contrast, are mentioned by Cetnarowska (2015).

6 Competition between compounds and juxtapositions

The conventionalisation of a given concept by means of a compound or a phrasal unit in Polish is to some extent arbitrary. For instance, while there exist the synthetic compounds proper *koni-o-krad* (horse+LV+steal+ø) 'horse thief' and (used rather rarely) *kur-o-krad* (hen+LV+steal+ø) 'chicken thief', N+N.GEN phrasal lex-

emes are used to denote a person who steals cars or bicycles, i.e. *złodziej samochodów* (thief.NOM.SG car.GEN.PL) 'car thief' and *złodziej rowerów* (thief.NOM.SG bicycle.GEN.PL) 'bicycle thief'.

Nevertheless, it is possible to come across synonymous compounds proper and juxtapositions in Polish. Let us look at the competition between (and coexistence of) subordinate synthetic compounds proper and N+N.GEN combinations (or N+A units).

There exist several institutionalised synthetic compounds which end in the constituent *-dawca* 'giver', e.g. *kredyt-o-daw-c-a* 'lender', *prac-o-daw-c-a* 'employer', *ustaw-o-daw-c-a* 'lawmaker, legislator', *spadk-o-daw-c-a* 'testator'. Jadacka (2001: 96, 99) observes that compounds terminating in *-dawca* represent a fairly numerous group of neologisms in the Polish vocabulary at the end of the twentieth century (i.e. after 1989).[23]

As shown in (21)–(22) below, the existence of synthetic compounds proper terminating in *-dawca*, such as *licencj-o-daw-c-a* 'licensor', does not block the formation (and use of) a synonymous N+N.GEN juxtaposition, i.e. *dawc-a licencj-i* 'licensor (lit. giver of licence)'.

(21) *licencj-o-daw-c-a*
licence+LV+give+SUFF+NOM.SG
'licensor'

(22) *daw-c-a* *licencj-i*
give+SUFF+NOM.SG licence+GEN.SG
'licensor'

(23a) *krwi-o-daw-c-a*
blood+LV+give+SUFF+NOM.SG
'blood donor'

(23b) *daw-c-a* *krw-i*
give+SUFF+NOM.SG blood+GEN.SG
'blood donor'

23 Nevertheless, the pattern of synthetic compounds with the constituent *-dawca* 'giver' shows many gaps. There are no attestations (in the National Corpus of Polish) of the potentially well-formed compounds ?*organodawca* (organ+LV+giver) 'organ donor', ?*szpikodawca* (marrow+LV+giver) '(bone) marrow donor' or ?*sercodawca* (heart+LV+giver) 'heart donor'. However, the anonymous reviewer points out that Google searches result in 17 hits for ?*organodawca* 'organ donor' (including some metaphorical uses of the word) and 9 hits for ?*szpikodawca* 'marrow donor'.

The comparison of the occurrence of the (various inflectional forms of the) lexemes in (21)–(23) in the National Corpus of Polish (NKJP) shows that the synthetic compound *licencjodawca* 'licensor' is more common in the corpus than the phrasal noun *dawca licencji* (giver.NOM.SG licence.GEN.SG) 'licensor': it occurs 167 times, while the equivalent phrasal noun is attested 9 times. In the case of the items in (23), both the synthetic compound *krwiodawca* 'blood donor' and the N+N.GEN phrasal noun *dawca krwi* 'blood donor' are fairly frequent.[24]

Jadacka (2001: 98) also points out the productivity of the pattern of interfixal-paradigmatic derivation of compounds, represented by such novel compounds as *diet-o-mierz* (diet+LV+measure+ø) 'dietometer', where the right-hand constituent is the verb stem *mierz-* (as in *mierzyć* 'measure.INF') and the nominalizing morpheme is the paradigmatic formative (i.e. the zero morpheme ø). There exist doublets or even triplets consisting of synonymous compounds terminating in *-mierz* or *-metr* and phrasal nouns consisting of the head *miernik* 'meter, gauge' followed by a noun in the genitive.

(24a) *głośn-ości-o-mierz*
loud+SUFF+LV+measure+ø
'volume unit meter'

(24b) *audio-metr*
audio+meter
'audiometer'

(24c) *mier-nik* *głośn-ośc-i*
measure+SUFF loud+SUFF+GEN.SG
'volume unit meter, volume indicator'

(25a) *wilgotn-ości-o-mierz*
wet+SUFF+LV+measure+ø
'moisture meter'

(25b) *higro-metr*
hygro+meter
'hygrometer'

(25c) *mier-nik* *wilgotn-ośc-i*
measure+SUFF wet+SUFF+GEN.SG
'hygrometer, moisture meter'

24 There is a difference in the occurrence of the nominative singular forms of both competing lexemes: the compound occurs 345 times and the phrasal noun 57 times, mainly in the expression *honorowy dawca krwi* 'honorary blood donor'.

The usage of N+N.GEN pattern allows the speaker to reach greater precision in denoting the kind of instrument. The genitive attribute can in turn be modified by another genitive, as is shown in (26)–(27).

(26) *mier-nik* *wilgotn-ość-i* *powietrz-a*
measure+SUFF wet+SUFF+GEN.SG air+GEN.SG
'air humidity meter'

(27) *mier-nik* *wilgotn-ość-i* *drewn-a*
measure+SUFF wet+SUFF+GEN.SG wood+GEN.SG
'wood moisture meter'

The N+N.GEN nouns in (26)–(27) above have no corresponding morphological compounds, since there is no pattern which would allow the name of the object (whose moisture is to be tested) to be included in a compound proper. The hypothetical lexemes **powietrz-o-wilgotności-o-mierz* (air+LV+moisture+LV+measure+ø) and **drewn-o-wilgotności-o-mierz* (wood+LV+moisture+LV+measure+ø) are ill-formed.

Another area where juxtapositions compete with compounds proper is the formation of coordinate composites. Jadacka (2001: 145) observes that juxtapositions, not morphological compounds proper, constituted previously (until the middle of the twentieth century) the recommended pattern employed in creating names of coordinate entities. On the other hand, coordinate juxtapositions (of the multifunctional type)[25] may evolve into compounds proper. While the N+N phrasal lexemes given in (28a) and (28c) are quoted in the literature (e.g. by Damborský 1966; Kallas 1980; Szymanek 2010), they have few (or no) attestations in the NKJP corpus. They were replaced by the corresponding coordinate compounds proper in (28b) and (28c).

(28a) *chłop-robotnik*
peasant+worker
'peasant farmer who works in a factory'

(28b) *chłop-o-robotnik*
peasant+LV+worker
'peasant farmer who works in a factory'

25 According to Renner/Fernández-Domínguez (2011: 876f.), a multifunctional coordinate compound denotes an entity which belongs to two categories simultaneously and can be paraphrased as 'an X + Y is an X who/which is also a Y'.

(28c) *klub-kawiarni-a*
club+café+NOM.SG
'café that hosts cultural events'
(28d) *klub-o-kawiarni-a*
club+LV+café+NOM.SG
'café that hosts cultural events'

In the case of the pairs of multifunctional coordinate phrasal nouns and compounds proper given in (29), both formations coexist (and compete).

(29a) *krem-żel*
cream+gel
'gel cream'
(29b) *krem-o-żel*
cream+LV+gel
'gel cream'
(29c) *barman-kelner*
bartender+waiter
'waiter-bartender'
(29d) *barman-o-kelner*
bartender+LV+waiter
'waiter-bartender'

Certain types of coordinate composites allow for one pattern only, i.e. either the creation of N+N juxtapositions or compounds proper. Multifunctional coordinate composites representing (among others) the following semantic types[26] cannot be expressed by synthetic compounds:

(30a) Sex+Profession: *kobieta tłumacz*
(woman translator) 'female translator',
not **kobiet-o-tłumacz*
(30b) Profession+Characteristic Activity: *tancerka szpieg*
(dancer spy) 'both female dancer and spy',
not **tancerk-o-szpieg*
(30c) Kinship+Profession: *żona aktorka*
(wife actress) 'actress wife',
not **żon-o-aktorka*

26 The semantic typology is based on that postulated for English by Olsen (2001).

Attributive juxtapositions, such as *wywiad-rzeka* (interview+river) 'extended interview', *kobieta anioł* (woman angel) 'angel of a woman', cannot be replaced by morphological compounds (with an interfix), i.e. **wywiad-o-rzeka* (interview+LV+river) or **kobiet-o-anioł* (woman+LV+angel).

On the other hand, hybrid coordinate compounds proper, which can be paraphrased as 'X is a blend of X and Y' (Renner/Fernández-Domínguez 2011), have no corresponding N+N juxtapositions, cf. *las-o-step* (forest+LV+steppe) 'forest-steppe', *gad-o-ptak* (reptile+LV+bird) 'archaeopteryx' and not **las-step* or **gad-ptak*.

Thus, juxtapositions not only compete with but also complement compounds proper in Polish.

7 The treatment of phrasal nouns in Construction Morphology

As noted by Grzegorczykowa (1982: 59) and Długosz-Kurczabowa/Dubisz (1999) and as mentioned in Section 2, in traditional accounts of Polish word-formation (e.g. Klemensiewicz 1939) phrasal nouns were treated as a subtype of composites (i.e. compounds in the broad sense of the term), namely as juxtapositions. In more rigorous descriptive grammars of Polish (e.g. those written in the structuralist paradigm), juxtapositions are excluded from the domain of morphology. Puzynina (1974) argues that multi-word expressions, such as *maszyna do szycia* (machine for sewing) 'sewing machine' and *szkoła podstawowa* (school elementary) 'primary school', should fall within the domain of phraseological research, and not morphological enquiry.[27] In their chapter on compound nouns in Polish, Grzegorczykowa/Puzynina (1984: 396) recognise only two types of compounds, i.e. compounds proper and solid compounds. They do not devote any attention to juxtapositions. Kallas (1980) treats coordinate multi-word units, such as *kobieta pilot* 'woman pilot' and *lalka-niemowlak* (doll baby) 'baby doll', as free syntactic combinations and analyses them in the same way as (regular) noun phrases in apposition, such as *mleko – cenny pokarm* 'milk – precious food'.

Nagórko (1997), in her brief but insightful account of Polish grammar, postulates a strict division between syntax, phraseology and the lexicon. Consequently,

27 Grzegorczykowa (1982: 59) mentions the existence of juxtapositions, such as *czarna jagoda* (black berry) 'bilberry' and *maszyna do pisania* (machine for typing) 'typewriter', yet she notes that they do not constitute the subject matter of word-formation proper.

in her chapter on Polish syntax (Chapter V), she notes the occurrence of conventionalised phraseological units but concludes that from the point of view of syntax such strings of words are indivisible (Nagórko 1997: 189).[28] Her conclusion refers both to idiomatic multi-word units, such as *kocie łby* (cat.RA head.NOM.PL) 'cobblestones' or *pies ogrodnika* (dog.NOM.SG gardener.GEN.SG) 'dog in the manger', as well as semantically regular juxtapositions, e.g. *kosz na śmieci* (bin for rubbish) 'rubbish bin' and *gwiazda polarna* (star polar) 'pole star, Polaris'. In a modular framework (such as the one assumed by Nagórko 1997) it is difficult to draw a rigid and uncontroversial border between lexical multi-word units and freely composed phrases. While such N+N combinations as *człowiek instytucja* (man institution) 'one-man-institution' or *kobieta szef* (woman boss) 'female boss' are regarded by Nagórko (1997: 190f.) as syntactic units (consisting of a head noun and a nominal attribute), other N+N juxtapositions, such as *lekarz pediatra* (physician pediatrician) 'pediatrician' and *szpital-pomnik* (hospital monument) 'memorial hospital', are recognised as lexical units.

Such a strict separation of modules of grammar, i.e. morphology, syntax and the lexicon, is characteristic both of structuralist linguistics and of generative framework.[29] Syntax and morphology do not interact, and the lexicon is treated as a collection of irregularities (Bloomfield 1933; Di Sciullo/Williams 1987), i.e. a list of items which carry unpredictable semantic information and/or exhibit other idiosyncratic properties.

A markedly different view of the lexicon and the architecture of grammar is postulated in Construction Grammar (Goldberg 2006), Parallel Architecture and Construction Morphology (Masini 2009; Booij 2010; Masini/Benigni 2012; Booij/Audring 2015; Booij/Masini 2015). The lexicon, referred to as the constructicon, is viewed as a network of construction schemas of varying degrees of abstractness. Schemas are instantiated by fully specified constructions, which are also stored in the lexicon. Such constructions can take the form of syntactic strings, words or units with an intermediate (i.e. both lexical and syntactic) status.

28 Phraseological units are treated as indivisible from the point of view of syntax as well as semantics also by Grochowski (1982). Cf., however, Lewicki (1976) and Węgrzynek (1998) for some discussion of the internal syntax of idioms in Polish.

29 N+A phrasal nouns are recognised as free syntactic combinations by, among others, Rutkowski/Progovac (2005), who are proponents of the Minimalist Program, and by Szymanek (2010), who advocates the lexicalist approach. Willim (2001) regards N+A and N+N multi-word units, such as *ogród zoologiczny* (garden zoological) 'zoo' and *kobieta-anioł* (lit. woman angel) 'angel of a woman' as syntactic constructs, basing her analysis on the discussion of Greek A+N combinations by Ralli/Stavrou (1998). Syntactic constructs are treated as syntactic compounds (i.e. phrasal lexemes) by Booij (2010).

In their cross-linguistic accounts of phrasal nouns, Booij (2010), Masini/Benigni (2012), Booij/Masini (2015), Booij/Audring (2015) formulate phrasal schemas which act both as redundancy statements, which are able to analyse the internal structure of conventionalised multi-word units, and as templates for forming novel multi-word expressions. Similar schemas, postulated for Polish phrasal nouns below, show that phrasal lexemes have the properties of both lexical and syntactic items. On the one hand, phrasal nouns show a complex internal structure analysable by means of phrasal schemas (which may be also employed in analysing the structure of freely composed syntactic units). On the other hand, they have a naming function, which is signalled by the element NAME in the statement of their meaning.

The phrasal schema in (31) can be employed to form novel N+A phrasal nouns, and analyse the structure of such conventionalised units as *kurier dyplomatyczny* (courier diplomatic) 'diplomatic courier' and *telefon komórkowy* (phone cellular) 'mobile phone'. The symbol "E" in (31) stands for the entity denoted by the nominal base of the relational adjective in a given multi-word unit, e.g. *dyplomata* 'diplomat' or *dyplomacja* 'diplomacy' (as the base of *dyplomatyczny* 'diplomatic'), and *komórka* 'cell' (as the base of *komórkowy* 'cellular').

(31) $[N^0_i\ A^0_j]_k \leftrightarrow$ [NAME for SEM_i with some relation R to entity E of $SEM_j]_k$

Since some N+A strings contain classifying adjectives which are not denominal, e.g. *panda wielka* (panda great) 'giant panda', the schema in (32) can account for their structure.

(32) $[N^0_i\ A^0_j]_k \leftrightarrow$ [NAME for SEM_i with property $SEM_j]_k$

A classifying adjective (be it relational or a non-derived one) can stand in the prehead position in a phrasal noun in Polish. Consequently, two more schemas are necessary, to account for the structure of RA+N phrasal nouns, e.g. *nocny dyżur* 'night shift' (where the relational adjective *nocny* is derived from *noc* 'night') and A+N units which contain a non-derived or deverbal adjective, e.g. *głuchy telefon* (deaf phone) 'Chinese whispers', *odżywczy krem na noc* (nourishing cream for night) 'nourishing night cream'.

(33) $[A^0_i\ N^0_j]_k \leftrightarrow$ [NAME for SEM_j with some relation R to entity E of $SEM_i]_k$

(34) $[A^0_i\ N^0_j]_k \leftrightarrow$ [NAME for SEM_j with property $SEM_i]_k$

Another phrasal schema, given in (35) below, can be postulated for N+N.GEN phrasal nouns, both transparent semantically and idiomatic ones, e.g. *prawa człowieka* (right.NOM.PL man.GEN.SG) 'human rights', and *pies ogrodnika* (dog. NOM.SG gardener.GEN.SG) 'dog in the manger'.

(35) $[N^0_i \text{ N-GEN}_j]_k \leftrightarrow [\text{NAME for SEM}_i \text{ with some relation R to SEM}_j]_k$

The schema for coordinate N+N juxtapositions, such as *kelner-barman* 'waiter-bartender', is shown below:

(36) $[N^0_i \ N^0_j]_k \leftrightarrow [\text{NAME for an entity which is both SEM}_i \text{ and SEM}_j]_k$

In the non-modular model of grammar, characteristic of Construction Morphology, the strict lexicon-syntax divide is abandoned. Syntax and morphology closely interact and compete with each other. Consequently, multi-word units which are lexical items "are an expected phenomenon within the constructionist view of the language architecture rather than an exception or a marginal case" (Masini/Benigni 2012: 448).

Another phenomenon which is expected within the model of Construction Morphology is the competition between phrasal patterns, which motivate phrasal lexemes, and morphological schemas, which motivate compounds proper or derivatives. The competition was illustrated above (in Section 6) for coordinate juxtapositions and coordinate compounds proper (with a linking vowel), such as *chłop-robotnik* and *chłoporobotnik*, both paraphrasable as 'peasant farmer who works in a factory'.

In Polish, as in other Slavonic languages (cf. Masini/Benigni 2012, Ohnheiser 2015 and the chapter on Russian, this volume), phrasal lexemes can undergo morphological condensation (i.e. univerbation) and act as (semantic) bases for suffixal derivatives. The derivative *budowlanka* (which contains the denominal adjective *budowlany* 'relating to building' and the nominalizing suffix *-ka*) is (roughly)[30] synonymous to the phrasal noun *szkoła budowlana* (school building. RA) 'secondary technical school of building'.

Interaction between phrasal lexemes and derivatives (or compounds proper), exemplified by univerbation, can be accounted for in Construction Morphology by means of second order schemas (as in Booij/Masini 2015, see also the chapter

30 Suffixal derivatives resulting from morphological condensation, such as *budowlanka* 'secondary technical school of building', are additionally marked as belonging to colloquial Polish (cf. Ohnheiser 2015).

on Dutch, this volume). Such schemas state paradigmatic relations between word-formation schemas and phrasal schemas.

(37) $<[N^0_i\ A^0_j]_k \leftrightarrow$ [NAME for SEM_i with some relation R to entity E of SEM_j $]_k>$
$\approx <[\ A\text{-}ka]_{Nz} \leftrightarrow [SEM_k\ [+familiar]]_z>$

The second order schema given above states that deadjectival nouns terminating in the suffix *-ka* can be motivated by (i.e. semantically related to) phrasal N+RA lexemes.

8 Conclusion

This chapter offered a brief overview of multi-word expressions in Polish, focusing on phrasal nouns (which are often referred to as "juxtapositions") and their interaction with compound nouns. The following subtypes of juxtapositions were discussed at greater length: N+N.GEN, N+A, A+N, and coordinate N+N phrasal lexemes. Juxtapositions do not meet the majority of the criteria for morphological compounds (as stated by Lieber/Štekauer 2009). A morphological compound in Polish, i. e. a compound proper, is written as one orthographic word and inflected like one morphological word (with the inflectional endings attached to the right-hand constituent). It carries one primary lexical stress (typically on the penultimate syllable). A juxtaposition, in contrast, consists of two or more orthographic words, each of which is inflected. Constituents of a juxtaposition can carry independent lexical stresses, e. g. *mĄż stAnu* (man.NOM state.GEN) 'statesman'. On the other hand, juxtapositions act as naming units, therefore they can be regarded as multi-word lexical items. It is important to emphasise here that phrasal nouns in Polish are far from being exclusively idiomatic and unanalysable multi-word expressions. While selected multi-word units are semantically non-compositional (and can be treated as figurative idioms), e. g. *biały kruk* (white raven) 'rare specimen', the majority of phrasal nouns in Polish show varying degrees of semantic transparency. They are also analysable syntactically, which results in some degree of their syntactic mobility, as is shown above for coordinate N+N juxtapositions and for phrasal nouns consisting of a head noun and a relational adjective. The syntactic analysability of phrasal nouns also tallies with the fact that their constituents are inflected as independent morphological words.

The approach of Construction Morphology allows the researcher to provide a proper account of the above-mentioned properties of phrasal nouns in Polish. Multi-word units inherit their syntactic structure from construction schemas. In

other words, phrasal construction schemas can be employed to analyse the internal structure of existing phrasal nouns. The construction schemas state that phrasal nouns are generally interpreted as "names of kinds" (i.e. as subtypes of entities), e.g. *droga dojazdowa* (road access.RA) 'access road', *miernik promieniowania* (meter.NOM radiation.GEN) 'radiation meter', *kierowca-dostawca* (driver.NOM supplier.NOM) 'delivery driver'. Phrasal schemas can be used not only as redundacy statements (to license conventionalised phrasal nouns), but also as patterns for creating novel multi-word units. The latter function of schemas is particularly important in Polish since the patterns for phrasal nouns discussed above are very productive. Novel phrasal lexemes abound in Polish, e.g. in the vocabulary associated with the Internet technology, as is illustrated by such multi-word units as *dostawca usług internetowych* (provider.NOM.SG service.GEN.PL Internet.RA.GEN.PL) 'Internet service provider', *pióro świetlne* (pen light.RA) 'light pen', *ekran dotykowy* (screen touch.RA) 'touch screen', *telefon z klapką* (phone with flip) 'clamshell phone'. Schemas for multi-word units in Polish both compete with and complement patterns of compounding. As was shown in Section 6, fairly numerous examples can be found of co-existence of synonymous compound nouns and phrasal nouns in Polish, such *licencjodawca* (licence+LV+giver) and *dawca licencji* (giver.NOM licence.GEN) 'licensor'. However, the formation of synthetic compounds appears to be more restricted than the coinage of N+N.GEN or N+A multi-word units. Moreover, some types of naming units can be formed only by using phrasal schemas, e.g. attributive N+N compounds, such as *człowiek-zagadka* (man mystery) 'mystery man', and coordinate phrasal nouns consisting of units denoting Kinship+Profession, e.g. *mąż prawnik* (husband lawyer) 'lawyer husband'. Finally, it was shown that multi-word units need to be accessible to affixation and compounding processes (i.e. to morphological construction schemas), as they undergo morphological condensation. Such evidence indicates that the study of both morphologically complex words (such as compounds proper) and multi-word units should be of interest to morphologists. Researchers should pay greater attention to the interaction between phrasal lexemes and morphologically complex words in Polish, which is the kind of phenomenon that can find an appropriate account within the framework of Construction Morphology.

References

Anderson, Stephen (1992): A-morphous morphology. Cambridge, UK: Cambridge University Press.

Bloomfield, Leonard (1933): Language. New York: Holt, Rinehart and Winston Inc.

Booij, Geert (2010): Construction morphology. Oxford: Oxford University Press.

Booij, Geert/Audring, Jenny (2015): Construction morphology and the parallel architecture of grammar. In: Cognitive Science 41, 2. 277–302.
Booij, Geert/Masini, Francesca (2015): The role of second order schemas in the construction of complex words. In: Bauer, Laurie/Körtvélyessy, Livia/Štekauer, Pavol (eds.): Semantics of complex words. Cham etc.: Springer. 47–66.
Buttler, Danuta (1976): Innowacje składniowe współczesnej polszczyzny. Warszawa: Państwowe Wydawnictwo Naukowe.
Cetnarowska, Bożena (2014): On pre-nominal classifying adjectives in Polish. In: Bondaruk, Anna/Dalmi, Gréte/Grosu, Alexander (eds): Topics in the syntax of DPs and agreement. Amsterdam/Philadelphia: Benjamins. 100–127.
Cetnarowska, Bożena (2015): The linearization of adjectives in Polish noun phrases: Selected semantic and pragmatic factors. In: Bondaruk, Anna/Prażmowska, Anna (eds.): Within language, beyond theories. Vol. 1: Studies in theoretical linguistics. Newcastle upon Tyne: Cambridge Scholars Publishing. 188–205.
Cetnarowska, Bożena (2016): Identifying (heads of) copulative appositional compounds in Polish and English. In: Körtvélyessy, Lívia/Štekauer, Pavol/Valera, Salvador (eds.): Word-formation across languages. Newcastle upon Tyne: Cambridge Scholars Publishing. 51–71.
Cetnarowska, Bożena (2018): Phrasal names in Polish: A+N, N+A and N+N units. In: Booij, Geert (ed.): The construction of words. Advances in construction morphology. Cham: Springer International Publishing AG. 287–313.
Cetnarowska, Bożena/Pysz, Agnieszka/Trugman, Helen (2011): Distribution of classificatory adjectives and genitives in Polish NPs. In: Dębowska-Kozłowska, Kamila/Dziubalska-Kołaczyk, Katarzyna (eds.): On words and sounds: A selection of papers from the 40th PLM, 2009. Newcastle upon Tyne: Cambridge Scholars Publishing. 273–303.
Cetnarowska, Bożena/Trugman, Helen (2012): Falling between the chairs: Are classifying adjective+noun complexes lexical or syntactic formations? In: Błaszczak, Joanna/Rozwadowska, Bożena/Witkowski, Wojciech (eds.): Current issues in generative linguistics: Syntax, semantics and phonology. Wrocław: University of Wroclaw. 138–154.
Damborský, Jiří (1966): Apozycyjne zestawienia we współczesnej polszczyźnie. In: Język Polski 46, 4. 255–268.
Di Sciullo, Anna-Maria/Williams, Edwin (1987): On the definition of word. Cambridge, MA: The MIT Press.
Długosz-Kurczabowa, Krystyna/Dubisz, Stanisław (1999): Gramatyka historyczna języka polskiego: Słowotwórstwo. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
Fellbaum, Christiane (2011): Idioms and collocations. In: Maienborn, Claudia/von Heusinger, Klaus/Portner, Paul (eds.): Semantics. An international handbook of natural language meaning. Vol. 1. Berlin/New York: De Gruyter. 441–456.
Goldberg, Adele (2006): Constructions at work. The nature of generalization in language. Oxford: Oxford University Press.
Granger, Sylviane/Paquot, Magali (2008): Disentangling the phraseological web. In: Granger, Sylviane/Meunier, Fanny (eds.): Phraseology. An interdisciplinary perspective. Amsterdam/Philadelphia: Benjamins. 27–49.
Grochowski, Maciej (1982): Zarys leksykologii i leksykografii. Toruń: Wydawnictwa Uniwersytetu Mikołaja Kopernika.
Grzegorczykowa, Renata (1982): Zarys słowotwórstwa polskiego. Słowotwórstwo opisowe. 5th edn. Warszawa: Państwowe Wydawnictwo Naukowe.

Grzegorczykowa, Renata/Puzynina, Jadwiga (1984): Słowotwórstwo rzeczowników. In: Grzegorczykowa, Renata/Laskowski, Roman/Wróbel, Henryk (eds.): Gramatyka współczesnego języka polskiego. Morfologia. Warszawa: Państwowe Wydawnictwo Naukowe. 332–407.
Hacken, Pius ten (2013): Compounds in English, in French, in Polish, and in general. In: SKASE Journal of Theoretical Linguistics 10. 97–113.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.). 450–467.
Jadacka, Hanna (2001): System słowotwórczy polszczyzny (1945–2000). Warszawa: Wydawnictwo Naukowe PWN.
Jadacka, Hanna (2005): Kultura języka polskiego. Fleksja, słowotwórstwo, składnia. Warszawa: Wydawnictwo Naukowe PWN.
Kallas, Krystyna (1980): Grupy apozycyjne we współczesnym języku polskim. Toruń: Wydawnictwo Uniwersytetu Milkołaja Kopernika.
Klemensiewicz, Zenon (1939): Gramatyka współczesnej polszczyzny kulturalnej w zarysie. Lwów/Warszawa: Książnica-Atlas.
Kolbusz-Buda, Joanna (2014): Compounding. A morphosemantic analysis of synthetic deverbal compound nouns in Polish in the light of parallel constructions in English. Lublin: Wydawnictwo KUL.
Lewicki, Andrzej Maria (1976): Wprowadzenie do frazeologii syntaktycznej. Katowice: Uniwersytet Śląski.
Lieber, Rochelle/Štekauer, Pavol (2009): Introduction: Status and definition of compounding. In: Lieber, Rochelle/Štekauer, Pavol (eds.). 3–18.
Lieber, Rochelle/Štekauer, Pavol (eds.) (2009): The Oxford handbook of compounding. Oxford/New York: Oxford University Press.
Linde-Usiekniewicz, Jadwiga (2013): A position on classificatory adjectives in Polish. In: Studies in Polish Linguistics 8. 103–125.
Masini, Francesca (2009): Phrasal lexemes, compounds and phrases: A constructionist perspective. In: Word Structure 2, 2. 254–271.
Masini, Francesca/Benigni, Valentina (2012): Phrasal lexemes and shortening strategies in Russian: The case for constructions. In: Morphology 22, 3. 417–451.
Müller, Peter O. et al. (eds.) (2015–2016): Word-formation: An international handbook of the languages of Europe. (= Handbooks of Linguistics and Communication Science (HSK) 40). Berlin/Boston: De Gruyter.
Nagórko, Alicja (1997): Zarys gramatyki polskiej. 2nd edn. Warszawa: Wydawnictwo Naukowe PWN.
Nagórko, Alicja (2016): Polish. In: Müller, Peter O. et al. (eds.). 2831–2852.
NKJP = National Corpus of Polish. Internet: http://nkjp.pl. (last access: 2.7.2018).
Nunberg, Geoffrey/Sag, Ivan A./Wasow, Thomas (1994): Idioms. In: Language 70. 491–538.
Ohnheiser, Ingeborg (2015): Compounds and multi-word expressions in Slavic. In: Müller, Peter O. et al. (eds). 757–779.
Olsen, Susan (2001): Copulative compounds: A closer look at the interface between syntax and morphology. In: Yearbook of Morphology 2000. 279–320.
Polański, Kazimierz (ed.) (1999): Encyklopedia językoznawstwa ogólnego. 2nd edn. Wrocław: Zakład Narodowy im. Ossolińskich.
Puzynina, Jadwiga (1974): Związki frazeologiczne a derywaty (na materiale języka polskiego). In: Prace Filologiczne 25. 441–446.

Ralli, Angela/Stavrou, Melita (1998): Morphology-syntax interface: A+N compounds and A+N constructs in modern Greek. In: Booij, Geert/van Marle, Jaap (eds.): Yearbook of morphology 1997. Dordrecht: Springer. 243–264.
Renner, Vincent/Fernández-Domínguez, Jesús (2011): Coordinate compounding in English and Spanish. In: Poznań Studies in Contemporary Linguistics 47. 873–883.
Rutkowski, Paweł/Progovac, Ljiljana (2005): Classification projection in Polish and Serbian: The position and shape of classifying adjectives. In: Franks, Steven/Gladney, Frank Y./ Tasseva-Kurktchieva, Mila (eds.): Formal approaches to Slavic linguistics: The South Carolina Meeting. Ann Arbor, MI: Michigan Slavic Publications. 289–299.
Scalise, Sergio/Bisetto, Antonietta (2009): Classification of compounds. In: Lieber, Rochelle/ Štekauer, Pavol (eds). 49–82.
Skorupka, Stanisław (1967): Słownik frazeologiczny języka polskiego. (2 Vols.). Warszawa: Wiedza Powszechna.
Sprenger, Simone (2003): Fixed expressions and the production of idioms. [Ph.D. dissertation, Max Planck Instituut voor Psycholinguïstiek]. Nijmegen: Max-Planck Institut fur Psycholinguistik.
Szerszunowicz, Joanna (2012): English-Polish contrastive phraseology. In: Rozumko, Agata/ Szymaniuk, Dorota (eds.): Directions in English-Polish contrastive research. Białystok: Wydawnictwo Uniwersytetu Białostockiego. 139–162.
Szumska, Dorota (2006): Przymiotnik jako przyłączone wyrażenie predykatywne. Analiza formalizacji struktur propozycjonalnych w warunkach predykacji niezdaniotwórczej. Kraków: UNIVERSITAS.
Szymanek, Bogdan (2010): A panorama of Polish word-formation. Lublin: Wydawnictwo KUL.
Węgrzynek, Katarzyna (1998): Składnia wyrażeń frazeologicznych w modelu gramatyki generatywno-transformacyjnej. In: Polonica 19. 67–74.
Willim, Ewa (2001): On NP-internal agreement: A study of some adjectival and nominal modifiers in Polish. In: Zybatow, Gerhild et al. (eds.): Current issues in formal Slavic linguistics. Frankfurt a.M. i.a.: Lang. 80–95.
Żmigrodzki, Piotr (2009): Wprowadzenie do leksykografii polskiej. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Irma Hyvärinen

Compounds and multi-word expressions in Finnish

1 Introduction

Most of the processes to expand the vocabulary of a language are based on a recycling principle: Instead of creating not yet occupied arbitrary sound sequences for new concepts, existing lexemes or morphemes are reused as material for new words. This can happen by borrowing a word from some other language or by altering the meaning and thus shifting the extension of an existing word. Yet, these means are fairly unsystematic. Instead, a system of word-formation offers productive models for expanding the lexicon in an economic way, and it is actually the most common way it happens.[1]

Word-formation types such as (1a–f) are usually regarded as a domain of morphology:

(1a) Composition (combining lexemes into a new lexical item):
kesä 'summer' + *yö* 'night' > *kesäyö* 'summer night'

(1b) Derivation (adding an affix):
kesä 'summer' + *-inen* (adjectival suffix) > *kesäinen* 'summery'

(1c) Backformation (removing an actual or supposed affix):
tarrata 'grab, stick' > *tarra* 'sticker'

(1d) Conversion, also called zero derivation (functional shift of a word or a stem[2] without adding morphological material):
minä 'I' (Pron) > *minä* 'ego' (N); *painia* 'wrestle': *paini-* (verb stem) > *paini* (N) 'wrestling'

1 Foreign influence can manifest itself in word formation, too, as calques of singular formations or by taking over a formation model from another language. Many Finnish compounds are loan translations from (or via) Swedish or German. Nowadays loan translations come increasingly from English, cf. *jakamis+talous* < *sharing economy*, *palvelu+muotoilu* < *service design*. In terminology, neoclassical compounds (with elements from Greek or Latin) as internationalisms play an important role.

2 The word stem is the form to which affixes can be attached. As for word stems in Finnish, cf. ISK (2004: 86–89).

https://doi.org/10.1515/9783110632446-011

(1e) Blending (merging parts of existing lexemes combining their semantic features):
kamraati 'comrade' + *toveri* 'companion, friend' > *kaveri* 'friend, mate'

(1f) Clipping (shortening a lexeme without changing the meaning):
akkumulaattori > *akku* 'accumulator'; *informaatioteknologia* > IT 'information technology'; *sosiaaliturva* > *sotu* 'social security'

However, also syntactic (phrasal) sequences can be lexicalized as nominations of specific concepts. Such multi-word expressions (MWEs) can be included in a discussion of word formation in a broad sense. MWEs are fixed word-groups with lexical, syntactic, semantic, pragmatic and/or statistical idiosyncrasies (Sag et al. 2002; Baldwin/Kim 2010; Hüning/Schlücker 2015). The term "multi-word expression" is established above all in computational linguistics; traditionally MWEs are called "phrasemes" or "idioms".[3] In this chapter, the term "idiom" is used for semantically idiosyncratic MWEs only, i.e. for cases where the meaning of an MWE cannot be concluded from the meanings of its components. MWEs can be fully idiomatic (2a–b), semi-idiomatic (2c) or non-idiomatic but statistically significant (institutionalized) (2d–e):[4]

(2a) *mennä mönkään* (lit. go UNIQUE COMPONENT) 'go wrong'[5]

(2b) *musta hevonen* (lit. black horse) 'dark horse' (a little known candidate or competitor who unexpectedly wins or succeeds)

(2c) *valkoinen valhe* 'white lie' (a harmless lie)

(2d) *rauhanomainen rinnakkaiselo* 'peaceful coexistence' (theory of the Soviet Union about relations between socialist and capitalist states during the Cold War)

(2e) *neoliittinen kausi* (altering with the compound *neoliitti*+*kausi*) 'Neolithic Period'

3 An overview of phraseology with examples from several European languages, e.g. German, English, French, Swedish and Finnish, is given by Korhonen (2018).

4 In Finnish, non-idiomatic MWEs have been studied primarily in terminology with a focus on nominal terms. It can be assumed that ongoing studies in computational linguistics will shed more light on the proportion of non-idiomatic MWEs in standard language, too.

5 In the examples, the compound constituent boundaries are marked with "+", if needed. Occasionally Finnish case form abbreviations are used as subscripts: ALL = allative, ELAT = elative, GEN = genitive, ILL = illative, INESS = inessive, PART = partitive.

The boundaries between different formation types are not always clear-cut: Compound nouns often compete with MWEs, for example as constructional synonyms in terminology, cf. (2e) above. Some Finnish compounds have internal inflectional elements, which is a syntactic feature (cf. Section 2.1). Moreover, there are hybrid formations, like the so-called "derived compounds" (Section 2.3.1.1, group 2). And finally, scholars have divergent views of certain structures, such as Finnish particle verbs that have been classified either as compounds, prefix derivations or MWEs (Section 3).

Compounds and MWEs share some characteristics: Both are complex lexical units and thus secondary signs for a specific concept, their constituents are words, and they can bear an idiomatic (figurative or opaque) or non-idiomatic (transparent) meaning. One instance of opaqueness is presented by unique components (isolates, cranberry morphemes), compare the MWE in (2a) with the cranberry-compound *puna+tulkku* (lit. red+UNIQUE COMPONENT) 'bullfinch' (cf. Nenonen 2002: 13, 15, 21f., 37–40; Stein 2012: 227f.). Both compounds and MWEs can express determinative, appositive and coordinative relations. The compound constituents occur in a fixed order; regarding MWEs this applies mainly to nominal, adjectival and adverbial expressions, whereas verbal MWEs are more flexible. In Finnish, the great majority of compounds are nouns (N), while among idiomatic MWEs verb idioms (V) are the predominant class.

In this chapter, the focus is on the characteristics of compounds, with remarks on differences and overlap in the structure and syntactic distribution of compounds and (fixed or free) phrasal units. Section 2 gives an overview of compounding in Finnish, mostly using examples of nouns and adjectives:[6] In Section 2.1 characteristics of prototypical compounds and their absence, making a compound less prototypical and bringing it nearer to an MWE, are discussed. Section 2.2 deals with the complexity of compounds, and in Section 2.3 the main semantic-hierarchical and morphosyntactic types of compounds are presented. Section 3 focuses on a word class that has been regarded as rather peripheral from the perspective of compounding in Finnish, namely complex verbs. They are interesting for two reasons: They are on the increase in modern Finnish, and they lie at the intersection of compounds (3.1), prefix derivatives (3.2) and MWEs (3.3). In the closing remarks (Section 4) observations on the blurred border between Finnish compounds and MWEs are gathered and suggestions for future research are presented.

6 Due to lack of space, a thorough description of all word classes is impossible.

2 Compounding in Finnish

2.1 Prototypical compounds

Finnish has an extensive system of word-formation: Both derivation and compounding are highly productive. In particular the diversity and productivity of suffix derivation is often regarded as a special characteristic of Finnish, but, actually, the majority of new words in modern Finnish are compounds (cf. Tyysteri 2015: 13, 223). Verbs, however, show a different profile: There is a rich and productive suffixation system, whereas compounding plays a marginal role. Yet, in the last decades the number of compound verbs has increased.

A compound is a combination of two or more lexemes constituting a new, complex word with a new lexical-conceptual meaning that is generally more specific than the additive meaning of its parts, e.g. *märkä+puku* 'wetsuit' (water sports garment) vs. *märkä puku* 'wet suit'. The constituents can be simplex lexemes, derivatives or even compounds, i.e. compounding is potentially recursive. In contrast to derivatives, vowel harmony (cf. Karlsson 2015: 16ff.) does not extent over the constituent boundary (Koivisto 2013: 170), i.e. the integration grade of compounds is lower, compare the suffix derivatives with vowel harmony in (3a) with the compounds in (3b):

(3a) Verb stem + suffix *-jA* → *juoja* 'drinker' vs. *syöjä* 'eater'
(3b) *yö+juna* (**yö+jynä*) 'night train', *varpus+pöllö* (**varpus+pollo*) (lit. sparrow+owl) 'pygmy owl'

The main characteristics of prototypical compounds in Finnish are: 1) The constituents occur also as autonomous lexemes, 2) the boundary between the constituents corresponds to a syntactic boundary, 3) the compound has only one main stress that – just as in simplex words – is on the first syllable, 4) a formally identical phrasal unit is not possible, 5) semantically, the compound has become estranged from the meanings of its constituents and lexicalized into a nomination of a concept of its own, 6) morphologically, the compound is internally invariable. Among new compounds in present-day Finnish the proportion of prototypical compounds is increasing, whereas non-prototypical features accumulate on one and the same words. However, counter to the trend towards prototypicality, formations with a non-autonomous pre-element (cf. criterion 1) are on the advance. Further, deriving new verbs and adjectives from already existing compounds, which leads to secondary "derived compounds" where the constituent boundary deviates from the logical syntactic-semantic structure (cf. criterion 2

and Section 2.3.1.1, group 2), has become more common than earlier (Tyysteri 2015).

As a general rule, Finnish compounds are written without space between the constituents, cf. (4) below. Hyphenation is obligatory in case of hiatus (5a) and to indicate the constituent boundary after a special sign (letter, number, acronym etc.) (5b). A compound differs also prosodically from a phrase: The main stress is on the first compound constituent (cf. criterion 3), while in a corresponding phrase both words have a stress of their own (Pääkkönen 1989: 371; Vesikansa 1989: 213; ISK 2004: 388), cf. (6). Yet, stress is not a reliable criterion: Adverbial and conjunctional units (7) bear only one main stress on the first part and show a strong tendency towards univerbation. Until the 1960s they could be written together or apart, today the orthographical norm requires separation and thus an MWE status for them, which is in contradiction with the stress pattern (Niinimäki 1992).

(4) *metsäyhtiö* (lit. forest+company) 'forestry company'

(5a) *öljy-yhtiö* 'oil company'
(5b) *A-vitamiini* 'vitamin A'

(6) *músta+rastas* (lit. black+thrush) 'blackbird' vs. *músta rástas* 'black thrush'

(7) *sítä vastoin* (lit. it$_{\text{PART}}$ against) 'by contrast'
níin ollen (lit. so being) 'thus, hence'
níin kuin (lit. so as) 'as, as if'

Generally, adjectival compounds can have descriptive, graduating or evaluative modifiers in the genitive; semantically relative adjectives like *kokoinen* 'of the size of', *näköinen* 'looking like' etc. even require a complement in the genitive. Here compounds and phrasal units of identical parts are often interchangeable (cf. criteria 4 and 5), as illustrated in (8). Generally, conventionalized (especially idiomatized) combinations and those with short components undergo univerbation, but a grey zone remains, cf. (9a) and (10a) vs. (9b) and (10b).

(8) *sydämen+muotoinen* ~ *sydämen muotoinen* 'heart-shaped'
vaalean+vihreä ~ *vaalean vihreä* 'light green'
hassun+kirjava ~ *hassun kirjava* 'kooky colorful'

(9a) *ruohon+vihreä* 'grass-green' (conventional)
(9b) *pinaatin+vihreä* ~ *pinaatin vihreä* 'spinach green' (occasional)

(10a) *kissan+kokoisin kirjaimin* 'in letters big as a cat, in huge letters' (conventional idiom)
(10b) *kissan kokoinen rotta* 'rat having the size of a cat' (concrete compositional meaning)

In contrast, the first constituent of similative adjectives that expresses an entity for which the property denoted by the head is typical, is – regardless of its nominative or genitive form – always unified with the head, cf. (11a–b). Here, alternation with multi-word similes depends on syntactic distribution: Similative compounds can be replaced by phrasal similes in predicative (12a) and adverbial function but not in attributive function (12b). They cannot always be exchanged, though: While similative compounds are mostly lexicalized stereotypes and the first constituent cannot have its own qualifiers, the expression potential of phrasal similes is broader: They are based on a productive phraseosyntactic pattern that is filled with conventionalized (lexicalized) or occasional word combinations, and the component that denotes point of comparison can have supplementary expansions, cf. (13a–b). Thus similative adjectives and phrasal similes – both typically used for intensification – are partly in a competitive, partly in a complementary relation to each other (ISK 2004: 411; Heinonen 2010).

(11a) *jää+kylmä* 'ice-cold'
(11b) *langan+laiha* (lit. thread$_{GEN}$+thin) 'sceletous'

(12a) *Koira oli salaman+nopea* (lit. lightning$_{GEN}$+quick) ~ *Koira oli nopea kuin salama* (lit. quick as lightning). 'The dog was as quick as a lightning.'
(12b) *salamannopea koira* ~ **nopea kuin salama koira*

(13a) *hidas kuin etana* 'slow as a snail' (conventional idiom) ~ *etanan+hidas* (lit. snail$_{GEN}$+slow)
(13b) *hidas kuin halvaantunut etana* 'slow as a paralyzed snail' (occasional expansion) ~ **halvaantuneen+etanan+hidas* (lit. paralyzed$_{GEN}$+snail$_{GEN}$ +slow)

Morphological integrity of prototypical compounds means that they are internally invariable (cf. criterion 6); the morphological head bears the inflectional elements. However, in some compounds the adjectival first constituent can (14a) or must (14b) agree in case and number with the nominal head, which indicates their phrasal origin (cf. Niemi 2009: 239f.):

(14a) *iso+sisko* 'big sister' – allative: *iso**lle**+sisko**lle** ~ iso+sisko**lle***
(14b) *nuori+pari* (lit. young+couple) 'newlyweds' – allative: *nuore**lle**+pari**lle***

Internal congruence is a recessive property. Of the 587 A+N compounds in "Suomen kielen perussanakirja" (Basic dictionary of Finnish, 1990–1994) 84 % do not allow congruence, while the remaining 16 % are distributed fairly equally among compounds with obligatory vs. optional congruence. In neologisms and occasional compounds non-congruent first constituents are almost exclusive (ISK 2004: 392, 406; Tyysteri 2015: 141–148). A compound without internal inflection underlines the term character: *oma+lääkäri* is a personal doctor nominated for a certain patient by public health care (15a). In contrast, the corresponding (congruent) attributive NP refers to a non-administrative private choice made by the patient (15b):

(15a) *oma+lääkäri* (lit. own+doctor) – allative: *oma+lääkäri**lle***
(15b) *oma lääkäri* 'own doctor' – allative: *oma**lle** lääkäri**lle***

However, a special class of compounds with internal inflection remains: compound numerals (16a). To avoid overlong compounds, numerals with hundreds, thousands etc. are "cut" into groups (ISK 2004: 756f.) so that they combine features of MWEs and compounds (16b).

(16a) *kolme**lle**+kymmene**lle**+neljä**lle*** '34_{ALL}'
(16b) *kahde**lle**+kymmene**lle**+tuhanne**lle** seitsemä**lle**+sada**lle** kolme**lle**+kymmene**lle**+neljä**lle*** '20734_{ALL}'

2.2 Complexity of compounds

The majority of Finnish compounds consist of nominal compounds. The most common type is a combination of two base (i.e. non-derived) nouns (N+N), the largest group being determinative compounds (cf. Section 2.3.1.1) with the first constituent in the (endingless) nominative case (Karlsson 2015: 282; Pitkänen-Heikkilä 2016: 3213). The typical base word structure in Finnish is bisyllabic,[7] so even compounds with two base words have mostly at least four syllables (17a), and since derivatives and compounds can function as compound constituents as well, Finnish compounds tend to be long (Karlsson 2004: 1329), cf. (17b). In principle, there is no upper limit on the complexity, but increasing complexity diminishes intelligibility. As a consequence of recursiveness, compounds with four or five components are not rare in languages for special purposes (e.g. administra-

7 There are less than 100 monosyllabic word roots in Finnish, whereas English has at least 7.000 (Karlsson 2004: 1329).

tion, medicine etc.), yet, (mostly occasional) polymorphemic compounds appear also in everyday language (17c). In Tyysteri's corpus[8] two-constituent compounds dominated with 83,6 %, whereas the ratio of three-constituent compounds ran into 15,5 % and that one of four-constituent compounds into 0,9 %; longer formations occurred only sporadically (Tyysteri 2015: 100–104; as for letter number in compounds, cf. ibid.: 104–108).

(17a) *vesi+pullo* 'water bottle'
(17b) *työ+ehto+sopimus+neuvottelut* (lit. work+condition+contract+negotiations) 'negotiations for collective bargaining'
(17c) *peruna+sose+hiutale+pakkaus* (lit. potato+mash+flake+package) 'package of mashed-potato flakes'

2.3 Main types of compounds

2.3.1 Semantic-hierarchical structure

Like in many other languages, Finnish compounds can be categorized as either determinative (subordinate) or copulative (co-ordinate) compounds.

2.3.1.1 Determinative compounds

In determinative compounds the final constituent is the morphosyntactic and semantic head: It bears the inflectional elements and expresses a general concept that is modified by the initial constituent so that the compound denotes a subordinate concept (hyponym) to the head (18a). Such compounds are called endocentric (Olsen 2015: 365f., 370). The modifier is not referential but has a general meaning, which makes the compound semantically different from a corresponding phrase (ISK 2004: 390), cf. (18b). Whether the first constituent is morphologically underspecified (18a) or has a case ending explicating the syntactic relation between head and modifier, cf. (18b), varies from compound to compound.

(18a) *kivi+talo* 'stone house' (a special kind of house: 'house made of stone')
(18b) *kirkon+kello* (lit. church$_{GEN}$+bell) 'church bell' vs. (*läheisen*) *kirkon kello* 'bell of the (nearby) church'

8 Tyysteri's material consists of more than 28.000 new compounds (types) in Finnish print media in the period 2000–2009, collected from *Nykysuomen sanastotietokanta* (Lexical database of modern Finnish) of *Kotimaisten kielten keskus* (*Institute for the Languages of Finland*) (Tyysteri 2015: 79–84).

In Finnish grammar, the following special types are regarded as subclasses of determinative compounds:

1) In **synthetic compounds** the first constituent is comparable with the subject (19a), object (19b) or some other argument (19c) of the verb from which the head is derived (cf. ISK 2004: 400f.; Olsen 2015: 370f.). The first constituent typically has a case ending, which is a syntactic feature transmitted by the verb. Nominalizations with *-minen* are not univerbated with the verb arguments (20a), while deverbal nouns with other suffixes form compounds as well as phrasal NPs (20b).

(19a) *auringo**n**+nousu* (lit. sun_{GEN}+rise) 'sunrise'
(19b) *kirja**n**+sitoja* (lit. book_{GEN}+binder) 'bookbinder'
(19c) *kirko**ssa**+kävijä* (lit. $\text{church}_{\text{INESS}}$+goer) 'churchgoer'

(20a) *pyyki**n** peseminen* (lit. $\text{laundry}_{\text{GEN}}$ washing) 'washing laundry' vs. **pyyki**n**+peseminen*
(20b) *pyyki**n**+pesu* (lit. $\text{laundry}_{\text{GEN}}$+wash) ~ *pyyki**n** pesu*

2) Words with characteristics of both compounds and derivatives are regarded as secondary "**derived compounds**" (Vesikansa 1989, 213; ISK 2004, 388; Koivisto 2013, 334f.; Pitkänen-Heikkilä 2016, 3211). They can be analyzed as derivatives from complex bases, i.e. compound nouns (21a), adjectives (21b) or phrasal items (21c). Yet, language users tend to reanalyze them, setting the morphological main boundary intuitively as if they were "normal" compounds, even if this does not correspond to the logical syntactic-semantic boundary. In (21a), *perus* 'base' does not modify the word *koululainen* 'pupil' (e.g. in the sense of 'typical pupil'). Here, the reanalysis from analogical derivation into analogical compounding (i) gives a kind of short cut to build compounds directly (ii). By generalization the right half of the equation in (ii) becomes model character also in cases where one member is missing in the left half, cf. (21b–c) where **mukaistaa* or **pukuinen* do not occur as autonomous words.

(21a) *perus+koululainen* 'comprehensive school pupil' < *perus+koulu* (lit. base+school) 'comprehensive school'
  (i) *koulu* (simplex) : > *koulu* + *-lainen* (suffixation)
  perus+koulu (compound) > (*perus+koulu*) + *-lainen* (suffixation)
  > ***perus+koululainen*** (reanalysis into a compound)
  (ii) *koulu* : *koululainen* = *perus+koulu* : *perus+koululainen*

(21b) *ajan+mukaistaa* 'modernize, update'< *ajan+mukainen* (lit. time$_{GEN}$+in accordance with) 'up to date'
(21c) *musta+pukuinen* 'dressed in black' < *musta puku* 'black dress'

3) **Possessive compounds** (*bahuvrīhi*) such as (22a–b) show a semantic modification similar to determinative compounds, but, due to a metonymic shift, instead of expressing a subcategory of the concept expressed by the final constituent they "identif[y] the intended referent as the possessor of the particularly salient property" they express; i.e. they are exocentric (Olsen 2015: 367; cf. Vesikansa 1989: 250–254; ISK 2004: 409).

(22a) *kalju+pää* 'boldhead'
(22b) *puna+rinta* (lit. red+chest) 'robin'

Schellbach-Kopra (1964) assumed that bahuvrīhis are decreasing in modern Finnish, but Heinonen (2001) and Malmivaara (2004) demonstrate their productivity: They are used creatively for example in journalistic texts and colloquial speech.

4) **Confix compounds**. It is disputable if Finnish has prefixes at all. That is why formations with "prefix-like elements" are subsumed under compounding and not treated as a subclass of affixation or as a word-formation type of its own (cf. Pitkänen-Heikkilä 2016: 3212).[9] The indigenous negation pre-elements *epä-* 'un-' and *ei-* 'non-', cf. (23a), are often called prefixes, but as they consist of a lexical stem (cf. derivatives like *evätä* 'refuse, deny', *epäillä* 'doubt, mistrust'; *eittämätön* 'undeniable'), the result is very compound-like (ISK 2004: 192). There are many further indigenous "prefix-like elements" that do not occur as autonomous lexemes but have a more or less lexical meaning (ISK 2004: 192f., 393, 402, 415), cf. (23b). Consequently, they can be classified as **confixes** (cf. Fleischer/Barz 2012: 63f., 107f., 172ff.). As for verbs with confixes, cf. Section 3.1.

(23a) *epä+onni* 'bad luck'
epä+suomalainen 'un-Finnish'
ei-eurooppalainen 'non-European'

9 As for the theoretical status of prefixation in the history of linguistics, cf. Olsen (2015: 364f.).

(23b) *etä+työ* 'remote work'
pika+ateria 'quick meal'
täsmä+ase 'smart weapon'
vähä+arvoinen 'of little value'

Similarly to *epä-*, *ei-*, foreign negation prefixes, such as *dis-*, *in-*, are treated as compound components in Finnish, as well as other foreign prefixes and confixes, e.g. *ex-/eks-*, *pre-*, *hyper-*, *mikro-*, *poly-*, *neo-*, *audio-*, *anglo-*, *bio-*, *geo-*, *psyko-* which occur in **neoclassical compounds** (see Olsen 2015: 374f.), cf. (24a). Some foreign pre-elements can also be combined with indigenous heads (24b) (cf. Sajavaara 1989: 76ff.[10]; ISK 2004: 192, 394, 402; Pitkänen-Heikkilä 2016: 3214).

(24a) *dis+harmonia* 'disharmony'
neo+nataalinen (med.) 'neonatal'
(24b) *anti+sankari* 'anti-hero'
ex-vaimo 'ex-wife'

5) **Appositive compounds** describe one particular referent from different perspectives. In contrast to additive (copulative) compounds (cf. Section 2.3.1.2), the constituents do not belong to the same conceptual category, cf. (25a). Even if the constituents are in an appositive relation to each other, a determinative interpretation is possible (ISK 2004, 407f.). In (25b) it is actually the second constituent that modifies semantically the first one, thus having the same function as a postponed apposition (Vesikansa 1989: 223). Further appositive compounds include subsumptive (explicative) compounds (25c) where the second constituent expresses the hyperonym of the first constituent (ISK 2004: 408).

(25a) *prinssi+puoliso* 'prince consort'
(25b) *puu+vanhus* (lit. tree+oldster) 'old tree'
(25c) *veli+mies* (lit. brother+man) 'brother'
perjantai+päivä (lit. Friday+day) '(the weekday) Friday'

6) **Iterative compounds** repeating the same lexeme are productive primarily in informal, playful style of young people; in standard language they are a marginal class. Their main function is emphasis. In N+N reduplications the first constituent expresses the real, prototypical or ideal character of the concept denoted by

10 Sajavaara (1989: 79f.) also gives an overview of bound second constituents of neoclassical compounds in Finnish.

the head and implies a contrast (26). As for adjectives, cf. (27), the first constituent is mostly in the genitive (A_{GEN}+A) and functions as an intensifier; the components can be combined as a compound or a phrase without an essential difference in meaning (ISK 2004: 410; Tyysteri 2015: 66f.), similar to (8) above.

(26) *ruoka+ruoka* (lit. food+food) 'real food' (in contrast to fast food or unhealthy food)
kirja+kirja (lit. book+book) 'printed book' (in contrast to e-book).

(27) *pienen+pieni* (lit. small$_{GEN}$+small) ~ *pienen pieni* 'tiny, minuscule, itsy-bitsy'

2.3.1.2 Copulative compounds

Copulative compounds consist of two or more parallel (coordinate) parts belonging to the same word class and the same conceptual category; the rightmost constituent is the morphological head.

Historically, **co-compound** (*dvandva*) is the oldest compound type in the Finno-Ugric languages (Vesikansa 1989: 214; Pitkänen-Heikkilä 2016: 3213). "Co-compounds denote a hyperonym of their constituents, or a superordinate concept" (Olsen 2015: 368); hence, they are exocentric. In Finnish, only *maa+ilma* (lit. earth+air) 'world' has remained until our days.

Additive compounds make up a productive subclass of copulative compounds. Their constituents represent the same conceptual category and stand semantically in an additive relation, similar to members in a syntactic coordination (ISK 2004: 416f.; Pitkänen-Heikkilä 2016: 3213). In Finnish, appositive compounds are dissociated from additive ones orthographically: The former are written as one word, cf. (25a–c) above; the latter are generally written with a hyphen, cf. (28).

(28) *laulaja-näyttelijä* 'actor-singer'
jääkaappi-pakastin 'fridge-freezer'
musta-puna-keltainen 'black-red-yellow'

2.3.2 Morphosyntactic classification

The primary morphosyntactic classification criterion of Finnish compounds is the word class of the head which determines the word class of the compound. There are no head-based categorical restrictions for the non-head constituent; it

can be a stem, case form or a specific combining form.[11] The first component is usually classified on grounds of its word class (if identifiable) and/or its form (nominative, genitive, other case form, combining form, indeclinable element or element with deficient paradigm). Subclasses that arise from the cross classification of the morphosyntactic types of both constituents are described semantically in detail in the research literature, but no hard and fast rules can be given.

It is a controversial question to what extent the meaning of a compound is influenced by the form of the first constituent. The most frequent first constituent form in Finnish is the nominative which is the base form without any inflectional elements. This base form, as well as the combining forms, leaves the constituent relation underspecified so that several interpretations are possible. Inherently ambiguous compounds can be interpreted semantically and pragmatically, such as world knowledge of prototypical (e.g. local, temporal, causal, instrumental, possessive etc.) relations, common ground and contextual inference (cf. Olsen 2015: 365f., 376 ff., 382; Pitkänen-Heikkilä 2016: 3213). Lexicalized and frequently used compounds can be understood holistically, without analytic compositional processing, but there is psycholinguistic evidence that some form of analysis is co-present (Mäkisalo 2000). Räisänen (1986) points out that lexicalized compounds can be reinterpreted on contextual grounds: In a football report, *maa+pallo* (lit. earth+ball) and *ilma+pallo* (lit. air+ball) with the lexicalized meaning 'globe' resp. 'balloon' are interpreted in a context-adequate way as occasionalisms describing the motion of the ball either along the ground or through the air.

If the first constituent is in the genitive or some other non-nominative case, the interpretation is more restricted. In such cases the head is usually a deverbal noun and the first component corresponds to an argument of the underlying verb (synthetic compounds, cf. Section 2.3.1.1, group 1). A first constituent in the genitive can indicate a (in a broad sense) subjective-possessive (29a) or objective relation (29b); the latter is more common (Saukkonen 1973: 338; cf. ISK 2004: 400). Locative cases are also current (29c). It is noteworthy, however, that case marking is not obligatory: Similar relations can also be expressed by compounds with morphologically unspecified modifiers (30a–c).

11 A combining form (*casus componens*) is a form of the non-head constituent that as such does not occur as an autonomous word form. Besides non-autonomous stem forms, such as *nais-* < *nainen* 'woman' (*nais+ryhmä* 'women's group') or *pien-* < *pieni* 'small' (*pien+teollisuus* 'small industries'), there are specific combining forms with additional morphological material. For example, verbal first constituents appear mostly in a combining form with *-ma-* or *-in-* (*istu**ma**+paikka* (lit. sitting+place) 'seat', *leiv**in**+uuni* 'baking oven' (cf. also Tyysteri 2015: 121, 131, 134f.).

(29a) *ilmasto**n**+muutos* ‘climate change’
*tie**n**+vieri* ‘roadside’
(29b) *puu**n**+hakkaaja* ‘woodcutter’
*ilma**n**+suodatin* ‘air filter’
(29c) *maa**sta**+muutto* (lit. country$_{\text{ELAT}}$+moving) ‘emigration’
*tili**lle**+pano* (lit. account$_{\text{ALL}}$+put) ‘deposit, payment into an account’

(30a) *terroristi+hyökkäys* ‘terrorist attack’ (subjective)
(30b) *oppilas+valinta* ‘student selection, selection of pupils’ (objective)
(30c) *koti+matka* ‘home journey’ (adverbial: goal)

There are pairs of compounds with a nominative vs. genitive first constituent where the case choice seems more or less arbitrary (31a–b), and others where the difference in meaning is minimal (32a–b). Yet, sometimes there is a clear semantic opposition: (33a) is a specific house, whereas in (33b) the head describes an action and the first constituent in the genitive is the object argument of the underlying verb (cf. Vesikansa 1989, 230–237; ISK 2004, 398–400).

(31a) *kulta+keräys* ‘gold collection, collecting gold’
(31b) *paperi**n**+keräys* (lit. paper$_{\text{GEN}}$+collection) ‘(waste) paper collection’

(32a) *juusto+pala* ‘cheese piece’ (the first component focuses on material)
(32b) *juusto**n**+pala* (lit. cheese$_{\text{GEN}}$+piece) ‘piece of cheese’ (whole to part relation)

(33a) *sauna+rakennus* ‘sauna building’ (a special type of building)
(33b) *sauna**n**+rakennus* (lit. sauna$_{\text{GEN}}$+building) ‘building of a sauna/saunas’

Case marking on the constituent boundary does not contradict the principle of world-knowledge and context-based interpretation, but in giving further information on the relation between the constituents it can exclude alternatives that are possible when the first constituent is unmarked: While the underspecified form *pöytä+tarjoilu* (lit. table+service) can be used in the meaning ‘buffet service, self-service from the table’ (source), the marked form *pöyt**iin**+tarjoilu* (lit. tables$_{\text{ILLAT}}$+service) precludes this interpretation because the illative ending makes the opposite direction (goal) explicit.

3 Complex verbs in Finnish at the intersection of compounds, prefix derivatives and MWEs

In Finnish, compound verbs are rare.[12] They belong to the category of determinative compounds;[13] the first constituent is a noun, adjective, numeral, pronoun, non-autonomous stem or particle (adverb/adposition) (Rahtu 1984: 409–412; ISK 2004: 414f.). Verbs with a particle as first constituent are often replaced by MWEs with the same elements. On the other hand, some first constituents come near to prefixes. Thus, complex verbs can be explored on a scale MWE – compound – prefix derivative.

Modern Finnish has about 250 lexicalized compound verbs with a full paradigm, but the number is increasing (ISK 2004: 414). Additionally, formations with a deficient paradigm (mostly participle forms) are in use, and occasionalisms occur. Compound verbs were banned by Finnish language planning as loan translations for a long time. In the last decades the norm has become more permissive, which can explain the increasing occurrence (cf. Rahtu 1984: 409; Vesikansa 1989: 254–258; Vaittinen 2003: 50; Tyysteri 2015: 40, 154, 220f.).

There are three historical layers of compound verbs in Finnish: The oldest compound verbs, with an adverb as first constituent, are loan translations from the time of the Reformation. In the end of the 19th century a new type, derived from compound nouns, appeared. In the beginning of the 20th century also adjective compounds became derivation bases of verbs (Häkkinen 1987: 10–19; Vaittinen 2003: 47). Also in modern Finnish most of the compound verbs are secondary "derived compounds", i.e. derivatives or backformations from compound adjectives or nouns, such as (34a–c) (Vesikansa 1989, 256ff.; Tyysteri 2015: 153; cf. also Section 2.3.1.1, group 2). According to ISK (2004: 414f.), most present-day compound verbs are derived from complex adjectives ending on the suffix *-inen*, cf. (34a). According to Tyysteri (2015: 158, 213), however, the majority of the newest compound verbs go back to compound nouns (34b–c). For the most part new compound verbs have a noun (N) as first component (Tyysteri 2015: 173), which is

12 According to Saukkonen (1973: 337f.), the proportion of verbs among all compounds in "Nykysuomen sanakirja" (1951–1961) remains at 0,3 %. In Tyysteri's (2015: 113) corpus their ratio (types) is 1,2 %.

13 Copulative compound verbs do not exist in Finnish. Compounds with a verb stem as first constituent are possible, cf. *riippu+liitää* 'hang-glide', but the constituent relation is determinative, not additive. In *itku+naurattaa* 'make cry and laugh' (Vesikansa 1989: 258) the semantic relation is similar to an additive compound, but the first constituent is a deverbal noun, i.e. the morphological structure is asymmetric.

unsurprising since compound nouns are the most common derivation base, and among these, the structure N+N is predominant.

(34a) *kaksi+kielistyä* 'become bilingual' < *kaksi+kielinen* (lit. two+lingual) 'bilingual'
(34b) *valo+kuvata* 'photograph' (V) < *valo+kuva* (lit. light+picture) 'photograph' (N)
(34c) *koe+lentää* (backformation) 'test fly' < *koe+lento* 'test flight'

Adverbs, particles and non-autonomous elements can combine directly with verbal heads (Vesikansa 1989: 254ff.). Such preverbs are often called "prefix-like elements" because they are in many respects similar to prefixes in other languages. In Finnish, however, prefixation is untypical (Häkkinen 1994: 488; Kolehmainen 2006: 111, 113). This is why word formation with bound "prefix-like elements" is subsumed under compounding in the Finnish grammar tradition, even if the notion of "prefix-likeness" varies (cf. Tyysteri 2015: 127ff.). In the following, the focus is on verbs with such prefix-like elements.

Kolehmainen (2006) makes a distinction between position fixed bound preverbs, divided into (a) confixes and (b) prefixes, and in contrast to them (c) separable particles in phrasal verbs. Consequently, in each group the word formation status of the verbs is different: in (a) compound (3.1), in (b) prefix derivative (3.2), and in (c) MWE (3.3). In the following, these groups are examined in detail in order to estimate their structural status and productivity.

3.1 Confix compounds

Complex words with a prefix-like first constituent that does not occur as an autonomous lexical unit (and thus has an unspecific word class status) are relatively common in modern Finnish. In Tyysteri's material, including all word classes, they make up 9,3% of all two-constituent compounds; indigenous and foreign pre-elements are roughly equally common. Yet, the word class distribution (e.g. the ratio of verbs) of such formations is not given (cf. Tyysteri 2015, 118ff., 125, 128). The examples in (35a) are lexicalized compounds (cf. Kolehmainen 2006: 115); neologisms and occasionalisms such as (35b) are being used more and more frequently.

(35a) *edes+auttaa* (lit. forth+help) 'help, assist, further'
jälki+kiillottaa (lit. after+polish) 'polish bright'

(35b) *etä+seurustella* (lit. distance+go together) 'have a long-distance relationship'
pika+syödä (lit. quick+eat), 'eat quickly, eat fast food'
täsmä+leikata (lit. precise+operate) 'operate/remove precisely'

The "prefix-likeness" of such elements is debatable. The term confix seems more suitable here because in contrast to semantically abstract prefixes, the pre-elements in question still have a more or less clear lexical-conceptual meaning. Historically, they go back to autonomous lexemes; some of them occur today only as bound elements (e.g. *epä-* 'un-, non-'; *esi-* 'pre-'; *etä-* 'long-distance, remote'), some are obsolete or archaic as autonomous words (e.g. *lähi-* 'near'; *taka-* 'back, rear'; *tasa-* 'even, equal'). Others have an autonomous homonym, but the semantic difference is so big that the common origin is not transparent (e.g. *edes-* 'further, forth'; *etu-* 'fore, forward, front'; *jälki-* 'post-, after-')[14] (Kolehmainen 2006: 113f., 128). Karlsson (1983: 192f.) points out that these elements are semantically similar to nouns and adjectives and calls them lexical "relic morphemes".[15]

Moreover, these elements differ from prefixes in their ability to function as derivation bases (cf. *esi-* 'pre-' in the derivate *esittää* 'present, put forth, perform' vs. *esi+katsella* 'preview'; more examples in Kolehmainen 2006, 119). They are somewhere between prototypical compound constituents and affixes (ibid.: 118–124). Confix verbs meet the prototypicality criterion 4) (cf. Section 2.1) according to which a form-identical phrase is not possible (**esi katsella*, **katsella esi*), but since the first constituent is not an autonomous lexeme (criterion 1), they count as non-prototypical compounds.

In spite of the fact that non-autonomous elements can in principle be combined regularly with verbal heads, many of the complex verbs in this group are actually secondary compounds, i.e. derivatives (36a) or backformations (36b) from already existing compounds (see above).[16] Many confix verbs have an incomplete paradigm: They are preferably used in infinite forms, especially as adjec-

14 In affirmative expressions the autonomous word *edes* means 'at least' in modern Finnish, with negation it has the meaning '[not] even'. The noun *etu* means 'advantage, benefit' and the noun *jälki* 'track, trace'. In spite of the common etymology, native speakers hardly associate these words with the corresponding pre-elements (Kolehmainen 2006: 114, 126).

15 ISK (2004: 192, 393, 414f.) and Rahtu (1984: 409) characterize them as "prefix-like nominal stems".

16 In Tyysteri's random sample of 300 two-constituent-compounds (100 nouns, 100 adjectives and 100 verbs), 75% of the compound verbs (including all kinds of first constituents) were formed by derivation or backformation and only 25% by regular compounding. The ratio of regular compounding is much lower than in previous studies (Tyysteri 2015: 154f., 158).

tive-like participles, which is a transitional phase on the way towards a full paradigm via analogy and generalization. Analogy plays a role in producing new verbs as well: When verbs with a given initial element, e.g. *ala-* 'sub-', become more frequent (e.g. *alaotsikoida* 'subtitle', *alaluokitella* 'subclassify' etc.), the word structure is reanalyzed such that the main constituent boundary is after the pre-element, and not after the complex nominal base, thus as if the verbs were formed regularly via combining *ala-* directly with the verb. In this way, an originally prenominal confix can develop into a preverbal confix, cf. (i), which leads to a symmetric compounding model (ii) that can be generalized, cf. (iii):

(36a) *ala+otsikoida* 'subtitle' (V) < *ala+otsikko* 'subtitle' (N)
  (i) *otsikko* 'title' (N) > *otsik-* + *-oida* (suffixation) 'title' (V)
  ala+otsikko 'subtitle' (compound N) > (*ala+otsik-*) + *-oida* (suffixation) 'subtitle' (V) > ***ala+otsikoida*** (reanalysis into a compound V)
  (ii) *otsikko* : *otsikoida* = *ala+otsikko* : *ala+otsikoida*
  (iii) N : V = *x*+N : *x*+V
(36b) *esi+pestä* 'prewash' (V) < *esi+pesu* 'prewash' (N)

In Kolehmainen's assessment (2006: 116f.), given the limited lexical variation in her research material (76 different verbs with 22 indigenous confixes)[17] the structure confix+verb plays a minimal role in modern Finnish, i.e. it is not productive. Yet, according to ISK (2004: 414f.), the number of different verbs with *epä-* 'un-', *esi-* 'pre-', *jälki-* 'post-', *pika-* 'quick, instant' is increasing, which means that at least these elements are productive. Among the new compounds from the first decade of the 21st century many more than the above-mentioned bound preverbs are in frequent use – to an extent that proves the productivity of this formation model (Tyysteri 2015: 130). Confix verbs are, however, often stylistically marked: They occur as terms in languages for special purposes; in everyday language and print media occasionalisms are often used playfully (Vesikansa 1989: 257f.; Kolehmainen 2006: 116; Tyysteri 2015: 88, 113, 213). Nevertheless, it is evident that the number of lexicalized confix verbs in standard language is increasing. The currently most popular indigenous and foreign verb confixes have a high communicative and cultural relevance: They reflect modern life with its hectic pace (*pika-*), green values (*bio-*, *eko-*) and technological innovations (*digi-*, *nano-*, *etä-*, *täsmä-*).

17 Kolehmainen collected her research material from dictionaries and authentic texts from the 1990es in SKTP (*Suomen kielen tekstipankki* / *Language Bank of Finland*).

3.2 Prefix verbs?

The question is whether adpositional and adverbial elements that are used as bound preverbs in Finnish can be regarded as prefixes. Kolehmainen (2006: 130–137) cautiously refers to them as "prefix-like elements" and underlines that they differ in some aspects from prefixes in Germanic languages. Firstly, they are not unstressed: The main word stress in Finnish is generally on the initial syllable, i.e. word stress does not apply as prefix criterion in Finnish. Secondly, the Finnish adpositions are mainly postpositions.[18] Thirdly, many of them are secondary adpositions, having developed from inflected forms of relative nouns,[19] and have therefore (fossilized) case endings; some of them have a restricted nominal paradigm in several (still existing or historical), mostly locative, especially directional cases. The same holds for adverbs: Many elements occur both as adpositions and as adverbs (ISK 2004: 664f.; Tyysteri 2015: 121). Consequently, there are hundreds of different adposition and adverb forms in Finnish, but not all of them function as preverbs.

Kolehmainen's research material from grammars, previous studies and dictionaries contains 70 of such elements (Kolehmainen 2006: 134–137). "Nykysuomen sanakirja" (Dictionary of modern Finnish, 1951–1961) mentions 251 complex verbs with these elements, but many of them are marked as archaic, e.g. *alas+astua* 'step down', and almost a half of them occur in univerbated form only as participles, cf. *yhteen+laskettu* (lit. together+counted) 'combined'. In both cases separated alternatives are recommended, cf. *astua alas*, *yhteen laskettu* (cf. Section 3.3). Thus the number of inseparable verbs in active use is much lower. Most of the elements combine only with one or two verbs (ibid.: 138). About ten elements show a somewhat broader spectrum, e.g. *irti-* 'loose, off', *läpi-* 'through, throughout', *sisään-* 'in, inside', *ulos-* 'out, outside', *yli-* 'over' (ibid.: 137ff.). All in all, Kolehmainen regards the model prefix+verb as unproductive.

Finnish inseparable verbs of this group are historical relics that go back to old loan translations from Germanic and classical languages resp. to an interference-based formation model (cf. Öhmann 1957: 33ff.; Vaittinen 2003; Toropainen 2017: 72). In Old Literary Finnish (1540–1810) the majority of printed texts were

18 In principle, postpositions (and adverbs) can develop into prefixes in SOV-languages where complements precede the verb. SOV is supposed to be the basic word order in Uralic languages; in Finnish, however, the order has changed into SVO. This is one possible explanation for the weak affinity to prefixes. As for typological theories of linearization in connection with prefixes, see the overview in Kolehmainen (2006: 149–156).

19 This is the first step of a gradual grammaticalization called "noun-to-affix-cline", cf. Lehmann (1985: 304); Hopper/Traugott (2003: 110); as for Finnish Jaakola (1997: 126f., 134).

translations of religious texts, following faithfully the formulations in the original (Häkkinen 1994: 11f.). For example Mikael Agricola (about 1510–1557), the "Finnish Luther", used 810 different compound verbs (including all first element categories)[20] in his texts, which makes up 32,5 % of all his compounds on type level (Toropainen 2017: 53, 55, 66, 74). In about 80 % of Agricola's verb compounds the first constituent was an adverb (Häkkinen 1987: 10). In the 17th century such compounds were often replaced with MWEs consisting of a verb and an adverb by Agricola's successors. In the 19th and 20th centuries compound verbs were combated by purist language planners as un-Finnish or ungrammatical (Häkkinen 1987: 7), resulting in a radical decline of use.

In modern Finnish, most combinations of adverb and verb, such as *pois* 'away' + *sulkea* 'close', are generally recommended to be formed as two separate words, i.e. as MWEs (e.g. by "Kielitoimiston sanakirja" (2006), a dictionary of Standard Finnish), where the adverb is postponed in case of neutral word order, cf. (37a). Yet, in attributive participles the only possible position for the adverb is before the verb. Although such a word order usually promotes univerbation, the norm of writing separately holds for most participles, cf. (37b), even if language users tend to write the parts together. However, when *pois* precedes an infinitive, the components are written together, in contrast to the reversed order, cf. (37c). The verb *irti+sanoa* (lit. off+say) 'discharge, fire; cancel, (fig.) break off' behaves in some details differently. As for the infinitive, the alternatives are the same (38a),[21] but in passive past participle, the preceding adverb is not separable (38b). In other words, the rules differ from verb to verb. Some lexicalized verbs cannot be separated at all (39). In some cases separation is combined with semantic difference: In a concrete meaning the adverb is separated (40a), whereas univerbation is preferred in an abstract meaning (40b) (as for orthographical norm, cf. Pääkkönen 1989: 375; Eronen 1996; Tyysteri 2015: 38).

(37a) *pois+sulkea*, better *sulkea pois* 'exclude, rule out'
(37b) *pois suljettu vaihtoehto* 'excluded alternative'
(37c) *Mitään vaihtoehtoa ei pitäisi pois+sulkea ~ *pois sulkea ~ sulkea pois*
'None of the alternatives should be excluded.'

20 According to Jussila (1988), about 61 % of Agricola's vocabulary has remained in use up to date, but as for compounds, the proportion is only 15,9 %; the strongest decline concerns compound verbs.

21 Although infinitive forms preceded by an adverb are normally written together (cf. **pois sulkea*, **irti sanoa*), the components must be separated, if an enclitic particle, e.g. *-kAAn* '[not] even, anyway, after all', is appended to the adverb, cf. *Ei sitä voi poiskaan sulkea* 'After all, it cannot be excluded'; *Ei heitä voi irtikään sanoa* 'Anyway, they cannot be fired'.

(38a) *irti+sanoa ~ *irti sanoa ~ sanoa irti* 'discharge, fire; cancel, (fig.) break off'
(38b) *irti+sanottu* vs. **irti sanottu*

(39) *jälleen+rakentaa* 'reconstruct, rebuild'
läpi+valaista (lit. through+lighten) 'scan, X-ray'
myötä+elää (lit. with+live) 'empathize'
perään+kuuluttaa (lit. after+announce) 'demand, claim, try to find'
ympäri+leikata (lit. round+cut) 'circumcise'

(40a) *ohi kiitävä auto* 'car speeding past'
(40b) *ohi+kiitävä hetki* 'fleeting moment'

In my opinion, these pre-elements are not prefixes. One reason is their obvious unproductivity, i.e. the restricted verb variation per pre-element – for affixes a far wider use is expected. The still existing bound forms are sporadic historical relics, based on calques from foreign languages with systematic prefixation, yet, in Finnish, a generalization never took place. The initial word stress protects the elements from phonological erosion typical of affixes. Above all, the fact that there are parallel phrasal forms, cf. (37a) and (38a), is a proof of the lexical autonomy of the elements in question – in that respect they show a higher autonomy than confixes (cf. Section 3.1). It follows that the univerbated forms are compounds. Here I agree with Tyysteri (2015: 119, 121) who, in contrast to Kolehmainen (2006), does not classify the above-mentioned elements as prefixes or "prefix-like elements" but as "indeclinable elements or elements with incomplete declination (adverbs, adpositions and particles)" in ordinary compounds. The advantage of this analysis is that the coexistence of occurrences with and without separation, i.e. MWEs vs. compounds, can be compared with similar cases in other word classes where both alternatives have (nearly) the same meaning, cf. (8) and (20b) above.

Whether the one-word and the two-word combination represent one and the same verb lexeme or two synonymous lexemes and whether the phrasal alternatives should be regarded as regular ("free") syntactic constructions or rather as phrasal verbs, i.e. MWEs, is discussed in the next section.

3.3 Phrasal verbs

In the linguistic literature the terms "phrasal verb" and "particle verb" are often used as synonyms. The former implies that the components are separate, while the latter refers to the functional category of the component the verb is connected

with. In English, for example, particle verbs are always phrasal verbs. In Finnish this need not be the case.

In traditional Finnish grammar phrasal verbs are not recognized as an established category, but several scholars refer to fixed sayings or idiomatic figures of speech in the form of MWEs, similar to separable particle verbs in Germanic languages (cf. Häkkinen 1997: 44; Nenonen 2002: 55), cf. (41a). They are semantically and structurally similar to verb idioms consisting of a verb and a non-particle component, for example a unique component (41b) or a nominal component in a locative case (41c) (cf. Nenonen 2002: 55f.; Kolehmainen 2006: 164).

(41a) *panna vastaan* (lit. put against) 'resist, struggle against'
(41b) *lyödä laimin* (lit. hit/beat UNIQUE COMPONENT) 'neglect, abdicate'
(41c) *ottaa huomioon* (lit. take account$_{ILL}$) 'take into account'

According also to ISK (2004: 447), particle verbs are "idiomatic predicates". Here "particle" refers to the functional category of the element co-occurring with the verb, regardless of univerbation or separation. In some cases "Kielitoimiston sanakirja" (2006) lemmatizes the univerbated form but refers to the phrasal one. From entries like (42a) it can be inferred that both forms are regarded as representations of the same lexeme; remarks such as 'mostly' or 'better' (42b) indicate that the MWE is generally the dominant form. Occasionally only the univerbated form is given although separated forms occur commonly, cf. (42c). However, as mentioned above, some verbs are used only in the univerbated form, cf. (39).

(42a) *irti+sanoa* = *sanoa irti*, cf. (38a)
laimin+lyödä = *lyödä laimin*, cf. (41b)
(42b) *ylen+antaa*, mostly *antaa ylen* 'throw up, vomit'
pois+sulkea, better *sulkea pois*, cf. (37a)
(42c) *ulos+liputtaa* 'flag out' vs.
Viking Line liputtaa ulos kaksi alusta. 'Viking Line is going to flag out two ships.'

Kolehmainen (2006: 170f.) sees the separation (i.e. the MWE structure) and the idiomaticity or metaphoricity of the combination to be key criteria; in her assessment particle verbs are either singular idioms or go back to phraseological patterns. This means that transparent (non-idiomatic) combinations, such as (43), are excluded from the class of particle verbs and regarded as products of free syntax; according to Kolehmainen (ibid.) native speakers do not perceive them as single semantic units.

(43) *muuttaa pois* 'move away'
kulkea edellä 'walk ahead'

Yet, it is not always easy to draw the line between idiomatic and free combinations because idiomaticity is a continuum. Kolehmainen (ibid.: 172–183) distinguishes between four grades of idiomaticity and compositionality:

(A) Fully idiomatic combinations that do not permit any component variation are obvious verb idioms, e.g. *lyödä laimin*, cf. (41b) above, where *laimin* is a unique (adverb-like) component and the verb *lyödä* does not bear its regular meaning 'hit, beat', cf. (44a) vs. (44b).

(44a) *He lyövät laimin lapsiaan*. 'They neglect their children.' (idiomatic meaning) vs.
(44b) *He lyövät lapsiaan*. 'They are beating their children.' (regular meaning)

Also a combination of verb and autonomous adverb can in principle become fixed as a single idiomatic MWE without component variation, e.g. (45a), where, however, the figurative meaning is compositional to some degree, as far as *ampua* is understood as a destructive action; the directionality of the adverb underlines telicity ('once for all'), and in up-down-metaphors 'down' means negative things, here (a change into) non-existence. A similar compositionality can be recognized also behind some other figurative expressions for resistance or undoing, consisting of a verb of destruction and *alas*, such as (45b) – i.e. the borderline between (A) to (B) is vague.

(45a) *ampua alas* 'shoot down (an idea, a plan)'
(45b) *repiä alas* (lit. tear down) 'break down (boundaries, conventional values etc.)'

(B) Serialization indicates compositionality. Rudiments of serialized formation occur as niches of a few similar particle verbs, i.e. there is some variation in the verb component, cf. (46a) where *yhteen* 'together' alludes to a confrontation, or (46b) where *kiinni* 'shut, fixed, closed' refers to a state that cannot be changed anymore.

(46a) *iskeä yhteen* (lit. hit together) 'clash, lock horns'
ottaa yhteen (lit. take together) 'clash, quarrel'
(46b) *iskeä kiinni* (lit. hit fixed) 'stabilize (e.g. a dominating position)'
naulata kiinni (lit. nail fixed) 'nail down, fix on (e.g. prizes)'

(C) MWEs consisting of verbs and the particles *ilmi* 'open(ly), apparent(ly)' and *julki* '(in) public, out' build productive phraseosyntactic patterns, expressing that information is made available or public. In contrast to adverbs like *ulos* 'out' or *kiinni* 'shut, fixed, closed' which occur both in concrete and in figurative combinations, *ilmi* and *julki* always have a constant abstract meaning, which can explain the stronger serialization. There are both intransitive and transitive series. The kernel verbs are so-called light verbs like *tulla* 'come', *antaa* 'give', *saada* 'get', *tuoda* 'bring', but they can be replaced with more specific verbs expressing for example that the publicity was not intended, cf. (47a) vs. (47b). In the transitive pattern, *antaa* 'give', *saattaa* 'put' or *tuoda* 'bring' can be replaced by various speech verbs and their descriptive and expressive variants, cf. (48a–c):

(47a) *tulla julki* 'come out, become public'
(47b) *vuotaa* ~ *lipsahtaa julki* 'leak ~ slip out'

(48a) *tuoda julki* 'bring into publicity'
(48b) *lausua* ~ *puhua* ~ *sanoa julki* 'express ~ speak ~ say publicly'
(48c) *kaakattaa* ~ *kiljua* ~ *möläytellä julki* 'cackle ~ scream ~ blurt out'

(D) Combinations of verb and directional adverb are often situated on the boundary between regular syntactic constructions and fixed MWEs. At first sight it seems controversial that, according to Kolehmainen (2006: 91, 97, 170), the German separable particle verbs in (49) are lexicalized phraseological (but not idiomatic) units, whereas the corresponding Finnish combinations are not. However, this is not necessarily controversial because the lexicalization strategies in two languages need not be identical. Yet, the difference in the language-specific affinity of such combinations to merge into one lexeme should be proved theoretically. A possible explanation could be related to the grade of semantic-structural autonomy of German and Finnish adpositions and adverbs. Different word order conditions could be relevant, too.

(49) *weg*/*ziehen* – *muuttaa pois* 'move away'
vor/*gehen* – *kulkea edellä* 'walk ahead'
auf/*blicken* – *katsoa ylös* 'glance up'
hinaus/*gehen* – *mennä ulos* 'go out'
nieder/*knien* – *polvistua alas* 'kneel down'

In any case it is obvious that lexicalization is mostly combined with semantic specificity. As for the directional adverb *ulos* 'out', for instance, the concrete

non-specific meaning is manifest in contexts where the locality inside of something that is left behind is explicated verbally (50a) or when the location is inferable by context and situation, like in (50b), assuming that 'being in a tunnel' is already a known fact (contextual ellipsis).

(50a) *ajaa ulos tunnelista* 'drive out of the tunnel, leave the tunnel'
(50b) *ajaa ulos* (Ø) 'leave'

Besides contextual ellipses there are conventionalized ellipses that are not figurative but bear some specific semantic features connected with a certain topic or text type. For example, in reports on road accidents or motor sports *ajaa ulos* has the conventional meaning 'drive off the track, swerve off the road' (51a). The noun *ulos+ajo* (51b) is used particularly in this specific meaning, yet it is difficult to say if it has been derived from the lexicalized phrasal verb. It could as well have been originated as a synthetic compound and then later specialized as a traffic term, of which the specific phrasal verb has been formed analogically, similar to backformation. This makes it difficult to use phrasal input for derivation as a criterion of lexicalizedness of the base, especially as there are synthetic compounds going back to fully transparent non-specific combinations, cf. (52) and (49) above – even if dictionaries codify primarily the idiomatized or spezialized compounds and leave the semantically self-evident ones out.

(51a) *Henkilöauto **ajoi ulos** sunnuntaina Räyringissä*. 'On Sunday, a passenger car drove off the road in Räyrinki.'
(51b) *ulos+ajo* (lit. out+driving) 'driving off the road' (*nomen acti*), cf. *Ulosajo tallentui videolle*. 'The accident [driving off the road] was videotaped.'

(52) *muuttaa pois* 'move away' – *pois+muutto* (N)
mennä ulos 'go out' – *ulos+meno* (N)

Components of lexicalized MWEs cannot be anaphorized. Consequently, if the particle in a combination with a verb is anaphorizable, the combination is free, cf. (53). However, many adverbs lack natural anaphors. For example *kiinni* 'fixed, shut, closed' is not anaphorizable regardless of whether it occurs in concrete or figurative meaning. So, anaphorizability can exclude a combination from phrasal verbs, but lacking anaphorizability cannot be used as evidence of lexicalizedness.

(53) *Anna meni ulos. – Menikö hän sinne yksin?* 'Anna went out. – Did she go there alone?'

Summa summarum: The concept of phrasal verbs deserves to be applied to Finnish, yet, further research is needed to define the limits of the category.

4 Concluding remarks

Compounding is the most common way to form new words in modern Finnish. Prototypical determinative nominal compounds with an underspecified first constituent (N+N) form the most common and still increasing type. Apart from this type many less prototypical compound models are productive, too. Among these, special attention has been paid above to formations showing syntactic features similar to MWEs and/or competing with MWEs. The essential findings can be summarized as follows:

1. In about one third of A+N compounds the adjective agrees in number and case with the head, which does not fulfil the criterion of morphological integrity. However, compound-internal congruence is a recessive feature; there are hardly any neologisms with internal congruence. Compounds with a non-congruent first constituent tend to have a term-like character.
2. Internal inflection also occurs in complex numerals. Numerals with hundreds, thousands etc. are grouped into smaller (still complex) units, thus combining characteristics of non-prototypical compounds and MWEs.
3. In synthetic compounds argument relations of the verb that underlies the head are explicated by case forms, which is a syntactic feature.
4. A prototypical compound cannot be replaced with a phrasal unit of formally identical components. Generally, if such pairs occur, they differ in meaning. Overlap occurs if the modifier is in the genitive, which is the situation for semantically relative adjectives and many deverbal nominalizations. Univerbation strengthens the conceptual unity, and vice versa, conceptualization furthers univerbation.
5. An opposite example of the correlation between conceptual unity and univerbation is represented by Finnish particle verbs. Compound verbs with an adverb or adposition as first constituent are not productive in modern Finnish, partly as consequence of normative language planning. This gap in the system is compensated by "phrasalization", i.e. keeping apart the components in particle verbs. However, the formation model is far less productive than in English or German, for instance. Apart from singular idioms, serialization, based on phraseosyntactic patterns, occurs in some amount. Drawing the line between lexicalized MWEs and syntactically free combinations requires further research.

6. When the syntactic distributions of a compound and a semantically equal MWE are different, their relation is complementary rather than competing. This applies to similative compound adjectives/adverbs and corresponding phrasal similes: The latter cannot occur as adjective attributes. Furthermore, while predicative and adverbial similative compounds can be transformed into phrasal similes, the opposite is not always possible: Only phrasal similes allow expansions in the part that expresses the point of comparison.

The following topics remain for further research: In Finnish, non-figurative MWEs such as fixed collocations and nominations for specific concepts have been so far studied mostly in terminology. In the future, more attention should also be paid to corresponding combinations in standard language. So far, MWEs have been excluded when working out the statistical distribution of different lexem structure types in the Finnish vocabulary. Another question deserving attention is the role of MWE patterns at the intersection of syntax and lexicon: Besides particle verbs and similes, e.g. light-verb constructions, binomials and serial modification of a specific idiom structure are topics worth of further attention. Several single studies to these areas have been carried out within contrastive phraseology and construction grammar but a systematic overview of MWE patterns is still outstanding.

References

Baldwin, Timothy/Kim, Su Nam (2010): Multiword expressions. In: Indurkhya, Nitin/Damerau, Fred J. (eds.): Handbook of natural language processing. 2nd ed. Boca Raton i.a.: CRC Press. 267–292.

Eronen, Riitta (1996): Monimuotoiset yhdyssanat. In: Kielikello 1996, 4. Internet: http://kielikello.fi.libproxy.helsinki.fi/index.php?mid=2&pid=11&aid=386 (last access: 15.9.2017).

Fleischer, Wolfgang/Barz, Irmhild (2012): Wortbildung der deutschen Gegenwartssprache. 4th ed. Berlin/Boston: De Gruyter.

Häkkinen, Kaisa (1987): Suomen kielen vanhoista ja uusista yhdysverbeistä. In: Sananjalka 29. 7–29.

Häkkinen, Kaisa (1994): Agricolasta nykykieleen. Suomen kirjakielen historia. Porvoo: WSOY.

Häkkinen, Kaisa (1997): Kuinka ruotsin kieli on vaikuttanut suomeen? In: Sananjalka 39. 31–53.

Heinonen, Tarja Riitta (2001): *Harmaaturkit herkkusuut* – bahuvriihit sanakirjassa ja kieliopissa. In: Virittäjä 105. 625–634.

Heinonen, Tarja Riitta (2010): *Kuin*-vertaukset. In: Virittäjä 114. 348–373.

Hopper, Paul J./Traugott, Elizabeth Closs (2003): Grammaticalization. 2nd ed. (= Cambridge Textbooks in Linguistics). Cambridge, UK: Cambridge University Press.

Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.). 450–467.
ISK (2004) = Hakulinen, Auli et al. (2004): Iso suomen kielioppi. (= Suomalaisen Kirjallisuuden Seuran Toimituksia 950.) Helsinki: SKS.
Jaakola, Minna (1997): Genetiivin kanssa esiintyvien adpositioiden kieliopillistumisesta. In: Lehtinen, Tapani/Laitinen, Lea (eds.): Kieliopillistuminen. Tapaustutkimuksia suomesta. (= Kieli 12). Helsinki: SKS. 121–156.
Jussila, Raimo (1988): Agricolan sanasto ja nykysuomi. In: Koivusalo, Esko (ed.): Mikael Agricolan kieli. (= Tietolipas 112). Helsinki: SKS. 203–288.
Karlsson, Fred (1983): Suomen kielen äänne- ja muotorakenne. Porvoo: WSOY.
Karlsson, Fred (2004): Finnish (Finno-Ugric). In: Booij, Geert et al. (eds.): Morphologie. Morphology. Ein internationales Handbuch zur Flexion und Wortbildung. An international handbook on inflection and word-formation. Vol. 2. (= Handbooks of Linguistics and Communication Science (HSK) 17.2). Berlin/New York: De Gruyter. 1328–1342.
Karlsson, Fred (2015): Finnish. An essential grammar. Translated by Andrew Chesterman. 3rd ed. London/New York: Routledge.
Kielitoimiston sanakirja (2006) = Grönros,Eija-Riitta et al. (2006): Kielitoimiston sanakirja. Osat 1–3. Helsinki: Kotimaisten kielten tutkimuskeskus.
Koivisto, Vesa (2013): Suomen sanojen rakenne. Helsinki: SKS.
Kolehmainen, Leena (2006): Präfix- und Partikelverben im deutsch-finnischen Kontrast. (= Finnische Beiträge zur Germanistik 16). Berlin i.a.: Lang.
Korhonen, Jarmo (2018): Fraseologia – Kiinteiden sanayhtymien tutkimus. Helsinki: Finn Lectura.
Lehmann, Christian (1985): Grammaticalization. Synchronic variation and diachronic change. In: Lingua e Stile 20, 3. 303–318.
Mäkisalo, Jukka (2000): Grammar and experimental evidence of finnish compounds. (= Studies in Languages 35). Joensuu: University of Joensuu.
Malmivaara, Terhi (2004): *Luupää*, *puupää*, *puusilmä*. Näkymiä sananmuodostuksen analogisuuteen ja bahuvriihiyhdyssanojen olemukseen. In: Virittäjä 108. 347–363.
Müller, Peter O. et al. (eds.) (2015–2016): Word-formation: An international handbook of the languages of Europe. (= Handbooks of Linguistics and Communication Science (HSK) 40). Berlin/Boston: De Gruyter.
Nenonen, Marja (2002): Idiomit ja leksikko. Lausekeidiomien syntaktisia, semanttisia ja morfologisia piirteitä suomen kielessä. (= Publications in the Humanities 29). Joensuu: University of Joensuu.
Niemi, Jussi (2009): Compounds in Finnish. In: Lingua & Linguaggio 8. 237–256.
Niinimäki, Anneli (1992): Sanaliittojen tiivistyminen yhdyssanoiksi. In: Virittäjä 96. 283–286.
Nykysuomen sanakirja (1951–1961) = Sadeniemi, Matti (1951–1961): Nykysuomen sanakirja. Osat I–VI. Porvoo: WSOY.
Öhmann, Emil (1957): Beobachtungen über feste Verbalzusammensetzungen im Finnischen. In: Ural-altaische Jahrbücher 29. 33–37.
Olsen, Susan (2015): Composition. In: Müller, Peter O. et al. (eds.). 364–386.
Pääkkönen, Irmeli (1989): Sanojen äänneasu ja oikeinkirjoitus. In: Vesikansa, Jouko (ed.). 357–382.
Pitkänen-Heikkilä, Kaarina (2016): Finnish. In: Müller, Peter O. et al. (eds.). 3209–3228.
Rahtu, Toini (1984): Suomen nominialkuiset yhdysverbit. In: Virittäjä 88. 409–430.
Räisänen, Alpo (1986): Sananmuodostus ja konteksti. In: Virittäjä 90. 155–163.

Sag, Ivan A. et al. (2002): Multiword expressions: A pain in the neck for NLP. In: Gelbukh, Alexander (ed.): Computational linguistics and intelligent text processing. Third International Conference, CICLing-2002, Mexico City, Mexico, February 17–23, 2002. (= Lecture Notes in Computer Science 2276). Berlin i.a.: Springer. 1–15.
Sajavaara, Paula (1989): Vierassanat. In: Vesikansa, Jouko (ed.). 64–109.
Saukkonen, Pauli (1973): Suomen kielen yhdyssanojen rakenne. In: Commentationes Fenno-Ugricae in honorem Erkki Itkonen. (= Suomalais-Ugrilaisen Seuran toimituksia 150). 332–339.
Schellbach-Kopra, Ingrid (1964): Die Bahuvrihi-Komposita in der alten finnischen Volksdichtung. In: Suomalais-ugrilaisen seuran aikakauskirja – Journal de la Société finno-ougrienne 65. 1–41.
Stein, Stephan (2012): Phraseologie und Wortbildung des Deutschen. Ein Vergleich von Äpfeln mit Birnen? In: Prinz, Michael/Richter-Vapaatalo, Ulrike (eds.): Idiome, Konstruktionen, „verblümte Rede". Beiträge zur Geschichte der germanistischen Phraseologieforschung. (= Beiträge zur Geschichte der Germanistik 3). Stuttgart: Hirzel. 225–240.
Suomen kielen perussanakirja (1990–1994) = Haarala, Risto (1990–1994): Suomen kielen perussanakirja. Osat 1–3. (= Kotimaisten kielten tutkimuskeskuksen julkaisuja 55). Helsinki: Valtion painatuskeskus.
Toropainen, Tanja (2017): Yhdyssanat ja yhdyssanamaiset rakenteet Mikael Agricolan teoksissa (= Turun yliopiston julkaisuja – Annales Universitatis Turkuensis C 439). Turku: University of Turku. Internet: https://doria.fi/bitstream/handle/10024/143331/AnnalesC%20439Toropainen.pdf?sequence=2 (last access: 25.8.2017).
Tyysteri, Laura (2015): *Aamiaiskahvilasta ötökkötarjontaan*. Suomen yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus. (= Turun yliopiston julkaisuja – Annales Universitatis Turkuensis C 408). Turku. Internet: https://doria.fi/bitstream/handle/10024/113113/AnnalesC408Tyysteri.pdf?sequence=2 (last access: 25.8.2017).
Vaittinen, Tanja (2003): Vanhan kirjasuomen yhdysverbit. In: Sananjalka 45. 45–66.
Vesikansa, Jouko (1989): Yhdyssanat. In: Vesikansa, Jouko (ed.). 213–258
Vesikansa, Jouko (ed.) (1989): Nykysuomen sanavarat. Porvoo: WSOY.

Ferenc Kiefer/Boglárka Németh

Compounds and multi-word expressions in Hungarian

The notion of compounding is notoriously difficult to define and there are hardly any universally accepted criteria for determining what a compound is. In the present chapter we will make a distinction between prototypical compounds and non-prototypical compounds. The latter but not the former are syntactically separable. All compounds are right-headed and are inflected as a whole. Moreover, according to the received view compounds express a conceptual unit though it is not easy to define what exactly this means. Finally, typically only the first syllable of a compound bears stress.

Compounding is a rather late development in the history of Hungarian. Though compounds can be found sporadically before the 18th century, during the language reform (end of 18th and beginning of 19th century) new compounds were massively created partly by using existing patterns and partly by loans mainly from German. This explains why productive patterns of root (endocentric) compounds are – as far as the categories involved are concerned – identical in Hungarian and German.[1]

The structure of our chapter is as follows: in the first part of the chapter we are going to provide an overview of productive compounding patterns, i.e. root compounds, morphologically marked compounds, deverbal compounds and coordinative compounds. Section 2 is devoted to the description of compound-like phrases in Hungarian, i.e. preverb + verb constructions and bare noun + verb constructions. Finally, Section 3 summarizes the main conclusions of the chapter.

1 Prototypical compounds

1.1 Root compounds

Let us first have a look at root compounds. A root compound is a compound whose head is not deverbal or whose non-head does not have the function of argument of the verb from which the head is derived. The productive patterns

1 Sections 1.1 through 1.3 and 2.2 are heavily based on our earlier works on the subject. Cf., in particular, Kiefer (1992, 1993, 2009) and Kiefer and Németh (2018).

https://doi.org/10.1515/9783110632446-012

involve nouns and adjectives only, there are no productive patterns with adverbs and/or verbs. All endocentric compounds in Hungarian are right-headed and are formed by juxtaposition of the relevant lexical items. No morphological markers appear between the constituents of root compounds. (1a–d) shows the chart of productive patterns.[2,3]

(1a) N + N
város+háza
'city hall'
tök+mag
'pumpkin seed'

(1b) A + N
kis+autó
'small car'
meleg+ágy
'hotbed'

(1c) N + A
kő+kemény
'stone hard'
oszlop+magas
'pillar high'

(1d) A + A
sötét+zöld
'dark green'
bal+liberális
'left-liberal'

Recently a fifth pattern seems to be gaining ground in addition to the ones shown in (1a–d), namely the pattern N + V. It can be argued, however, that the corresponding compounds are (at least in the majority of cases) backformations from the corresponding deverbal compounds. For some examples, cf. (2a–c).[4]

2 In Hungarian compounds are usually written as one word. In the examples the constituents are written separately for the sake of clarity.

3 1 = first person; 3 = third person; ACC = accusative; COM = comitative; COND = conditional; DAT= dative; DEF = definite; INF = infinitive; INSTR = instrumental; INTR = intransitive; LOC = locative; NMLZ = nominalization; PL = plural; POSS = possessive; PREV = preverb; PST = past; PTCP = participle; RES = resultative; SG = singular; TEMP = temporal (terminative).

4 Cf. also Ladányi (2007: 64 f.).

(2a) N + V
gép+ír
machine write
'write on a typewriter'
from *gép+ír-ás*[5]
machine writing
'typing'

(2b) *ház+kutat*
house search (verb)
from *ház+kutat-ás*
house search (noun)

(2c) *tömeg+közlekedik*
mass run
from *tömeg+közleked-és*
mass/public transportation

Similar examples are legion. It should be noted, however, that compounds such as (2a–c) are more frequent in everyday and newspaper language than in literary language.

1.2 Morphologically marked compounds

Compounds in Hungarian may be morphologically marked or morphologically unmarked. In the first case the morphological marker may appear either on the first or on the second member of the compound, e.g. *újjá+épít* (*új* 'new' + *-já* 'translative case suffix' + *épít* 'build') 'reconstruct', *tévét néz*[6] (*tévé* 'television' + *t* 'accusative case suffix' + *néz* 'look, watch') 'watch television'. In such cases the head of the compound is always a V and the nonhead is a syntactic or semantic argument of the verb. Note that neither *újjá* nor *tévét* are independent lexical items. Moreover syntactic rules may manipulate the internal structure of such

5 *ás/és* is a nominalizing suffix, the choice between the two forms is determined by vowel harmony. The usual phonological notation is *-Vs* where V denotes the harmonizing vowel, i.e. *-ás* or *-és*.

6 In contrast to phrases such as *könyv-et néz* book acc look 'look at a book, on books', *kép-et néz* 'look at a picture on pictures', which are not compound-like since they don't share any property of compounds. Cf. Section 2.2 for a more detailed discussion of 'bare object noun + verb' constructions.

compounds, in other words these compounds must be considered non-prototypical.

The morphological marker appears on the second member of the compound if it is derived from a possessive construction, e.g. *város+háza* (*város* 'city' + *ház* 'house' + *-a* 'possessive suffix') 'city hall', *tojás+fehérje* (*tojás* 'egg' + *fehér* 'white' + *-je* 'possessive suffix') 'egg-white'. Neither can the members of such compounds be separated by syntactic rules. In this sense they belong to prototypical rather than to non-prototypical compounds. Note that the second member of such compounds is not an independent word: **háza*, **fehérje*.[7] Though such compounds are rather frequent, it is unclear to what extent the pattern is productive and/or rule-governed.

Another case where the second member of the compound is morphologically marked are N+A compounds in which the head is derived from a past participle. In such compounds the participle is suffixed by the 3P personal suffix and the nonhead is interpreted as a kind of causer, i.e. of being the cause of the eventuality, normally referred to as Natural Force.

(3a) *vihar+ver-t-e*
storm+beat-PTCP-3SG
'storm-beaten'

(3b) *víz+mos-t-a*
water+wash-PTCP-3SG
'water-lashed'

Once again the participial head adjective of the compound is not an independent word: **verte*, **mosta*.[8] At first sight it would seem that in these compounds the first member satisfies the subject argument of the deverbal head. However, such an analysis would run counter the received view that subject arguments cannot be satisfied in compound structure (cf., for example, Di Sciullo/Williams 1987). The analysis of N+A constructions with participial heads as verbal compounds is not mandatory, however. It can be argued that these constructions are participial constructions rather than genuine compounds (cf. Kenesei 1986). Productive participial constructions must be distinguished from frozen ones, while the former can freely be modified, modification is impossible in the latter case. Compounds

7 Though it is often used in certain contexts as a shortened form of *tojásfehérje* 'egg-white'.
8 Note that *verte* and *mosta* are identical with the 3P Sg Past Tense forms of the verbs *ver* 'beat' and *mos* 'wash', respectively.

such as *víz+mosta* 'water-lashed', *por+lepte* 'covered with dust' are frozen expressions. In contrast, an expression such as (4),

(4) *munkás+lak-t-a*
worker+inhabit-PTCP-3SG
'inhabited by workers'

can be modified: it is possible to say *sok/kevés munkás lakta* 'inhabited by many/few workers'. Since modification of the nonhead is not possible in the case of genuine compounds we must conclude that the participial constructions such as (4) are not compounds.

1.3 Deverbal compounds

Deverbal compounds are special and have received much attention in the pertinent literature because there is a clear argument-head relationship between the elements of the compound. In this case two questions need to be answered: (i) what kind of arguments can the head inherit from its base; (ii) which arguments can be satisfied by the nonhead.

Nouns can be derived from verbs by means of the suffix *-ás* and in a considerable number of cases the derived nouns can be interpreted as event nouns, e.g. *ír-ás* 'writing' (from the verb *ír* 'write'), *olvas-ás* 'reading' (from the verb *olvas* 'read').[9] If such an event noun occurs as the head of a compound the nonhead can be interpreted as an argument of the verb. Apparently in the case of a deverbal noun derived from a transitive verb the only argument which can occur in nonhead position is the object argument:

(5a) *levél+ír-ás*
letter+write-NMLZ
'letter writing'
(5b) *könyv+olvas-ás*
book+read-NMLZ
'book reading'

9 In the case of resultative verbs the derived nominal may be ambiguous between the action and result reading. The deverbal noun Italics may mean the activity of writing but also the result of writing.

(5c) *levél+ír-ás-a Péternek*
letter+write-NMLZ-POS Peter.DAT
'writing a letter to Peter'
(5d) **levél Péternek+írás-a*
letter Peter.DAT+write-NMLZ-POS

In (5c) the dative form *Péternek* can never occur in compounds.

The situation is similar in (6) where *Péterrel* 'with Peter' is the comitative form of the noun:

(6a) *találkoz-ás Péterrel*
meet-NMLZ Peter.COM
'meeting with Peter'
(6b) **Péterrel találkoz-ás*
Peter.COM meet-NMLZ

The following generalizations hold:

(7a) If the deverbal head of a compound is derived from a transitive verb the only argument which can occur in nonhead position is the object argument.
(7b) No other internal argument can occur in compounds.

The subject argument is normally considered to be an external argument and it is claimed that external (subject) arguments can never occur in nonhead position. In Hungarian the following examples seem to contradict this generalization.

(8a) *hó+es-és*
snow+fall-NMLZ
'snowfall'
(8b) *motor+zúg-ás*
engine+buzz-NMLZ
'hum of the engine'
(8c) *dió+ér-és*
walnut+ripen-NMLZ
'ripening of walnuts'

(9a) *liba+gágog-ás*
goose+gaggle-NMLZ
'gaggling of a goose'

(9b) *kutya+ugat-ás*
dog+bark-NMLZ
'barking of a dog'
(9c) *gyermek+sír-ás*
child+cry-NMLZ
'crying of a child'

In the theory of thematic roles normally a distinction is being made between an intentionally acting (normally human) agent and an unintentionally acting actor. In both cases the nonhead is not an agent who acts intentionally in order to change the world, the event is rather brought about by natural force or an unintentionally acting actor. This means that the generalization (7b) can be saved if we restrict it to agent arguments, i.e. it can be claimed that agent arguments cannot occur in nonhead position. On the other hand, actor arguments are not excluded from this position. Notice furthermore that the compounds in (8a–c) and (9a–c) seem to fall into two semantic classes: (8a–c) describe phenomena of nature, while (9a–c) describe events of unintentional sound production.

Next consider the following examples. The verb *csökken* 'decrease' is intransitive, its transitive counterpart is *csökken-t*. Prices can decrease transitively and intransitively as shown by (10a–b).

(10a) *ár+csökken-és*
price+decrease.INTR-NMLZ
'drop in prices'
ár+drágul-ás
price+go.up-NMLZ
'rise of prices'
(10b) *ár+csökken-t-és*
price+decrease-ACC-NMLZ
'reduction of prices'
ár+drágít-ás
price+raise-NMLZ
'raising of prices'

The examples in (10a–b) demonstrate the difference between a head derived from an intransitive and a head derived from a transitive verb. In (10a) the nonhead can only be interpreted as the actor argument of the verbal base. In contrast the head in (10b) is derived from a transitive verb, hence the nonhead is interpreted as the object argument of the verbal base.

There are a number of compounds in which the nonhead looks very much like an actor argument but it can be shown that the relation between nonhead and head can only be interpreted conceptually but not syntactically. Consider:

(11a) *bolha+csíp-és*
flea+sting-NMLZ
'flea-bite'
(11b) *kutya+harap-ás*
dog+bite-NMLZ
'dog-bite'
(11c) *disznó+túr-ás*
pig+root-NMLZ
'rooting of pigs'

In the examples in (11) the head noun is a result nominal (referring to the result of biting or rooting) which has not inherited the argument structure of the base verb, hence argument satisfaction does not arise. The properties of result nominals are well-known from the relevant literature which we will not repeat here. Suffice it to mention that result nominals are incompatible with durative temporal adverbials while action nominals are.

Before embarking on the discussion of coordinative compounds it should be made clear that deverbal compounds can also be formed by means of the participial suffixes *-ó*[10] (present participle) and *-t* (past participle). E.g. *dió+darál-ó* 'nut grinder' and *sertés+sül-t* 'roast pork' (from *sül* 'roast').

1.4 Coordinative compounds

Formally, there are two main categories of coordinative compounds in Hungarian: actual coordinatives and compounds derived by lexical reduplication.

As Kiefer (2000: 525) points out, actual coordinative compounds are derived from free lexemes, as shown in (12a) below.

(12a) *ad-vesz* (from *ad* 'give' + *vesz* 'buy') 'mart, buy and sell'
jön-megy (from *jön* 'come'+ *megy* 'go') 'come and go, fidget'
üt-ver (from *üt* 'hit'+ *ver* 'beat') 'beat, pound'

10 Denoting the suffixes *-ó* or *-ő* where once again the choice is determined by vowel harmony.

jár-kel (from *jár* 'walk'+ *kel* 'traverse') 'go about, shuttle'
él-hal (from *él* 'live' + *hal* 'die') 'be overfond of sth'
eszik-iszik (from *eszik* 'eat' + *iszik* 'drink') 'eat and drink, regale oneself'

(12b) */?*rohan-szalad* (rush + run)
*/?*szeret-imád* (love + adore)
*/?*sír-bőg* (cry + bellow)
*/?*esik-zuhan* (fall + dive, tumble)
*/?*nyomtat-szkennel* (print + scan)

The ill-formed examples in (12b) above are meant to demonstrate the limited productivity of the construction type: the compounds in (12a) are all fully lexicalized, frozen items, while derivation from other non-bound elements seems to be rather problematic.

Another type of coordinative compounds is derived by lexical reduplication, which has several subcategories, as shown in (13).

(13a) *alig-alig* (hardly + hardly) 'hardly, with great difficulty'
sok-sok (many + many) 'very many'
olykor-olykor (sometimes + sometimes) 'rarely, seldom'

(13b) *egyszer-egyszer* (once + once) 'sometimes, rarely'
ki-ki (who + who) 'each'

(13c) *tarka-barka* (from *tarka* 'colourful, spotty')
'very colourful, spotty'
csiga-biga (from *csiga* 'snail') '(tiny, sweet) snail'
cica-mica (from *cica* 'kitten') '(tiny, sweet) kitten'

(13d) *dimbes-dombos* (from *domb* 'hill' + *-os* 'adjectivizing suffix')
'hummocky, full of hills'
girbe-görbe (from *görbe* 'curved') 'full of curves, sinuous'
rissz-rossz (from *rossz* 'bad') 'very bad'

(13e) *irul-pirul* (from *pirul* 'blush') 'blush, be blushful'
izeg-mozog (from *mozog* 'move') 'fidget, wiggle'
ici-pici (from *pici* 'tiny') 'very tiny'

The examples in (13a–b) demonstrate the case of total lexical reduplication, where the base is copied without modification. Semantically, the derivation serves the purpose of intensification, i.e. the meaning of the compound is analogous with that of the reduplicated base, which means that the derivation only adds the feature of intensification to the base (cf. 13a). However, in some lexicalized cases the meaning of the compound is totally different from that of the base (cf. 13b) (cf. Kiefer 2000: 524 f.; Brdar/Brdar-Szabó 2014).

Another type of lexical reduplication is when the base is copied with some kind of modification: either an initial consonant of the base is replaced by another one (cf. 13c), or there is a vowel alternation pattern similar to ablaut (cf. 13d). Brdar/Brdar-Szabó (2014: 39 f.) label the former phenomenon as *inexact total reduplication* or *rhyming*(*-motivated*) *reduplication*, and the latter as *ablaut-motivated reduplication*. Finally, the examples in (13e) are instances of *partial reduplication*, where only a segment of the base is copied (ibid.: 39).

Note that in these cases, too, the semantic feature added to the base is intensification, and the compounds mainly serve as stylistic versions of their bases: they mostly express the endearing attitude of the speaker, thus they should be dealt with in a morphopragmatic framework as well.

2 Compound-like phrases

We have already mentioned some cases of non-prototypical compounds; in the present section a more detailed analysis of such constructions will be provided.

2.1 Preverb + verb constructions

In Hungarian preverbs (particles attached to the verb base) are all separable and can fulfil various functions. If fully grammaticalized they express telicity, the most typical being the preverb *meg* which has completely lost its original meaning and has become an aspectual marker. Among other things, it can express the resultative Aktionsart as in the case of *főz* 'cook' – *meg+főz* 'cook.RES', *varr* 'sew' – *meg+varr* 'sew.RES' or the semelfactive Aktionsart as in *vakar* 'scrape' – *meg+vakar* 'scrape once', *csóvál* 'wag' – *meg+csóvál* 'wag once'.

Most preverbs are less grammaticalized yet they can be used to derive an Aktionsart. For example, the preverb *el* (whose original directional meaning is 'away') can be used to express inchoativity if it is accompanied by the reflexive pronoun *magát* 'self', e.g. *ordít* 'shout, cry' – *el+ordítja magát* 'cry out' or *nevet* 'laugh' – *el+neveti magát* 'burst out laughing'. In addition to *meg* some other originally directional preverbs can be used to express resultativity: *takarít* 'tidy, clean' – *ki+takarít* 'clean up', *gereblyéz* 'rake' – *fel+gereblyéz* 'rake up', *kaszál* 'scythe' – *le+kaszál* 'scythe.RES', *költ* 'spend' – *el+költ* 'spend.RES'.

At first sight Aktionsart-formation may seem to belong to derivational morphology. This would, however, contradict several generalizations concerning derivational morphology in Hungarian. First, derivational affixes harmonize with

the verbal stem (*szép-ség* 'beauty', *jó-ság* 'goodness'), in contrast, preverbs never harmonize.[11] Second, derivational affixes may change the part of speech category of the base which is not the case with preverbs. Third, derivational affixes are bound morphemes. On the other hand, preverbs can be detached from their base. First, they can be used in short answers to a question without their base as in (14–15) below.

(14a) *Meg+írtad a levelet?*
'Have you written the letter?'
(14b) *Meg*.
'Yes.'

(15a) *Ki+mentél a kertbe?*
'Have you gone out into the garden?'
(15b) *Ki*.
'Yes.'

Moreover, preverbs can freely be moved to various positions in the sentence, cf. the variants of (15a) in (16a–c).

(16a) *A kertbe mentél ki?*
(16b) *Ki a kertbe mentél?*
(16c) *Mentél ki a kertbe?*

We may thus conclude that the formation of complex verbs cannot be part of derivational morphology. On the other hand, preverb+verb constructions are not prototypical compounds either, at least not with respect to their behavior vis-a-vis syntax. In other words, their internal structure is accessible to syntactic rules. Yet they are compounds semantically as testified, among other things, by the large number of lexicalized forms. It should also be noted that a large number of preverbs are undistinguishable from the formally identical adverbs.

An interesting property of the Hungarian preverbs is that they can be reduplicated to express iterativity.[12] Consider:

11 Preverbs with a front vowel such as *ki* can easily be attached to back vowel stems as in *ki+mar* 'corrode', *ki+old* 'undo', *ki+rúg* 'kick out'.

12 Iterativity can also be expressed by the verbal suffix *-gat* which is, however, semantically radically different from the iterativity expressed by preverb reduplication.

(17a) *Ki-ki+megy a kertbe*.
PREV-PREV+go the garden.LOC
'From time to time he/she goes out into the garden.'
(17b) *Meg-meg+ír egy levelet*.
PREV-PREV+write the letter.ACC
'From time to time he/she writes a letter.'

The type of iterativity is one of the Aktionarten in Hungarian which, however, is not expressed by a particular preverb or suffix but by reduplicating the preverb. Note that reduplicated preverbs cannot be separated from the verb base by another constituent and they cannot be moved after the verbal base either. From this property it follows that reduplicated verbs cannot be negated since the negative particle *nem* must immediately precede the verbal base, cf. (18). External negation is, of course, possible (19).

(18a)	**Nem*		*meg-meg+ír*			*egy*	*levelet*.
	not		PREV-PREV+write			a	letter.ACC
(18b)	**Nem*		*ír*	*egy*	*levelet*		*meg-meg*.
	not		write	a	letter.ACC		PREV-PREV

(19)	*Nem*	*igaz*,	*hogy*	*meg-meg+ír*		*egy*	*levelet*.
	not	true	that	PREV-PREV+write		a	letter.ACC
	'It is not true that he always (repeatedly) writes a letter.'						

These properties seem to suggest that reduplicated forms are not only semantically but also syntactically words. First they have a specific meaning (to do something repeatedly), second syntactic rules cannot change their internal structure.

Preverb reduplication is not possible across the board: it must obey a phonological and several semantic constraints. The phonological constraint refers to the length of the preverb in terms of the number of syllables: preverbs longer than two syllables cannot be reduplicated, as shown by (20).

(20a) **utána-utána+megy* 'go after, follow' (lit. after-after go)
(20b) **keresztül-keresztül+vág* 'cut through' (lit. through-through cut)

As far as the semantic constraints are considered, apparently activities if pushed to the extreme cannot be reduplicated. The preverbs *túl* 'over', *agyon* 'over', *tönkre* 'over' are used to express the extreme degree of an activity, therefore it does not come as a surprise that such preverbs cannot be reduplicated. Consider:

(21a) **túl-túl+hangsúlyoz* 'over stress' (lit. over-over stress)
(21b) **agyon-agyon+hajszol* 'over-fatigue, work to death' (lit. over-over work)
(21c) **tönkre-tönkre+dolgozza magát* 'work oneself to death' (lit. over-over work)

2.2 Bare noun + verb constructions

According to the literature (Kiefer 1990; Farkas/de Swart 2003), Hungarian bare noun + verb constructions (in short, BNV constructions) are instances of type I noun incorporation in terms of Mithun (1984). Mithun describes the phenomenon as a type of compounding where a verb and a noun with the semantic function of patient, location or instrument combine to form a new complex verb. The eventuality designated by the BNV construction is not just a random co-occurrence of an entity and an eventuality, but it is perceived as a recognizable, unitary concept worth labelling (cf. Mithun 1984: 848 f.).

We consider the Hungarian BNV construction type as a special case of compounding by juxtaposition, the general characteristics of which are briefly captured by Mithun as follows:

> A number of languages contain a construction in which a V and its direct object are simply juxtaposed to form an especially tight bond. The V and N remain separate words phonologically; but as in all compounding, the N loses its syntactic status as an argument of the sentence, and the VN unit functions as an intransitive predicate. The semantic effect is the same as in other compounding: the phrase denotes a unitary activity, in which the components lose their individual salience. (ibid.: 849)

The examples in (22)–(23) below demonstrate some of the commonly recognized features of the Hungarian BNV construction type.

(22a)	*Péter*	*újságot*	*olvas*.
	Peter	newspaper.ACC	read
	Péter	*zenét*	*hallgat*.
	Peter	music.ACC	listen
	Péter	*tanulmányt*	*ír*.
	Peter	article.ACC	write
	Péter	*keresztrejtvényt*	*fejt*.
	Peter	crossword.ACC	solve
	Péter	*ruhát*	*próbál*.
	Peter	outfit.ACC	try on

'Peter is reading (a) newspaper(s) / listening to music / writing an article / solving (a) crossword puzzle(s) / trying on (an) outfit(s).'

(22b)

Péter	*olvassa*	*az*	*újságot*.
Peter	read.3SG.DEF	the	newspaper.ACC
Péter	*hallgatja*	*a*	*zenét*.
Peter	listen.3SG.DEF	the	music.ACC
Péter	*írja*	*a*	*tanulmányt*.
Peter	write.3SG.DEF	the	article.ACC
?Péter	*fejti*	*a*	*keresztrejtvényt*.
Peter	solve.3SG.DEF	the	crossword.ACC
?Péter	*próbálja*	*a*	*ruhát*.
Peter	try on.3SG.DEF	the	outfit.ACC

'Peter is reading the newspaper / listening to the music / writing the article / solving the crossword puzzle / trying on the outfit.'

(23)

*/?Péter	*újságot*	*olvas*,	*és*	*elégedett*	*vele*.
Peter	newspaper.ACC	read	and	content	INSTR
*/?Péter	*zenét*	*hallgat*,	*és*	*elégedett*	*vele*.
Peter	music.ACC	listen	and	content	INSTR
*/?Péter	*tanulmányt*	*ír*,	*és*	*elégedett*	*vele*.
Peter	article.ACC	write	and	content	INSTR
*/?Péter	*keresztrejtvényt*	*fejt*,	*és*	*elégedett*	*vele*.
Peter	crossword.ACC	solve	and	content	INSTR
*/?Péter	*ruhát*	*próbál*,	*és*	*elégedett*	*vele*.
Peter	outfit.ACC	try on	and	content	INSTR

'Peter is reading (a) newspaper(s) / listening to music / writing an article / solving (a) crossword puzzle(s) / trying on (an) outfit(s), and he is content with it.'

As pointed out by Kiefer (1990: 153 f.) and shown in (22) above, Hungarian BNVs form one single phonological unit from the point of view of stress assignment (i.e., only the subject and the incorporated object bear stress on their first syllable, cf. 22a), while their V + DP counterparts show the opposite pattern (i.e., the subject, the verb and the direct object all bear separate stress on their first syllable, cf. 22b). The ill-formedness of some of the constructions in (23) is due to the fact that some of these BNVs, namely *keresztrejtvényt fejt* 'solve crossword puzzles' and *ruhát próbál* 'try on outfits' seem to be lexicalized units without exact syntactic paraphrases, e.g. V + DP counterparts.

One of the key semantic features of direct object incorporation, often mentioned in the literature (cf. Mithun 1984; Kiefer 1990; Farkas/de Swart 2003), is the non-referentiality of the bare object noun, which means that the nouns in these BNV constructions do not denote any specific, identifiable entity in the

world. This feature can be tested by adding an anaphoric pronominal constituent to the sentence, as in (23) above. The examples in (23) are ill-formed because the nouns in each construction have a type referring function, i.e. they only add a specific classificatory feature/component to the eventuality expressed by the verb.

(24a) *Péter érdekes újságot olvas, és elégedett vele.*
Peter interesting newspaper.ACC read and content INSTR
Péter érdekes tanulmányt ír, és elégedett vele.
Peter interesting article.acc write and content INSTR
'Peter is reading an interesting newspaper / writing an interesting article, and he is content with it.'[13]

(24b) *Péter egy érdekes újságot olvas, és elégedett vele.*
Peter a interesting newspaper.ACC read and content INSTR
Péter egy érdekes tanulmányt ír, és elégedett vele.
Peter a interesting article.ACC write and content INSTR
'Peter is reading an interesting newspaper / writing an interesting article, and he is content with it.'

The constructions in (24a) above are meant to demonstrate the effects of modification on BNV constructions. The inserted adjective overrides the non-referentiality property of the object noun and – as a consequence – the complex eventuality meaning of the BNVs. This means that we are dealing with at least two different construction types from the point of view of semantics and discourse transparency, as shown by the fact that, contrary to the case of (23), the modified version of the construction admits the insertion of an anaphoric pronominal constituent into the sentence. As noted in Kiefer (1990: 152), the constructions like those in (24a) seem to be some kind of stylistic variants of the full-fledged construction types shown in (24b).

The number neutrality of the singular incorporated noun is another important characteristic of BNVs, and it is strongly connected to the above mentioned non-referentiality feature. As Farkas/de Swart (2003: 13 f.) point out, morphologically singular incorporated nouns are compatible with both atomic and non-atomic interpretations. Most of the examples in (22a) above are underspecified regarding the number of objects involved in the eventualities described by the BNVs. The singular noun in the BNV *újságot olvas* 'read (a) newspaper(s)', for

13 Similar things were discussed in considerable detail in Maleczki (1994).

instance, allows for both an atomic (singular) and a non-atomic (plural) interpretation, i.e. the BNV does not specify whether Peter is reading one newspaper or several newspapers one after the other. As shown by the examples in (25) below, the varying interpretations are influenced by pragmatic (contextual) information. The BNV in (25a) triggers an atomic interpretation due to extra linguistic knowledge about marriage related customs (though it would allow for a non-atomic interpretation in the context of legal bigamy), the one in (25b) clearly triggers an atomic interpretation (without any cultural variation), finally, the one in (25c) unambiguously triggers a non-atomic interpretation.

(25a) *Feri* *feleséget* *keres*. (Farkas/de Swart 2003: 14)
Feri wife.ACC search
'Feri is looking for a wife.'

(25b) *Anna* *napfelkeltét* *néz* *az* *erkélyen*.
Anna sunrise.ACC watch the balcony.LOC
'Anna is watching the sunrise on the balcony.'

(25c) *Mari* *bélyeget* *gyűjt*. (ibid.: 13)
Mari stamp.ACC collect
'Mari is collecting stamps.'

As far as plural bare objects are concerned, the following generalization holds: plural bare object nouns form grammatical BNVs, however, as shown in (26) below, their discourse transparency properties are similar to the ones of modified singular objects, as shown in (25a) above.

(26a) *Anna* *leveleket* *ír*, *és* *elküldi* *őket*.
Anna letter.PL.ACC write and PREV.send.3SG.DEF them
'Anna writes letters and sends them.'

(26b) *Az* *orvos* *betegeket* *vizsgál*, *és* *megpróbál* *segíteni* *rajtuk*.
The doctor patient.PL.ACC examine and PREV.try help.INF LOC.3PL
'The doctor examines patients and tries to help them.'

Finally, a distinction must be made between fully productive and idiomatic cases. As pointed out in Kiefer (1990), the meaning of idiomatic BNVs cannot be derived from a corresponding free construction (cf. the examples in (27)–(28) below), while fully productive BNVs generally have matching syntactic paraphrases as already demonstrated by the examples in (23a–b) above.

(27a) A *behaviorista szemlélet gyökeret vert a*
the behaviorist approach root.ACC beat.PST the
nyelvészetben is.
linguistics.LOC too
'The behaviorist approach invaded linguistics as well.'

(27b) *Péter bocsánatot kért a barátjától*.
Peter forgiveness.ACC ask.PST the friend.3SG.POSS.LOC
'Peter apologized to his friend.'

(27c) *Az autó tegnap gazdát cserélt*.
the car yesterday owner.ACC change.PST
'The car changed owners yesterday.'

(27d) *Mari gyereket vár*.
Mari child.ACC wait
'Mari is pregnant.'

(28a) **A behaviorista szemlélet verte a*
the behaviorist approach beat.PST.3SG.DEF the
gyökeret a nyelvészetben is.
root.ACC the linguistics.LOC too

(28b) **Péter kérte a bocsánatot a barátjától*.
Peter ask.PST.3SG.DEF the forgiveness.ACC the friend.3SG.POSS.LOC

(28c) **Az autó tegnap cserélte a*
gazdá(já)t.
the car yesterday change.PST.3SG.DEF the
owner.(3SG.POSS.)ACC

(28d) *Mari várja a gyereket* /
Mari wait.3SG.DEF the child.ACC /
vár egy gyereket.
wait a child.ACC
'Mari is waiting for the / a kid.'

The difference between the lexicalized BNVs in (27a–c) and (27d) is that the former type cannot be grammatically matched with a syntactic paraphrase (cf. (28a–c)), while the latter construction type has a well-formed syntactic paraphrase, however, (synchronically) this paraphrase has nothing to do with the meaning of its BNV counterpart (compare (27d) and (28d)).

As mentioned above, the most prominent and universal semantic and pragmatic feature of BNVs is that the eventuality designated by the construction has to be perceived as a recognizable, unitary concept worth separately labelling. This 'institutionalized' character of the complex activity expressed by the BNV seems to be a strong criterion regarding the derivation of the construction type. Thus it does not come as a surprise that not all bare objects are admitted in BNV constructions with equal ease. Consider the examples in (29b) and (29d) which, as opposed to those in (29a) and (29c), are odd on their generic reading.

(29a)	*Mari*	(*épp*)	*újságot*	*olvas*	*a*	*szobájában*.
	Mari	just	newspaper.ACC	read	the	room.3SG.POSS.LOC
	'Mari is reading the newspaper in her room.'					
(29b)	*Mari*	(*épp*)	*csomagolást*	*olvas*	*a*	*húsrészlegen*.
	Mari	just	package.ACC	read	the	meat aisle.LOC
	'Mari is reading (a) package(s) in the meat aisle.'					
(29c)	*Virágék*	(*épp*)	*vendéget*	*várnak*.		
	Virág.PL	just	guest.ACC	wait.3PL		
	'The Virágs are waiting for (a) guest(s).'					
(29d)	*Virágék*	(*épp*)	*világvégét*	*várnak*.		
	Virág.PL	just	apocalypse.ACC	wait.3PL		
	'The Virágs are waiting for the end of the world.'					

The oddness of (29b) is caused by the fact that, generally speaking, reading packages is not considered a recognizable, re-occurring complex eventuality, however, the BNV in question becomes acceptable if matched with a proper context: if, for example, the participants of the speech situation know that Mari has a habit of reading the package of meat products trying to avoid certain ingredients. The same holds true for (29d) as well: waiting for the end of the world is generally not perceived as an 'institutionalized' activity, nevertheless, the use of the BNV is justified in the context of knowing that the Virágs have prepared for the end of the world on several occasions in the past due to false predictions.

These types of marginal examples show that, although there may be some pragmatic factors that influence the derivation of BNVs, if the contextual factors match the corresponding pragmatic criteria, even seemingly odd BNVs will be considered well-formed.

Finally, mention must be made of the aspectual restrictions filtering the range of input verbs. The generalization seems to be as follows: activity/process verbs, i.e. [+dynamic, −telic] verbs potentiate well-formed BNVs, while accomplishment and achievement verbs, i.e. [+dynamic, +telic] verbs as well as stative,

i.e. [–dynamic, –telic] verbs do not tend to form grammatical constructions (cf. Kiefer 1990), as shown by the examples in (30) below.[14]

(30a) *Péter újságot elolvasott.
Peter newspaper.ACC PREV.read.PST
*Péter zenét meghallgatott.
Peter music.ACC PREV.listen.PST
*Péter keresztrejtvényt megfejtett.
Peter crossword.ACC PREV.solve.PST
'Peter read the newspaper / listened to music / solved a cross-word puzzle.'

(30b) ?István keze autót érintett az utcán.
István hand.3SG.POSS car.ACC touch.PST the street.LOC
'István's hand touched a car on the street.'

(30c) ?Anna barátot hívott, mert egyedül nem
Anna friend.ACC call.PST because alone not
tudta megoldani a problémát.
can.PST solve.INF the problem.ACC
'Anna called (for) a friend, as she could not solve the problem alone.'

(30d) *Tamás poharat tört a konyhában,
Tamás glass.ACC break.PST the kitchen.LOC
és rögtön bocsánatot kért.
and immediately forgiveness.ACC ask.PST
'Tamás broke a glass in the kitchen and immediately apologized for it.'

(30e) *Éva fiút szeretett, de nem lett jó vége.
Eva boy.ACC love.PST but not become good end.3SG.POSS
'Eva loved a boy, but it did not end well.'

(30f) *Laci hegyet látott a kiránduláson.
Laci mountain.ACC see.PST the trip.LOC
*Laci hegyet látott, amikor fölhívtam.
Laci mountain.ACC see.PST when call.PST.1SG.DEF
'Laci saw a mountain on the trip / when I called him.'

14 We use the terms *activity*, *achievement*, *accomplishment* and *state* according to the Vendlerian tradition well known in the literature on aspect. Vendler (1967) isolated four situation types: states (e.g. *love*, *know*, etc.), activities (e.g. *run*), achievements (e.g. *reach the summit*) and accomplishments (e.g. *draw a circle*). For more on these aspectual categories, cf. Smith (1991), Tenny (1994), Kiefer (2006), etc.

(30g) **Matyi titkot tudott, és hosszú*
Matyi secret.ACC know.PST and long
ideig nem mondhatta el senkinek.
time.TEMP not tell.COND.PST PREV nobody.DAT
'Matyi knew a secret, and he was not allowed to tell it to anyone for a long time.'

According to these examples, the above generalization seems to hold true for Hungarian BNVs. The constructions in (30a–d) derived from telic verbs are ungrammatical, although a distinction should be made between prefixed and unprefixed telic verbs, as the latter are invariably ungrammatical in these constructions, while in some cases the former may serve as acceptable input verbs (as shown in (31a–b) below).[15] The ungrammatical BNVs like those in (30e–g) lead to the conclusion that stative verbs are indeed excluded from the range of possible input verbs, however, as shown in (31d–e), we may find some grammatical BNVs derived from stative verbs as well.

(31a) *István keze labdát érintett, és*
István hand.3SG.POSS ball.ACC touch.PST and
a biró észrevette.
the referee observe.PST
'István's hand touched the ball, and the referee saw it.'

(31b) *Anna mentőt hívott, mert egyedül nem*
Anna ambulance.ACC call.PST because alonenot
tudta megoldani a problémát.
can.PST solve.INF the problem.ACC
'Anna called an ambulance, as she could not solve the problem alone.'

(31c) *Tamás diót tört a kalákán.*
Tamás nut.ACC break.PST the group work.LOC
'Tamás was cracking nuts at the group work.'

15 The distributional properties of these verb classes are captured in Kiefer (1990: 169) as follows: "Syntactically, both the bare noun and the prefix belong to the same class of elements, often referred to as preverb since under normal circumstances an element of this class occupies the position immediately preceding the verb. Consequently, two preverbs can never co-occur."

(31d) *Mari* *fájdalmat* *érzett* *a* *bal* *lábában*,
Mari pain.ACC feel.PST the left foot.3SG.POSS.LOC
ezért *orvoshoz* *ment*.
hence doctor.LOC go.PST
'Mari felt pain in her left leg, so she went to the doctor.'

(31e) *Az* *éjjeliőr* *zajt* *hallott*, *ezért*
the night-watchman noise.ACC hear.PST hence
újra *ellenőrizte* *a* *folyosókat*.
again check.PST the hallway.PL.ACC
'The night watchman heard noise, so he checked the hallways again.'

The well-formed examples in (31) violate the aspectual criteria formulated above, so we need to take a closer look at the semantic and pragmatic features of these BNVs. The sentences in (31a–b) contain BNVs derived from telic verbs, while the ones in (31d–e) contain stative verbs. The example in (31c), contrasted with (30d), is meant to demonstrate how contextual non-atomicity entailments induce aspectual coercion in the case of punctual verbs (the BNV triggers an iterative interpretation, otherwise, with an atomic interpretation, it would be considered ill-formed, like the one in (30d) above; and reversely: the BNV *poharat tör* 'break glasses' becomes well-formed with an iterative and habitual interpretation).

The common feature of these BNVs is that they all denote institutionalized, re-occurring eventualities. The institutionalized nature of the eventualities expressed by (31a–b) is also shown by their contrast with the constructions in (30b–c) above: in football, touching the ball with one's hand is a frequent, punishable occurrence. The same institutionalized character holds true for the eventuality of calling an ambulance and for the stative predicates in (31d–e).

Based on these observations, we conclude that the aspectual criterion described above should be reduced to a remark regarding the prevalency of process verbs in BNVs, as the range of verbs which (potentially) denote institutionalized eventualities strongly overlaps with the category of process verbs, however, some telic and stative verbs also describe eventualities which satisfy the pragmatic criterion controlling BNV formation.

3 Summary

In the present paper we have summarized the most important facts concerning compounds and compound-like phrases (= non-prototypical compounds) in Hungarian. We have concentrated on the productive, or at least regular patterns of compounding and derivation of compound-like constructions. In particular,

we have stressed the features which deviate from "Standard Average European". Some of such features can be found in the case of deverbal compounds as well, e.g. that the subject argument can be satisfied in compounds which does not seem to be the case in Germanic or Romance. However, the most striking feature of Hungarian compounding is the existence of bare noun constructions and their relation to verbal aspect.

References

Brdar, Mario/Brdar-Szabó, Rita (2014): Syntactic reduplicative constructions in Hungarian (and elsewhere): Categorization, topicalization and concessivity rolled into one. In: Rundblad, Gabriella et al. (eds.): Selected Papers from the 4th UK Cognitive Linguistics Conference. London: UK Cognitive Linguistics Association. 36–51.

Di Sciullo, Anna Maria/Williams, Edwin (1987): On the definition of word. Cambridge, MA: The MIT Press.

Farkas, Donka/de Swart, Henriëtte (2003): The semantics of incorporation: From argument structure to discourse transparency. Stanford, CA: CSLI Publications.

Kenesei, István (1986): On the role of the agreement morpheme in Hungarian. ALH 86, 1–4. 104–120.

Kiefer, Ferenc (1990): Noun incorporation in Hungarian. In: Acta Linguistica Hungarica 40, 1–2. 149–177.

Kiefer, Ferenc (1992): Compounding in Hungarian. Rivista di Linguistica 4, 1. 45–55.

Kiefer, Ferenc (1993): Thematic roles and compounds. Folia Linguistica 27, 1–2. 25–55.

Kiefer, Ferenc (2000): A szóösszetétel [Compounds]. In: Kiefer, Ferenc (ed.): Strukturális magyar nyelvtan 3. Morfológia. Budapest: Akadémiai Kiadó. 519–568.

Kiefer, Ferenc (2006): Aspektus és akcióminőség – különös tekintettel a magyar nyelvre [Aspect and Aktionsart – with special emphasis on Hungarian]. Budapest: Akadémiai Kiadó.

Kiefer, Ferenc (2009): Compounding in Hungarian. In: Lieber, Rochelle/Stekauer, Pavol (eds.): Oxford handbook of compounding. Oxford: Oxford Handbooks. 527–541.

Kiefer, Ferenc/Németh, Boglárka (2018): Aspectual constraints on noun incorporation in Hungarian. In: Zoltán, Huba Bartos/den Dikken, Marcel/Váradi, Tamás (eds.): Boundaries crossed: Studies of the crossroads of morphosyntax, phonology, pragmatics, and semantics. Berlin: Springer. 21–32.

Ladányi, Mária (2007): Produktivitás és analógia a szóképzésben [Productivity and analogy in word formation]. Budapest: Tinta Könyvkiadó.

Maleczki, Márta (1994): Bare common nouns and their relation to the temporal constitution of events in Hungarian. In: Dekker, Paul/Stokhof, Martin (eds.): Proceedings of the Eighth Amsterdam Colloquium. Amsterdam: Institute for Logic, Language and Computation, University of Amsterdam. 347–365.

Mithun, Marianne (1984): The evolution of noun incorporation. Language 60. 847–895.

Smith, Carlota (1991): The parameter of aspect. Dordrecht: Springer.

Tenny, Carol L. (1994): Aspectual roles and the syntax-semantics interface. Dordrecht: Springer.

Vendler, Zeno (1967): Verbs and times. Linguistics in philosophy. Ithaca/New York: Duke University Press. 97–121.