

## Differential Object Marking in Romance

# **Beihefte zur Zeitschrift für romanische Philologie**



Herausgegeben von  
Éva Buchi, Claudia Polzin-Haumann, Elton Prifti  
und Wolfgang Schweickard

## **Band 459**

# Differential Object Marking in Romance



The third wave

Edited by

Johannes Kabatek, Philipp Obrist and Albert Wall

**DE GRUYTER**



An electronic version of this book is freely available, thanks to the support of libraries working with Knowledge Unlatched. KU is a collaborative initiative designed to make high quality books Open Access. More information about the initiative and links to the Open Access version can be found at [www.knowledgeunlatched.org](http://www.knowledgeunlatched.org).

ISBN 978-3-11-064656-6

e-ISBN (PDF) 978-3-11-071620-7

e-ISBN (EPUB) 978-3-11-071623-8

ISSN 0084-5396

DOI <https://doi.org/10.1515/9783110716207>



This work is licensed under the Creative Commons Attribution 4.0 International License. For details go to <https://creativecommons.org/licenses/by/4.0/>.

**Library of Congress Control Number: 2021940289**

**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>.

© 2021 with the authors, editing © 2021 Johannes Kabatek, Philipp Obrist, Albert Wall, published by Walter de Gruyter GmbH, Berlin/Boston.

The book is published open access at [www.degruyter.com](http://www.degruyter.com).

Typesetting: Integra Software Services Pvt. Ltd.

Printing and binding: CPI books GmbH, Leck

[www.degruyter.com](http://www.degruyter.com)

# Contents

## I Introductory remarks

Johannes Kabatek, Philipp Obrist and Albert Wall

**The third wave of studies on DOM in Romance: An introduction to this volume — 3**

Georg Bossong

**DOM and linguistic typology: A personal view — 21**

## II Research perspectives

Chantal Melis

**From topic marking to definite object marking: Focusing on the beginnings of Spanish DOM — 39**

Alessia Cassarà and Sophie Mürmann

**Role-semantic parameters for DOM in Italian — 65**

Elisabeth Mayer and Liliana Sánchez

**Emerging DOM patterns in clitic doubling and dislocated structures in Peruvian-Spanish contact varieties — 103**

Albert Wall and Philipp Obrist

**Multilingualism effects in an elicitation study on Differential Object Marking in Cusco (Peru) and Misiones (Argentina) — 139**

Alina Tigău

**Differential Object Marking in Romanian and Spanish: A contrastive analysis between differentially marked and unmarked direct objects — 173**

Niklas Wiskandt

**Scale-based object marking in Spanish and Portuguese: *leísmo*, null objects and DOM — 213**

Anna Pineda

**The development of DOM in the diachrony of Catalan: (Dis)similarities with respect to Spanish — 243**

Senta Zeugin

**DOM in Modern Catalan varieties: An empirical study based on acceptability judgment tasks — 279**

Diego Romero Heredero

**Telicity and Differential Object Marking in the history of Spanish — 315**

Javier Caro Reina, Marco García García and Klaus von Heusinger

**Differential Object Marking in Cuban Spanish — 339**

**Index — 369**

---

## I Introductory remarks





Johannes Kabatek, Philipp Obrist and Albert Wall

# The third wave of studies on DOM in Romance

An introduction to this volume

## 1 Preliminary remarks

Recent years have seen increased research interest in Differential Object Marking (DOM, Bossong 1982; 1985; cf. also Bossong, this volume). DOM is not only the subject of specialized workshops but also figures frequently at conferences on language typology, generating an enormous body of published work, so that it becomes more and more difficult to maintain an overview. Rather than adding new complexity to this picture, the present volume brings DOM back to its origins, the Romance languages, and aims to deliver fresh insights arising from new data, new methods, and new theoretical approaches.

If we consider the history of studies on DOM in Romance, we can identify three different phases or “waves” of studies (the allusion to Eckert 2012 is intended). The first could be called the “prehistorical” one, referring as it does to studies about the phenomenon prior to the terminologization proposed by Bossong. These studies refer to DOM in the Romance languages generally as “prepositional accusative”, and they should not be ignored by contemporary research: some pioneering studies (such as the explanations of Spanish DOM in Bello 1847 or Lenz 1920) are still relevant today and contain important intuitions. In this phase, DOM is considered mainly as a typical Romance evolution emerging from the Latin preposition *ad* (with the particular and well-known deviation of Romanian), and research focuses on the functional status or on diachronic evolutions in one language or in the language family. The second wave could be called the phase of “typologization”, after Bossong’s claims about DOM in the early 1980s, and opens the discussion to the languages of the world. DOM is not only a Romance phenomenon but can be found widely, and DOM in Romance is not necessarily only an internal feature but may also be conditioned or catalyzed by contact with languages such as Hebrew. Typologization also means that

---

**Johannes Kabatek**, University of Zurich, e-mail: kabatek@rom.uzh.ch

**Philipp Obrist**, University of Zurich, e-mail: philipp.obrist@uzh.ch

**Albert Wall**, University of Vienna, e-mail: albert.wall@univie.ac.at

research on DOM moves from a semasiological, language-specific view to an onomasiological perspective and to general principles of grammaticalization. Thus it was recognized that DOM grammaticalizes similarly in very different languages, with the two factors of animacy and definiteness being the most determinative. And DOM as a general marking strategy does not depend on a prepositional marker, but may be expressed through different devices.

What, then, is the third wave like? In Romance linguistics, there are two main directions for recent studies on DOM. On the one hand, researchers go back to apparently established issues and take a closer, more fine-grained look at the data. Typological generalizations have opened up new perspectives, but in doing so they have also broadened the horizon towards related phenomena. The interplay of different means of object differentiation within the same language is studied (e.g. clitic doubling or “indexing” vs. traditional “flagging” DOM, i.e. marking by an element that immediately stands by the object). Not only objects are considered; instead, whole constructions are also taken into account. And newly identified factors such as agentivity, telicity and affectedness are considered when looking closely at language variation. On the other hand, new methodological perspectives have been opened. In historical studies, the availability of large corpora makes it possible to work with comprehensive databases and to identify statistically relevant factors, while in studies concerning contemporary usage, experimental methods allow for controlled studies with fine-grained manipulations.<sup>1</sup> In the next Section, we will introduce current issues that characterize the third wave of DOM studies in Romance: the debate about the general concept (2.1), the conditioning factors (2.2), more fine-grained and comparative approaches to variation (2.3) and to contact (2.4), and methodological innovations (2.5). Section 3 then relates the contributions of this volume to the issues introduced in Section 2.

---

<sup>1</sup> To mention just one example, the research projects “Differential Object Marking in Spanish. Emergence and tendencies of the current system” and “Experimental morphosyntax of Romance languages”, both funded by the Swiss National Science Foundation at the University of Zurich, investigate DOM systems in Romance languages with corpora and experimental methods. Among other things, the projects collect empirical data on the variation of DOM in the Spanish-speaking world and in Catalan, they investigate contact scenarios of the DOM languages Spanish and Romanian, and they look at rudimentary DOM systems in Brazil and Portugal. Furthermore, they provide an original description of the use of DOM in Rhaeto-Romance varieties.

## 2 Current issues in the studies of Romance DOM

### 2.1 What is DOM and what is it not?

In his contribution to this volume, Bossong points out that a crucial moment in his attempt to understand the Spanish “prepositional accusative” (1) was to look at it from a rather abstract perspective: differentiation of elements within one syntactic function or argument.

- (1) a. *Veo*            \*Ø/a *la*            *guitarrista*.  
       see.PRS.1SG    DOM    the.FEM    guitar player
- b. *Veo*            Ø/\*a *la*            *guitarra*.  
       see.PRS.1SG    DOM    the.FEM    guitar

This differentiation is conditioned by a number of factors, but rather than getting lost in the complexity of the interaction of the different factors, Bossong highlights the availability of a differentiation device. From this general perspective, the device marks prominent (or non-prototypical) objects. Such devices exist in a series of Romance languages whose grammatical characteristics overlap to different degrees. This group consists of varieties of Portuguese, Spanish, Catalan, Sardinian, Italian, Rhaeto-Romance and Romanian. With the exception of Rhaeto-Romance, all these languages are discussed in the present volume and exemplification can be found in the respective chapters.

Bossong’s abstract conception at the same time serves quite well as a comparative concept for typology (Haspelmath 2010), and soon many other languages with such devices were “discovered”, by Bossong and other researchers. It also was pointed out that such differentiation also exists for subjects in some languages (Differential Subject Marking) and both notions were brought together under the term Differential Argument Marking (DAM, cf. Witzlack-Makarevich/Seržant 2018 for an overview). In typology, the notion is usually understood in a broader sense than in Bossong’s work, where the device is characterized by traditional conditioning factors (above all, animacy and definiteness), a morphological marker (so-called “grammemic marking”) and certain grammaticalization paths. The success of the broader application of the term in typology, and the discussion of more and more grammatical configurations as instances of this very same sort of device, made it necessary at some point to introduce entire DAM typologies (Witzlack-Makarevich/Seržant 2018). Eventually, the broader approach made its way back into Romance linguistics. Bossong himself had already identified clitic doubling as a second DOM system in Spanish. More recently, other phenomena from the object domain have been added and discussed in terms of DOM, includ-

ing Spanish *leísmo* (Flores/Melis 2007, among others), null vs. overt objects in Portuguese (Schwenter 2014), and most recently, Differential Goal Marking in Spanish (Melis/Rodríguez Cortés 2017). All these phenomena feature splits in the object domain and can be related to the semantic hierarchies from the seminal work of Silverstein (1976). As it turns out, most of these phenomena, when occurring in other Romance DOM languages, could also be related to the notion of DOM in one way or another.

However, these phenomena are also different in many ways and it is not always clear what insights are gained for the description/analysis of language-specific grammars when subsuming all such phenomena under one (umbrella) term. If we accept the idea that all the aforementioned phenomena instantiate DOM, only a small number of Bossong's constituent features for the definition of DOM remain valid. The idea of a common grammaticalization pathway must be excluded and therefore also the idea that DOM systems may arise when the case system of a language collapses, as occurred in Latin, since this idea applies to flagging DOM in Romance, but not necessarily to other types. The condition of the split being realized by a privative opposition with a grammatical morphemic marker is obviously also obsolete under the umbrella view. The common remaining core for a definition or the insight, therefore, is the manifestation of some sort of split in the object domain localized on the Silverstein hierarchies.

Since many (if not all) of these phenomena exist in different Ibero-Romance languages, and also differ from language to language, the same question applies at the level of the language family. For instance, if all the phenomena are instances of DOM, the question arises as to whether this multiple and extraordinarily subtle differentiation is a special feature of (Ibero-)Romance or whether other DOM languages, once subjected to a comparable fine-grained analysis, will also show multiple differentiations, maybe of different kinds, in the object domain. As a further abstraction or generalization, it has been observed that the marking of non-prototypicality is not restricted to the arguments of a verb. DOM or DAM, on this view, could be seen as a consequence of the universal tendency in grammar to mark prominent elements with longer forms (Haspelmath 2020). However, the more abstract this unifying typological account becomes, the more phenomena it leaves unaccounted for in comparison to Bossong's more restricted definition and explanation. Thus, DOM seems not only to be conditioned by multiple factors, but it is also a concept applied at different levels of linguistic description. The typologist's DOM is not always identical to what is understood as DOM in Romance linguistics, and the same holds for the application of the concept to individual Romance grammars.

## 2.2 “Local” and “global” factors

Depending on how widely the notion of DOM is understood, the factors to be taken into account when studying it will be different. However, what will remain relevant is an overall distinction between “local” and “global” factors (Laca 2006, 430). “Local” factors refer to the characteristics of the marked or unmarked object NP, whereas “global” factors include a larger context within the sentence or even beyond. As far as this distinction is concerned, the history of DOM studies is not linear, and there seems to be a certain back and forth in terms of focus, or a coexistence of studies centring on different aspects. If we look at studies of the first wave, the principal interest lies in the characteristics of the objects. However, there is also a long tradition which claims that the main purpose of DOM is distinguishing subjects and objects in languages which lack other morphological devices for this purpose; thus, analyzing it will involve the whole sentence. A pioneering view in this sense can be seen in Lenz (1920), where it is claimed that the semantic characteristics of the object NP derive from the higher principle of differentiation. The fact that animate objects are more frequently marked than inanimate ones would then not be primarily due to the need to mark animates, but rather to the higher frequency of animate objects in constructions with animate subjects. Frequency-based accounts of typologically dominant construction patterns seem to confirm this view (Jäger 2004; Hogeweg/de Hoop 2010 *apud* García García 2014). Of course, this will affect the emergence rather than the later distribution of DOM. Recent studies, most notably by García García (2014) (cf. also Kabatek 2016) pick up these traditional ideas and reinterpret them in new theoretical frameworks. Whereas local factors were considered primarily due to their easier cross-linguistic comparability during the second wave, the third wave returns to more contextual views, including factors that surpass the sentence level and look at the informational status of the object referent within a larger context. The contributions in this book consider both local and global factors, including combinations of both.

## 2.3 Languages and varieties

While the tendency of the first wave was to look at languages such as Spanish, Portuguese or Romanian as a whole, and to consider variation basically from a diachronic point of view, the second wave broadened its scope to embrace a global perspective and found DOM languages throughout the world. At the same time, within the Romance area, DOM was identified in several diatopic varieties. The third wave looks not only at the coexistence of DOM devices within

the same language or dialect, but also at variation phenomena within one historical language. The most salient example is probably Spanish, where DOM is described in synchronic grammars as a generalized phenomenon without much local differentiation. There is a tradition that in a very general way claims DOM to be more frequent or to function differently in American Spanish in comparison to Peninsular Spanish (Kany 1951; Company Company 2002; von Heusinger/Kaiser 2005); however, general judgments about “two Spanishes” remain very vague, and variational patterns of DOM appear to be more complex than generally assumed. Of course, a distinction is necessary between the core functions of DOM and the possible zones of variation. Once we enter more marginal zones, it becomes more difficult to access enough comparable data. This is why new methodological approaches have recently addressed marginal phenomena comparing diatopic variation with an objectively comparative experimental setting (cf. Section 2.5.).

If we look at the other dimensions of variation, we must recognize that less is known about DOM and diastratic variation. Generally speaking, DOM in Romance is generally not a “marker” or a “stereotype” in a Labovian sense (Labov 1972), and only rarely do we find metalinguistic comments on DOM outside of linguistics, in more general discussions of language use. On the diaphasic level, DOM (and particularly flagging DOM) might acquire the value of a stylistic marker, especially in those contexts where it is clearly optional and where it seems to add an expressive surplus (Pottier 1968; Kabatek 2016). This seems to be the case, for instance, in the preference of marking salient toponyms (such as *a España*). However, such idiosyncratic cases should be separated from the functional view on the system of a language or variety.

## 2.4 Contact

A still rather poorly explored aspect of DOM is whether it is sensitive to language contact, and if so which of the dimensions elucidated above are receptive to such influence. In the context of this volume, it seems reasonable to make a distinction between DOM as a factor of convergence (or divergence) within the closely related Romance languages on one hand, and contact across language families on the other.

Comparison between Romance languages being the essence of what we have called the first wave of DOM studies, the question of contact and the internal typology of Romance languages naturally comes into play. Although the scattered distribution of DOM within Romance – particularly the rather isolated instance of Romanian – does not necessarily suggest a specific geographic point of origin,

some of the earlier accounts did suggest that Romance DOM might lead back to a pre-Latin substratum, either Iberic (Criado de Val 1954) or of an supposed Mediterranean origin (Meier 1948; Niculescu 1959). However, this explanation is obviously problematic due to its notoriously speculative nature. Instead, the attention of Romance scholars shifts to contact between Romance languages, where the Iberian Peninsula with the dominant presence of Spanish and its highly grammaticalized DOM alongside “smaller” languages represents a most promising area of research. In a comprehensive diachronic corpus study on DOM in Portuguese, Delille (1970) aligns the intensity of Spanish-Portuguese language contact and the evolution of object marking in Portuguese. Initially lagging behind its twin language as DOM expands in medieval Castilian, Portuguese DOM shows a sudden rise of marked objects during the Iberian Union (16<sup>th</sup> to 17<sup>th</sup> century), only to return largely to the initial prevalence after Portuguese independence was restored, thus suggesting that DOM is easily amenable to contact-induced language change. Although this hypothesis raises many questions in the light of new accounts on DOM, it remains largely unchallenged (cf. Aldon/Della Costanza 2013; Döhla 2014; Pires 2017). East of the Castilian domain, the case of Catalan is a more controversial one, since it is less proximate to Spanish from a typological perspective, but prone to longer and more intense contact. Hence, Catalan DOM has long been considered as a “barbarism”, not only by (mainly Catalan) grammarians, but also by linguists. However, recent studies have shed light on its own, autochthonous structural evolution. In this volume, the papers by Anna Pineda and Senta Zeugin give a differentiated account of the matter, the former from a diachronic, the latter from a synchronic point of view.

With the second phase of DOM studies, contact across language families came to the fore. In this field, results are even more inconclusive. Several studies suggest that DOM is difficult to acquire in L2 settings and that interference between languages with and without DOM occurs even in early childhood bilinguals, such as heritage speakers in the US or Quechua natives fluent in Spanish (Montrul/Bowles 2009; Ticio 2015; Wall/Obrist, this volume). On the other hand, the presence of an original DOM system in L1 appears to facilitate the acquisition of another in L2, even if the respective splits present structural divergences (Montrul/Gürel 2015). Whether or not social bilingualism can facilitate the development of DOM in a contact language is a largely unexplored issue (cf. Döhla 2011), and completing the picture will require further empirical fieldwork, such as the contribution of Mayer/Sánchez to this volume and the experimental study reported by Wall/Obrist.

## 2.5 Methodological innovations

At approximately the same time that the term DOM was coined, an overview article on the prepositional accusative in Spanish (Pensado 1985, 18) noted that no sociolinguistic study on this topic had yet appeared. The author suspected that the reason for this might be that researchers were unsure whether it was indeed possible to find any clear sociolectal distributions or even to simply document the generalization of the marker with this methodology. Things have changed considerably since then. Not only do we have a series of sociolinguistic studies from different regions for Spanish (Tippets 2011; Balasch 2011, among others), diachronic research has also documented the contribution of several of the factors in the generalization of Spanish DOM (Laca 2006; Della Costanza 2015) and, to a lesser degree, for Catalan (Pineda, this volume) and Portuguese (Delille 1970; Pires 2017). Furthermore, Ibero-Romance DOM has more recently begun to be studied with experimental approaches, including off-line methods such as controlled acceptability judgment tasks, picture verification tasks, elicitation tasks (Montrul 2013; Hoff 2018; Zeugin 2018; Bautista-Maldonado/Montrul 2019, among others), but also on-line processing studies, for instance making use of EEG (Nieuwland et al. 2013). However, many of these studies of the third wave are pioneering work and the new methods are still in an exploratory phase.

One issue in all corpus-related research – either diachronic or sociolinguistic – is the highly skewed distribution of DOM markers. As a marker of non-prototypical objects, its general frequency is not expected to be very high. More problematic, though, is that most examples that occur represent a limited set of configurations, and for many of the relevant factors contemporary corpora are not able to provide sufficient occurrences. This is not likely to change as long as DOM cannot be annotated automatically in very large corpora. Even in the largest “manual” diachronic studies, only the impact of the major factors can be traced across the centuries (Laca 2006; Della Costanza 2015). A somewhat different approach is to restrict the search in diachronic corpora to a specific factor and to analyze a fixed number of occurrences of relevant examples per century. Von Heusinger/Kaiser (2011) have shown that it is possible to trace the impact of more subtle factors this way, but of course it is then difficult to relate these results to other factors. As long as different approaches are explored as a means of identifying different and more subtle factors, the collected datasets will inevitably remain heterogeneous. In the longer term, a standardization of the approaches will be necessary in order to bring the different isolated findings into perspective.

The variational and sociolinguistic literature has shown that it is indeed possible to obtain robust results. Many of the factors discussed in the theoretical literature have been found to be empirically relevant, and different impacts for



these factors have been discovered in different samples of speakers. For instance, marking seems to be more frequent in certain places of the Americas, such as Mexico and the River Plate, than in Spain (Tippets 2011, among others), and less frequent in others, such as Cuba and perhaps the Caribbean more generally (Alfaraz 2011; Caro Reina/García García/von Heusinger, this volume). However, the corpus problems mentioned above also persist in this field to a certain degree. First, there are no standards for corpus selection, corpus cleaning or annotation. This will be necessary in the future in order to obtain comparable datasets. Second, infrequent DOM-sensitive constructions are very difficult to document. Finally, even if larger quantities of occurrences of rather rare configurations are made available at some point, another problem arises, namely that of the statistical treatment of variables when one of them is very dominant in absolute numbers. Balasch (2011), for instance, discusses the problem of inanimates being so rarely marked in her corpus that it is statistically questionable to group them with animate objects for a variational analysis, the latter being almost categorically marked.

Experimental approaches seem to be especially useful for the study of DOM because they provide better controlled access to data collection on many of the relevant phenomena. As mentioned above, different experimental paradigms have proven to be applicable and the rate of new publications is increasing. Several papers in this volume present new experimental findings. However, it is still too early to derive more general conclusions from these approaches. Even for acceptability judgments, which are the most frequently used technique, the experimental designs and stimuli differ considerably, and the degree to which particular choices affect participants' ratings has not yet been assessed. It is well known that the choice of the rating scale, the number of distractors included in the study, the use of written or auditive stimuli, presentation mode, presentation of the same item in different manipulations to one participant, etc., all affect the judgments (Schütze 2016). Since the acceptability experiments on DOM vary strongly in their design choices, we need a comprehensive comparative discussion to understand to what extent such methodological choices affect the outcomes, this in order to compare findings from different experiments. The same, obviously, holds for the other techniques, once there are enough studies available. Furthermore, it is also clear that experimental studies rely crucially on replication. So we have still to see which of the effects found in the studies of this first wave of experiments will be confirmed by future investigations.

While different empirical approaches are being applied to Spanish and the number of studies is growing, the situation is much more precarious for the other Romance languages. The experimental studies on Catalan and Italian published in this volume are pioneer contributions. Despite all the possible improvements

and unsolved issues for better empirical procedures mentioned above, empirical studies on other DOM languages should be strongly encouraged and linguists working experimentally as well as corpus linguists interested in those languages can both profit from the experiences reported in the development of the methods for the investigation of Spanish DOM.

### 3 The contributions to this volume

The first study is a text which might serve as an opening statement for the whole volume: a personal statement by Georg **Bossong** on the genesis of the term and the concept of Differential Object Marking. Bossong explains how the personal experience of comparing Spanish and Hebrew and the striking similarities between Hebrew *et* and Spanish *a* as object marker led him to recognize DOM as a more general phenomenon. The subsequent discovery of parallel elements in different Semitic languages (Aramaic, modern Arabic dialects, Akkadian, Classical Ethiopian) and in typologically very distant languages like Persian and Guaraní inspired him to dub this phenomenon *Differential Object Marking*, a term first introduced at the end of the 1970s. Bossong offers a list of arguments illustrating the fascinating nature of DOM, and gives several examples of characteristics of DOM in very different languages of the world.

Chantal **Melis**' article "From topic marking to definite object marking. Focusing on the beginnings of Spanish DOM" is based on a broad review of the typological literature on DOM. After laying out the various theories on the origins and grammaticalization paths of DOM, she assesses them with regard to their relevance for the history of the phenomenon in Spanish. Drawing heavily on examples taken from the *Cantar de Mio Cid*, Melis presents Spanish DOM as originating from topicalization, gradually evolving into a marker of prominence and individuation, before grammaticalizing to become primarily a marker of human objects. In light of the historical evidence provided, other motivations for DOM, such as convergence with the homonymous dative marker or its use as a strategy to resolve syntactic ambiguity, would play at most a secondary role.

Alessia **Cassarà**/Sophie **Mürmann** investigate instances of DOM in colloquial Italian, pointing out that Italian is generally not considered to be a DOM language. Based on observations in the literature, they investigate the use of *a*-marking with OBJECT-EXPERIENCER PSYCH-VERBS in left-dislocated structures and different referring expressions used as direct objects. Their hypotheses are based on García García's (2014) extension of Dowty's proposal of thematic macro-roles. They report an acceptability study conducted with speakers from

Northern Italy in which they tested the aforementioned factors. The findings largely support their hypotheses – left-dislocated OBJECT-EXPERIENCER VERBS are the most acceptable verb class tested with a DOM marker, and pronominal objects are favoured over proper names and definite NPs. A surprising finding, namely that indefinite NPs with a generic reading are even more acceptable than all other referential expressions, is explained by assuming that generic NPs are more similar to proper names and that sentences with such expressions might be easier to accept in an out-of-the blue judgment.

Elisabeth **Mayer**/Liliana **Sánchez** provide a qualitative and quantitative analysis of corpus data which they collected during fieldwork in bilingual communities in the Peruvian Andes and Amazonas regions and compare them with longitudinal data from monolingual speakers in Lima. They argue for different emerging DOM patterns of clitic-doubling and dislocated structures in the Spanish varieties spoken by these communities, depending on the configuration of the morphology in the given contact languages. These languages are Huánuco Quechua, Asháninka and Shipibo and arguably lead to different preferences that can be captured by aligning the properties of the object NPs on differently arranged scales. The relevant features, it is argued, are definiteness, animacy, PATIENT and THEME. According to the authors, it is easier for speakers of Huánuco Quechua than for those of the Amazonian languages to acquire the patterns of DOM and clitic doubling in Spanish, hence the much more frequent occurrence of DOM in their production, which is also much more similar to that of monolingual speakers. The authors also mention other factors which seem to influence the contact varieties, such as the more nuanced marking of thematic roles in Asháninka and the lack of a definite determiner in Shipibo. However, they also note that these differences do not block the acquisition of DOM in clitic doubling structures but rather lead to different results, depending on the feature pool to which speakers of a certain contact scenario have access.

Albert **Wall**/Philipp **Obrist** present an elicitation experiment with highly comparable data for six DOM-sensitive configurations in four varieties. They investigate the variation of DOM in Cusco, where Spanish is in contact with Quechua, and in Misiones, where Portuguese and Guaraní are contact languages. The data from the contact varieties are compared with data from predominantly monolingual speakers from Lima and Montevideo. Their results suggest that the different contact scenarios had different effects on the DOM system of the speakers from those regions. While Quechua L1 speakers from Cusco show a rather rudimentary DOM system in general, Misiones differs from predominantly monolingual varieties in terms of greater inter-speaker variability. The study highlights the importance of collecting controlled and comparable data for DOM in comparative studies and shows that different constructions need to be treated differently

in quantitative studies, because even superficially very similar sentences may produce very different rates of *a*-marking.

The contrastive study by Alina **Tigău** discusses the main commonalities and differences of DOM in Spanish and Romanian, focusing in great detail on syntactic as well as semantic questions, such as specificity, scope, small clause complements and object control, among others. It is shown that marked DOs behave similarly, whereas major differences can be found in unmarked DOs. In light of these findings, it is proposed that the two languages differ in the setting of a parameter which is sensitive to the syntactic type of the direct object: in Spanish, a crucial distinction between KPs and DPs is made, whereas in Romanian the distinction is between KPs and DPs on the one hand and “smaller” nominals on the other. Following López (2012), this distinction is related to the availability of scrambling for the respective nominals. Furthermore, the author concludes that in Spanish, DPs, as well as NumPs and NPs, can incorporate into the verb, whereas in Romanian DPs cannot do so.

In his paper, Niklas **Wiskant** discusses various proposals that seek to unify *a*-marking of objects, the distribution of null objects, and *leísmo* under the same label of Differential Object Marking, using data from Spanish and Portuguese. While both the distribution of null objects and *leísmo* have been associated with DOM separately, this had not previously been done for the three phenomena together. The author argues that the three phenomena are all based on the same semantic scale and hence proposes to subsume them under the label *scale-based object marking*. This scale-based object marking is argued to be typical of Ibero-Romance DOM.

Anna **Pineda**'s contribution departs from the normative view on Catalan since the early 20<sup>th</sup> century, confronting the prescriptive view with a historical analysis of data from the *Corpus Informatizat del Català Antic*, from the earliest written medieval Catalan texts to those of the 16<sup>th</sup> century. She shows that the norms propagated by Pompeu Fabra in the first half of the 20<sup>th</sup> century contrast not only with what can be observed in spoken Catalan but even with Fabra's own description of the language. His view attributes the strong presence of DOM in spoken Catalan to the influence of Spanish and considers it to be erroneous in several cases. DOM is only tolerated for clear disambiguation. Pineda's detailed analysis of diachronic data shows that DOM cannot be ascribed exclusively to the influence of Spanish. If present at all, this influence has more quantitative than qualitative effects: the contact language acts, in those periods and regions where its influence increases, as a catalyzer for DOM. DOM is already present in Old Catalan, but increases substantially in the 16<sup>th</sup> century, when Spanish influence is attested. The detailed analysis shows, among other things, that it is not only important to distinguish diatopic biases in a historical corpus analysis, but that

even the level of the individual text can be relevant: the strong relative weight of a 15<sup>th</sup>-century Valencian chivalric novel in the corpus, with few occurrences of DOM, has a potentially distorting effect on the overall picture.

Senta **Zeugin**'s paper also takes its starting point in the apparent mismatch between the denial of an autochthonous Catalan DOM by normative grammar and several descriptive reports on DOM in Catalan dialects. Her contribution, however, is synchronic and experimental, which, in contrast to the limited access to systematically comparable diachronic dialect data, allows for a design that can provide controlled sets of data. She suggests that Catalan DOM be described as an independent phenomenon, paying special attention to variation between four major regions (Central, North-Western, Valencian, Majorcan). The relevant variables for her experimental approach are animacy and, to a minor degree, the syntactic position of the object, both of which are treated in their own experiments. Results reveal that in general, -DOM is equally acceptable with all degrees of animacy, while +DOM is favoured with higher degrees; with humans, +DOM even attains similar scores as -DOM in Majorcan and Valencian. Unsurprisingly, unmarked canonical sentences score better than DOM-marked dislocations. In general, DOM is preferred in topicalized constructions over dislocations to the right, but this preference is significantly more strongly marked in Central Catalan than elsewhere. The author concludes that Catalan DOM appears to resemble (Peninsular) Spanish DOM in the differentiated acceptability of +DOM with different types of objects, while -DOM is, unlike in Spanish, always acceptable. Each step of the experimental design is carefully documented, which makes this article a methodological model of the implementation of claims from the literature in an experimental design.

Diego **Romero** discusses a possible “global” influence of DOM, the factor of telicity and its impact in DOM in the history of Spanish. His study is based on a diachronic corpus-based analysis with data from the *Corpus del nuevo diccionario histórico del español* (CDH). Romero filters data from three diachronic cuts (14<sup>th</sup>, 16<sup>th</sup> and 20<sup>th</sup> centuries) through a “telicity test”, looking at the telic or atelic verbal aspect and at several properties of the object. Against Torrego's (1999) claim that telic verbs favour DOM and DOM favours a telic interpretation of atelic verbs, no such correlation is confirmed by the analysis of the data; there is, however, a rather weak effect on the subset of indefinite direct objects. Even if the general result is negative and the initial hypothesis is thus not confirmed, the author admits that the apparent correlation between telicity and DOM observed in literature might be an indirect effect of other global influences triggered by the relationship of the semantics of the verb and the object, such as agentivity or affectedness.

The principal hypothesis of Javier **Caro Reina**/Marco **García García**/Klaus **von Heusinger** is that the lower incidence of DOM marking in Cuban Spanish might be the consequence of DOM retraction. Their empirical study is twofold: first, a corpus study based on selected verbal lemmata in CORDE (equally distributed between verbs denoting more or less affectedness), limited to human NPs (both definite and indefinite), examines the evolution of DOM from the 19<sup>th</sup> to the 20<sup>th</sup> century. Only “canonical” SVO-type sentences were taken into account. Second, a judgment experiment conducted in Cuba and Spain investigates acceptability for both of the aforementioned categories. The authors derive evidence for retraction from two findings: on the one hand, whereas the frequency of DOM in 20<sup>th</sup>-century Cuban Spanish is similar to that in 16<sup>th</sup>-century European Spanish, according to works of other authors, their own CORDE-based diachronic study shows a decrease in DOM from the 19<sup>th</sup> to the 20<sup>th</sup> centuries, at least in indefinite NPs. Comparing the CORDE data to Alfaraz’s 2011 oral corpus study, the decrease seems even more plausible, assuming that oral registers are less conservative. On the other hand, the acceptability experiment reveals that unmarked human objects are considerably more acceptable in Cuba than in Spain. The gradings mirror DOM expansion, both unmarked definites and unmarked indefinites being more acceptable in Cuba, while there is no dialectal difference for the respective marked conditions. In general, the article raises several methodological questions (complementarity of experimental and corpus approaches, comparability of different corpora). Even within the corpora, inter-author variation is considerable.<sup>2</sup>

## 4 Outlook

In her overview article, Pensado (1985, 19) characterizes the field as a series of pioneering approximations to the phenomenon that deal with similar problems and propose similar solutions, but she regrets the lack of a common ground for research on DOM. She expresses the hope that in future work, once the different proposals are “unified in a critical perspective”, there will be a major advance in knowledge. With regard to methodological innovations thus far, it still seems accurate to say that there is a lot of pioneering work towards approximation, but

---

<sup>2</sup> Most of the papers presented in this book are revised versions of contributions to a symposium held in Zurich in summer 2018. We would like to thank the Swiss National Science Foundation for its generous support. Thanks also to the anonymous reviewers for their useful comments and to John Barlow, Barbara Reynoso and Larissa Klose for proofreading and preparing the manuscript.

that a full convergence of results is still out of reach. However, despite the lack of a unified account, we observe the accumulation of knowledge through empirical work with different data types. New problems are being identified, and proposals for solutions are being developed. Hence, it seems that an accumulation of knowledge is possible even before the big unifying breakthrough. It may in fact be necessary first to build up a more solid foundation of knowledge with different data types that can compensate for the shortcomings of each other in order to complete the picture of what the unifying proposal might ultimately look like.

## Bibliography

- Aldon, Jean-Pierre/Della Costanza, Mario, *DOM en portugués. ¿Proceso propio o influencia del español? Estudio preliminar*, in: Manzano Rovira, Carmen/Schlumpf, Sandra (edd.), *Traspasando fronteras. Selección de trabajos presentados en el X Encuentro Hispano-Suizo de Filólogos Noveles*, Basel, Seminar für Iberoromanistik der Universität Basel, 2013, 71–88.
- Alfaraz, Gabriela G., *Accusative object marking. A change in progress in Cuban Spanish?*, *Spanish in Context* 8:2 (2011), 213–234.
- Balasz, Sonia, *Factors determining Spanish Differential Object Marking within its domain of variation*, in: Michnowicz, Jim/Dodsworth, Robin (edd.), *Selected proceedings of the 5<sup>th</sup> workshop on Spanish sociolinguistics*, Somerville, Mass., Cascadilla, 2011, 113–124.
- Bautista-Maldonado, Salvador/Montrul, Silvina, *An experimental investigation of Differential Object Marking in Mexican Spanish*, *Spanish in Context* 16:1 (2019), 22–50.
- Bello, Andrés, *Gramática de la lengua castellana destinada al uso de los americanos*, ed. Trujillo, Ramón, Santa Cruz de Tenerife, Instituto Andrés Bello, 1847/1981.
- Bosson, Georg, *Historische Sprachwissenschaft und empirische Universalienforschung*, *Romanistisches Jahrbuch* 33 (1982), 17–51.
- Bosson, Georg, *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen, Narr, 1985.
- Company Company, Concepción, *El avance diacrónico de la marcación prepositiva en objetos directos inanimados*, in: Bernabé Pajares, Alberto, et al. (edd.), *Presente y futuro de la lingüística en España. La Sociedad de Lingüística, 30 años después. Actas del II Congreso de la Sociedad Española de Lingüística*, vol. 2, Madrid, Consejo Superior de Investigaciones Científicas, 2002, 146–154.
- Criado de Val, Manuel, *Fisionomía del idioma español*, Madrid, Gredos, 1954.
- Delille, Karl Heinz, *Die geschichtliche Entwicklung des präpositionalen Akkusativs im Portugiesischen*, Bonn, Romanisches Seminar der Universität Bonn, 1970.
- Della Costanza, Mario A., *La marcación diferencial del objeto (DOM) en español. ¿Una construcción con varios significados?*, PhD dissertation, University of Zurich, 2015.
- Döhla, Hans-Jörg, *Differential Object Marking (DOM) in some American Indian languages. Contact induced replication and convergence or internal development?*, in: Mendoza, Imke/Pöll, Bernhard/Behensky, Susanne (edd.), *Sprachkontakt und Mehrsprachigkeit als Herausforderung für Soziolinguistik und Systemlinguistik*, München, Lincom Europa, 2011, 27–45.

- Döhla, Hans-Jörg, *Diachronic convergence and divergence in Differential Object Marking between Spanish and Portuguese*, in: Braunmüller, Kurt/Höder, Steffen/Kühl, Karoline (edd.), *Stability and divergence in language contact. Factors and mechanisms*, Amsterdam/Philadelphia, John Benjamins, 2014, 265–289.
- Eckert, Penelope, *Three waves of variation study. The emergence of meaning in the study of sociolinguistic variation*, *Annual Review of Anthropology* 41 (2012), 87–100.
- Flores, Marcela/Melis, Chantal, *El léísmo desde la perspectiva del “mercado diferencial de objeto”*, *Revista de Historia de la Lengua Española* 2 (2007), 83–107.
- García García, Marco, *Differentielle Objektmarkierung bei unbelebten Objekten im Spanischen*, Berlin/Boston, de Gruyter, 2014.
- Haspelmath, Martin, *Comparative concepts and descriptive categories in crosslinguistic studies*, *Language* 86:3 (2010), 663–687.
- Haspelmath, Martin, *Role-reference associations and the explanation of argument coding splits*, *Linguistics* 59:1 (2020), 123–174.
- Hoff, Mark R., *Are Argentines “a”-blind? Acceptability of “a”-marked inanimate direct objects*, in: MacDonald, Jonathan E. (ed.), *Contemporary trends in Hispanic and Lusophone linguistics. Selected papers from the Hispanic Linguistic Symposium 2015*, Amsterdam/Philadelphia, John Benjamins, 2018, 121–142.
- Hogeweg, Lotte/de Hoop, Helen, *Children and transitivity. The subject-object asymmetry in a natural setting*, in: Brandt, Patrick/García García, Marco (edd.), *Transitivity. Form, meaning, acquisition, and processing*, Amsterdam, John Benjamins, 2010, 143–160.
- Jäger, Gerhard, *Learning constraint sub-hierarchies. The bidirectional gradual learning algorithm*, in: Blutner, Reinhard/Zeevat, Henk (edd.), *Optimality theory and pragmatics*, Basingstoke, Palgrave Macmillan, 2004, 251–287.
- Kabatek, Johannes, *Wohin strebt die differentielle Objektmarkierung im Spanischen?*, *Romanistisches Jahrbuch* 67 (2016), 211–239.
- Kany, Charles, *American-Spanish Syntax*, Chicago, University of Chicago Press, 1951.
- Labov, William, *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania Press, 1972.
- Meier, Harri, *Sobre as origens do acusativo preposicional nas línguas românicas*, in: Meier, Harri (ed.), *Ensaio de filologia românica*, Lisboa, Edição da ‘Revista de Portugal’, 1948, 115–164.
- Melis, Chantal/Rodríguez Cortés, Diego, *El mercado diferencial de la función “meta” en español*, *Verba* 44 (2017), 195–230.
- Montrul, Silvina/Bowles, Melissa, *Back to basics. Incomplete knowledge of Differential Object Marking in Spanish heritage speakers*, *Bilingualism: Language and Cognition* 12:3 (2009), 363–383.
- Montrul, Silvina, *Differential Object Marking in Argentine Spanish. An experimental study*, in: Colantoni, Laura/Rodríguez Louro, Celeste (edd.), *The handbook of Argentine Spanish*, Frankfurt, Vervuert Iberoamericana, 2013, 207–228.
- Montrul, Silvina/Gürel, Ayşe, *The acquisition of Differential Object Marking in Spanish by Turkish speakers*, in: Judy, Tiffany/Perpiñán, Silvia (edd.), *The acquisition of Spanish in understudied language pairings*, Amsterdam, John Benjamins, 2015, 281–308.
- Niculescu, Alexandru, *Sur l’objet direct prépositionnel dans les langues romanes*, Bucarest, Editura Academiei Republicii Populare Romîne, 1959.
- Nieuwland, Mante S./Martin, Andrea E./Carreiras, Manuel, *Event-related brain potential evidence for animacy processing asymmetries during sentence comprehension*, *Brain and Language* 126:2 (2013), 151–158.



- Laca, Brenda, *El objeto directo. La marcación preposicional*, in: Company Company, Concepción (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, vol. 1, Mexico DF, UNAM/Fondo de Cultura Económica, 2006, 423–475.
- Lenz, Rodolfo, *La oración y sus partes*, Madrid, Centro de Estudios Históricos, 1920.
- Pensado, Carmen, *El complemento directo preposicional. Estado de la cuestión y bibliografía comentada*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 11–59.
- Pires, Aline Jéssica, *A marcação diferencial de objeto no português. Um estudo sintático-diacrônico*, MA Thesis, Campinas, 2017.
- Pottier, Bernard, *L'emploi de la préposition "a" devant l'objet en espagnol*, Bulletin de la Société de Linguistique 63 (1968), 83–95.
- Schütze, Carson T., *The empirical bases of linguistics. Grammaticality judgments and linguistic methodology*, Berlin, Language Science Press, 2016.
- Schwenter, Scott, *Two kinds of Differential Object Marking in Portuguese and Spanish*, in: Amaral, Patrícia/Carvalho, Ana Maria (edd.), *Portuguese-Spanish interfaces. Diachrony, synchrony, and contact*, Amsterdam, John Benjamins, 2014, 237–260.
- Silverstein, Michael, *Hierarchy of features and ergativity*, in: Dixon, Robert M.W. (ed.) *Grammatical categories in Australian languages*, Canberra, Australian Institute of Aboriginal Studies, 1976, 112–171.
- Ticio, M. Emma, *Differential Object Marking in Spanish-English early bilinguals*, Linguistic Approaches to Bilingualism 5:1 (2015), 62–90.
- Tippets, Ian, *Differential Object Marking. Quantitative evidence for underlying hierarchical constraints across Spanish dialects*, in: Ortiz-López, Luis A. (ed.), *Selected proceedings of the 13<sup>th</sup> Hispanic Linguistics Symposium*, Somerville, Mass., Cascadilla, 2011, 107–117.
- Torrego, Esther, *El complemento directo preposicional*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española. Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*, vol. 2, Madrid, Espasa Calpe, 1999, 1779–1805.
- von Heusinger, Klaus/Kaiser, Georg A., *The evolution of Differentiated Object Marking in Spanish*, in: von Heusinger, Klaus/Kaiser, Georg A./Stark, Elisabeth (edd.), *Proceedings of the workshop "Specificity and the Evolution/Emergence of Nominal Determination Systems in Romance"*, Konstanz, Fachbereich Sprachwissenschaft der Universität Konstanz, 2005, 33–69.
- von Heusinger, Klaus/Kaiser, Georg A., *Affectedness and Differential Object Marking in Spanish*, Morphology 21:3 (2011), 593–617.
- Witzlack-Makarevich, Alena/Seržant, Ilja A., *Differential Argument Marking. Patterns of variation*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 1–40.
- Zeugin, Senta, *La marca diferencial del objeto en las variedades catalanas. Animación y posición del objeto como factores de influencia en la aceptabilidad de la MDO*, MA Thesis, University of Zurich, 2017.



Georg Bossong

# DOM and linguistic typology

A personal view

## 1 The discovery of Differential Object Marking

The discovery of what I later baptized as Differential Object Marking (DOM) was a crucial moment in my life. In 1977 I had completed my thesis on the *Problems of translation of scientific works from Arabic into Old Spanish in the age of Alfonso el Sabio* (Bossong 1979) and was looking for new challenges. I was deeply fascinated by the relationship between universal and particular structures in human language, and hence I was keen on linguistic typology, which had a rather marginal status in international linguistics at the time. I had the strong feeling that this kind of research was exactly where the future lay: empirically based research into language universals. How to describe the astounding diversity of languages and at the same time capture their underlying unity?

For me, the first instance of what later on was to be described as DOM was the typological parallelism between Hebrew and Spanish. This similarity was not apparent at first sight, and it was not explicitly discussed in descriptions at the time. Spanish was described according to the traditional patterns of Latin and Romance grammars; and Hebrew grammar had, in the West, a longstanding tradition of its own. So, in Spanish grammars a strange phenomenon was to be found, called the *prepositional accusative*, something unknown in the Latin tradition; and on the other hand, Hebrew grammars showed something equally strange, traditionally called *nota accusativi* (Hebrew grammars in the West were written in Latin until the 19th century). After Latin and Greek, I had learned Hebrew, a long time before Spanish, and so I had become at an early stage in my life what in the Renaissance was called a *homo trilinguis*. The similarity between the Hebrew *nota et* and the specific use of the Spanish preposition *a* struck me as particularly interesting, although the use of *et* was restricted to definite objects and the use of *a* to animate ones. Nevertheless, I very soon became aware that the common factor was that in both cases a differentiation was made inside a grammatical category, namely the object, according to certain semantic rules.

In 1978 I lived in Paris with my wife and our first-born son. For my research I relied essentially on the *Bibliothèque Nationale*, the old one, of course, in the Rue de Richelieu, with all its splendour and all its shortcomings. I began to delve

---

Georg Bossong, University of Zurich, e-mail: georg@bossong.de

into the grammars of numerous languages, beginning with Romance. The articles of Harri Meier on the *acusativo preposicional* were particularly revealing (Meier 1948). He had worked as a lecturer in Portugal in the early 1950s, just as I worked as a lecturer in the late 1970s in France. I soon discovered Bodo Müller's *morphemmarkiertes Satzobjekt* (Müller 1971) and many other studies on the preposition *a* in Romance and *pe* in Romanian. On the Semitic side, I discovered that Hebrew *et* was by no means an isolated phenomenon, but that it had etymological and grammatical parallels in Aramaic, in modern Arabic dialects, and in older and more remote Semitic languages such as Akkadian and Classical Ethiopian. From that moment on, my curiosity was definitively aroused, and I was soon overwhelmed by a never-ending stream of new discoveries. I had not simply found a theme, but rather the theme had found me! It was an exciting experience.

My excitement grew when I discovered that two of the languages I was studying at that moment showed similar phenomena, namely Persian and Guaraní, two languages geographically, typologically, and culturally as distant from one another as can be imagined. At that time, I was deepening my study of Persian in the research group of Gilbert Lazard at the *École Pratique des Hautes Études*, and was studying Guaraní in connection with Bernard Pottier's research group on American Indian languages. Soon I realized that traditional descriptions were hopelessly insufficient, insofar as they stuck to the surface of morphological marking, without reaching the deeper layers of universal structures. There was a deeply rooted commonality between all these languages, but nobody had become aware of it. Traditional descriptions remained totally superficial.

So, the first step was discovery. The next step had to be naming. Finding a name for a newly discovered reality is of crucial importance. With an appropriate name a vaguely imagined idea gets a clear-cut identity, a mental shape which permits one to go beyond initial empirical limitations. At some moment between 1978 and 1980 a decisive intuition struck me: the common factor is differentiability. In all languages I had considered thus far, there was a differentiation made within a given syntactic category, namely the object. Some objects were marked, whereas others were not. Very soon this fundamental point became crystal-clear to me. The important point was to make abstraction from all superficial differences and to work out the basic structure. The term to be applied had to be sufficiently abstract and general to cover all individual variations. Only when this level of abstraction had been reached did the basic underlying plans have a chance to reveal themselves from beneath the overwhelming variety of individual language structures.

It was in this way that the term *Differential Object Marking* came to my mind. This creation was not only abstract enough to cover all the various individual

realizations of a general principle, but had also the advantage of being international, to be easily pronounceable in most European languages, and to be presented in an easily manageable abbreviated form (DOM). I had created the term in English, and there was no problem with German (*Differentielle Objektmarkierung*). In the Romance languages the only difficulty is the well-known difference of the position of the adjective (*Marquage Différentiel de l'Objet*, *Marca Diferencial del Objeto*, both abbreviated as MDO, a bit like UNO vs. ONU). But this minimal lack of internationality was no obstacle at all. The term was there, and it made its way into scholarship. The initial empirical discoveries had helped to create the term, and once created, the term itself helped to foster new discoveries in an ever-increasing number of languages and language families, all around the world.

Differentiability refers to differentiations according to what at the time was called *animacy* (as in Spanish and Guaraní) and/or according to *definiteness* (as in Hebrew and Persian). In the higher semantic domains, there is frequently a certain overlap between animacy and definiteness, especially in personal pronouns and in proper names, which are necessarily both animate and definite. The term *animacy* has certain shortcomings: while it can easily be transferred into German 'Belebtheit' or Italian 'animatezza', it cannot be easily rendered in French or in Spanish. I proposed to call it *inherence* since it refers to semantic features inherent to the noun or noun phrase. Gilbert Lazard proposed the beautiful neologism *humanitude* ( $\neq$  *humanité*), which works well in French, but not in other languages. As for definiteness, I prefer the more general term of *reference*. However, none of these terms had the success of the term DOM, and they did not impose themselves universally.

My findings on DOM were first presented in public at the 16<sup>th</sup> International Congress of Romance Philology, held in Palma de Mallorca in 1980. I attended this congress in the company of my former supervisors, Kurt Baldinger and Klaus Heger, as well as Gerold Hilty and Eugenio Coseriu. My article began to circulate in manuscript form, as a kind of *samizdat*, but it was never published in printed form, because, after several years of waiting, the publication of the congress acts was cancelled due to a lack of funds. Nevertheless, the idea made its way into the world.

In 1981 we moved to Munich. I had my office just opposite the *Bayerische Staatsbibliothek* (StaBi). Umberto Eco once said that "paradise must be something like a big library". Indeed, the StaBi was for me a paradise where I found plenty of information on the most remote and exotic languages. So, I went ahead and continued on my voyage of discovery.

One of my first adventures was to work through Grierson's *Linguistic Survey of India*, whose 11 large volumes contain detailed information on 364 languages

belonging to the Indo-Aryan, Dravidian, Sino-Tibetan, and Munda families (Grierson 1967–1968). The difficulties of the postposition *ko* in Hindi are well-known, indeed almost proverbial, but this is only the tip of an iceberg. When I penetrated the jungle of Indian languages, I soon realized that DOM was omnipresent on the subcontinent, in all language families. Grierson's survey is extremely useful insofar as he quotes the parable of the prodigal son in hundreds of languages – a text which contains many object constructions, animate and inanimate as well as definite and indefinite.

However, my research on Indian languages never took the form of a book. Instead, I began an intense study of the Iranian (Irano-Aryan) language family, this for several reasons: because of my long-standing love for the Persian language (and poetry), because of my friendship with Gilbert Lazard, and because of the rich internal variety of forms and functions in that language family. In 1985 my book on *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen* appeared (Bossong 1985). It did not have the international echo I had hoped, mainly, I think, because it was written in German, a language which was slowly dying out as a medium of international scientific communication. Nevertheless, it helped me to climb the academic ladder. I have several more books of this kind sleeping in my drawers, especially one on the Semitic language family, but as time went on, I focused on other topics. In Zurich the strong local traditions brought me back to my first love, namely Al-Andalus, the linguistic and cultural relations between Arabic/Hebrew and the Ibero-Romance languages. Several disciples, and also my successor in Zurich, Johannes Kabatek, have taken up the challenge and continue the research on DOM in Romance and beyond. The topic is now widespread in international research.

## 2 Short outline of selected topics

Why is DOM so fascinating? I think it is because of its “squishiness”. The boundary between the presence and absence of object marking is fluid. In most languages there are transitional zones where marking and non-marking are both possible. There follows a short collection of quotations describing these difficulties (from my 1985 book):

Spanish:

“Les notions d'animation et de particularisation étant essentiellement subjectives, il est parfois malaisé de décider de l'emploi ou de l'omission de la préposition 'A.'”

(Coste/Redondo 1976, 321)

Syro-Aramaic:

“In den meisten Fällen herrscht hinsichtlich der Wahl oder der Weglassung eines Objectzeichens beim Determinierten vollständiges Schwanken.” (Nöldeke 1898, 220)

Hindi:

“The correct use of these two alternative forms and constructions [ko vs. Ø] is perhaps the most difficult thing in the Hindi language. Only by extensive and continual reading of native books and by intercourse with the people can the foreigner become able to use them with idiomatic accuracy.” (Kellogg 1893, 397)

Finnish:

“Den Ausländer, der sich in den Bau der finnischen Sprache hineinzuarbeiten versucht, überkommt angesichts des Partitivs ein Gefühl der Hilflosigkeit, so wie einem Stoßtruppführer vor einem feindlichen Minenfeld zumute sein mag: bei jedem Schritt lauern Gefahren, und selbst der guten, verlässlichen Mutter Erde, die uns allen einen festen Standort gibt, darf er nicht mehr trauen.” (Raible 1976, 10)

Such categories are interesting for the linguist. Clear-cut boundaries, such as grammatical gender in European languages, are uninteresting, in that they may be difficult in second language learning, but they do not present a theoretical challenge. To say *el mesa* or *die Tisch* is simply wrong. But to say *he visto al perro* or *he visto el perro* may be equally correct according to the context and the meaning intended by the speaker, and so the use or non-use of the preposition becomes a challenging problem for linguistic research.

One of the fundamental discoveries concerning DOM is that it can be nominal and/or verbal. A grammatical relationship between nominal actant and verbal predicate can be expressed by grammemes added to one of the two terms, or to both. The interplay between verbal clitics and nominal affixes is particularly complex and therefore interesting for linguistic research. In Spanish, *lo he visto al perro* is in more or less free alternation with *he visto al perro*. In some languages, such as Hungarian (9) or the Bantu language Zulu (10), the alternation is exclusively verbal (compare the well-known “object conjugation” of Hungarian); in other languages, such as in Hungarian’s Ob-Ugric relatives Ostyak and Vogul, all four variants occur: the marking can be nominal alone, verbal alone, or nominal plus verbal, and of course marking can be lacking altogether. The result is a complex and subtle interplay of factors which allows for the expression of a great variety of semantic nuances.

An important chapter, although up to now a rather neglected one, is that dealing with incorporation vs. excorporation. The natural place for an object is close to the verb, with which it forms a semantic unit. Verb and object tend to undergo coalescence. According to DOM, marking may serve to interrupt this close relationship by rendering the object more independent and autonomous. This process can be called *excorporation*: instead of melting with the verb, the

object stands by itself. So, it can be topicalized, for instance, or can otherwise occupy a prominent position. Nahuatl is a classic example, already occurring in the work of Wilhelm von Humboldt, author of the term *Einverleibung* which later spread internationally in its Latinized form *incorporation*. Of particular interest are those cases where an incorporated object forms with the verb a basic unit, to which a marked, and thereby excorporated object is then added. Compare the example from Udi (7) where the Arabism *xabar* ‘message, information’ forms with the verb a phraseological unit meaning ‘to ask’, and then the unit of verb plus the incorporated object as a whole takes an independent object, to which the postposition *-ox* is added. (Note that in the Udi example we observe the simultaneous presence of the ergative marker *-en* for the subject.) Such constructions are extremely widespread in Iranian, Turkic, and Indo-Aryan languages.

By *configurational typology* I understand the distinction between the three basic types of marking the fundamental relationship of subject and object: the configuration can be accusative, ergative, or “active” (or, as Gilbert Lazard has felicitously termed it, “dual”). The combination of these three basic configurations with DOM yields very interesting results, which cannot be presented here in detail (cf. Bossong 1985, 116–121).

In terms of natural iconic markedness (Mayerthaler 1981), DOM is “iconic” insofar as natural combinations of features are left unmarked, whereas non-natural combinations are marked. It is “natural” for a prototypical object to be inanimate and/or indefinite, and the specific marking of animate/definite objects is natural and thus iconic. DOM is of particular interest for the theory of natural markedness.

Synchronic analysis has its logical counterpart in diachrony. Languages are always in movement; systems never stand still. In this sense, the distinction of synchrony and diachrony is useful, but artificial. With respect to DOM this has important implications. The pathways of diachronic change are universal, but the particular stretch of road on which a given language finds itself is individual. It is always the same street, but the point that a language has reached is variable.

In recent years, contact linguistics has generated much interest. Even on remote islands or in mountain valleys languages are continuously in contact with other languages. DOM is a phenomenon which spreads easily across borders of languages or language families. DOM can arise on the base of universal tendencies, or from internal factors, but it can also arise due to the influence of neighbouring languages. To give just one example: modern standard Guaraní has the postposition *pe* which works more or less like Spanish *a*. Thanks to the research on missionary grammars we know that this state of affairs is relatively recent. Older descriptions and texts clearly show that traditional (“tribal”) Guaraní did not show any trace of DOM. The emergence of this construction must be explained



as a contact induced change, namely contact with Spanish. Similar observations can be made with respect to Aymara, for instance (see (8), also compare Papia Kristang (12)).

### 3 Some examples

The short collection of examples which follows is intended to give a – necessarily superficial – impression of the enormous formal, genetic, and geographic variety of DOM. It occurs across all continents, and in many linguistic families. There is an astonishing diversity of structures, but the underlying structural principles are uniform. It is this combination of unity and diversity which makes the study of DOM so rewarding.

First, a few illustrative examples from Hindi (1) are given. As in most other Indo-Aryan languages, the postposition *ko* not only has the function ACC', but also DAT. In this respect its behaviour resembles the use of Spanish *a*. Such a constellation is frequent in the world's languages, although by no means universal. The boundary between marked and unmarked objects lies in the domain of animals. Also note that Hindi, like many other languages of the region, has an ergative configuration for the past tenses (what I call *preterital ergativity*). We observe here the simultaneous presence of overt ergative marking (*ne*) and the marking of the animate object (*ko*).

#### (1) Hindi (Indo-Aryan/Indo-European)

*Maĩ laṛkē kō dēkhtā hū*  
I boy ACC' seeing am  
'I see the boy.'

*laṛkī ghoṛ-ō kō dēkhā*  
girl horse-OBLPL ACC' saw  
'The girls saw the horses.'

*maĩ-nē cīṛi-yā dēkhī*  
I-ERG bird-OBL PL seen-FEM PL  
'I have seen birds.'

*us-nē brāhman kō dān diyā*  
he (OBL)-ERG Brahman DAT gift gave  
'He gave the brahman a gift.'

The Iranian language Ormuri (2), spoken in the north-western frontier of Pakistan, shows a classical pattern. The preposition *ku* (etymologically related to Slavic *k*) is used differentially, in combination with simple or periphrastic verbs (*xalas kon* ‘make free’ = ‘liberate’).

### (2) Ormuri (Iranian/Indo-European)

*Az ku tū bē-nas-am, ku mūn kara dī xalās kōn*  
 I ACC’ you PRS-take-1 SG ACC’ me from him free make  
 ‘I take you, but liberate me from him!’

*Ku boz-am dek*  
 ACC’ goat-1 SG saw  
 ‘I have seen the goat.’

*yāsp bu nāl ka-m*  
 horse PRS iron make-1 SG  
 ‘I am just shoeing a horse.’

The Munda language Sora (3) is semantically interesting. The noun *ad’ong* is used for both ACC’ and DAT. *He has seen the child* is expressed as *he has seen the child’s body*.

The same construction is used for the DAT function: *he has given money to the child* is rendered as *he has given money the child’s body*. From a typological perspective, this is a rare construction, and I have not found parallels elsewhere, but it is clearly in line with the general semantic rules of DOM based on inherency factors.

### (3) Sora (Austroasiatic/Munda)

*Anin pəsij-ən ad’ong gij-le*  
 He child-GEN “body” = ACC’ see-PRT  
 ‘He has seen the child.’

*Anin kənsim [-ən ad’ong] tīb-le*  
 he chicken-GEN ACC’ carve-PRT  
 ‘He has carved the chicken.’

*Anin pəsij-ən ad’ong lebun tiy-le*  
 he child-GEN “body” = DAT money give-PRT  
 ‘He gave money to the child.’

Great parts of Eurasia are occupied by languages which belong to a hypothetical macro-phylum, termed Altaic. The genetic relationship between the families of this phylum is a matter of dispute, but the structural similarity concerning DOM is striking. We consistently find the same pattern everywhere, namely nominally marked DOM on the base of referentiality: definite objects are marked, whereas indefinite objects remain unmarked, from the earliest documents until today. This pattern is universally attested in Turkic, Mongolic, and Tungusic languages, see the examples of Bashkir (4), Mongolian (5), and Mandchu (6) below.

#### (4) Bashkir (Altaic/Turkic)

*čay-di ičtim* vs. *čay ičtim*  
 tea- ACC' I.drunk tea I.drunk  
 'I have drunk the tea – I have drunk tea.'

#### (5) Mongolian (Altaic/Mongolic)

*Bi nom-yg avlaa* vs. *bi nom avlaa*  
 I book- ACC' bought I book bought  
 'I bought the book – I bought a book.'

#### (6) Manchu (Altaic/Tungusic)

*Bi dengjan-be mukiyebuhe* vs. *bi bithe arambi*  
 I lamp- ACC' I-extinguished I letter I-wrote  
 'I switched off the lamp – I wrote a letter.'

This pattern is so pervasive that it has penetrated into Udi (7), an outlier of the Lezghian language family, situated in Turkic speaking Azerbaijan. Among the fifty-odd languages of the Caucasus, which was called the *mons linguarum* in Antiquity, DOM is absent. This is especially noteworthy in the Eastern language families, the Daghestanian area, which is well-known for being the realm of extremely rich case marking systems. Nevertheless, no trace of DOM is found elsewhere in this “mountain of languages”. The Lezghic branch of this group has some of the most extended case marking systems of the world, including Tabassaran with between 48 and 52 cases. Nevertheless, DOM is completely absent from this area, except in Udi, which is spoken in an otherwise Turkic region. It stands to reason that DOM in Udi can be assumed to be due to this adstratal influence. This example shows that DOM follows universal rules and tendencies, but that it is also sensitive to contact influences.

**(7) Udi (Daghestanian/Lezghian)**

*pāčay-en armuy-ox xabarre aqsa vs. Arcen śum uken*  
 king-ERG son- ACC'-PL information he-asked let's-sit bread let's-eat  
 'The king asked his sons. – Let's sit down and eat some bread!'

The development of DOM in the American Indian languages Aymara and Guaraní (8) can be shown to be due to contact with Spanish. Marking is purely nominal, the postposition for ACC' is identical to the marker for DAT, just like in Spanish. DOM has developed in times of Spanish influence; according to early missionary descriptions, it was absent in the original states of these languages. Below, there are a few examples from these two languages, which are probably genetically unrelated, but strongly influenced each other through adstratal contact. In the examples below (8), it becomes evident that the inherent feature of definiteness prevails over the inherent feature of animateness.

**(8) Aymara/Guaraní (American Indian)**

*k'usilu-χa qamaqhi-ru nac'antaj-na*  
 monkey-THEM fox- ACC' bind-PRT  
*ka'I o-mosã aguara-pe*  
 monkey 3SG-bound fox- ACC'  
 'The monkey bound the fox.'

*juqasa-χa katxatã-na-wa mā lunthata*  
 son-THEM catch-PRT-ASSERT one thief  
*ñande ra'y o-ipyhy peteĩ mondaha*  
 we son 3SG-caught one thief  
 'Our son caught a thief.'

The purely verbal marking of DOM occurs in many parts of the world, although on the whole it is less frequent than purely nominal marking. Hungarian (9) is well-known for its so-called object conjugation. Here, DOM takes the following form: the nominal object is always marked, without any differentiability (here glossed as ACC = 'non-differential accusative'); in contrast, the verbal conjugation has two forms, depending on the referential status of the object. If the object is definite, the verb obligatorily refers to it by a combined subject + object conjugation; if it is indefinite, there is no verbal agreement with the object, only with the subject.

**(9) Hungarian (Finno-Ugric/Ugric)**

*az-t az újságo-t kér-i vs. újságo-t kér*  
 ACC-ART newspaper-ACC ask-for-SBJ+OBJ newspaper-ACC asks-for-SBJ  
 ‘He asks for the newspaper. – He asks for a newspaper.’

The structure of Bantu languages is comparable in certain respects to the structures found in Ugric. The numerous nominal classifiers so typical of Bantu are integrated in the verbal predicate whenever the object is definite, yet they are not when the object is indefinite, a classic case of verbal DOM. The object noun itself remains unchanged, and differentiability appears only in the verb. Here, an example from Zulu (10) is given for illustration.

**(10) Zulu (Niger-Kongo/Bantu)**

*ngiya-m-bona umuntu vs. ngiya-bona umuntu*  
 1SG-CLASS OBJ-see man 1SG-see man  
 ‘I see the man. – I see a man.’

Ritharngu (11) is spoken in Arnhem Land in Northern Australia. It belongs to Pama-Nyungan, the most widespread genetic phylum of Australia. In this language, several structural particularities appear in combination. As in most Australian languages, one form or another of ergativity prevails in the noun. In Ritharngu, we observe the simultaneous presence of ergative and accusative markers. The presence or absence of the accusative marker in the noun follows the rules of DOM. Moreover, there is congruence between nominal and verbal object marking; in consequence, it can be said that both the noun and the verb are organized according to DOM. The semantic differentiation is purely inherent, the boundary being situated between animate and inanimate beings (*kangaroo* vs. *spear*). Such a combination of nominal and verbal DOM is not uncommon in Australian languages.

**(11) Ritharngu (Pama-Nyungan)**

*nāwala-ñā-ŋay garčambal-na guya-du vs. nāwala-ŋay wartambal*  
 see-OBJ-PERF kangaroo-ACC’ ish-ERG see-PERF spear  
*ding’-du*  
 woman-ERG  
 ‘The fish has seen the kangaroo. – The woman has seen the spear.’

The final Section analyzes two special phenomena related to DOM in Ibero-Romance varieties.

The first is Papia Kristang (12), the Portuguese based creole of Melaka, also spoken by a few people in Singapore, Kuala Lumpur, and other places. In this variety, we find a special formal realization of DOM, namely the Portuguese preposition *com*, which appears phonetically as *ku/kung/kong*. As usual, this preposition is obligatory with personal pronouns and proper names; it is optional with nouns denoting human beings, unusual with higher animals (dogs and the like) and impossible with lower animals, such as insects, and with abstract nouns. *Kung* also stands for the dative, for the instrumental and the comitative function. Semantically and functionally, it is a prototypical DOM preposition, which corresponds in almost every detail to Ibero-Romance *a*. But its form comes from a different Portuguese word. Ultimately it stems from Hokkien *kǎp* which has roughly the same range of meanings. Hokkien, in Mandarin Fújiàn, is a Sinitic language (popularly classified as a Chinese dialect) spoken in south-eastern China and widely used as a lingua franca among overseas Chinese in the region. Note that the Hokkien word *kǎp* has no relationship whatsoever to any standard Chinese preposition or particle. If there is anything like DOM in Mandarin Chinese, it takes the form of a particle *bǎ*, whose original meaning is ‘to take, to seize’. This particle differs considerably from Hokkien *kǎp*, not only in etymology, but also in function.

It seems that this Hokkien particle was copied semantically in a pidginized variety of Malay, the so-called Bazaar Malay used in the market places of Malaysia. There the preposition *sama*, meaning ‘with’, takes the functions of a prototypical DOM particle: marked accusative, dative, and comitative. From Bazaar Malay it spread to Papia Kristang, where it was functionally copied. This example shows that DOM may even arise in a creole language, despite the radical elimination of almost all inflectional morphology. Functionally, such newly formed DOM markers follow the universal patterns, even if morphologically they show a different pattern.

### (12) Papia Kristang (Portuguese Creole of Melaka)

*Eli konesé kung Mary*  
 He knows ACC' Mary  
 ‘He knows Mary.’

*Eli ja matá Ø/kung eli sa kachoru kontu ja mudré kung*  
 He PRF kill- ACC' he GN dog because PRF bite ACC'  
*eli sa krensa*  
 he GN child  
 ‘He killed his dog because it has bitten his child.’

*Eli ja matá bichu*  
 He PRF kill insect  
 ‘He killed the insect.’

*Eli ja kotrá aké kandri ku faka*  
 He PRF cut that meat with knife  
 ‘He cut that meat with a knife.’

Way of influence: Papia Kristang *ku* ← Bazar Malay *sama* ← Hokkien *kăp*

### (13) Hokkien (with the Mandarin and standard Malay versions in brackets)

*Guà kăp î khuă*  
 I with (=ACC’) he see  
 ‘I see him.’

[compare Mandarin:  
*Wǒ kànjiàn tā*  
 I see he  
 ‘I see him.’]

### (14) Bazaar Malay

*gua tengok sama lu (saya tengok kamu)*  
 I see with (=ACC’) you  
 ‘I see you.’

[compare Standard Malay:  
*saya tengok kamu*  
 I see you  
 ‘I see you.’]

### (15) Papia Kristang

*yo olá ku eli/ yo olá ku bos*  
 I see with (=ACC’) he/ I see with (=ACC’) you

In conclusion, some short observations on the languages of Jewish Bible translations are given. In several Jewish communities, there exist specific language forms for Bible translations. These translation forms copy the original Hebrew

with extreme literalism. They do not stand for themselves, but have a serving function, which is to help the reader to better understand the original Hebrew. They were formed in Antiquity after the model of the Aramaic Targumim: since Hebrew and Aramaic are relatively closely related Semitic languages, it was easy to copy the Hebrew original in Aramaic, without forcing or straining the language structure of the target language. The Greek translation of the Septuaginta is literal, but it follows the rules of standard Greek. The translation of Aquila, transmitted only in fragments, is quite different. If we compare the beginning of Genesis in these two Greek translations, we find that Aquila's version contains a strange construction for rendering the Hebrew DOM preposition *et*.

*ἐν κεφαλαίῳ ἔκτισεν ὁ θεὸς σὺν τὸν οὐρανὸν καὶ σὺν τὴν γῆν ...  
καὶ εἶδεν ὁ θεὸς σὺν τὸ φῶς ὅτι καλόν*

Here the preposition *σὺν* is used in a way which contradicts two elementary rules of Greek: first, *σὺν* is never used for the object of a transitive verb; and second, it is never construed with the accusative, but always with the dative. Apparently, the translator wanted to render the Hebrew particle *et*, which indeed also means 'with', besides its function as ACC'. Nevertheless, it is a mistake to take this as a grammatical synonym, since in Hebrew *et* has two functions, whose difference becomes clear when personal suffixes are added: 'me' is *oti*, 'with me' is *itti*. The translator must have been aware of this, and nevertheless he makes this strange and shockingly wrong use of the Greek preposition *σὺν*. However, the later translation into *dhimotiki* is grammatically correct.

Ladino is another case in point. This particular form of Jewish Spanish is exclusively used in Bible translation and in liturgy. Ladino is found often in open contradiction of the rules of standard Spanish. Let us turn again to the beginning of Genesis:

*en prensipyo kriyó el dyo alos syeloš i ala tyera ...  
i vido el dyo ala luz ke buena*

Here, many features copy directly the Hebrew original: the lack of the article in *en prensipiyo*; the plural of *syeloš*, which imitates the plurale tantum *šamayim*; the lack of the copula verb *to be* in *ke bueno*. However, the most striking feature is related to DOM. In Hebrew, the preposition *et* is used only with definite objects, and it follows a referential semantic pattern. In Spanish, the preposition *a* is used with animate objects, and it follows an inherent pattern. Both prepositions are differential, although with divergent semantics. What is noteworthy is that the Ladino translation follows the semantic pattern of Hebrew, not that of Spanish.



The differentiability is kept in the translation, but otherwise the grammar of Spanish is violated, because the inherent semantics of Spanish is replaced by the referential semantics of Hebrew. Thus, the translation uses *a* where according to the rules of Spanish there should be no preposition at all. It is remarkable that the translator seems to have become aware of the fundamental similarity of Hebrew and Spanish, which despite their different semantic patterns, both have differential marking of the object.

**(16) Bible translations (*Genesis 1, 1/4*)**

*bě-re'sit bara' 'ēlohim' et ha-šamayim wě-'et ha-'areš ...*  
 in-beginning created God ACC' ART-heaven and- ACC' ART -earth  
*wa-yar' 'ēlohim' et ha-'or ki tob*  
 and-saw God ACC' ART -light that good  
 [Hebrew original]

*bě-qadmin bēra' YY yāt šēma-yā wē-yāt 'ar'-ā ...*  
 in-beginning created God ACC' heaven- ART and- ACC' earth- ART  
*wa-ħazā YY yāt nēhor-ā' are tīb*  
 and-saw God ACC' ART that good  
 [Aramaic translation (Targum Onqelos)]

*Ἐν ἀρχῇ ἐποίησεν ὁ θεὸς τὸν οὐρανὸν καὶ τὴν γῆν...*  
*καὶ εἶδεν ὁ θεὸς τὸ φῶς ὅτι καλόν*  
 [Koiné Greek translation (Septuaginta)]

*ἐν κεφαλαίῳ ἔκτισεν ὁ θεὸς σὺν τὸν οὐρανὸν καὶ σὺν τὴν γῆν ...*  
*καὶ εἶδεν ὁ θεὸς σὺν τὸ φῶς ὅτι καλόν*  
 [Judeo-Greek, Aquila, in Origenes' Hexapla]

*εις ἀρχὴ ἐπλασεν ο θεός τον ουρανó και την ιγη ...*  
*και ειδαν ο θεός το φώς στι καλό*  
 [Dhimotiki, Polyglot Bible Constantinople 1547]

*en prensipyo kriyó el dyo alos syeloš i ala tyera ...*  
*i vido el dyo ala luz ke buena*  
 [Ladino, Polyglot Bible Constantinople 1547]

## Bibliography

In order not to overload this personal contribution, the bibliographical references have been kept to a minimum.

Bossong, Georg, *Probleme der Übersetzung wissenschaftlicher Werke aus dem Arabischen in das Altspanische zur Zeit Alfons des Weisen*, Tübingen, Niemeyer, 1979.

Bossong, Georg, *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen, Narr, 1985.

Coste, Jean/Redondo, Auguste, *Syntaxe de l'espagnol moderne*, nouvelle éd. corrigée, Paris, Société d'Édition d'Enseignement Supérieur, 1976.

Grierson, George Abraham (ed.), *Linguistic survey of India*, 11 vol., Delhi et al., Motilal Banarsidass, 1967–1968.

Mayerthaler, Willi, *Morphologische Natürlichkeit*, Wiesbaden, Akademische Verlagsgesellschaft Athenaion, 1981.

Meier, Harri, *Sobre as origens do acusativo preposicional nas línguas românicas*, in: Meier, Harri (ed.), *Ensaaios de Filologia Românica*, Lisboa, Revista de Portugal, 1948, 115–164.

Müller, Bodo, *Das morphemmarkierte Satzobjekt der romanischen Sprachen (Der sogenannte präpositionale Akkusativ)*, *Zeitschrift für romanische Philologie* 87 (1971), 477–519.

Nöldeke, Theodor, *Kurzgefasste syrische Grammatik*, Leipzig, Tauchnitz, 1898.

Raible, Wolfgang, *Zum Objekt im Finnischen. Eine sprachwissenschaftliche Fallstudie*, Hamburg, Christoph von der Ropp, 1976.



## **II Research perspectives**



Chantal Melis

# From topic marking to definite object marking

Focusing on the beginnings of Spanish DOM

**Abstract:** The present work is a diachronic study of Differential Object Marking (DOM) in Spanish which focuses on the early stages of the grammaticalization process undergone by the *a*-marker. Following Pensado (1995), I relate DOM *a* to the topicalizing function of Latin AD ‘with regard to, as to’, and then proceed to examine the step-by-step expansion of the grammaticalizing item, leading from personal pronouns to proper names and from these to definite common nouns of human reference. The central aim of the paper is to shed light on the specific properties associated with the distinct subsets of direct objects which were reached by DOM, under the assumption that these properties served to mediate shifts between categories and facilitated the development of a system governed by the semantic dimensions of animacy and definiteness, whose origins were tied to a parameter of information structure.

**Keywords:** Differential Object Marking (DOM), animacy, definiteness, grammaticalization, information structure, topicality

## 1 Introduction

Modern Spanish is a language with an extensive, dynamic and complex system of Differential Object Marking (DOM).<sup>1</sup> At present, DOM in Spanish displays a basic distinction between human and non-human referents (Torrego Salcedo 1999, 1782; Leonetti 2004, 82; de Swart 2007, 132), according to which human-referring objects are regularly preceded by the case form *a*, while inanimate entities are generally unmarked (Torrego Salcedo 1999, 1781–1782):<sup>2</sup>

---

1 I owe thanks to two anonymous reviewers for invaluable comments on a previous version of this paper.

2 The distinction is usually couched in terms of an animate vs. inanimate opposition. The status of animate non-human direct objects, however, is not clear in Spanish (von Heusinger 2008, 4). Unlike the human-referring objects, which trigger DOM on a regular basis, the poorly document-

---

**Chantal Melis**, National Autonomous University of Mexico, e-mail: [cme@unam.mx](mailto:cme@unam.mx)

- (1) a. *Traj-eron a un amigo con ellos.*  
 bring-PFV.3PL ACC a friend with them  
 ‘They brought a friend with them.’
- b. *Traj-eron una maleta con ellos.*  
 bring-PFV.3PL a suitcase with them  
 ‘They brought a suitcase with them.’

The current situation is the result of an extended process of grammaticalization, whereby *a*-marking, much more restricted at earlier stages of the language, was allowed to spread to an increasing range of direct objects along a path usually held to have been governed by the familiar animacy and definiteness scales (Bossong 1985; Aissen 2003; Laca 2006). In general terms, *a*-marking, with a strong bias towards human entities throughout the history of Spanish, developed along the definiteness scale from higher-ranked definite objects to lower-ranked indefinite noun phrases until all human referents were reached.

Of particular interest in this diachronic scenario is the way in which definiteness and animacy were kept closely bound. By contrast, in many languages where definiteness is crucial for DOM, all higher-ranked entities on this dimension trigger marking, irrespective of animacy (Bossong 1991, 161; Aissen 2003, 53). This suggests that definiteness and animacy are independent parameters (de Swart 2007, 187). Indeed, according to Anagnostopoulou (1999, 789), the animacy restrictions imposed on the property of definiteness, as observed in Spanish, appear to relate to a “general mystery of natural languages”, whereby “*animacy* plays a role in determining the ‘degree’ of *definiteness* of NPs for reasons that are not entirely clear”. On this view, then, an intriguing question raised by Spanish DOM has to do with the origin of the close relationship established between definiteness and humanhood.

Another issue of interest concerns the strong personal pronouns of Spanish, identified as the first set of direct objects that generalized an obligatory use of *a*-marking (Meier 1948; Ramsden 1961; Fernández Ramírez 1964; Rohlf’s 1971). To explain the leading role of these forms in the grammaticalization process of DOM, scholars invoke the animacy scale (e.g. Comrie 1989, 128; Bossong 1991, 159) or the definiteness scale (Aissen 2003, 444; von Heusinger 2008, 5), on which the personal pronouns are located at the respective top ends. In neither case, of course, the pronouns can be regarded as more “animate” or more “definite” than lower-ranked sets of direct objects, such as, for example, definite common

---

ed category of animals shows highly optional marking (cf. García García 2018, 217–218 for discussion).

nouns of human reference. In fact, reflecting on the cross-linguistic preference of DOM for pronouns, Comrie (1989, 198) wonders whether the relevant criterion might not rather be conceived of in terms of a “hierarchy of topic-worthiness”. This hypothesis concurs with a number of recent proposals in which topicality is claimed to be the driving force behind the emergence and development of DOM systems (Escandell Vidal 2009; Iemmolo 2010; Dalrymple/Nikolaeva 2011; Iemmolo/Klumpp 2014; Witzlack-Makarevich/Seržant 2018). The notion of topicality has also been brought up in relation to Spanish DOM, particularly when the oldest instances of *a*-marking are examined (Melis 1995; Pensado 1995; Laca 2006). However, as noted in García García (2018, 216), the extent to which topicality may satisfactorily account for the entire evolution of *a* is not clear, although attempts in this direction have been made: one could argue, for instance, that the basic function of *a* is, and has always been, to mark direct objects as internal (secondary) topics (Leonetti 2004), or *a*-marking could be interpreted as a strategy for highlighting discourse prominent and, in this sense, highly topical objects (Laca 1995). Nevertheless, assuming that topicality played a role in the evolution of Spanish DOM, what remains to be outlined is the diachronic path along which the referential properties of animacy and definiteness were brought to interact with the pragmatic dimension of topicality.

Another matter deserving of attention relates to the commonly held view that systems of DOM arise in response to the need to disambiguate grammatical relations in transitive clauses. This is the sense in which the so-called “distinguishing” or “discriminatory” function of case marking (Malchukov/de Swart 2009; Siewierska/Bakker 2009) is applied to DOM: direct objects which, owing to some of their associated properties, run the risk of being confused with subjects trigger marking. Ambiguity avoidance is also claimed to have inspired the original recourse to DOM in Spanish, according to a long-standing hypothesis (Müller 1971) which continues to draw support in the literature (e.g. de Swart 2007, 142–147). The trouble with this claim is that it does not suit the oldest data of *a*-marking, since the personal pronouns, first targeted by DOM, displayed unmistakable object forms (*mí*, *ti*, etc. vs. nominative *yo*, *tú*, etc.) and were very unlikely to generate confusion. This has been observed in a number of studies (Pensado 1995, 191; Bossong 1998, 223; cf. Aissen 2003, 437), in light of which the initial motivation for differential marking in Spanish seems worth examining in greater detail.

Finally, the *a*-marker itself has to be looked at more closely. On the surface, the device used for DOM is identical to the case marker which obligatorily precedes all indirect object noun phrases in Spanish. The coincidence in form between DOM *a* and dative *a* lies at the bottom of the prevailing consensus that the dative preposition was employed for Differential Object Marking. This consensus has been strengthened by cross-linguistic findings regarding similar ten-

dencies in a variety of DOM systems (Bossong 1991). Nonetheless, an alternative view does exist, according to which the *a*-marker in Spanish (and other Romance languages) goes back to the topicalizing use of the Latin preposition AD meaning ‘with regard to, as to’, as argued in Iemmolo (2010), with reference to Pensado (1995), who developed this proposal in an important paper a couple of decades ago. Exploring the history of Spanish DOM from the vantage point of this challenging hypothesis is a task which is worth to be taken up.

The pending questions concerning Spanish DOM bring to mind recent typological publications which show that the factors conditioning split case alternations are far more diverse across and within linguistic families than generally supposed (Bickel/Witzlak-Makarevich 2008; Bickel/Witzlak-Makarevich/Zakharko 2015; cf. also Sinnemäki 2014). For a deeper understanding of the attested heterogeneity, it is concluded, fine-grained diachronic studies should be carried out exploring the idiosyncratic ways in which individual systems of Differential Object Marking arise, develop and mature (Bickel/Witzlak-Makarevich 2008, 33).

My aim in the present work is to shed light on the peculiarities of Spanish DOM, focusing on the early stages of its development and watching for details of the unfolding evolutionary path of *a*-marking. I will situate the source of the marker in the topicalizing use of the Latin preposition AD ‘with regard to, as to’, following Pensado (1995), and I will then proceed to examine the beginnings of the process of grammaticalization. As is customary in functional approaches, a gradual type of expansion will be expected, moving through a sequence of step-by-step changes associated with the spread of the original form to new environments (Traugott/Trousdale 2010). In this process, the element undergoing grammaticalization develops new functions which imply more and more distancing with respect to the lexical value of the source item (García/van Putte 1987, 373; Lichtenberk 1991, 76). Grammaticalizing *a* will give us the opportunity to plot a movement of this nature, departing from contexts with features of meaning closely related to the sense of the original topic-marker and extending to contexts where topicality fades out and new functions linked to referential properties emerge.

The chapter is organized as follows. In Section 2 I summarize Pensado’s (1995) proposal regarding the topicalizing construction that gave rise to DOM in Spanish. Section 3 discusses the grammaticalization of *a* with the strong personal pronouns and postulates a shift from a notion of topicality to a feature of discourse prominence. Section 4 deals with the extension of obligatory DOM to proper names of human reference, which had the effect of turning individuation into a crucial factor for marking. In Section 5 I examine the incipient spread of *a* to human common nouns, as reflected in the epic poem *Cantar de mio Cid*, written around 1200, where singularity is found to outweigh definiteness and



topicalized objects verify the origin of the *a*-marker. Section 6 sketches the subsequent diffusion of *a* to the entire class of human objects and briefly touches upon the interacting parameter of verbal semantics. The paper ends with a short conclusion in Section 7.

## 2 The source of Spanish DOM: topicalizing Latin AD

As mentioned in the Introduction, DOM-marked objects in Spanish carry a case form which is identical to the preposition introducing indirect object noun phrases. The tendency for DOM and dative markers to converge is a recurrent phenomenon cross-linguistically (Bossong 1991, 157), susceptible of being explained if one considers that “direct and indirect objects are structurally similar in being non-subject arguments”, in conjunction with the fact that “indirect objects are overwhelmingly human (or animate) and definite, exactly the properties which favor DOM for direct objects” (Aissen 2003, 446–447, note 10). With regard to Spanish, it is important to underline that the similarly marked arguments behave differently from a syntactic point of view.<sup>3</sup>

DOM markers, evidently, may have other sources (Bossong 1991, 167, note 41). Within the Romance family, for example, Romanian *pe* ‘on’ (< Lat. *PER*) used for DOM is a preposition of locative origin which bears no resemblance to the morphological dative case form, while in other languages, discussed in Iemmolo (2010, 262–265), the marker appearing on a subset of direct objects (and covering some locative and dative functions as well) conveys a basic sense of “aboutness”. The topical features associated with the latter type are particularly relevant to our study, considering that a couple of decades ago Pensado (1995) put forward a hypothesis as to the origin of the *a*-marker of Spanish which likewise appeals to a notion of topicality. Most significantly, her historical reconstruction, to be summarized in a moment, leads to the establishment of a clear distinction between dative *a* – developed from the directional uses (‘to, towards’) of the Latin preposition *AD* – and DOM *a* – deriving from the topicalizing function (‘with regard to, as to’) of the same Latin etymon –, which imposes a new perspective on the way in

---

<sup>3</sup> The differences mentioned in Bossong (1991, 155) include the differential use of DOM *a* versus the compulsory use of dative *a*, the occurrence of DOM-marked objects with transitive verbs, and the pronominalization of the two types of object with distinct clitics. Moreover, DOM-marked objects and dative objects in Spanish contrast with respect to phenomena such as passivization, nominal modification and depictive secondary predicates (Bárány 2016).

which a system of Differential Object Marking was introduced into the language. Studies on Spanish DOM now often cite Pensado's work to express agreement with her proposal (Torrego Salcedo 1999; Leonetti 2004; Escandell Vidal 2009; Iemmolo 2010), although the full implications of this novel and challenging view have not received due attention.

According to Pensado (1995), the marker recruited for DOM in Spanish goes back to the use of Latin *AD* in contexts where the preposition, meaning 'with regard to, as to', served to indicate a shift of topic.<sup>4</sup> Basing her proposal on a careful examination of late Latin and early Romance data, Pensado reconstructs an evolutionary path along which the topicalizing function of *AD* was passed on to some of the Romance dialects, where the inherited structure appears to have been initially reserved for the personal pronouns of first and second person, that is, for the speech act participants. In classical Latin, *AD* governed accusative case, but due to the gradual erosion of the case marking system which is known to have taken place in the transition period between Latin and Romance, and as a result of the merging processes which it occasioned, in late Latin combinations of *AD* and the dative case form of the pronouns began to emerge (*AD* + *MIHI<sub>DAT</sub>*).<sup>5</sup> This use is well attested in Late Latin textual sources (Müller 1971, 494–495), and Pensado conjectures that it was introduced in the topicalizing structures, irrespective of whether the dislocated pronominal constituent fulfilled the role of direct or indirect object in the main clause (Pensado 1995, 203):

- (2) *Ad mihi, (mihi) dixit* 'As to me, he told (me)'  
*Ad mihi, (me) amat* 'As to me, he loves (me)'

The hypothesis that the differential object marker *a* of Spanish and other Romance languages developed from the topicalizing function of *AD* with the personal pronouns of first and second person singular illuminates the contrast existing between DOM languages, like Spanish, and non-DOM languages,

---

<sup>4</sup> The topicalizing function of Latin *AD* manifests itself in examples where the topic shifter forms part of a larger phrase (*quod ad me attinet, X* 'as far as I am concerned, X'; *quod ad Xenonem, X* 'as for Xenon, X'), and in contexts where it is used alone (*ad ea autem, quae scribis de testamento, X* 'with regard to what you write about the will, X') (Pensado 1995, 200–201).

<sup>5</sup> In the nominal area, the dative case was replaced by *AD* + accusative for the expression of the indirect object function (*ad filios tuos* 'to, for your children'), and the nominal pattern appears to have acted as the model for the extension of the preposition to the personal pronouns, in spite of the fact that these were marked as dative and did not need the preposition to display their case (*Dixit spiritus ad mihi* 'the holy spirit said to me') (Pensado 1995, 184–185).

like French. The difference in question has to do with the well-known divergence of the Romance pronominal system into stressed – strong or tonic – and unstressed – weak or atonic – forms. Under Pensado’s analysis, the dialects which developed the use of *a mí* (< AD MIHI) in pragmatically marked constructions retained the dative ancestors for a longer time and made them serve as exponents of the stressed series of pronominal forms (ACC/DAT *a mí* and *a ti*), as opposed to the unstressed forms (ACC/DAT *me* and *te*), which were derived from the Latin accusative.<sup>6</sup> By contrast, in the dialects where topicalizing AD does not appear to have taken root, dative MIHI and TIBI were lost at an earlier date and the accusative paradigm of Latin provided the source for both the weak (Lat. ME > Fr. *me*) and the strong (Lat. ME > Fr. *moi*) pronouns (Pensado 1995, 180–181; cf. Harris 1978, 102–103).

Additionally, the conjecture that the topicalizing structures with *a* were initially limited to the speech act participants gains support from looking at the stressed forms of the third person pronouns (*él* ‘him’, *ella* ‘her’, etc.), which for the most part descend from Latin accusatives (of the paradigm of demonstrative pronouns) and, unlike *mí* and *ti*, are not dative. In other words, DOM marking on the Spanish third person stressed object pronouns (ACC/DAT *a él*, *a ella*, etc.) must have spread at some later point in history, after the split between stressed and unstressed forms (Pensado 1995, 194–195).<sup>7</sup>

The hypothesis advanced by Pensado is very much in line with the growing body of research on the fundamental role topicality is assumed to play in the emergence and development of DOM systems across the world (Escandell Vidal 2009; Iemmolo 2010; Dalrymple/Nikolaeva 2011; Iemmolo/Klumpp 2014; Witzlack-Makarevich/Seržant 2018). Worthy of note is the supporting evidence her hypothesis receives from a group of Romance dialects, where a presently incipient system of DOM is targeting topical objects appearing in left- or right-dislocation structures (Escandell Vidal 2009; Iemmolo 2010).

When DOM arises in these pragmatically marked constructions, future events may take different courses: in some languages DOM remains confined to topical objects, whereas in other languages DOM is extended to non-topical objects (Iemmolo 2010, 247; cf. Dalrymple/Nikolaeva 2011). If extended, DOM loosens

---

<sup>6</sup> Scholars assume that the ACC/DAT syncretism in the unstressed pronouns was created by analogy with the stressed forms (Folgar 1993, 52).

<sup>7</sup> The personal pronouns of first and second person plural show a similar phenomenon: their stressed and unstressed forms jointly derive from the Latin nominative/accusative (Pensado 1995, 194). This corroborates Pensado’s hypothesis that the topicalizing source of DOM *a* was linked to the pronouns of the singular first and second persons.

its ties to information structure and becomes sensitive to the presence of “topic-worthiness” features, which often translate into semantic factors such as animacy and definiteness (Iemmolo 2010, 257). In what follows, an attempt will be made to spell out the details of this general scenario in relation to Spanish.

### 3 Personal pronouns at the first stage of the grammaticalization process

In the literature on Spanish DOM it has long been noted that the first targets of grammaticalizing *a* were the stressed object personal pronouns (Meier 1948; Ramsden 1961; Fernández Ramírez 1964; Rohlf's 1971). The oldest sources manifesting a regular use of DOM with these pronouns are Mozarabic texts of the Iberian Peninsula dating back to the 11<sup>th</sup> and 12<sup>th</sup> centuries (Bossong 1998, 223–224). Additional evidence emanates from the epic poem *Cantar de mio Cid* (early 13<sup>th</sup> century), in which the referred pronouns are always *a*-marked, regardless of their position in the clause. Some topicalized pronouns occur initially (3a), but others occupy the postverbal slot typical of direct object noun phrases (3b):<sup>8</sup>

- (3) a. *a ti adoro e credo de toda voluntad*  
 ACC 2SG worship.PRS.1SG and believe.PRS.1SG of all goodwill  
 ‘I worship you and believe in you with all my heart.’ (v. 362)
- b. *Oí-d a mí, Álbar Fáñez e todos los cavalleros*  
 listen-IMP ACC 1SG Alvar Fañez and all the knights  
 ‘Alvar Fañez and all the knights, listen to me.’ (v. 616)

Our discussion of the grammaticalization process undergone by *a* has to start by stressing that DOM was not implemented for the purpose of disambiguating participant roles. This seems clear since the objects encoded in *mí* and *ti* are overtly marked as accusative forms and are not liable to being confused with their subjective counterparts (*yo* and *tú*, respectively). Ambiguity avoidance defines one of the oldest hypotheses invoked in the literature to explain the origin of Spanish DOM (Müller 1971) and continues to draw scholarly approval (de Swart 2007,

<sup>8</sup> The examples of the *Cid* are cited from Montaner's (1993) edition.

142–147; Malchukov/de Swart 2009, 349).<sup>9</sup> As noted in various studies, however, the pronominal data associated with the oldest phase of the grammaticalization process militate against this hypothesis (Pensado 1995, 191; Bossong 1998, 223; cf. Aissen 2003, 437).

From a different perspective, the personal pronouns are natural candidates to attract differential marking at the initial stage, given their position in the hierarchies of animacy and definiteness which are held to regulate DOM cross-linguistically (Comrie 1979; Bossong 1985; Lazard 1998; Aissen 2003). These hierarchies, as is well known, organize discourse entities in accordance with the likelihood of their appearance as subjects (high end) or objects (low end) in bivalent clauses. Independently of whether the personal pronouns are situated on the animacy (Comrie 1989, 128; Bossong 1991, 159) or definiteness scale (Aissen 2003, 444; von Heusinger 2008, 5), they are treated as conforming the highest-ranked category of elements, optimally predisposed to function as subjects. Consequently, if the iconic motivation for DOM is to *mark* objects sharing properties with prototypical subjects, that is, semantically *marked* objects (Comrie 1989, 128; Bossong 1991, 162; Aissen 2003, 438), personal pronouns cast in the direct object role have to be viewed as prime triggers of differential marking.

On a third approach, embodied in studies which relate the genesis of DOM systems to topicalizations in dislocated constructions, the extension to the personal pronouns at the beginning stage of the grammaticalization process is justified on the grounds that the members of this category are “highly topical” (Iemmolo 2010, 258). It is also pointed out that the notion of topicality meshes with the referential dimensions of animacy and definiteness insofar as “prototypical topics are usually definite, specific and animate” (Escandell Vidal 2009, 836).

---

<sup>9</sup> Actually, the disambiguating hypothesis is supposed to account for the emergence of DOM systems across languages (Malchukov 2008). In this view, DOM arises with a context-dependent (global) function of discrimination, when subject and object require disambiguation, and develops into a processing-wise less costly system of context-independent (local) discrimination, which no longer relies on a comparison of the object with the subject of the sentence, but shows a generalized type of marking appearing on certain classes of referents. In accordance with this scenario, it can then be explained why there do not exist many differential marking systems that seem to perform a purely disambiguating function (Malchukov 2008; Witzlack-Makarevich/Seržant 2018): disambiguation captures what sets DOM systems in motion and continues to operate as “a weak universal force”, surfacing on occasion in some contexts of use, but subordinated to other processes (Seržant 2019). As said, the hypothesis does not concur with the data of incipient DOM in Spanish, but this does not rule out the possibility that, once introduced into the language, the new device was called upon to distinguish objects from subjects in specific contexts where clarification was necessary or highly desirable.

The relevant question here is what it means to be “topical”. If we interpret topic as characterizing the entity “which the proposition expressed by the sentence is ABOUT” (Lambrecht 1994, 118), the definition applies to constituents which have topic status in particular syntactic environments. But personal pronouns are not necessarily topical in this sense, because they may also appear in the focus domain of an utterance, as part of the comment or as the sole constituent in focus (Lambrecht 1994, 128–130). And it is reasonable to assume that, in some instances, the DOM-marked personal pronouns of Spanish stood in a focal relationship to the proposition.

At the same time, theories of grammaticalization predict that the first context to which the grammaticalizing element is extended will show a high degree of compatibility with the original value of the lexical source (García/van Putte 1987, 373; Lichtenberk 1991, 76). Our task thus consists in identifying meaning properties which could have facilitated the passage from topicalizing *a(d)* to *a*-marking on all personal pronouns, regardless of whether the initial targets of DOM functioned as clausal topics or not.

I suggest that these properties may be localized in the general condition of prominence that is attributed to the pronouns (Anagnostopoulou 1999, 770). The condition is defined as involving both “familiarity” and being in the current centre of attention, that is, being “active” in Chafe’s (1987) terms. Since both speech act participants (*a mí, a ti*) and anaphoric third parties (*a él, a ella*, etc.) conform to this description, the functional change experienced by *a* can be specified by means of a concept of prominence: the topicalizing marker severed its ties to information structure and moved in the direction of a more broadly conceived dimension of communicative importance, i.e. being familiar and active at the moment of utterance.

For the future development of grammaticalizing *a*, two additional pieces of information need to be taken into account. The first one bears on the emphatic character of the strong personal pronouns of Spanish. The weak (clitic) object pronouns (*me, te, lo*, etc.) predominate in discourse, because they are given the task of encoding the personal referents in normal circumstances. The strong pronouns, on the other hand, are used occasionally in contexts where some kind of special contrastive effect is intended; they signal an explicit or implicit attempt on the part of the speaker to compare or oppose the individual in question to other referents (Luján 1999). This leads to the recognition that Spanish DOM initiates its trajectory in association with a special class of marked forms, to which an appeal is made when the distinct identity of the highlighted participant matters to a considerable degree. We will see below that issues of referential identity will soon be moved into the foreground.

Secondly, one needs to be aware of the fact that the stressed personal pronouns of Spanish in (subject and) object function only refer to persons and never to inanimate entities (Luján 1999, 1294; Ramsden 1961). Pronouns in general are not restricted in this way, but this feature will turn out to be instrumental in maintaining the preference of *a*-marking for human beings throughout the evolution of Spanish DOM.

In sum, the first step in the process under analysis reveals a pathway of change moving from pragmatically marked constructions, at the information structure level, towards marked pronominal forms of special communicative importance. As topicalized objects in the source constructions, the pronouns were assigned the role of acting as subjective “viewpoints” (DeLancey 1981) from which speakers chose to report the event (cf. Section 5 below).<sup>10</sup> The first step towards the establishment of a grammaticalized system of differential marking consisted in extending topicalizing *a* to prominent – familiar, activated and, in the case of Spanish, strictly human – entities embedded in discourse contexts in which their unique identity, as opposed to that of others, turned out to be exceptionally relevant.

## 4 From personal pronouns to personal names

Human-referring proper names were next in motivating a compulsory use of *a* (Müller 1971; Folgar 1993), that is, DOM was extended from the pronominal domain to the nominal category. This represents an important shift, interpretable as indicative of the fact that topical values are fading out. Indeed, compared with pronouns, lexically coded objects are less likely to enjoy the status of active discourse participants at the moment of utterance and hence less likely to function as aboutness topics (according to the topic acceptability scale: Givón 1983, 18; Lambrecht 1994, 165).

Evidence for the second stage of the grammaticalization process is provided by the *Cantar de mio Cid*, in which, beyond the personal pronouns, proper names of human reference are alone in showing obligatory marking wherever they occur:<sup>11</sup>

---

<sup>10</sup> DeLancey (1981, 637, note 17) alludes to the relation existing between the category of viewpoint and topicality.

<sup>11</sup> Toponyms, by contrast, are only optionally marked (Monedero Carrillo de Albornoz 1978). It should be pointed out that human names yield one pragmatically motivated omission of *a* discussed in García/van Putte (1987).

- (4) *los brazos abiertos, recibe a Minaya*  
 the arms open receives ACC Minaya  
 ‘With open arms, he welcomes Minaya.’ (v. 489)

Why proper names were reached before common nouns by grammaticalizing *a* is often explained with an appeal to the hierarchies of animacy and definiteness. On these hierarchies, proper names are placed at the higher end (following pronouns and preceding common nouns) to reflect the fact that DOM-based splits across languages show a tendency to treat proper names as more animate (Comrie 1989, 186) or more definite (Aissen 2003, 443) than common nouns.<sup>12</sup>

Bearing in mind that *a* was extended to definite common nouns at the next stage (cf. Section 5), we are justified in exploring to what extent the definiteness dimension can account for the expansion of Spanish DOM to proper names. Definiteness is a complex notion, in which a component of familiarity – the referent is assumed to be identifiable to the addressee – interacts with an idea of uniqueness – there exists one entity (or one set of entities) which satisfies the description. As stated in the literature, definite noun phrases do not always carry these values (cf. Belloro 2007, 110), but it can still be argued that, in the clearest or more typical case, definite expressions encode uniquely identifiable referents.

Proper names are definite in this sense. They are characterized by a feature of uniqueness, “since they involve only a single individual, a class of one member”, and may therefore be considered as “the epitome of individuation” (Kliffier 1984, 199). However, referring uniquely to an individual is not a property associated with proper names alone. This function is shared by (the singular set of) personal pronouns, as well as by definite singular common nouns, with which *a* generalized at a much later date. Why the separate treatment then? Proper names appear to have played a mediating role in the shift from pronouns to common nouns for some reason to be elucidated.

The plausible explanation points to a couple of features which link proper names with personal pronouns and oppose them to definite nouns. From a semantic point of view, personal pronouns and proper names are alike in carrying little information beyond the designation of an individual (Laca 1995, 82); common nouns have categorial meaning. On the formal level, the personal pronouns are inherently definite forms, unlike the common nouns, which, by contrast, depend

---

<sup>12</sup> Aissen (2003, 444) suggests that the ordering of elements on the definiteness scale (PRONOUN > NAME > DEFINITE > INDEFINITE SPECIFIC > NON-SPECIFIC) has to do with the extent to which the value assigned to the discourse referent is fixed. It is fixed by the speech situation in the case of pronouns and by convention in the case of proper names. Definite descriptions, by comparison, rely on previous discourse, and indefinites allow for greater freedom in fixing their value.



on the presence of some determiner to transform into definite referential expressions (Siewierska 2004, 10),<sup>13</sup> and from this point of view, proper names side with pronouns, against common nouns, given their capacity of introducing definite referents without the support of overt modification (Lambrecht 1994, 87).

In grammaticalization processes, new functions develop from older functions one after the other, in a sequence of small steps that imply more and more distancing from the original value of the source item (García/van Putte 1987, 373; Lichtenberk 1991, 76). Recent work on grammaticalization has brought analogy back into the limelight to explain how such movements from one context to another take place. Analogy thrives on relations of similarity, in form, meaning or grammatical function, which are perceived by speakers and operate as major mechanisms in determining pathways of change (De Smet 2012; Fischer 2013, among others). Note how analogy likewise enables us to clarify the role played by proper names in the expansion of grammaticalizing *a*: DOM proceeded to definite noun phrases which by themselves, without determiners, like the personal pronouns, designate unique referents.

The new targets of DOM have brought individuation to the fore as a crucial factor for marking. The continuity with the previous stage is easy to see if we recall that the emphatic pronominal contexts were exceptionally concerned with the individuality of the referent, understood in the sense of having a distinct identity, of being that person and no other (Laca 1995, 82). But concomitantly, as expected, the functional load of *a* has been shifted. The marked objects, indeed, are no longer required to encode active participants; the degree of necessary prominence has been lowered. Neither are they pragmatically marked forms in the way the emphatic pronouns were. These changes signal that discourse-related conditions are losing their grip on the marking, while opening the way for the interaction of DOM with a hierarchy of individuation, which, following Comrie (1989, 199), may be equated with a hierarchy of salience, since “[s]alience relates to the way in which certain aspects present in a situation are seized upon by humans as foci of attention, only subsequently attention being paid to less salient, less individuated objects”. Hence, at this second stage, saliency in terms of standing out as a distinct, single, separate entity is now sufficient to induce *a*.

---

**13** According to Siewierska (2004, 124), a primary feature of the personal pronouns is “necessary referentiality and even definiteness”, reflected in the fact that these forms “typically cannot occur with definite determiners, or indefinite articles, be construed as bound variables or receive a non-specific or generic interpretation”.

## 5 Extension to definite common nouns

The *Cantar de mio Cid* affords a picture of the incipient expansion of DOM to human-referring common nouns. In contrast to personal pronouns and proper names, which are always preceded by *a*, common noun phrases are optionally marked. Melis (1995) reports a global frequency of 36% of marking on these objects (68/187) and illustrates the variation with examples such as in (5):

- (5) a. *Casastes sus fijas con ifantes de Carrión*  
 marry.PFV.2SG his daughters with infantiles of Carrion  
 ‘You married his daughters to the infantiles of Carrion.’ (v. 2939)
- b. *reciba a míos yernos como él pudier*  
 receive.IMP.2SG ACC my sons-in-law as he can.prs.cond.3sg  
*mejor*  
 better  
 ‘Let him give my sons-in-law the best possible welcome.’ (v. 2637)

As discussed in diachronic studies on Spanish DOM, the diffusion of *a* among common nouns was a gradual process which took many centuries before culminating in near obligatory marking on all human objects, with definite nouns being clearly favoured during the earlier phases of this process (Company Company 2002b; Laca 2006; von Heusinger/Kaiser 2011; García García 2018).

The sensitivity of DOM to definite nouns at the beginning of the third stage can be verified in the *Cid*, where definite human objects trigger *a* in 46% (50/109) of their uses, as compared to 22% (19/88) of marked indefinite human nouns (García/van Putte 1987, 376; cf. Laca 2006, 441, with results based on a partial revision of the poem). Moreover, when a number distinction is taken into account, it is observed that singular definite human objects motivate DOM with the highest frequency (14/19 = 74%) (García/van Putte 1987, 376). This comes as no surprise, given the proximity of this subgroup to proper names; in both cases, the object refers to a uniquely identifiable human being.

What the epic poem additionally reveals, somewhat unexpectedly, is that plural definite nouns (36/90 = 40% of *a*) and singular indefinite nouns (7/20 = 36% of *a*) share a similar distribution of optional marking (García/van Putte 1987, 376). These findings suggest that the parameter of singularity competes in a significant way with the definite dimension. In fact, as García/Van Putte (1987) demonstrate via a calculus of relative weights, singular number in the *Cid* turns out to be more influential than definiteness as the relevant factor for DOM.

Cross-linguistically, DOM systems have a tendency to introduce a singular-plural opposition in noun phrases higher in animacy and not in lower ranked

(inanimate) categories. Comrie (1989, 189) proposes that the correlation between number distinctions and animacy may be seen as “reflecting greater human concern with entities of higher animacy as individuals, therefore countable, while entities of lower animacy are more readily perceived as an indeterminate class”. According to Bickel/Witzlak-Makarevich/Zakharko (2015, 17), the selection of singular over plural referents in differential marking systems “is based on the assumption that singular is more indexible than nonsingular and therefore ranks higher: singular items can be better pointed at than other referents”. As is clear, in one way or another, these explanations lead back to the hierarchy of salience, presented above, in regard to which Comrie (1989, 199) brings to attention that “work on salience indicates that singular entities are more salient than plural entities”.

From this perspective, the precedence taken by singularity over definiteness in the *Cid*'s optional markings on human common nouns is easy to interpret. What underlies these uses is the criterion of individuation ushered in by proper names (cf. Section 4 above). We are witnessing a grammaticalization process that is unfolding step by step through the mediation of similarities and continuities: in the expansion from proper names to definite common nouns, singularity has had the role of acting as a factor of transition.

At the same time, there is evidence that grammaticalizing *a* has not been fully bleached of its original lexical sense in the *Cid*. Significantly, indeed, in approximately half of the DOM marked examples, the human common nouns occur in topicalizing structures, with the object detached to the left periphery and usually accompanied by a coreferential weak pronoun (Melis 1995; 2018; Laca 2006). Note how a paraphrase of (6) in terms of ‘as to his daughters, he took them in his arms’ is still available:

- (6) *a las sus fijas en braços las prendía*  
 ACC the his daughters in arms ACC.FEM.PL took.3SG  
 ‘He embraced his daughters.’ (v. 275)

Left-and right-dislocated structures are widely used in the *Cid* (Menéndez Pidal 1964, 323). It is possible that they should be viewed as lingering traces of the oral tradition that is assumed to have given shape to the epic poem, considering that in later medieval texts dislocations become extremely rare and objects affected by DOM are no longer topicalized in this way (Laca 2006, 471). One could therefore argue that the frequency with which common noun objects appear in topicalizations has more to do with general principles governing the architecture of the epic poem than with the original function of the preposition that gave rise to DOM. However, if this were the case, one would expect a similar distribution of topi-

calized entities across the object category. But this does not happen: whereas the obligatorily marked personal pronouns and proper names appear left-dislocated in a quarter of their uses, the human common nouns marked with *a*, as I said, motivate the topicalizing strategy in half of the registered examples.<sup>14</sup> The inference to be drawn from this contrast is that the common nouns are still very much dependent on pragmatically marked constructions to license a DOM-driven split.

I will use example (6) for a brief reflection on the informative status of the topicalized object. This matter deserves a moment of attention, because studies on DOM, in which the origin of differential marking has been explicitly related to mechanisms of topicalization, disagree as to whether the promoted object in these constructions assumes the role of primary (Iemmolo 2010) or secondary (Dalrymple/Nikoaleva 2011) topic. My proposal is to look at (6) to see if it can help us resolve the debate. We begin by noting that the participants present in the depicted scene are the *Cid* and his daughters. The father appears as the grammatical subject of the clause, anaphorically encoded in the verbal suffix and clearly functioning as primary topic, in light of which, the daughters, notwithstanding their communicative importance, have to be interpreted as secondary topic. Sentences with two topics are not uncommon. According to Lambrecht (1994, 148), a sentence of this nature “in addition to conveying information about the topic referents, conveys information ABOUT THE RELATION that holds between them as arguments in the proposition” (cf. Nikolaeva 2001). The relation holding between two topical participants can be expressed in a variety of sentential forms (Lambrecht 1994). What explains the marked character of structures like (6) is that the secondary topic is given the highest degree of pragmatic prominence among the participants, by being chosen as the viewpoint (DeLancey 1981) from which the event – a loving embrace between father and daughters – is described.

It is important to realize that the hypothesis of secondary topic in no way collides with the role Iemmolo (2010) attributes to this type of dislocated structures in his proposal about the emergence of DOM systems. Certainly, there will be cases in which the dislocated object does indeed function as primary topic, with the subject pertaining to the focal portion of the clause, but more commonly there will be a topical subject in relation to which the object defines a secondary topic.

The dislocated examples of the *Cid* are precious data lending full support to the topicality origins of Spanish DOM. Although the situation reflected by the epic poem displays a phenomenon of grammaticalization already in process, accom-

---

<sup>14</sup> In my data sample of the *Cid*, personal pronouns functioning as direct objects appear topicalized in 27% of their occurrences (4/15) and proper names of human beings in 26% of the extracted tokens (14/54). DOM-marked common nouns of human reference yield 46% of topicalizations (31/68).

panied by the development of new conditioning factors such as discourse prominence and semantic individuation, the topicalizations involving common nouns are contexts in which the source sense of *a(d)* ‘with regard to, as to’ remains visible and active. The co-existence of the preposition’s older and newer functions within the poem exemplifies the principle of layering discussed in Hopper’s (1991) work on grammaticalization and can easily be accounted for.

## 6 Evolution of DOM

In medieval texts posterior to the *Cantar de mio Cid*, marking on human-referring common nouns continues to be optional, with the difference, as mentioned above, that objects triggering DOM are rarely topicalized. The diachronic data also suggest that the criterion of singularity, prevailing in the epic poem, eventually gave way to that of definiteness.

Historical studies of Spanish DOM have outlined an evolutionary path roughly divisible into three major phases: optional *a* favours human definite objects during the Middle Ages; a fairly regular use of DOM with these objects is established in Renaissance Spanish, at a time when indefinite human objects are marked in about half of their occurrences; *a* becomes the norm with the indefinite group in the modern period of the language (for statistical data supporting this division, cf. von Heusinger/Kaiser 2005, 45; 2011, 602; Laca 2006, 442–443; von Heusinger 2008, 14; García García 2018, 214–215).<sup>15</sup>

What García/van Putte (1987) have added to this picture is a proof of the shift from singularity to definiteness that took place during the medieval era. This is achieved by means of a comparison between the *Cantar de mio Cid* and Cervantes’ *Quijote* (early 17<sup>th</sup> century), which shows, via a similar calculus of relative weights (cf. Section 5), that in the later text definiteness has come to weigh more heavily on the operation of DOM than singular number. That is to say, objects encoding identifiable human referents have gained the upper hand as preferred attractors of DOM, regardless of whether the objects denote individuals or correspond to definite sets of plural membership.

---

<sup>15</sup> As explicitly set forth in the cited papers of von Heusinger and colleagues, a more accurate description of this evolution has to take the parameter of specificity into account, introducing a division between specific indefinite and non-specific indefinite human objects. The latter continue to license optional marking in modern Spanish (Leonetti 2004, 80; von Heusinger 2008, 5). For a good discussion of specificity effects on Spanish DOM, cf. von Heusinger/Kaiser 2003).

The attraction of plural definite nouns to the domain of DOM evidences an increase in the schematic value of *a*, to the extent that the feature of salience associated with individuated entities is growing more obscure and is giving way to a less restrictive condition stipulating that the referent solely be identifiable to the addressee. Within the framework of a markedness approach to DOM – objects resembling subjects have marked properties which motivate iconic patterns of marked encoding (Bossong 1991; Aissen 2003), the changes we have analyzed appear to have evolved on a par with a progressively wider conception of what it means to be “subject-like”, as reflected in the outlined path of grammaticalization: CLAUSAL TOPIC > PROMINENT DISCOURSE PARTICIPANT > UNIQUELY SALIENT INDIVIDUAL > IDENTIFIABLE REFERENT. In each step, as we notice, the marking value of *a* is expanded to embrace successively larger sets of properties and is thereby brought closer and closer to profiling a more general type of entity, capable of filling the grammatical slot of subject in accordance with speakers’ expectations.

In present-day Spanish, the category of objects viewed as subject-like now comprises all human beings, as a result of the downward movement of DOM along the definiteness scale. By contrast, the lower-ranked objects on the animacy scale – non-human animate and inanimate objects – do not seem to have participated in the diachronic expansion of *a*. Optional uses of marking are registered throughout the centuries, but they remain marginal, with no sign of perceptible growth. Of the scarcely documented non-human animate objects nothing significant can be stated (for some discussion, cf. García García 2018, 217–218). Inanimate entities, on the other hand, frequent in all textual sources, yield a clear panorama of uncommon marking. The suggestion (Company Company 2002a; 2002b) that DOM has begun to extend to inanimate objects does not appear to encounter much support in empirical data (cf. García García 2018 and references therein; cf. also von Heusinger/Kaiser 2005).

The distinction between human and non-human objects characterizing Spanish DOM in its present state determines case marking splits in other languages (Comrie 1989, 195). Drawing on the markedness theory, it may be sufficient to observe that in prototypical transitive clauses the subject is high in animacy and the object is lower in animacy (Comrie 1989, 128) in order to explain why DOM prefers human (subject-like) referents. But in the case of Spanish, matters are not straightforward, because definiteness was crucially involved in the evolution of *a* and could have affected inanimate objects as well. As mentioned in the Introduction, indeed, animacy and definiteness are independent dimensions that need not work hand in hand. The fact that in Spanish they did so may be attributed to a couple of motives.

One of these points to a phenomenon of persistence (Hopper 1991), according to which features of the lexical source of a grammaticalizing morpheme may

continue to influence and restrict the grammatical distribution of the item at later stages of its development. As to *a*, persistence is detected in the binding relationship between the object marker and human entities which derives from the original use of topicalizing *a(d)* with human-referring personal pronouns.

Secondly, the differential treatment of human beings has to be situated within the context of the language system as a whole. Spanish syntax is highly sensitive to semantic distinctions (Melis/Flores 2013) and exhibits a whole range of phenomena that are similarly regulated by a human vs. non-human opposition. The division between persons and things permeates Spanish grammar (Narbona Jiménez 1989, 106–107) to such an extent that it has come to be regarded as an essential and defining property of the language (Lapesa 1968).<sup>16</sup>

Before I close this paper, I have to emphasize that Spanish DOM has functioned as a dynamic and complex system of split case alternations throughout its history, fundamentally dependent on properties of the direct object, but simultaneously influenced by additional parameters in a secondary way (for an overview, cf. Fábregas 2013). Among these, the effect of verbal semantics on the patterns of *a*-marking has attracted special attention (García García 2018).<sup>17</sup> In particular, when the verbal parameter is explored, the focus comes to be placed on the involvement of the object in the designated event. Role properties, however, are approached from two distinct, if not opposite, vantage points.<sup>18</sup>

On the one hand, DOM is assumed to mark highly affected direct objects. This view relates to the so-called “indexing” or “characterizing” function of case marking (Malchukov/de Swart 2009; Siewierska/Bakker 2009). Case forms of this nature encode semantic roles and are normally associated with oblique arguments. When the indexing function is extended to DOM, with an appeal to Hopper/Thompson’s (1980) model of transitivity, it is argued that DOM has a preference for objects which conform to the canon of high transitivity, the property of high affectedness being what defines the semantic role of such objects (Næss 2004).<sup>19</sup> The relevance of affectedness for *a*-marking in Spanish has been

---

**16** It is worth pointing out that modern Spanish exhibits a phenomenon of Differential Goal Marking (Kittilä 2008), associated with the use of a whole series of special markers serving to contrast human with non-human landmarks (Melis/Rodríguez Cortés 2017).

**17** The connections between DOM and verbal semantics are motivating a growing body of cross-linguistic research (Malchukov/de Hoop 2011; Lemmolo/Klumpp 2014; Witzlack-Makarevich/Seržant 2018).

**18** But cf. de Swart (2006) for an attempt to reconcile the two approaches in terms of a principle of “minimal semantic distinctness”.

**19** Under this proposal, the feature of high affectedness is expected to correlate with definiteness, to the extent that “[a]n action can be more effectively transferred to a patient which is individuated than to one which is not” (Hopper/Thompson 1980, 253; cf. Næss 2004, 1191), and

noted in various studies (Torrego Salcedo 1999; cf. García García 2018 for more references) and has been the topic of fine grained diachronic analyses (von Heusinger/Kaiser 2007; 2011; von Heusinger 2008), which show, via correlations established between degrees of affectedness and particular verbal classes, that the role dimension has a certain impact on the uses of DOM, subordinated to the workings of the definiteness scale.

On the other hand, the role features of the direct object are inspected through the lens of the distinguishing or discriminatory view of case marking (Malchukov/de Swart 2009; Siewierska/Bakker 2009). Here the assumption is that DOM selects objects whose properties resemble those of subjects and sets them apart from canonical patients, for example, when the direct object referent – a human being and sometimes an inanimate entity – demonstrates a certain level of agent-like activity or is conceived of as a relatively autonomous participant (Hatcher 1942; Weissenrieder 1985; 1991; Delbecque 2002; García García 2007; 2014; 2018; Primus 2012). That is to say, instead of highlighting prototypical transitive objects, as claimed under the former approach, on this view DOM signals atypical, deviant, and in this sense marked objects.

Independently of the theory one chooses to adhere to, it is safe to conjecture that Spanish DOM might be sensitive to event semantics. The influence of role distinctions on Spanish syntax is a pervasive phenomenon (e.g. García-Miguel 2015), and other instantiations of DOM in Spanish, which for reasons of space could not be treated in the present paper, have been argued to depend on evaluations of the role of the object participant in the verbal situation.<sup>20</sup>

## 7 Conclusions

Research on DOM in Spanish has generated a multiplicity of proposals seeking to identify the driving forces behind *a*-marking within a system in which unquestionable regularities coexist with a host of variable choices. From a diachronic perspective, the spread of *a* to an ever-widening range of principally human objects, observed in textual materials, has been described in terms of a gradual

---

further carries an implication of animacy because “effects on human or animate entities are perceived as more dramatic, more significant, than effects on inanimates” (Næss 2004, 1202). From these correlations the explanation of why DOM prefers animate and definite objects follows naturally (Næss 2004, 1203).

<sup>20</sup> I am referring to the phenomena known as *leísmo* (Flores 2002; Flores/Melis 2007) and clitic doubling (Melis 2018).



movement downward along the hierarchy of definiteness, whose outcome was a situation of obligatory marking on nearly all human objects. A notion of topicality is sometimes introduced into the historical panorama, especially when scholars concentrate on the older stages of the grammaticalization process, but it is fair to say that the issue of how topical values gave way to referential properties in determining DOM has not properly been addressed.

The central aim of the present chapter was to gain deeper insight into the early uses of *a*-marking which were instrumental in orienting subsequent developments. We adopted the well-founded hypothesis that the origin of Spanish DOM had to be tied to a topicalizing structure inherited from Latin and we examined the expansion of grammaticalizing *a* toward personal pronouns, then proper names, and finally common nouns of human reference. The small steps involved in this progression were subjected to critical analysis; pinpointing the specific properties the *a*-marker was growing sensitive to while moving from one context to another was our primary concern. This enabled us to trace a sequence of functional changes, grounded in analogical relationships, which led from a topic-marker to a definiteness-marker through mediating features of discourse prominence, salient individuation, and singularity.

The idea defended in this paper is that appealing to a notion of topicality may turn out to be necessary to account for the operation of DOM in some or perhaps many languages. As argued in the literature, pragmatic constraints linked to information structure will have a tendency to weigh more heavily in early phases, giving place to semantic factors later in time. Such a scenario is consistent with what is known about the role communicative goals and subjective perspectives play in shaping grammars.

In the case of Spanish, the strong personal pronouns supplied clear evidence that DOM did not arise as a strategy to resolve syntactic ambiguity. But this fact does not invalidate the view held by many scholars that the basic function of DOM has to do with distinguishing objects. The entities set apart by DOM have special properties which oppose them to the regular exponents of the direct object category and approximate them to participants expected to appear as clausal subjects. The discriminating task of DOM consists in enhancing the marked profile of the items in question. Evidently, languages will vary considerably with respect to how these marked properties are defined. The bundle of characteristics commonly associated with subject arguments – topical, animate, individuated, definite, agentive, autonomous, etc. – guarantees the manifestation of variable DOM patterns within and across languages. Of particular interest in this regard was the opportunity the evolutionary history of *a* gave us to watch the integration of a growing spectrum of features conceived of as subject-like and hence distinguished by DOM.

## Bibliography

- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21 (2003), 435–483.
- Anagnostopoulou, Elena, *Conditions on clitic doubling in Greek*, in: van Riemsdijk, Henk (ed.), *Clitics in the languages of Europe*, vol. 20, Berlin/New York, de Gruyter, 1999, 761–798.
- Bárány, András, *DOM and dative. Morphology or syntax*, Invited talk at SyntaxLab, 26 April, University of Cambridge, Cambridge, 2016.
- Belloro, Valeria, *Spanish clitic doubling. A study of the syntax-pragmatics interface*, PhD Thesis, Buffalo, State University of New York, 2007.
- Bickel, Balthasar/Witzlak-Makarevich, Alena, *Referential scales and case alignment. Reviewing the typological evidence*, in: Malchukov Andrej L./Richards, Marc (edd.), *Scales*, Leipzig, Institut für Linguistik, 2008, 1–37.
- Bickel, Balthasar/Witzlak-Makarevich, Alena/Zakharko, Taras, *Typological evidence against universal effects of referential scales on case alignment*, in: Bornkessel-Schlesewsky, Ina/Malchukov, Andrej L./Richards, Marc D. (edd.), *Scales and hierarchies. A cross-disciplinary perspective*, Berlin/Boston, de Gruyter, 2015, 7–43.
- Bossong, Georg, *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen, Gunter Narr, 1985.
- Bossong, Georg, *Differential Object Marking in Romance and beyond*, in: Wanner, Dieter/Kibbee, Douglas A. (edd.), *New analyses in Romance linguistics. Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, April 7–9, 1988*, Amsterdam/Philadelphia, John Benjamins, 1991, 143–170.
- Bossong, Georg, *Le marquage différentiel de l'objet dans les langues d'Europe*, in: Feuillet, Jack (ed.), *Actance et valence dans les langues de l'Europe*, vol. 20, Berlin/New York, de Gruyter, 1998, 193–258.
- Chafe, Wallace, *Cognitive constraints on information flow*, in: Tomlin, Russell S. (ed.), *Coherence and grounding in discourse*, Amsterdam/Philadelphia, John Benjamins, 1987, 21–52.
- Company Company, Concepción, *El avance diacrónico de la marcación prepositiva en objetos directos inanimados*, in: Bernabé Pajares, Alberto, et al. (edd.), *Presente y futuro de la lingüística en España. La Sociedad de Lingüística, 30 años después. Actas del II Congreso de la Sociedad Española de Lingüística*, vol. 2, Madrid, Consejo Superior de Investigaciones Científicas, 2002, 146–154 (= 2002a).
- Company Company, Concepción, *Grammaticalization and category weakness*, in: Wischer, Ilse/Diewald, Gabriele (edd.), *New reflections on grammaticalization*, Amsterdam/Philadelphia, John Benjamins, 2002, 201–215 (= 2002b).
- Comrie, Bernard, *Definite and animate direct objects. A natural class*, *Linguistica Silesiana* 3 (1979), 13–21.
- Comrie, Bernard, *Language universals and linguistic typology*, Oxford, Blackwell, <sup>2</sup>1989.
- Dalrymple, Mary/Nikolaeva, Irina, *Objects and information structure*, Cambridge, Cambridge University Press, 2011.
- De Smet, Hendrik, *The course of actualization*, *Language* 88:3 (2012), 601–633.
- de Swart, Peter, *Case markedness*, in: Kulikov, Leonid/Malchukov, Andrej/de Swart, Peter (edd.), *Case, valency and transitivity*, Amsterdam/Philadelphia, John Benjamins, 2006, 249–267.

- de Swart, Peter, *Cross-linguistic variation in object marking*, PhD Thesis, Radboud University Nijmegen, LOT Publications, 2007.
- DeLancey, Scott, *An interpretation of split ergativity and related patterns*, *Language* 57:3 (1981), 626–657.
- Delbecque, Nicole, *A construction grammar approach to transitivity in Spanish*, in: Davidse, Kristin/Lamiroy, Béatrice (edd.), *The nominative & accusative and their counterparts*, Amsterdam/Philadelphia, John Benjamins, 2002, 81–130.
- Escandell Vidal, Victoria, *Differential Object Marking and topicality. The case of Balearic Catalan*, *Studies in Language* 33:4 (2009), 832–885.
- Fábregas, Antonio, *Differential Object Marking in Spanish. State of the art*, *Borealis* 2:2 (2013), 1–80.
- Fernández Ramírez, Salvador, *Un proceso lingüístico en marcha*, in: *Presente y futuro de la lengua española. Actas de la Asamblea de Filología del I Congreso de Instituciones Hispánicas*, vol. 2, Madrid, Cultura Hispánica, 1964, 277–285.
- Fischer, Olga, *An inquiry into unidirectionality as a foundational element of grammaticalization. On the role played by analogy and the synchronic grammar system in processes of language change*, *Studies in Language* 37 (2013), 515–533.
- Flores, Marcela, *Leísmo, laísmo, loísmo. Sus orígenes y evolución*, Mexico, UNAM, 2002.
- Flores, Marcela/Melis, Chantal, *El leísmo desde la perspectiva del “marcado diferencial del objeto”*, *Revista de Historia de la Lengua Española* 2 (2007), 83–107.
- Folgar, Carlos, *Diacronía de los objetos directo e indirecto (del latín al castellano medieval)*, Santiago de Compostela, Universidade de Santiago de Compostela, 1993.
- García García, Marco, *Differential Object Marking with inanimate objects*, in: Kaiser, Georg A./Leonetti, Manuel (edd.), *Proceedings of the workshop “Definiteness, Specificity and Animacy in Ibero-Romance Languages”*, Arbeitspapier 122, Konstanz, Fachbereich Sprachwissenschaft der Universität Konstanz, 2007, 63–84.
- García García, Marco, *Differentielle Objektmarkierung bei unbelebten Objekten im Spanischen*, Berlin/Boston, de Gruyter, 2014.
- García García, Marco, *Nominal and verbal parameters in the diachrony of Differential Object Marking in Spanish*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 209–242.
- García-Miguel, José María, *Variable coding and object alignment in Spanish. Some corpus-based evidence*, *Folia Linguistica* 49 (2015), 205–256.
- García, Érica/van Putte, Florimon, *Forms are silver, nothing is gold*, *Folia Linguistica Historica* 8:1 (1987), 365–384.
- Givón, Talmy, *Introduction*, in: Givón, Talmy (ed.), *Topic continuity in discourse. A quantitative cross-language study*, Amsterdam/Philadelphia, John Benjamins, 1983, 5–41.
- Harris, Martin, *The evolution of French syntax. A comparative approach*, London/New York, Longman, 1978.
- Hatcher, Anna G., *The use of “a” as a designation of the personal accusative in Spanish*, *Modern Language Notes* 57 (1942), 421–429.
- Hopper, Paul J./Thompson, Sandra A., *Transitivity in grammar and discourse*, *Language* 56:2 (1980), 251–299.
- Hopper, Paul J., *On some principles of grammaticization*, in: Traugott, Elizabeth C./Heine, Bernd (edd.), *Approaches to grammaticalization*, vol. 1, Amsterdam/Philadelphia, John Benjamins, 1991, 17–35.

- Iemmolo, Giorgio, *Topicality and Differential Object Marking. Evidence from Romance and beyond*, *Studies in Language* 34 (2010), 239–272.
- Iemmolo, Giorgio/Klumpp, Gerson, *Introduction to the special issue “Differential Object Marking. Theoretical and empirical issues”*, *Linguistics* 52 (2014), 271–279.
- Kittilä, Seppo, *Animacy effects on differential goal marking*, *Linguistic Typology* 12 (2008), 245–268.
- Kliffner, Michael D., *Personal “a”, kinesic and individuation*, in: Baldi, Philip (ed.), *Papers from the XII Linguistic Symposium on Romance Languages*, Amsterdam/Philadelphia, John Benjamins, 1984, 195–216.
- Laca, Brenda, *Sobre el uso del acusativo preposicional en español*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 61–91.
- Laca, Brenda, *El objeto directo. La marcación preposicional*, in: Company Company, Concepción (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, vol. 1, México DF, UNAM/Fondo de Cultura Económica, 2006, 423–475.
- Lambrech, Knud, *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*, Cambridge, Cambridge University Press, 1994.
- Lapesa, Rafael, *Evolución sintáctica y forma lingüística interior en español*, in: Quilis, Antonio/Carril Blanco, Ramón/Cantarero, Margarita (edd.), *Actas del XI Congreso Internacional de Lingüística y Filología Románicas (Madrid, 1–9 septiembre 1965)*, vol. 1, Madrid, Revista de Filología Española, 1968, 131–150.
- Lazard, Gilbert, *Actancy*, Berlin/New York, de Gruyter, 1998.
- Leonetti, Manuel, *Specificity and Differential Object Marking in Spanish*, *Catalan Journal of Linguistics* 3:1 (2004), 75–114.
- Lichtenberk, Frantisek, *On the gradualness of grammaticalization*, in: Traugott, Elizabeth C./Heine, Bernd (edd.), *Approaches to grammaticalization*, vol. 1, Amsterdam/Philadelphia, John Benjamins, 1991, 37–80.
- Luján, Marta, *Expresión y omisión del pronombre personal*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española*, vol. 1, Madrid, Espasa Calpe, 1999, 1275–1315.
- Malchukov, Andrej, *Animacy and asymmetries in differential case marking*, *Lingua* 118 (2008), 203–221.
- Malchukov, Andrej/de Swart, Peter, *Differential case marking and actancy variations*, in: Malchukov, Andrej/Spencer, Andrew (edd.), *The Oxford handbook of case*, Oxford, Oxford University Press, 2009, 339–355.
- Malchukov, Andrej/de Hoop, Helen, *Tense, aspect, and mood based differential case marking*, *Lingua* 121 (2011), 35–47.
- Meier, Harri, *Sobre as origens do acusativo preposicional nas línguas românicas*, in: Meier, Harri (ed.), *Ensaio de filologia românica*, Lisboa, Edição da “Revista de Portugal”, 1948, 115–164.
- Melis, Chantal, *El objeto directo personal en “El Cantar de Mio Cid”. Estudio sintáctico-pragmático*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 133–163.
- Melis, Chantal/Flores, Marcela, *On the historical expansion of non-canonically marked “subjects” in Spanish*, in: Seržant, Ilja A./Kulikov, Leonid (edd.), *The diachronic typology of non-canonical subjects*, Amsterdam/Philadelphia, John Benjamins, 2013, 163–184.
- Melis, Chantal/Rodríguez Cortés, Diego, *El marcado diferencial de la función “meta” en español*, *Verba* 44 (2017), 195–230.

- Melis, Chantal, *Spanish indexing DOM, topicality, and the case hierarchy*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 97–128.
- Menééndez Pidal, Ramón, *Cantar de mio Cid*, vol. 1, Madrid, Espasa Calpe, 1964.
- Monedero Carrillo de Albornoz, Carmen, *El objeto directo preposicional y la estilística épica (Nombres geográficos en el “Cantar de Mio Cid”)*, *Verba* 5 (1978), 259–303.
- Montaner, Alberto (ed.), *Cantar de mio Cid. Estudio preliminar de Francisco Rico*, Barcelona, Crítica, 2<sup>a</sup>1993.
- Müller, Bodo, *Das morphemmarkierte Satzobjekt der romanischen Sprachen (Der sogenannte präpositionale Akkusativ)*, *Zeitschrift für romanische Philologie* 87 (1971), 477–519.
- Næss, Åshild, *What markedness marks. The markedness problem with direct objects*, *Lingua* 114 (2004), 1186–1212.
- Narbona Jiménez, Antonio, *Sintaxis española. Nuevos y viejos enfoques*, Barcelona, Ariel, 1989.
- Nikolaeva, Irina, *Secondary topic as a relation in information structure*, *Linguistics* 39 (2001), 1–49.
- Pensado, Carmen, *La creación del complemento directo preposicional y la flexión de los pronombres personales en las lenguas románicas*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 179–233.
- Primus, Beatrice, *Animacy, generalized semantic roles, and Differential Object Marking*, in: Lamers, Monique/de Swart, Peter (edd.), *Case, word order, and prominence. Interacting cues in language production and comprehension*, Dordrecht, Springer, 2012, 65–90.
- Ramsden, Herbert, *The use of “a” + personal pronoun in Old Spanish*, *Bulletin of Hispanic Studies* 38 (1961), 42–54.
- Rohlf, Gerhard, *Autour de l'accusatif prépositionnel dans les langues romanes*, *Revue de Linguistique Romane* 35 (1971), 312–334.
- Seržant, Ilja A., *Weak universal forces. The discriminatory function of case in Differential Object Marking systems*, in: Schmidtke-Bode, Karsten, et al. (edd.), *Explanation in typology. Diachronic sources, functional motivations and the nature of evidence*, Berlin, Language Science Press, 2019, 149–178.
- Siewierska, Anna, *Person*, Cambridge, Cambridge University Press, 2004.
- Siewierska, Anna/Bakker, Dik, *Case and alternative strategies. Word order and agreement marking*, in: Malchukov, Andrej/Spencer, Andrew (edd.), *The Oxford handbook of case*, Oxford, Oxford University Press, 2009, 290–303.
- Sinnemäki, Kaius, *A typological perspective on Differential Object Marking*, *Linguistics* 52 (2014), 281–313.
- Torrego, Esther, *El complemento directo preposicional*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española. Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*, vol. 2, Madrid, Espasa Calpe, 1999, 1779–1805.
- Traugott, Elizabeth C./Trousdale, Graeme, *Gradience, gradualness and grammaticalization*, in: Traugott, Elizabeth C./Trousdale, Graeme (edd.), *Gradience, gradualness and grammaticalization*, Amsterdam/Philadelphia, John Benjamins, 2010, 19–44.
- von Heusinger, Klaus/Kaiser, Georg A., *The interaction of animacy, definiteness and specificity in Spanish*, in: von Heusinger, Klaus/Kaiser Georg A. (edd.), *Proceedings of the workshop “Semantic and Syntactic Aspects of Specificity in Romance Languages”*, Konstanz, Fachbereich Sprachwissenschaft der Universität Konstanz, 2003, 41–65.

- von Heusinger, Klaus/Kaiser, Georg A., *The evolution of Differential Object Marking in Spanish*, in: von Heusinger, Klaus/Kaiser, Georg A./Stark, Elisabeth (edd.), *Proceedings of the workshop "Specificity and the Evolution/Emergence of Nominal Determination Systems in Romance"*, Konstanz, Fachbereich Sprachwissenschaft der Universität Konstanz, 2005, 33–69.
- von Heusinger, Klaus/Kaiser, Georg A., *Differential Object Marking and the lexical semantics of verbs in Spanish*, in: Kaiser, Georg A./Leonetti, Manuel (edd.), *Proceedings of the workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*, Konstanz, Fachbereich Sprachwissenschaft der Universität Konstanz, 2007, 83–109.
- von Heusinger, Klaus, *Verbal semantics and the diachronic development of DOM in Spanish*, *Probus* 20:1 (2008), 1–31.
- von Heusinger, Klaus/Kaiser, Georg A., *Affectedness and Differential Object Marking in Spanish*, *Morphology* 21:3 (2011), 593–617.
- Weissenrieder, Maureen, *Exceptional uses of the accusative "a"*, *Hispania* 68 (1985), 393–398.
- Weissenrieder, Maureen, *A functional approach to the accusative "a"*, *Hispania* 74 (1991), 146–156.
- Witzlack-Makarevich, Alena/Seržant, Ilja A., *Differential Argument Marking. Patterns of variation*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 1–49.

Alessia Cassarà and Sophie Mürmann

## Role-semantic parameters for DOM in Italian

**Abstract:** Italian is said to be a no-DOM language. Some studies have shown, however, that under certain conditions occurrences of DOM-marking can be found in the colloquial spoken variety (Berretta 1989; 1991; Belletti 2018). In a corpus of spontaneous speech of Italian collected by Berretta 1991, DOM seems to appear with verbs associated with “non-prototypical objects”, such as OBJECT-EXPERIENCER PSYCH-VERBS (e.g. *preoccupare* ‘worry’, *spaventare* ‘frighten’) or INTERACTION VERBS (e.g. *salutare* ‘greet’, *chiamare* ‘call’). While Berretta’s study provides an accurate description of such cases, we investigated possible motivations for the marker to appear. We postulate a hypothesis in terms of role-semantic parameters, based on Dowty’s (1991) proto-role model, assuming that the appearance of the *a*-marker is triggered by the proto-agent properties that specific verb types assign to their objects. In order to assess a possible effect of agentivity on DOM in Italian, we carried out an acceptability judgment task in which verb type and NP type of the object have been manipulated. As our results show, verbs with an agentive object are more likely to be accepted with DOM than objects bearing exclusively patient properties. Such hypothesis seems to hold particularly for OBJECT-EXPERIENCER PSYCH-VERBS, where the proto-agent property sentience is entailed for the direct object rather than only presupposed as in the case of INTERACTION VERBS. At the same time, the interaction between verb semantics, referential and syntactic prominence proves to be very relevant, suggesting that DOM in Italian is far from being grammaticalized. Overall, this study is a further contribution supporting the role-semantic model for DOM.

---

**Acknowledgements:** The present paper has been presented as a talk in the Linguistic Colloquium of the University of Cologne and at the Workshop “Differential Object Marking in Spanish (and beyond) – diachronic change and synchronic variation” at the University of Zurich. We thank all participants for their valuable comments and suggestions. Our thanks also go to Javier Caro Reina and Marco García García for support and helpful discussion during the preparation and evaluation of our study. Furthermore, we would like to thank the SFB 1252 “Prominence in language” at the University of Cologne for support, in particular Maximilian Hörl for his help with the statistical analysis of our data, and two anonymous reviewers for their comments on a previous version of this paper.

---

**Alessia Cassarà**, University of Cologne, e-mail: [alessia.cassara@uni-koeln.de](mailto:alessia.cassara@uni-koeln.de)

**Sophie Mürmann**, University of Cologne, e-mail: [sophie.muermann@uni-koeln.de](mailto:sophie.muermann@uni-koeln.de)

Open Access. © 2021 Alessia Cassarà and Sophie Mürmann, published by De Gruyter. 

This work is licensed under the Creative Commons Attribution 4.0 International License.

<https://doi.org/10.1515/9783110716207-004>

**Keywords:** agentivity, animacy, colloquial Italian, Differential Object Marking (DOM), judgment tasks, proto-roles, referentiality, role semantics

## 1 Introduction

Italian, as well as French, is usually reported to be a Romance language that does not exhibit the phenomenon of Differential Object Marking (DOM) (cf. e.g. Bossong 1998). Unlike Southern Italian varieties, Standard Italian is said to lack morphological marking of prominent, i.e. animate and definite, direct objects. This view seems to be confirmed if we consider the pair of examples in (1):

- (1) a. *Questi argomenti non hanno convinto  $\emptyset$ /*a* Gianni.*  
 these arguments NEG have.PRS.3PL convinced  $\emptyset$ /DOM Gianni  
 ‘These arguments haven’t convinced Gianni.’  
 (Belletti 2018, 446)
- b. *La polizia non ha ferito  $\emptyset$ /*a* Gianni.*  
 the police NEG have.PRS.3SG injured  $\emptyset$ /DOM Gianni  
 ‘The police did not injure Gianni.’

In both (1a) and (1b), *a*-marking of the direct object *Gianni* would be ungrammatical, independently of the type of verb that accompanies it. In (1a), we deal with the OBJECT-EXPERIENCER (OE) PSYCH-VERB *convincere* ‘convince’, whereas (1b) shows the highly transitive verb *ferire* ‘injure’. However, once the object moves from its clause-internal position to the left periphery of the sentence,<sup>1</sup> DOM gets broadly acceptable for speakers of Standard Italian in (2a) but stays ungrammatical in (2b):<sup>2</sup>

<sup>1</sup> We adopt Belletti’s (2018) syntactic analysis of these constructions stating that *a*-marked direct objects in Standard Italian occur in the TopP position within the CP of the sentence. Yet, while Belletti labels these objects *a*-Topics, we refer to them as instances of the more general phenomenon of DOM. It will be stressed below, though, that Belletti’s syntactic analysis cannot capture all instances of *a*-marked direct objects in Italian: especially with OE-psych-verbs and causative constructions, *a*-marked objects seem to predominantly occur clause-internally (as in example 6 below).

<sup>2</sup> Examples (2b), (3b) and (4a–b) are based on the judgment of a native speaker.



- (2) a.  $\emptyset^{\rho}A$  Gianni, questi argomenti non l' hanno  
 $\emptyset/DOM$  Gianni these arguments NEG CL.3SG have.PRS.3PL  
*convinto*  
 convinced  
 'As for Gianni, these arguments have not convinced him.'  
 (Belletti 2018, 445)
- b.  $\emptyset/*A$  Gianni, la polizia non l' ha ferito.  
 $\emptyset/DOM$  Gianni the police NEG CL.3SG have.PRS.3SG injured  
 'As for Gianni/him, the police did not injure him.'

Even more striking, DOM seems to be rather obligatory than optional with the verb *convincere* 'convince' if we change the NP type from a personal name to a first- or second-person pronoun (3a). The verb *ferire* 'injure', on the contrary, is considerably less acceptable with DOM also with a left-dislocated first- or second-person pronoun (3b):

- (3) a.  $*\emptyset/A$  me/te, questi argomenti non m'/t' hanno  
 $\emptyset/DOM$  1SG/2SG these arguments NEG CL.1/2SG have.PRS.3PL  
*convinto*  
 convinced  
 'As for me/you, these arguments have not convinced me/you.'
- b.  $\emptyset^{??}A$  me/te, la polizia non m'/t' ha ferito.  
 $\emptyset/DOM$  1SG/2SG the police NEG CL.1/2SG have.PRS.3SG injured  
 'As for me/you, the police did not injure me/you.'

The reason for the *a*-marker being not completely ruled out in (3b) might be due to the fact that tonic pronouns generally cannot occur in peripheral position in Italian without being *a*-marked. As far as definite human NPs are concerned, we also see a difference of acceptability between the two verbs in question. Whereas DOM is optional with *la ragazza* 'the girl' with *convincere* (4a), it is ungrammatical with *ferire* (4b):

- (4) a.  $\emptyset/A$  (l) la ragazza, questi argomenti non l'  
 $\emptyset/DOM$  the girl these arguments NEG CL.3SG  
 hanno convinta.  
 have.PRS.3PL convinced  
 'As for the girl, these arguments have not convinced her.'

- b.  $\emptyset$ /\*A (l) la ragazza, la polizia non l' ha  
 $\emptyset$ /DOM the girl the police NEG CL.3SG have.PRS.3SG  
*ferita.*  
 injured  
 'As for the girl, the police did not injure her.'

Restrictions of DOM in Italian which seem to be due to the verb type have been already put forward by Berretta (1989; 1991). In a corpus of spontaneous speech, she identifies three groups of verbs or verbal constructions that are attested with DOM in Italian (Berretta 1991, 137–138):<sup>3</sup>

1. OBJECT-EXPERIENCER PSYCH-VERBS: *affascinare* 'fascinate', *attrarre* 'attract', *colpire* 'impress, strike', *confortare* 'comfort', *consolare* 'cheer up', *convincere* 'convince', *deludere* 'disappoint', *disturbare* 'disturb', *divertire* 'entertain', *eccitare* 'excite', *entusiasmare* 'excite', *incantare* 'enchant', *ingannare* 'deceive', *innervosire* 'make nervous', *invitare* 'tempt', *mettere (di buon umore)* 'cheer', *persuadere* 'persuade', *preoccupare* 'worry', *rattristare* 'sadden', *spaventare* 'frighten',
2. CAUSATIVE CONSTRUCTIONS: with *fare* 'make'/*lasciare* 'let' + infinitive: *fare dormire* 'make sleep', *fare ingrassare* 'fatten', *fare morir dal ridere* 'make die laughing', *fare ridere* 'make laugh', *fare piangere* 'make cry', *fare sentir male* 'make feel bad', *lasciare salire* 'pick up', *lasciare stare* 'let be/go', and
3. OTHER VERBS: *accompagnare* 'accompany', *aspettare* 'wait for', *chiamare* 'call', *coccolare* 'cuddle', *conoscere* '(get to) know', *danneggiare* 'damage, harm', *fermare* 'stop', *fregare* 'wipe', *fucilare* 'shoot', *graffiare* 'scratch', *incastare* 'trap', *informare* 'inform', *incolpare* 'blame', *lasciare (in pace)* 'leave sb alone', *mandare* 'send', *mettere (in galera)* 'jail', *pagare* 'pay', *picchiare* 'beat', *portare* 'carry', *prendere* 'take', *proteggere* 'protect', *pungere* 'offend', *ringraziare* 'thank', *rovinare* 'ruin', *salutare* 'greet', *seppellire* 'bury', *sposare* 'marry', *stancare* 'tire sb out', *stendere* 'hang', *svegliare* 'wake sb up', *temere* 'fear', *toccare* 'touch', *trattare (male)* 'misuse', *vedere* 'see'

While the first two groups of verbs are precisely classified as (1) OE-PSYCH-VERBS and (2) CAUSATIVE CONSTRUCTIONS, the third group is not further specified, being simply labelled (3) OTHER VERBS. As we will argue in chapter 3.2, a considerable number of verbs of class (3) overlap with a role-semantically defined category called INTERACTION VERBS by Blume (1998). These verbs share the property of not select-

<sup>3</sup> All relevant examples of the corpus are listed in the appendix of her article (Berretta 1991, 143–148).

ing a prototypical patient object but one that bears an equivalent degree of agentivity as the subject.

Importantly, as we have seen in the examples (1)–(4) above, DOM with the listed verbs and verbal constructions faces further semantic and syntactic constraints: its degree of acceptability depends on (i) whether the direct object is expressed by a strong personal pronoun, a personal name or a human definite NP, and it is restricted to (ii) direct objects appearing in (left-)<sup>4</sup> peripheral sentence position but not in canonical SVO word order. Both points require further specifications. As for (i), it has to be added that also indefinite generic NPs can bear DOM. These instances of DOM in Italian have not received attention in the literature so far. An example from Berretta’s corpus is given in (5):

- (5) *Ad un linguista possono colpire particolarmente [...]*  
 DOM a linguist can.PRS.3PL impress particularly  
*frasi del seguente tenore*  
 phrases of the following tenor  
 ‘A linguist may be particularly impressed by [...] phrases of the following tenor.’

(Berretta 1991, 143f.)

Note that the verb *colpire* is used in its OE-PSYCH reading ‘impress, strike’ here. In contrast to the previous examples, it is questionable if we deal with an instance of left-dislocation in (5) or rather with a clause-internal object in OVS word order. Since the sentence lacks a resumptive clitic, we would argue in favour of the latter analysis. This leads us to the syntactic constraint in (ii), which needs to be refined. Contrary to the syntactic analysis given by Belletti (2018), Berretta’s corpus data reveal that for a majority of cases *a*-marked objects of OE-PSYCH-VERBS and CAUSATIVE CONSTRUCTIONS are not dislocated but rather clause-internal direct objects:

- (6) a. OE-PSYCH-VERB  
*A me non convince.*  
 DOM 1SG NEG convince.PRS.3SG  
 ‘I am not convinced.’

<sup>4</sup> In Berretta’s corpus, marked objects appear in the left periphery (or sentence-initially) in 80.4% (74/92) of the cases in contrast to 19.6% (18/92) in the right periphery.

## b. CAUSATIVE CONSTRUCTION

*A me fanno piangere.*

DOM 1SG make.PRS.3PL cry

‘They make me cry.’

(Berretta 1991, 139)

The lack of a resumptive clitic in (6a) and (6b) is a strong indicator that we deal with a pre-posed direct object in these cases. As remarked by Berretta (1991, 139), *a*-marked objects like in (6) also differ in register from their clitic-left-dislocated counterparts (cf. 7a and 7b). While the former can be characterized as socio-linguistically unmarked, appearing also in more formal registers, the latter are clearly confined to colloquial registers.

## (7) a. OE-PSYCH-VERB

*A me non mi convince.*

DOM 1SG NEG CL.1SG convince.PRS.3SG

‘As for me, I am not convinced.’

## b. CAUSATIVE CONSTRUCTION

*A me mi fanno piangere.*

DOM 1SG CL.1SG make.PRS.3PL cry

‘As for me, they make me cry.’

(Berretta 1991, 139)

A similar point had been made by Benincà (1986, 232), though for OE-PSYCH-VERBS only. She states that DOM with OE-PSYCH-VERBS also appears in the written language where even the lack of *a*-marking is perceived as ungrammatical. With other verbs, such as *invitare* ‘invite’ in contrast, DOM is limited to colloquial contexts. In these cases, the object is always dislocated as in (8a), a variant with the object occurring sentence-initially like in (8b) would not be of a higher register but simply pragmatically odd (cf. Berretta 1991, 139).

(8) a. *A me non mi hanno invitato.*

DOM 1SG NEG CL.1SG have.PRS.3PL invited

‘As for me, they did not invite me.’

(Benincà 1986, 232)

b. *#A me non hanno invitato.*

DOM 1SG NEG have.PRS.3PL invited

‘They did not invite me.’

The particular status of OE-PSYCH-VERBS and CAUSATIVE CONSTRUCTIONS is borne out by the distribution of the resumptive clitic among the three verbs classes differentiated by Berretta: while OE-PSYCH-VERBS (75%, 30/40 occurrences) and CAUSATIVE CONSTRUCTIONS (62.5%, 5/8) are for the most part attested without a resumptive clitic,<sup>5</sup> the third class of OTHER VERBS overwhelmingly occurs with a clitic (present in 86.4% of the occurrences, 38/44, cf. Berretta 1991, 139). Another closely related tendency is that the former two classes strongly prefer the left position, while the latter also occur in contexts with right-dislocated marked objects. In these cases, the resumptive clitic is nearly always present, as in *non t'ho visto a te* 'I have not seen, DOM you' (cf. Berretta 1991, 128–132). For our purposes, both subtypes of the syntactic constraint (ii) can be reconciled: what both pre-posed and dislocated *a*-marked objects have in common is a deviation from canonical SVO word order with the direct object occurring in a non-prototypical syntactic position. Since only *a*-marked dislocated direct objects occur across verb classes, most typically as Clitic Left Dislocations (CLLDs), we will focus on this phenomenon which is arguably confined to colloquial registers of Italian. The aim of the present paper is thus to obtain a more detailed picture on the acceptability of *a*-marked direct objects in CLLDs by testing the structure with predicate classes showing different role-semantic configurations as well as with different NP types of the object.

Besides Berretta (1989; 1991), several other authors have discussed the occurrences of DOM in colloquial Standard Italian (cf. Benincà 1986; Iemmolo 2010; in preparation; Belletti 2018). Whereas these approaches mainly seek explanations for the phenomenon in terms of information structure and argue in favour of a topic-marking function of DOM in Italian, the present article wants to elaborate on the role-semantic properties that characterize *a*-marked objects. This is of interest for mainly two reasons: first, research on DOM in Spanish has revealed that the consideration of role-semantic factors is fruitful for the investigation of DOM in general offering a more adequate explanation for the phenomenon than purely nominal-based or information-structural approaches (cf. Weissenrieder 1991; García García 2007; 2014; Primus 2012; Kabatek 2016; García García/Primus/

---

5 Certainly, a further analysis would be required in order to separate the cases in which the clause-internal clitic is indeed absent from those in which it is not overtly expressed but present in the structure. As for the latter cases, it must be considered that the lack of an overt clitic might also be attributed to a matter of register since prescriptive grammars of Italian ban clitic-doubling. Thus, overtly expressed clitics are expected to be used in less formal registers whereas they are more likely to be absent in formal settings. However, reconsidering that Berretta's corpus consists of data of spontaneous speech recorded in informal settings, it does not seem satisfactory to reduce the lack of the clitic to a question of register.

Himmelmann 2018). Second, the fact that DOM in Italian, in contrast to Spanish, has not been grammaticalized yet, allows for a fine-grained analysis of semantic, pragmatic and syntactic constraints that have to interact systematically for the marker to occur. These initial triggering factors are difficult to disentangle in Modern Spanish, since the *a*-marking of nearly every human definite direct object blurs the underlying concepts, such as individuation or agentivity.

The paper is structured as follows: Section 2 introduces the theoretical framework of generalized semantic roles employed, exemplified for Spanish DOM. Section 3 offers a role-semantic analysis of two DOM-sensitive verb classes in Italian, namely OE-PSYCH-VERBS and INTERACTION VERBS. Likewise, we will sketch additional constraints on DOM which concern NP type and syntactic position. Section 4 presents our online acceptability judgment task through which we tested the effect of role semantics in interaction with NP type for left-dislocated objects in colloquial Standard Italian. Section 5 draws the conclusions arguing that among the factors favouring DOM in Italian, agentivity should be treated on par with referential prominence and topicality.

## 2 DOM and generalized semantic roles

In order to examine the influence of role-semantic factors for Italian DOM, we shall motivate a role-semantic approach on DOM in general in a first step. This can be illustrated by Spanish which is usually said to have an animacy- and definiteness-based DOM system (cf. e.g. Aissen 2003). Such interpretation is confirmed by (9) and (10):

(9) *Conozco*         $\emptyset$ /*\*a*    *esta* *película*.  
 know.PRS.1SG  $\emptyset$ /DOM this film  
 'I know this film.'

(10) *Conozco*        *\* $\emptyset$ /a*    *este* *actor*.  
 know.PRS.1SG  $\emptyset$ /DOM this actor  
 'I know this actor.'

(García García 2007, 63)

Whereas *a*-marking of the inanimate, though definite object NP *esta película* 'this film' in (9) would be ungrammatical, the marker must occur obligatorily with the human definite NP *este actor* 'this actor' in (10). However, the following examples

challenge a purely animacy- and definiteness-based approach on DOM and, as we will further argue, support a role-semantic analysis.

(11) *Un artículo precede*                    \* $\emptyset/a$     *un sustantivo.*  
 a    article    precede.PRS.3SG     $\emptyset$ /DOM    a    noun  
 ‘An article precedes a noun.’

(12) *La euforia caracteriza*                    \* $\emptyset/a$     *la situación.*  
 the    euphoria    characterize.PRS.3SG     $\emptyset$ /DOM    the    situation  
 ‘Euphoria characterizes the situation.’

(García García 2018, 226)

As examples (11) and (12) illustrate, animacy is not a necessary criterion for DOM to appear in Spanish. In both examples, we find an inanimate object which in (11) must be obligatorily *a*-marked to obtain a grammatical sentence and in (12) is highly preferred with the marker. Instead, the occurrence of DOM in both examples could be explained by role-semantic factors: considering the semantic roles of both subject and object, we see that in (11) and (12) the prototypical AGENT-PATIENT-asymmetry of a transitive sentence is not respected. García García (2007; 2014; 2018) captures this in his generalization of THEMATIC DISTINCTNESS:

(13) Generalization of thematic distinctness:

DOM in Spanish is required with inanimate objects when the subject does not outrank the direct object in terms of agentivity.

(García García 2007, 71; 2014, 145; 2018, 227)

The approach of García García is based on Dowty’s (1991) proto-role model. This model has originally been established to account for lexicalization patterns of predicates, exemplified with English transitive verbs. Later research mainly used Dowty’s model in order to make predictions on argument realization phenomena, i.e. to explain the mapping from lexical semantics to syntax as well as morphosyntactic linking. We will shortly introduce the main ideas of the model before motivating its applicability to differential morphosyntactic argument realization. Dowty elaborates the concept of two generalized semantic roles, one proto-agent and one proto-patient. Each of these two cluster roles consists of five different properties listed below:

(14) AGENT PROTO-ROLE:

- a. volitional involvement in the event or state
- b. sentience (and/or perception)

- c. causing an event or change of state in another participant
- d. movement (relative to the position of another participant)
- (e. exists independently of the event named by the verb)

(Dowty 1991, 572)

## (15) PATIENT PROTO-ROLE:

- a. undergoes change of state
- b. incremental theme
- c. causally affected by another participant
- d. stationary relative to movement of another participant
- (e. does not exist independently of the event, or not at all)

(Dowty 1991, 572)

Based on these proto-properties, which are understood as verbal entailments in the strictly logical sense, Dowty formulates an argument selection principle and two corollaries that predict how subject and direct object selection can be deduced from the accumulation of proto-agent and proto-patient properties of a predicate. It is predicted that the argument bearing a higher number of proto-agent properties is lexicalized as the subject and the argument having a higher number of proto-patient properties is lexicalized as the direct object (cf. Dowty 1991, 576). Thus, the model correctly predicts that in a sentence like *Peter ate an apple* or *Peter wrote a letter*, *Peter* bearing all the given proto-agent properties is realized as the subject and *an apple/a letter* having all the indicated proto-patient properties is realized as the direct object, respectively. One great advantage of the model, which is essential for the present approach, is the possible assignment of proto-agent properties to both arguments as well as the possible combination of proto-agent and proto-patient properties for each argument. While Dowty's model is restricted to English and does not provide corollaries for morphosyntactic argument realization, it has been successfully adopted and refined in later work to make generalizations about universal preferences in morphosyntactic case selection (cf. Blume 1998; 2000; Primus 1999a; 1999b; 2006; Ackerman/Moore 2001).

Coming back to the systematic cases of *a*-marked inanimate objects in Spanish, we can now elaborate on García García's (2007, 71; 2014, 145; 2018, 227) relational understanding of agentivity, as stated in the generalization in (13), more precisely. Taking into account the distribution of proto-role entailments of a transitive predicate, two scenarios could lead to a lack of thematic distinctness, that is either



- (i) when both subject and object bear the same number of proto-agent properties (e.g. REVERSIBLE-CONVERSE PREDICATES, such as *preceder* ‘precede’, *seguir* ‘follow’ or REVERSIBLE-SYMMETRICAL PREDICATES, such as *igualar* ‘equal’)<sup>6</sup> or
- (ii) when the object has more proto-agent properties than the subject (e.g. VERBS OF ATTRIBUTION, such as *caracterizar* ‘characterize’ or *definir* ‘define’)

The configuration in (i) is reflected in examples such as *Los días siguen \*ø/a las noches* ‘The days follow the nights’ (or in 11 above) where we find a REVERSIBLE PREDICATE with two participants bearing each the proto-agent property of independent existence. This needs further specification: as Dowty (1991, 572) himself mentions, the status of independent existence as a proto-role entailment is unclear since the term covers various dimensions. On the one hand, it expresses a *de re* (vs. *de dicto*) reading of the noun phrase in question, given e.g. for the subject but not for the object in *John needs a new car*. Note that this dimension concerns rather the semantic domain of specificity than the semantics of the verb. On the other hand, the criterion expresses that a referent “is not brought into being or destroyed by the event named by the verb but is presumed to exist before and after the event” (Dowty 1991, 573). Independent existence is implied by all other proto-agent properties (9a–d). That it can be interpreted as a core criterion for the causal relation between the arguments has been shown by Primus’ (1999a; 1999b; 2006) modified version of Dowty’s proto-role model. Here, we can only outline her innovation in a simplified way: in her model, Primus defines the proto-patient role by its dependency from the proto-agent role. This co-argument dependency relation is based on a very broad understanding of causality. The degree of involvement of the proto-patient is thus dependent from the involvement of the proto-agent, characterized by a set of proto-properties which is comparable to the ones presented above in (14) and (15).<sup>7</sup> Dealing with REVERSIBLE PREDICATES, such as the ones mentioned in (i), none of the proto-agent (14a–d) and proto-patient properties (15a–d) are assigned to the arguments, so that there is no involvement dimension between them. That means that a co-argument dependency relation between subject and direct object cannot be established. What follows from

---

<sup>6</sup> While for REVERSIBLE-SYMMETRICAL PREDICATES, subject and object are interchangeable without imposing a change in truth-conditions, in case of REVERSIBLE-CONVERSE PREDICATES, the predication can be reversed through a lexical doublet (cf. García García 2014, 147–170).

<sup>7</sup> The main modification in comparison to Dowty’s (1991, 572) set of entailments is that Primus’ proto-properties are defined as primitive predicates and that for each proto-agent predicate a converse proto-patient predicate can be derived. Her list of primitive predicates includes the following notions: CONTROLLER VS. CONTROLLED, CAUSER VS. CAUSALLY AFFECTED, MOVER VS. MOVED, EXPERIENCER VS. EXPERIENCED, and POSSESSOR VS. POSSESSED (cf. Primus 1999b, 141; 2012, 73).

that, is that both arguments qualify as weak proto-agents (cf. Primus 2006, 56–59; García García 2014, 144–149). So, in the configuration in (i), subject and direct object are both (weak) proto-agents and are thus not distinguishable in terms of their semantic roles.

The configuration in (ii) involves an even more dramatic deviation from the thematic distinctness we would expect in a prototypical transitive sentence: in sentences such as *La seriedad caracteriza \*ø/a su atuendo* ‘Seriousness characterizes his outfit’ (or in 12 above), the subject merely denotes a property of the object and does not have argument status (cf. García García 2014, 171–172), whereas the object can be qualified as agentive since it bears the proto-agent property of independent existence.<sup>8</sup> The role-semantic explanation to DOM with inanimate objects has also been proven promising for DOM with animate objects (Primus 2012). Whereas the former case shows an *actual* need for disambiguation of subject and direct object, it has been argued that in the case of animate objects it is rather their *potential* agentivity in the given event that could blur the distinctness of the two arguments.

To sum up, we have seen how a model of generalized semantic roles can be used to account for examples of DOM in Spanish with inanimate objects which at first sight contradict traditional – animacy-based – approaches. Putting it more precisely, DOM in these cases seems to be triggered by a lack of thematic distinctness between subject and direct object. In order to see whether a similar hypothesis can be developed for the instances of DOM in colloquial Standard Italian, we will continue by a role-semantic analysis of two relevant verb classes as a next step.

---

<sup>8</sup> As has been pointed out by a reviewer, the configuration in (ii) *does* imply distinctness of subject and direct object. While this holds true, the crucial point is that we deal with a sharp deviation from the canonical distinctness of arguments in a transitive sentence, where the subject outranks the object in terms of agentivity. In the case of VERBS OF ATTRIBUTION, the deviation from the prototypical AGENT-PATIENT-asymmetry is even more remarkable since the subjects do not even have argument status. Normally, in such case, linking theories would predict the agentive object to be realized as the subject. As argued by García García (2014, 176–177), the opposite linking pattern in Spanish can be accounted for by lexical economy: ATTRIBUTION VERBS like *caracterizar* ‘characterize’ also have another – more frequent – reading in which the subject outranks the direct object in terms of agentivity (e.g. *Ana ha caracterizado la situación* ‘Ana characterized the situation’, cf. García García 2014, 177). Hence, for reasons of economy, the valency frame is not changed in the attributive – less usual – reading and the non-prototypicality of the object is indicated by DOM instead.

### 3 A role-semantic account on DOM in Italian

The present Section puts forward the argument that two verb classes that are attested with DOM in Italian, namely OE-PSYCH-VERBS (3.1) and INTERACTION VERBS (3.2),<sup>9</sup> can be grasped within a role-semantic account using the framework introduced in the previous Section. Our theoretical part is completed by a sketch of two additional constraints for DOM in Italian, i.e. NP type and syntactic position of the object, in Section 3.3.

The following examples (16) and (17) illustrate the requirement of the *a*-marker with OE-PSYCH-VERBS and INTERACTION VERBS if the object is a tonic pronoun and situated in the left periphery of the sentence. Note that although the label INTERACTION VERB has not been employed by Berretta (1991), we will prove its adequacy in Section 3.2 by role-semantic criteria that allow for a systematic classification of a number of verbs of her group of OTHER VERBS under this notion.

(16) OE-PSYCH-VERB

\* $\emptyset$ /A *me preoccupa Torino: è una città difficile.*  
 DOM 1SG worry.PRS.3SG Torino be.PRS.3SG a city difficult  
 ‘Torino worries me: it is a difficult city.’

(Berretta 1991, 147)

(17) INTERACTION VERB

\* $\emptyset$ /A *loro le aspettava Adone in doppio petto blu.*  
 DOM 3PL CL.3PL wait.for.PST.3SG Adone in double-breasted blue  
 ‘Adone waited for them in a blue double-breasted (suit).’

(Berretta 1991, 143)

In both (16) and (17), the omission of DOM would turn the sentence ungrammatical. For HIGHLY TRANSITIVE VERBS such as *ferire* ‘injure’, in contrast, the acceptability of the marker considerably decreases:

(18) HIGHLY TRANSITIVE VERB

$\emptyset$ /\*A *me mi ha ferito la polizia due anni*  
 DOM 1SG CL.1SG have.PRS.3SG injured the police two years  
*fa*  
 fare.PRS.3SG  
 ‘The police injured me two years ago.’

<sup>9</sup> CAUSATIVE CONSTRUCTIONS which, as identified by Berretta (1991), also show a high preference for DOM in Italian are not considered in the present work.

In what follows, we will elaborate on the role-semantic properties of the two DOM-sensitive verb classes of OE-PSYCH-VERBS and INTERACTION VERBS and point out what distinguishes them from HIGHLY TRANSITIVE VERBS.

### 3.1 OE-PSYCH-VERBS

Transitive OE-PSYCH-VERBS are a construction type of psychological predicates in which the STIMULUS is encoded as the subject and the EXPERIENCER as the direct object (e.g. *annoy, frighten, impress*). They can be contrasted with transitive SUBJECT- EXPERIENCER (SE)-PSYCH-VERBS which exhibit the reverse linking pattern taking an EXPERIENCER subject and a STIMULUS object (e.g. *adore, love, hate*) (cf. e.g. Verhoeven 2014, 130). As argued by Kutscher (2009, 27–40) and many others (e.g. Croft 1993; cf. Kailuweit 2005; 2015 for Romance languages), against Dowty (1991, 580), OE-PSYCH-VERBS are an aspectually heterogeneous class. Furthermore, they vary in their causal structure. A distinction is usually made between causative (or agentive) and non-causative (or non-agentive) OE-PSYCH-VERBS. It is language-specific if non-causative PSYCH-VERBS form a separate lexical class or if a language only has one lexical class of  $\pm$  causative PSYCH-VERBS (cf. Verhoeven 2014, 131–132). Italian, as well as the other Romance languages, has both classes as illustrated in (19) vs. (20):

(19) a. NON-CAUSATIVE

*Maria ha affascinato Pietro (\*con intenzione).*  
 Maria have.PRS.3SG fascinated Pietro (intentionally)  
 ‘Maria fascinated Pietro (\*intentionally).’

b. *Le domande di Maria hanno affascinato Pietro*  
 The questions of Maria have.PRS.3PL fascinated Pietro  
 (\*con intenzione).  
 (intentionally)  
 ‘Maria’s questions fascinated Pietro (\*intentionally).’

(20) a.  $\pm$  CAUSATIVE

*Maria ha disturbato Pietro (con intenzione).*  
 Maria have.PRS.3SG disturbed Pietro (intentionally)  
 ‘Maria disturbed Pietro (intentionally).’

- b. *Le domande di Maria hanno disturbato Pietro*  
 The questions of Maria have.PRS.3PL disturbed Pietro  
 (\*con intenzione).  
 (intentionally)  
 ‘Maria’s questions disturbed Pietro (\*intentionally).’

The non-causative PSYCH-VERB *affascinare* ‘fascinate’ cannot obtain an agentive, i.e. volitional reading, regardless of whether the subject is animate (19a) or inanimate (19b). The causative PSYCH-VERB *disturbare* ‘disturb’, conversely, can be interpreted as agentive with an animate subject (20a) but stays non-agentive with an inanimate subject (20b). Hence, the causative reading can only arise with an animate (or better human) subject which is typically interpreted to act volitionally in the event.

Table 1 shows how OE-PSYCH-VERBS can be analyzed in terms of Dowty’s proto-role entailments.

**Table 1:** Distribution of proto-properties for OE-PSYCH-VERBS.

Verb class	Subject P-agent	Direct Object	
		P-agent	P-patient
OE-PSYCH			
(A) NON-CAUSATIVE (e.g. <i>affascinare</i> ‘fascinate’, <i>deludere</i> ‘disappoint’, <i>incantare</i> ‘enchant’)	+ independent ex.	+ sentience + independent ex.	± change of state
(B) ± CAUSATIVE (e.g. <i>disturbare</i> ‘disturb’, <i>divertire</i> ‘entertain’, <i>spaventare</i> ‘frighten’)	± volition ± sentience ± causation + independent ex.	+ sentience + independent ex.	+ causally affected ± change of state

As for non-causative OE-PSYCH-VERBS and ± CAUSATIVE VERBS in their non-causative reading, the subject is solely assigned the proto-agent property independent existence, while the direct object bears sentience and independent existence. In addition, the object can undergo a change of state and hence also exhibit a proto-patient property. In their causative reading, the subject bears the proto-agent properties of causation and independent existence. Likewise, the object has two proto-agent properties, namely sentience and independent existence. As a proto-patient property, the object entails causally affected and, optionally, also change of state. Moreover, as seen above, human subjects can be interpreted as volitionally acting participants (and since volition implies sentience, also

as sentient participants).<sup>10</sup> So, while for non-causative PSYCH-VERBS the object outranks the subject in terms of agentivity, in case of causative PSYCH-VERBS the subject is more agentive than the object. However, there is one criterion which distinguishes causative OE-PSYCH-VERBS from prototypical TRANSITIVE VERBS, namely the object bearing the proto-agent property of sentience. Sentience is defined by Dowty (1991, 573) in the following way: “Sentience means more than a presupposition that an argument is a sentient being; it is rather sentience with respect to the event or state denoted by the verb”. This proto-agent property overlaps with one or two proto-patient properties for the object. Hence, one could argue that due to this role overlap, also for causative OE-PSYCH-VERBS a clear-cut co-argument dependency is blurred (cf. Primus 2012, 73).

In (21) and (22), we list again the OE-PSYCH-VERBS attested with DOM by Berretta (1991, 137f.), now divided into NON-CAUSATIVE and ± CAUSATIVE VERBS:

- (21) NON-CAUSATIVE: *affascinare* ‘fascinate’, *attrarre* ‘attract’, *colpire* ‘strike’, *confortare* ‘comfort’, *consolare* ‘cheer up’, *convincere* ‘convince’, *deludere* ‘disappoint’, *entusiasmare* ‘excite’, *incantare* ‘enchant’, *ingannare* ‘deceive’, *persuadere* ‘persuade’, *preoccupare* ‘worry’
- (22) ± CAUSATIVE: *disturbare* ‘disturb’, *divertire* ‘entertain’, *eccitare* ‘excite’, *innervosire* ‘make nervous’, *invitare* ‘tempt’, *mettere (di buon umore)* ‘cheer’, *rattristare* ‘sadden’, *spaventare* ‘frighten’

Benincà (1986, 239, fn. 14) makes an interesting remark with regard to the above-made distinction suggesting that the reading of an OE-PSYCH-VERB (causative or non-causative) affects the acceptability of DOM in Italian. She takes the ambiguous sentence *Giorgio non mi ha convinto* ‘Giorgio did not convince me’, which can have the following two interpretations depending on the subject’s agentivity:

- (i) *Il comportamento di Giorgio non è risultato per me convincente.*  
 ‘Giorgio’s behaviour was not convincing to me.’ [- CAUSATIVE]
- (ii) *Giorgio non è riuscito a convincermi (a prestargli un milione).*  
 ‘Giorgio couldn’t convince me (to lend him a million).’ [+ CAUSATIVE]

If in such context, the sentence *A me, Giorgio non mi ha convinto* ‘DOM me, Giorgio did not convince me’ was uttered, it would be generally accepted and unmarked with respect to register with the reading in (i), while it would be con-

<sup>10</sup> Strictly speaking, volition and sentience are conveyed in these cases via conversational implicature rather than via lexical entailment (cf. Primus 1999a, 51).

finned to colloquial registers with the reading in (ii). If this holds true, it might indicate that the degree of thematic distinctness between subject and object has an impact on the general acceptability of DOM in Italian. Note, however, that the assumption that the *a*-marked CLLD structure above can be interpreted as either sociolinguistically unmarked or marked is somewhat contradictory to the claim of Berretta (1991, 139), who ascribes a general colloquial flavour to structures with *a*-marked in CLLDs.

It must be noted that Italian has another construction type of non-causative PSYCH-VERBS which must be clearly differentiated from the OE-PSYCH-VERBS in question. This type is most prototypically represented by VERBS OF LIKING (cf. Kailuweit 2005), such as *piacere* ‘like’ which select for an indirect object (e.g. *A Gianni piace la pittura* ‘Gianni likes the painting’). The presented examples of *a*-marked objects with OE-PSYCH-VERBS can be disproved to be indirect objects for the following two reasons: first, if the dislocated object is coindexed by a clitic, we always find the direct object clitic (though only visible for the 3<sup>rd</sup> person: *A lui lo preoccupa* ‘It worries him’ vs. colloquial *A lui gli piace* ‘He likes it’). Second, and even more convincing, we cannot have an *a*-marked object of OE-PSYCH-VERBS in canonical word order: *??∅/A lui lo preoccupa la situazione* ‘The situation worries him’ vs. *La situazione preoccupa \*a lui* (correct only: *La situazione lo preoccupa*).<sup>11</sup> When we deal with an indirect object, in contrast, the insertion of *a* in the canonical sentence is possible: *A lui non (gli) piace Gianni* ‘He does not like Gianni’ vs. *Gianni non piace a lui/Gianni non gli piace*. So, we deal indeed with two different structures here and cannot treat the cases of DOM as instances of indirect objects.

### 3.2 INTERACTION VERBS

As will be argued in the following, there is a second semantically defined verb class that shows a preference for DOM in Italian. This class is labelled INTERACTION VERBS by Blume (1998, 254) and comprises (typically) two-place predicates which denote complex events of social interaction. Representative examples are verbs such as ‘help’, ‘greet’ or ‘thank’.

<sup>11</sup> There are instances, though, in which tonic pronouns can appear in canonical object position with transitive verbs, namely when they appear in emphatic contexts (e.g. accompanied by focus particles such as *solamente* ‘only’ or *anche* ‘too’), as in *La situazione preoccupa solamente (\*a) lui* ‘The situation only worries him’ (cf. also Benincà 1986, 231). As expected, DOM would be ungrammatical in these cases. We thank a reviewer for this remark.

In terms of role semantics, the only property that holds for members of this class is the proto-agent property of independent existence which is entailed both for the subject and the direct object. The implication of volition (and sentience) for the subject varies from verb to verb. It is given e.g. for the subject of *chiamare* ‘call’ and *salutare* ‘greet’ but only pragmatically inferred by conversational implicature for *aiutare* ‘help’ which can therefore also be referred to as being “semantically underspecified for volition” (Primus 2012, 85). What is important here is that the object does not bear any proto-patient property. There is thus no co-argument dependency relation between the two arguments, which both bear at least one proto-agent property.

**Table 2:** Distribution of proto-properties for INTERACTION VERBS.

Verb class	Subject P-agent	Direct Object	
		P-agent	P-patient
INTERACTION (e.g. <i>aspettare</i> ‘wait for’, <i>chiamare</i> ‘call’, <i>salutare</i> ‘greet’)	± volition ± sentience ± movement + independent ex.	[± volition] <sub>PRESUPPOSED</sub> [± sentience] <sub>PRESUPPOSED</sub> [± movement] <sub>PRESUPPOSED</sub> + independent ex.	

We say *at least* one proto-agent property since, in addition to the entailed property independent existence, the object of a social interaction event also bears one or more presupposed proto-agent properties (cf. Table 2). Our point of departure for introducing presupposed proto-agent properties is Blume’s (1998) modified version of Dowty’s (1991) proto-role model, which she uses in order to account for the morphosyntactic linking of INTERACTION VERBS cross-linguistically. In her model, proto roles are understood as relations of participants to subevents. She assumes the object of verbs like ‘thank’, ‘answer’ and ‘call’ to bear proto-agent properties in a presupposed subevent. This subevent is temporally prior to the entailed subevent in which the subject participant is acting. To be more precise, we seem to deal with a very general presupposition of sentience for the object argument, which is of the following kind: “y is a sentient being, able to perceive (and react) in the given event” or, even more simple, “y is autonomously active”. As a proper presupposition, and unlike a predicate entailment, it is kept constant under negation: Thus, it would be also true for the object of a sentence like *Peter did not greet Maria* which carries the presupposition that Maria would have been able to perceive Peter’s greeting and react to it. There are some pieces of evidence that presupposed sentience of the object, probably together with the lack of an AGENT-PATIENT-asymmetry, has an impact on morphosyntactic argument reali-



zation. First, as revealed by Blume (1998; 2000), INTERACTION VERBS display the cross-linguistic tendency to select marked case frames, e.g. NOMINATIVE/DATIVE. Second, there are examples like (23) in Spanish where DOM occurs with an inanimate object that can be ascribed presupposed sentience in the event denoted by the predicate:

- (23) *¡Hans, puñeta, llám-a al ascensor!*  
 Hans, damn, call.PRS.3SG DOM.the elevator  
 ‘Hans, damn, call the elevator!’

(García García 2014, 189; García García/Primus/Himmelman 2018, 30)

In this case, the insertion of an inanimate *a*-marked object can be explained as follows: since an elevator is programmed to perceive a certain signal and to react to it, it can be qualified as sentient, and hence agentive, in the given context (cf. García García/Primus/Himmelman 2018, 32). But, unlike for human beings, the sentience of elevators is restricted to the programmed stimulus, that is why they cannot context-independently fill the slots of other predicates entailing or presupposing sentience (e.g. ‘love’, ‘cuddle’, ‘be jealous’), while humans can. Hence, the example in (23) nicely illustrates the relative dimension of presupposed proto-agent properties and its possible impact on DOM, at least for Spanish.

Among Berretta’s (1991, 138) third group of OTHER VERBS, we can identify a dozen of verbs which match the role-semantic criteria shown above and that can thus be subsumed under the class of INTERACTION VERBS:<sup>12</sup>

- (24) INTERACTION VERBS:

*accompagnare* ‘accompany’, *aspettare* ‘wait for’, *chiamare* ‘call’, *coccolare* ‘cuddle’, *informare* ‘inform’, *lasciare (in pace)* ‘leave sb alone’, *mandare* ‘send’, *ringraziare* ‘thank’, *salutare* ‘greet’, *sposare* ‘marry’<sup>13</sup>

<sup>12</sup> Certainly, in the mixed group that Berretta provides, there are also verbs which entail proto-patient properties for the object (e.g. *graffiare* ‘scratch’, *incolpare* ‘blame’, *mettere (in galera)* ‘jail’). However, we made sure that the ones used to build our claim and in the judgment task (Section 4) assigned a balanced degree of agentive properties to subject and object. A further analysis must show if the remaining verbs share similarities. Interestingly, a number of them prototypically select a human (or at least animate) direct object or denotes an event of physical contact.

<sup>13</sup> With difference to all other INTERACTION VERBS listed, the verb *sposare* ‘marry’ does not presuppose but entails the proto-agent properties volition and sentience to its object argument.

Table 3 compares the two classes of OE-PSYCH-VERBS and INTERACTION VERBS to HIGHLY TRANSITIVE VERBS, such as *uccidere* ‘kill’, *ferire* ‘injure’ and *arrestare* ‘arrest’. The latter verbs show a prototypical AGENT-PATIENT-asymmetry with the subject bearing at least the proto-agent properties of causation, movement and independent existence and the object having the proto-patient properties causally affected, stationary participant and change of state. So, in comparison to the objects of the previously mentioned verb classes, HIGHLY TRANSITIVE VERBS show a clear-cut distinction of subject and direct object in terms of their semantic roles.

**Table 3:** Distribution of proto-properties for OE-PSYCH-VERBS, INTERACTION VERBS and HIGHLY TRANSITIVE VERBS.

Verb class	Subject P-agent	Direct Object	
		P-agent	P-patient
OE-PSYCH			
(A) NON-CAUSATIVE (e.g. <i>affascinare</i> ‘fascinate’, <i>deludere</i> ‘disappoint’, <i>incantare</i> ‘enchant’)	+ independent ex.	+ sentience + independent ex.	± change of state
(B) ± CAUSATIVE (e.g. <i>disturbare</i> ‘disturb’, <i>divertire</i> ‘entertain’, <i>spaventare</i> ‘frighten’)	± volition ± sentience ± causation + independent ex.	+ sentience + independent ex.	+ causally affected ± change of state
INTERACTION	± volition ± sentience ± movement + independent ex.	[± volition] <sub>PRESUPPOSED</sub> [± sentience] <sub>PRESUPPOSED</sub> [± movement] <sub>PRESUPPOSED</sub> + independent ex.	
HIGHLY TRANSITIVE	± volition ± sentience + causation + movement + independent ex.		+ change of state ± incremental theme + causally affected + stationary part. ± dependent ex.

To summarize, the role-semantic analysis of OE-PSYCH-VERBS and INTERACTION VERBS has revealed that both classes deviate from the prototypical AGENT-PATIENT-asymmetry of a transitive sentence. We would thus suggest that the affinity of these classes to take DOM in Italian can be accounted for by their lack of thematic distinctness. However, only the class of non-causative OE-PSYCH-VERBS satisfies the generalization of thematic distinctness established for Spanish in (13) above

in the strict sense: for this class, the object outranks the subject in terms of agentivity. For  $\pm$  causative OE-PSYCH-VERBS, in contrast, an AGENT-PATIENT-asymmetry can be established in their causative reading. Yet, thematic distinctness may be blurred since proto-patient and proto-agent properties overlap for the direct object. The class of INTERACTION VERBS does not suffice the generalization of thematic distinctness in a strict sense either since the subject might outrank the object in number of proto-agent properties. Crucially, though, no co-argument dependency relation between subject and object can be established since both participants are independently agentive in two different subevents of which one is presupposed.

### 3.3 Interacting constraints in NP type and word order

Although the focus of the present article clearly lies on the contribution of verbal semantics to the occurrence of DOM in colloquial spoken Italian, it is important to keep in mind that the agentive properties of the verb alone can never trigger DOM: as already mentioned in the introduction, they must always co-occur with (i) an object NP that is highly ranked in the Referentiality Scale and with (ii) a direct object situated in the left (or right) periphery of the sentence. Both additional requirements match typological tendencies: the first constraint (i), often associated with the semantic concept of definiteness, can be identified as a factor which cross-linguistically determines DOM (cf. e.g. Bossong 1991; Aissen 2003; Witzlack-Makarevich/Seržant 2018). The ranking of NP types can be exemplified by a version of the Referentiality Scale by von Heusinger (2008) in (25),

- (25) Referentiality scale: PERSONAL PRONOUN > PROPER NAME > DEFINITE NP > INDEFINITE SPECIFIC NP > NON-SPECIFIC NP > NON-ARGUMENTAL

We have already pointed out in the Introduction that personal pronouns constitute a special case since, in Italian, they generally cannot occur in peripheral position without being *a*-marked. Interestingly for our purposes, common versions of the Referentiality Scale like the one in (25) do not include indefinite generic NPs. While many studies on indefinite NPs focus mainly on specificity (or its absence), the combination between DOM and genericity seems to be a rather understudied phenomenon. Evidence from the few studies on generics shows that in Spanish (Leonetti 2004) and Mandarin Chinese (Iemmolo/Arcodia 2014) the marker is required, while it is rejected in Romanian (Mardale 2008), Neo-Aramaic (Coghill 2014) and Tucano (Iemmolo 2010). In Italian, the marker is rejected with generic

objects situated in a canonical postverbal position (as it is the case for other NP types in SVO structures):

- (26) *La matematica non affascina molto  $\emptyset$ /*a un filosofo.*  
 the mathematics NEG fascinate.PRS.3SG much DOM a philosopher  
 ‘Maths does not fascinate a philosopher that much.’*

However, as already mentioned in the Introduction, its acceptability increases considerably if the generic object is pre-posed as the example in (27) illustrates:

- (27) \* $\emptyset$ /*Ad un filosofo la matematica non affascina*  
 DOM a philosopher the mathematics NEG fascinate.PRS.3SG  
*molto*  
 much  
 ‘To a philosopher maths does not fascinate that much.’

The second constraint (ii), non-canonical object position, has been intensively explored by Iemmolo (2010; in preparation) who provides cross-linguistic evidence for the hypothesis that differential object markers, such as Romance *a*, at least initially occurred in left (or right) detached position carrying a topic marking function. He proposes the following grammaticalization path of *a(d) + object NP* in Romance (Iemmolo in preparation, 266):

- (28) LOCATIVE, ALLATIVE > (TOPIC) > DATIVE > (DIFFERENTIAL) DIRECT OBJECT MARKER

A similar origin of DOM in Romance had been claimed by Pensado (1995, 202–203) who takes the topicalization of indirect and direct personal object pronouns as the starting point of the phenomenon. To put it briefly, the interaction of syntactic and NP-type constraints seems to reflect cross-linguistic tendencies in initial stages of DOM. It is one of the purposes of the present paper to shed light on the question how verb type constraints, more precisely role-semantic properties, enter the picture.

Based on the theoretical assumptions made in this Section, we designed an acceptability judgment task in order to systematically test the impact of role-semantic and referential properties on DOM with CLLD objects in colloquial Standard Italian.

## 4 Judgment task

We designed a questionnaire in order to assess the acceptability grade regarding the presence of DOM with peripheral objects exhibiting different proto-properties and different degrees of referentiality. Section 4.1 formulates the hypotheses based on the theoretical considerations made so far. Section 4.2 describes the study design in detail. Section 4.3 presents the results of the acceptability judgments of the marker according to the proto-properties and the referentiality level of the object. Finally, Section 4.4 discusses the results.

### 4.1 Hypotheses

In Berretta's corpus (1991), the overwhelming majority of *a*-marked objects was situated in the left periphery (*Al gatto, io lo coccolo* 'I cuddle the cat'), some sporadic cases of right dislocations were also attested (*Vi aspetto più tardi, a tutt'e due* 'I'll see you both in a while'), while not even one single case of objects in canonical position (SVO) has been reported. Moreover, *a*-marking only occurred with animate objects. In line with these results, Belletti (2018) points out that the *a*-marking of the object is a property of the left periphery and hence possible only with peripheral DOs. She also clearly states that the presence of the same *a*-marked objects in a canonical SVO structure would lead to the ungrammaticality of the sentence. To formulate our hypotheses, we start from the consideration that for the marker to occur, the object must be both animate and located in a peripheral position of the sentence.

In Section 3, it has also been pointed out that the types of verbs accompanying the object in Berretta's corpus assign non-prototypical properties to their DOs. That means, that, in the examples reported, the object exhibits at least one agentive property. This motivates our first hypothesis, according to which, in colloquial Italian, DOM is more likely to be accepted with agentive objects than with prototypical patients. Additionally, there is enough evidence for the claim that, cross-linguistically, the marker is more likely to occur with objects having a high level of referentiality, such as personal pronouns or proper names, than with, say, indefinite NPs (Aissen 2003; Dalrymple/Nikolaeva 2011). We built on this claim to put forward a second hypothesis, according to which DOM in colloquial Italian is more accepted with objects ranking high in the Referentiality Scale (cf. Aissen 2003) than with objects showing a low level of referentiality. The two hypotheses can be summarized as follows:

H1: In colloquial Italian, DOM is more accepted with objects showing agentive properties than with prototypical patients.

H2: In colloquial Italian, DOM is more accepted with objects ranking high in the Referentiality Scale proposed by Aissen (2003) than with objects ranking low in such scale.

To assess if H1 is on the right track, we tested two classes of verbs where the object possesses agentive properties (OE-PSYCH-VERBS and INTERACTION VERBS) and one class of verbs where the object only exhibits proto-patient properties (HIGHLY TRANSITIVE VERBS). To test H2, we built items containing different NP types (pronouns, proper names, definite NPs and indefinite generics) and hence different degrees of individuation. The details of the study design are provided in the next section.

## 4.2 Study design

For our questionnaire, a group of 43 Italian-speaking adults (MA = 26,0), mostly coming from Lombardy and Piedmont (Northern Italy) was recruited online. Speakers with possible interferences from any Southern-Italian dialects exhibiting DOM (e.g. Sicilian or Sardinian) have been excluded, as their judgments could have been biased by dialectal interferences.

Non-canonical structures, such as left dislocations, right dislocations and topicalizations, are generally used to mark information-structural functions (e.g. topicality) and are much more frequently found in the spoken, rather than the written, variety of the language. To guarantee a more natural effect, sentences were thus recorded by a native speaker of Italian. Participants were asked to listen to the sentences and judge their well-formedness, giving their intuition in a 1-to-5 estimation scale, where 1 represented “unacceptable” and 5 “totally acceptable”. They were instructed to imagine an informal, spoken situation among friends or relatives. To facilitate the process of recruitment, all questionnaires were distributed online.

As stated in Section 4.1, within the items we manipulated both verb type and NP type, keeping the syntactic construction stable. That means that, except for the fillers, our critical items were only constructions containing a direct object situated in the left periphery of the sentence. We tested 3 verb classes (OE-PSYCH, INTERACTION, HIGHLY TRANSITIVE) and 4 NP types (pronouns, proper names, definite NPs, indefinite generic NPs). The questionnaire contained 20 critical items (cf. annexe) and 40 fillers which resulted in a dataset of 60 (items) x 43 (speakers), so that we obtained 2,580 judgments altogether. The presentation order of the entire material was randomized. The items are presented in detail in the following Section.

### 4.2.1 Items

Within our critical items, we manipulated both verb type and NP type, testing 3 classes of verbs (OE-PSYCH-VERBS, INTERACTION VERBS, HIGHLY TRANSITIVE VERBS), 4 types of NPs (pronoun, proper name, definite NP, generic NP), and looking at their interaction. These interactions result in a total of 20 items (15 CLLD and 5 OVS), among which 4 combinations for OE-PSYCH-VERBS (2 lexicalizations + pronouns; 2 lexicalizations + proper names; 1 lexicalization + definite NPs; 4 lexicalizations + generics), 4 combinations for INTERACTION VERBS (2 lexicalizations + pronouns; 1 lexicalization + proper names; 2 lexicalizations + definite NPs; 1 lexicalization + generics) and 3 combinations for HIGHLY TRANSITIVE VERBS (1 lexicalization + pronoun; 2 lexicalizations + proper names; 2 lexicalizations + definite NPs). The selection of OE-PSYCH and INTERACTION VERBS was based on the inventory of lexical items contained in the corpus collected by Berretta (1991).

#### Verb type:

- 6 OBJECT-EXPERIENCER-PSYCH-VERBS (*spaventare* ‘frighten’; *affascinare* ‘fascinate’; *innervosire* ‘make nervous’; *convincere* ‘to convince’; *entusiasmare* ‘to enthusiasm’; *infastidire* ‘to annoy’). An example of the test item is provided in (29):

(29) *Alla ragazza, la matematica non l' ha*  
 DOM.the girl the mathematics NEG CL.3SG have.PRS.3SG  
*mai affascinata molto*  
 never fascinated much  
 ‘Maths never fascinated the girl that much.’

- 5 INTERACTION VERBS (*aspettare* ‘wait’; *ringraziare* ‘thank’; *salutare* ‘greet’; *accompagnare* ‘to accompany’; *sposare* ‘to marry’) such as in (30):

(30) *A Maria, i suoi compagni di classe non*  
 DOM Maria the POSS.3M.PL classmates NEG  
*l' aspettano mai.*  
 CL.3SG wait.PRS.3PL never  
 ‘Her classmates never wait for Maria.’

- 5 HIGHLY TRANSITIVE VERBS (*ferire* ‘injure’; *arrestare* ‘arrest’; *uccidere* ‘kill’; *prendere* ‘to catch’; *fermare* ‘to stop’), as (31) shows:

- (31) ??A *Giacomo, l' hanno arrestato ieri notte.*  
 DOM Giacomo CL.3SG have.PRS.3PL arrested yesterday night  
 'They arrested Giacomo last night.'

NP type:

- 3 *a*-marked personal pronouns (e.g. *a lui* 'him'; *a lei* 'her')
- 3 *a*-marked personal names (e.g. *a Maria* 'Maria'; *a Gianni* 'Gianni')
- 3 *a*-marked definite NPs (e.g. *alla ragazza* 'the girl')
- 3 *a*-marked generic NPs (e.g. *ad un filosofo* 'a philosopher')

To prevent participants from easily identifying the phenomenon under investigation, a considerable number of fillers (40 in total) was included. The group of filler sentences was divided in "good" and "bad" fillers. The good fillers included 20 fully grammatical SVO utterances, while the bad ones were composed by 20 ill-formed sentences that included agreement mismatches, wrong auxiliary-choice, incorrect use of preposition, etc.

Additionally, all OE-PSYCH-VERBS include an inanimate subject. The lack of animacy of the subject with this verb type guarantees its non-causative reading.

#### 4.2.2 Predictions

We predicted the good fillers to be rated with 5 (totally acceptable) and the bad ones with 1 (totally unacceptable). As for the critical items, we expected participants to use the middle values (2; 3; 4).

Predictions for the grammaticality judgments tasks based on H1 and H2:

- H1: If, in a transitive sentence, the prototypical AGENT-PATIENT-asymmetry cannot be established unequivocally because the DO bears any proto-agent property that is not expected in a typical transitive sentence, the AGENT-PATIENT-asymmetry is blurred, and the marker becomes more acceptable.
- P1: Based on the distribution of proto-agent vs. proto-patient properties of the verbs tested in our study (cf. Section 2.2), we predict the following scale of acceptability:

OE-PSYCH & INTERACTION > HIGHLY TRANSITIVE

H2: In colloquial Italian DOM is better accepted with NP types ranking high in the Referentiality Scale identified by Aissen (2003).

P2: We expect the following ranking for the NP types we investigated:

PERSONAL PRONOUNS > PERSONAL NAMES > DEFINITE NPs > GENERICS



Regarding generics, the lack of robust previous literature makes it difficult to make predictions. For Spanish, various authors have pointed out that the occurrence of the marker with an indefinite determiner is rather marginal and often constrained by particular requirements (i.e., specificity, cf. Leonetti 2004). For some varieties of Spanish, like Cuban, acceptability judgments have shown that the lack of DOM with indefinite specific NPs is even more accepted than in the peninsular variety (cf. Caro Reina/García García/von Heusinger, this volume). However, the case of indefinites with a generic reading is different. According to Lambrecht “NPs whose referents identify either the whole class of all entities singled out by it or some representative set of members of this class, can be assumed to be identifiable” (1994, 82). Thus, although morphologically indefinites, their level of identifiability seems to be rather strong. In this case, they should be fairly accepted. The results of this study can provide indications on whether speakers rely more on the morphological form or the semantics of the NP in evaluating the occurrence of the marker.

### 4.3 Results

Figure 1 shows the ratings of 43 participants for the acceptability between DOM and different classes of verbs. OE-PSYCH-VERBS receive the highest rating (3.3), followed by INTERACTION VERBS (2.4) and HIGHLY TRANSITIVE VERBS (1.8). Interestingly, the ratings for INTERACTION VERBS is closer to the one of HIGHLY TRANSITIVE than OE-PSYCH-VERBS. The difference in the acceptability means of the three verb classes is, however, significant.

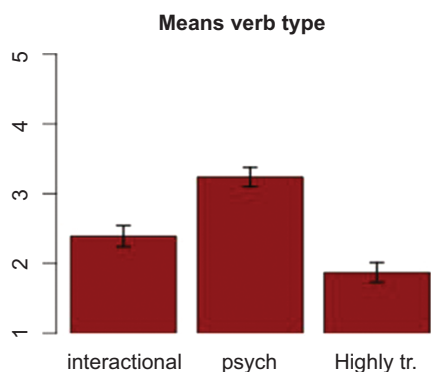
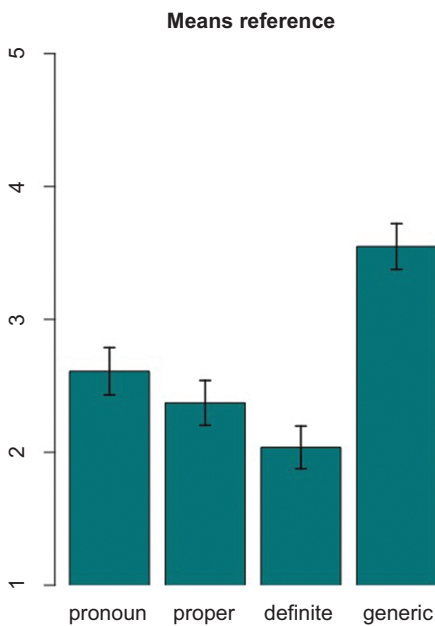


Figure 1: Means verb types.

Figure 2 shows the participants' evaluations for different types of NP types. Pronouns display an acceptability rate of 2.6, followed by proper names (2.4) and definite NPs (2.2). Contrary to our expectations, indefinite generics seem to perform better than the other types, with a rating of 3.5.<sup>14</sup> Besides generics, the remaining NP types seem to follow the Referentiality Scale so far predicted (Aissen 2003), where *a*-marked pronouns are better accepted than *a*-marked proper names, which, in turn, are better evaluated than less prominent NP types, like *a*-marked definite NPs. However, as confidence intervals show, while the difference between pronouns and proper names is significant (p-value = 0.03), the one between proper names and definite NPs is not (p-value = 0.09).



**Figure 2:** Means NP types.

The evaluations for the interaction between different NP types and OE-PSYCH, INTERACTION and HIGHLY TRANSITIVE VERBS are presented in Figures 3–5, respectively.

Both OE-PSYCH-VERBS and INTERACTION VERBS seem to perform better with indefinite generics (their interaction has received an evaluation of 3.6 and 3.2,

<sup>14</sup> It is important to recall that indefinite generics have only been tested with OE-PSYCH-VERBS and one INTERACTION VERB, while the other NP types have been tested with all three verb types.

respectively), a category that we had predicted as being situated low in the Referentiality Scale and, hence, less likely to be acceptable with the marker.

Apart from generics, our predictions are borne out only when it comes to the combination between INTERACTION VERBS and different NP types (Figure 4), where pronouns receive an average rating of 2.7, followed by proper names (2.2) and definite NPs (1.9). Confidence intervals show that the difference between proper and definites in Figure 5 can still be considered significant ( $p$ -value = 0.05).

In the case of HIGHLY TRANSITIVE VERBS (Figure 5), there is no significant difference in the ratings of personal pronouns and proper names (2.09 and 2.04 respectively;  $p$ -value = 0.9), while both differ significantly from definite NPs, that receive the lowest score (1.4).

OE-PSYCH-VERBS (Figure 4) even show a reverse referentiality effect, with definites (3.1) showing higher acceptability values than the other NP types (2.7 for pronouns, 2.8 for proper names). Confidence intervals display, however, that the difference in the rating of the three NP types is not significant (pronoun and proper:  $p$ -value = 0.7; pronouns and definites:  $p$ -value = 0.6; proper and definites:  $p$ -value = 0.4).

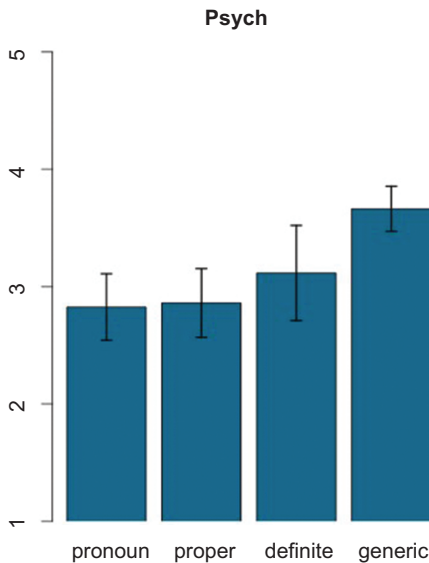


Figure 3: Means OE-PSYCH-VERBS (e.g. *spaventare* ‘frighten’) and NP type.

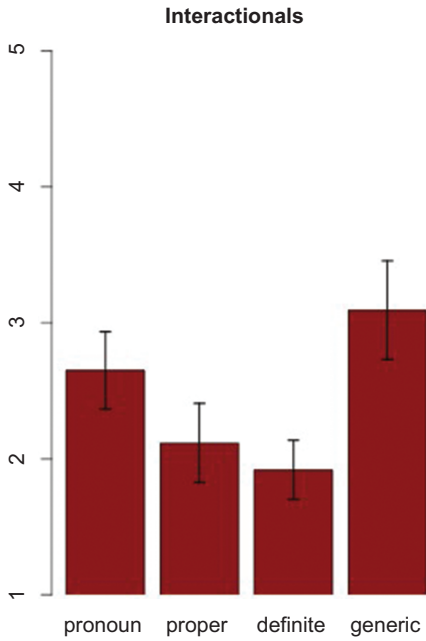


Figure 4: Means INTERACTION VERBS (e.g. *aspettare* 'wait') and NP type.

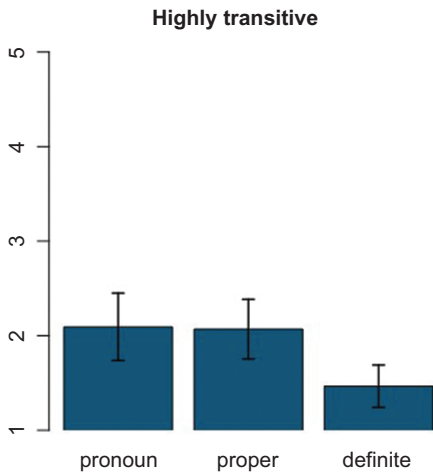


Figure 5: Means OE-PSYCH-VERBS (e.g. *spaventare* 'frighten') and NP type.

## 4.4 Discussion

The results of the pilot study show that the acceptability of the tested items is generally rather low. It is possible to see, however, that there are differences in the judgments reported for each verb type. Recall that, according to H1, DOM is likely to occur when the traditional AGENT-PATIENT-asymmetry cannot be established unequivocally because the direct object bears some proto-agent properties.

OE-PSYCH-VERBS, where the direct object receives the thematic role of EXPERIENCER, are a clear example of a case where such AGENT-PATIENT-asymmetry is not as clear-cut, as the object possesses [+sentience], [+independent existence] and the subject, being the STIMULUS, exhibits [+independent existence]. The validity of H1 is also supported by the comparison between the evaluation of OE-PSYCH-VERBS (3.3) and the low ratings reported for HIGHLY TRANSITIVE VERBS (1.8), where the object bears only proto-patient properties. Thus, when the thematic role of the object is the prototypical patient, the marker is less (or even not at all) acceptable.

In the case of INTERACTION VERBS, both subject and object bear proto-agent properties while proto-patient properties are not assigned. In this case again, an AGENT-PATIENT-asymmetry cannot be established. Nevertheless, the results reported for INTERACTION VERBS (2.4) show an acceptability rate closer to HIGHLY TRANSITIVE VERBS (1.8) than to OE-PSYCH-VERBS (3.3) and, as such, considerably low. This result goes against our expectations. Our hypothesis was based on the assumption that, whenever a verb assigns at least one proto-agent property to its direct object, the marker should become more acceptable. Both INTERACTION and OE-PSYCH-VERBS fall into this category and should therefore perform in a similar way. However, this is not reflected in the ratings.

One possible explanation for the difference in the behaviour of the two verb classes could lie in the subtle distinctions at the level of their proto-properties. In Section 3, we mentioned the fact that verbs could either entail specific proto-agent properties for their arguments or presuppose them. While in the case of OE-PSYCH-VERBS proto-agent properties are entailed for the object, INTERACTION VERBS only presuppose them. The two verb classes also exhibit a difference at the level of proto-agent properties exhibited by the subject. As Table 3 shows, the subject of an OE-PSYCH-VERB bears one proto-agent property in its non-causative reading, while the subject of an INTERACTION VERB bears up to four, including [+volition], a property typically found in subjects of HIGHLY TRANSITIVE VERBS. Whether this explanation could account for the difference in the evaluation of the two verb classes is yet to be assessed. A follow-up study, testing more lexicalizations for each verb type, would certainly reveal further insights.

As for the NP type, our prediction according to which participants' evaluations would follow the Referentiality Scale identified by Aissen (2003) (cf. H2 P2, Section 3.2.2) is not borne out. Set aside dislocated pronouns, for which the marker is obligatory, proper names and definites don't show significant differences in their evaluations. Looking at their interaction with different verb types, pronouns perform better only with INTERACTION VERBS, while OE-PSYCH-VERBS even show an anti-referentiality effect, where definites override the other two NP types. To account for this result, one might assume an ongoing process of grammaticalization, for which the degree in referentiality of different NPs doesn't play a role anymore.

A noteworthy finding that emerges from our acceptability judgment task is the particular behaviour of indefinite generics. Differently from what we would have expected, their rating (3.5) is considerably higher than that of other NP types. A closer look at the semantic interpretation of genericity (Krifka 1987; Carlson 1995) suggests that generics, although introduced by an indefinite determiner, tend to behave like proper names, being interpreted as unique entities and having, therefore, a rigid reference. Leonetti (2004) notices that both specific and generic interpretations of indefinites belong to the family of strong interpretations, while non-specific interpretations are typically weak. This suggests that their classification within the Referentiality Scale might need to be rethought. Moreover, the overall low acceptability rates of our experimental items and the fact that indefinite generic NP-objects have better scores might be amenable to the same phenomenon: topicality. Low acceptability may be partially explained by the fact that object-preposing is a syntactically-marked construction, usually associated to backgrounded, topical information. Syntactically-marked constructions usually require additional contextual support, which is lacking in our items. This might have led to infelicity in context and low acceptability rates. Similarly, higher acceptability of indefinite-generic NPs may derive from the fact that for an indefinite NP to be interpreted as generic, it has to belong to the topical part of the utterance. Thus, while on the one hand, topicality lowers the acceptability of certain items, on the other hand, it increases the scores of generics of our indefinite NPs.

Such opposite effect might be due to the fact that, in out-of-the-blue contexts, generics are more felicitous than personal pronouns, personal names or definite NPs. Sentences containing a personal pronoun and deprived of context are somewhat artificial for speakers. On the contrary, generic statements, such as "philosophers are not attracted by mathematics" can possibly sound more natural and be, therefore, more acceptable.

Whether this explanation is on the right track or not, the literature on the behaviour of generic objects and their interaction with DOM seems to be rather scarce and our results suggest that the phenomenon deserves further investigation.

The main aim of the current study was to give indications on whether different types of verbs might have an influence in the degree of acceptability of the marker with their respective DOs. Our results suggest that the type of verb might be one of the factors responsible for occurrence of DOM in Italian, given that the marker is better accepted with objects of OE-PSYCH-VERBS, exhibiting the thematic role of EXPERIENCER than with HIGHLY TRANSITIVE VERBS, where the object is the prototypical patient. The design of the task at this stage, however, presents several limitations and a more controlled experiment is still necessary. For instance, set aside the fillers, the test presented participants only with *a*-marked DOs. A condition where sentences with the same verbs are presented without the marker is still needed and will be the object of a more complete follow-up experiment.

Moreover, syntactic position has not been tested. Although the literature agrees on the claim that only peripheral objects are likely to be marked, it is necessary to assess such claim empirically and verify to what extent the acceptability rate of *a*-marked objects in SVO position differs from their peripheral correlates. Finally, the behaviour of generics still needs to be better assessed by testing more interactions with different verb types and possibly in different syntactic positions.

## 5 Conclusions

With the present paper, we have adopted an explanation in terms of role-semantic parameters for justifying the occurrence of Differential Object Marking in spoken colloquial Italian, a phenomenon that had been reported only from a descriptive perspective so far. We have investigated the well-formedness of sentences containing DOM in Italian by means of an acceptability judgment test, advancing the hypothesis that non-prototypical, agentive objects are more likely to be *a*-marked than typical, “patient-like” ones. Previous research on DOM has often explained such deviation from the prototypicality of the object in terms of animacy. With our study, we add a piece to this puzzle showing that, in colloquial Italian, the non-prototypicality of the DO also concerns its thematic role.

Moreover, the present article is a further contribution to the view that morphosyntactic phenomena like DOM are not sensitive to animacy exclusively but rather to agentivity, represented e.g. through the proto-agent property of sentence (cf. e.g. García García/Primus/Himmelmann 2018). This becomes evident

when we have a look at Italian verbs like *spaventare* ‘frighten’ and *uccidere* ‘kill’: while both verbs select for an [+animate] direct object, only the former but not the latter allows for the *a*-marker in Italian. This constraint can be only explained in an agentivity-based but not in an animacy-based account. Whereas *spaventare* ‘frighten’ entails sentience for the object argument, *uccidere* ‘kill’ only assigns proto-patient properties to its object. Furthermore, the case of DOM in Italian is worth being investigated within a theory of grammaticalization: The fact that DOM underlies multiple constraints, namely syntactic construction, NP type as well as verb type, suggests that the phenomenon in Italian could be at an early stage of development and deserves to be further investigated in time to observe a possible expansion. Likewise, the robustness of the present account can be proved by a systematic analysis of (rare) instances of DOM in French put forward by Fagard/Mardale (2014), another Romance language where DOM is usually said to be absent. Their examples suggest that in French, like in Italian, different constraints must be fulfilled in order to allow DOM to appear. These comprise “inherent factors” of the object NP as well as so-called “global factors” which include not only topicality but also the verb type. A comparative account of French and Italian could reveal if the same role semantic properties lead to a preference for DOM in both languages.

## Annexe

### OE-PSYCH-VERBS

---

*Visibilmente, a lei, certi argomenti non l'hanno convinta.*

‘Visibly, certain arguments didn’t convince her.’

---

*A lui, questa favola l'ha sempre spaventato.*

‘This fairy tale has always scared him.’

---

*Alla ragazza, la matematica non l'ha mai affascinata molto.*

‘Maths never fascinated the girl that much.’

---

*Ad Elena, i film di Tarantino non l'hanno mai entusiasmata.*

‘Elena was never thrilled by Tarantino’s films.’

---

*A Pietro, l'atteggiamento di Maria l'ha sempre innervosito.*

‘Mary’s attitude has always made nervous Peter.’

---



## Interaction verbs

---

*A Maria, i suoi compagni di classe non l'aspettano mai.*

'Her classmates never wait for Maria.'

---

*A me di sicuro Laura non mi saluta.*

'Laura for sure doesn't greet me.'

---

*Al professore, devi ringraziarlo sempre dopo la discussione della tesi.*

'You should always thank the professor after the discussion of the thesis.'

---

*A causa del suo atteggiamento, a lei, non la sposa nessuno.*

'Because of her attitude, nobody marries her.'

---

*Di solito, alla sposa, il padre l'accompagna all'altare.*

'Usually, the father walks the bride down the aisle.'

---

## HIGHLY TRANSITIVE VERBS

---

*A Giacomo, l'hanno arrestato ieri notte.*

'They arrested Giacomo last night.'

---

*A lei, l'ha ferita Paolo.*

'Paolo injured her.'

---

*A Luca, l'hanno preso alla guida ubriaco.*

'They caught Luca driving drunk.'

---

*Alla vittima, l'ha sicuramente uccisa il marito.*

'The victim was definitely killed by her husband.'

---

*A tuo fratello, l'ha fermato ieri la polizia.*

'The police stopped your brother yesterday.'

---

## GENERIC NPS

---

*Ad una madre, certe cose infastidiscono.*

‘Some things annoy a mother.’

---

*Ad un bambino questa favola generalmente spaventa.*

‘This fairy tale generally frightens a child.’

---

*Ad un filosofo la matematica non affascina molto.*

‘Maths does not fascinate a philosopher that much.’

---

*Ad un perenne ritardatario, alla fine, non lo aspetta più nessuno.*

‘A perpetual latecomer, in the end, is no longer expected.’

---

*Di solito, ad un ragazzo, l’atteggiamento di Maria innervosisce.*

‘Usually, Mary’s attitude makes a boy nervous.’

---

## Bibliography

- Ackerman, Farrell/Moore, John, *Proto-properties and grammatical encoding. A correspondence theory of argument selection*, *Linguisticæ Investigationes* 29:2 (2006), 318–321.
- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21 (2003), 435–483.
- Belletti, Adriana/Rizzi, Luigi, *Psych-verbs and  $\theta$ -theory*, *Natural Language and Linguistic Theory* 6 (1988), 291–352.
- Belletti, Adriana, *On a-marking of object topics in the Italian left periphery*, in: Petrosino, Roberto/Cerrone, Pietro/van der Hulst, Harry (edd.), *Beyond the veil of Maya. From sounds to structures*, vol. 135, Berlin/Boston, de Gruyter, 2018, 445–466.
- Benincà, Paola, *Il lato sinistro della frase italiana*, *Balkan-Archiv* 11 (1986), 213–243.
- Berretta, Monica, *Sulla presenza dell’acusativo preposizionale in italiano settentrionale. Note tipologiche*, *Vox Romanica* 48 (1989), 13–37.
- Berretta, Monica, *Note sulla sintassi dell’acusativo preposizionale in italiano*, *Linguistica* 31 (1991), 211–232.
- Blume, Kerstin, *A contrastive analysis of interaction verbs with dative complements*, *Linguistics* 36:2 (1998), 253–280.
- Blume, Kerstin, *Markierte Valenzen im Sprachvergleich. Lizenzierungs- und Linkingbedingungen*, Tübingen, Niemeyer, 2000.
- Bossong, Georg, *Differential Object Marking in Romance and beyond*, in: Wanner, Dieter/Kibbee, Douglas A. (edd.), *New analyses in Romance linguistics. Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, April 7–9, 1988*, Amsterdam/Philadelphia, John Benjamins, 1991, 143–170.

- Bossong, Georg, *Le marquage différentiel de l'objet dans les langues d'Europe*, in: Feuillet, Jack (ed.), *Actance et valence dans les langues de l'Europe*, vol. 20, Berlin/New York, de Gruyter, 1998, 193–258.
- Carlson, Gregory N./Pelletier, Francis J. (edd.), *The generic book*, Chicago, University of Chicago Press, 1995.
- Coghill, Eleanor, *Differential Object Marking in Neo-Aramaic*, *Linguistics* 52:2 (2014), 335–364.
- Croft, William, *Case marking and the semantics of mental verbs*, in: Pustejovsky, James (ed.), *Semantics and the lexicon*, Dordrecht/Boston, Kluwer Academic Publishers, 1993, 55–72.
- Dalrymple, Mary/Nikolaeva, Irina, *Objects and information structure*, Cambridge, Cambridge University Press, 2011.
- Dowty, David, *Thematic proto-roles and argument selection*, *Language* 67:3 (1991), 547–619.
- Fagard, Benjamin/Mardale, Alexandru, *Non, mais tu l'as vu à lui? Analyse(s) du marquage différentiel de l'objet en français*, *Verbum* 36:1 (2014), 143–168.
- García García, Marco, *Differential Object Marking with inanimate objects*, in: Kaiser, Georg A./Leonetti, Manuel (edd.), *Proceedings of the workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*, Arbeitspapier 122, Konstanz, Fachbereich Sprachwissenschaft der Universität Konstanz, 2007, 63–84.
- García García, Marco, *Differentielle Objektmarkierung bei unbelebten Objekten im Spanischen*, Berlin/Boston, de Gruyter, 2014.
- García García, Marco, *Nominal and verbal parameters in the diachrony of Differential Object Marking in Spanish*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 209–242.
- García García, Marco/Primus, Beatrice/Himmelman, Nikolaus P., *Shifting from animacy to agentivity*, *Theoretical Linguistics* 44:1–2 (2018), 25–39.
- Iemmolo, Giorgio, *Topicality and Differential Object Marking. Evidence from Romance and beyond*, *Studies in Language* 34:2 (2010), 239–272.
- Iemmolo, Giorgio/Arcodia, Giorgio F., *Differential Object Marking and identifiability of the referent. A study of Mandarin Chinese*, *Linguistics* 52:2 (2014), 315–334.
- Iemmolo, Giorgio, *Differential Object Marking*, Oxford, Oxford University Press (in preparation).
- Kabatek, Johannes, *Wohin strebt die differentielle Objektmarkierung im Spanischen?*, *Romanistisches Jahrbuch* 67:1 (2016), 211–239.
- Kailuweit, Rolf, *Linking. Syntax und Semantik französischer und italienischer Gefühlsverben*, Tübingen, Niemeyer, 2005.
- Kailuweit, Rolf, *Romance object-experiencer verbs. From aktionstart to activity hierarchy*, in: Barrajón López, Elisa/Cifuentes Honrubia, José Luis/Rodríguez Rosique, Susana (edd.), *Verb classes and aspect*, Amsterdam/Philadelphia, John Benjamins, 2015, 312–333.
- Krifka, Manfred, *An outline of generics*, in: *Forschungsberichte des Seminars für natürlich-sprachliche Systeme*, Tübingen, Universität Tübingen, 1987.
- Kutscher, Silvia, *Kausalität und Argumentrealisierung. Zur Konstruktionsvarianz bei Psychverben am Beispiel europäischer Sprachen*, Tübingen, Niemeyer, 2009.
- Lambrecht, Knud, *Information structure and sentence form. Topic, focus, and the mental representations of discourse referents*, Cambridge, Cambridge University Press, 1994.
- Leonetti, Manuel, *Specificity and Differential Object Marking in Spanish*, *Catalan Journal of Linguistics* 3:1 (2004), 75–114.
- Mardale, Alexandru, *Microvariation within Differential Object Marking*, *Revue Roumaine de Linguistique* 53:4 (2008), 449–467.

- Mürmann, Sophie, *Differential Object Marking and role semantics in Romance*, University of Cologne, Ph. D. thesis, 2021.
- Pensado, Carmen, *La creación del complemento directo preposicional y la flexión de los pronombres personales en las lenguas románicas*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 179–233.
- Primus, Beatrice, *Cases and thematic roles. Ergative, accusative and active*, Tübingen, Niemeyer, 1999 (= 1999a).
- Primus, Beatrice, *Rektionsprinzipien*, in: Wegener, Heide (ed.), *Deutsch kontrastiv. Typologisch-vergleichende Untersuchungen zur deutschen Grammatik*, Tübingen, Stauffenburg, 1999, 135–170 (= 1999b).
- Primus, Beatrice, *Hierarchy mismatches and the dimensions of role semantics*, in: Bornkessel, Ina, et al. (edd.), *Semantic role universals and argument linking. Theoretical, typological, and psycholinguistic perspectives*, Berlin/Boston, de Gruyter, 2006, 53–88.
- Primus, Beatrice, *Animacy, generalized semantic roles and Differential Object Marking*, in: Lamers, Monique/de Swart, Peter (edd.), *Case, word order, and prominence. Interacting cues in language production and comprehension*, Dordrecht, Springer, 2012, 65–90.
- Verhoeven, Elisabeth, *Thematic prominence and animacy asymmetries. Evidence from a cross-linguistic production study*, *Lingua* 143 (2014), 129–161.
- von Heusinger, Klaus, *Verbal semantics and the diachronic development of DOM in Spanish*, *Probus* 20:1 (2008), 1–31.
- von Heusinger, Klaus/Kaiser, Georg A., *Affectedness and Differential Object Marking in Spanish*, *Morphology* 21:3 (2011), 593–617.
- Weissenrieder, Maureen, *A functional approach to the accusative “a”*, *Hispania* 74:1 (1991), 146–156.
- Witzlack-Makarevich, Alena/Seržant, Ilja A., *Differential Argument Marking. Patterns of variation*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 1–40.

Elisabeth Mayer and Liliana Sánchez

# Emerging DOM patterns in clitic doubling and dislocated structures in Peruvian-Spanish contact varieties

**Abstract:** In this chapter we explore the expression of Differential Object Marking (DOM) in monolingual and bilingual Spanish in contact with typologically different languages. We focus on how DOM patterns are expressed in bilingual and monolingual clitic doubling and dislocated structures by investigating the effect of typological differences in case marking and the effects of definiteness, animacy and thematic role. Our findings show that while definiteness, animacy and thematic structure are relevant factors in the production of DOM across bilingual groups, there are also differences in DOM frequency related to typological distinctions in the L1 of the bilingual groups. Finally, our findings also show variability in the monolingual data of three individuals raised in a contact situation as well as the raise of topicality as a possible factor that contributes to DOM in the L2 varieties under study.

**Keywords:** Differential Object Marking (DOM), clitic doubling, bilingual acquisition, typological differences, Andean and Amazonian Spanish

## 1 Introduction

Differential Object Marking (DOM) is a variable and widespread argument marking system (Bossong 1991; 2003; Aissen 2003; Dalrymple/Nikolaeva 2011; Witzlack-Makarevich/Seržant 2017; a.o.). It is a widely attested strategy across many genetically unrelated languages used to mark a select range of specific and/or topical direct objects (Bossong 1991; 2003; Aissen 2003; Dalrymple/Nikolaeva 2011; a.o.). In the emergence of DOM in Romance, three different but interacting facts have been identified (Laca 2006; Mayer 2017, 101). One is the need to differ-

---

**Acknowledgements:** We would like to thank all participants in the study and acknowledge the contribution of the following research assistants: Yoshidaira Garcia for help with Huánuco Quechua, Caleb Cabello Chirisente for Asháninka, and Enrique Espinoza for Shipibo, and Julio López Otero for help with the statistics [R].

---

**Elisabeth Mayer**, Griffith University Queensland, e-mail: e.mayer@griffith.edu.au  
**Liliana Sánchez**, University of Illinois at Chicago, e-mail: lesanche@uic.edu

entiate between subject and object arguments (Bossong 1991). Secondly, separation of dative and accusative marking is relevant, and finally variability in verbal argument structure in certain verbs (Laca 2006). These factors have given rise to multiple competing theories on how DOM emerged. The emergence of a new accusative form (Givón 1997; Bossong 2003) versus a single plus/minus dative case system (Alsina 1996) and the transition from neutralization of dative and accusative marking (Lapesa 2000) to the “cannibalistic” dative theory (Company 2001; 2003). The third and last point refers to topic marking (Givón 1976) and its extension to secondary topic marking (Dalrymple/Nikolaeva 2011). While DOM preferably marks the argument itself, Differential Object Indexing (DOI), another related and very similar strategy, preferably marks the verb. Notably, both strategies can cooccur in one language (Iemmolo 2011).

Once DOM emerges in the grammar, it may be conditioned by a wide range of semantic and syntactic properties of the marked direct object. Factors that have been argued to trigger DOM are the interaction of the semantic-pragmatic properties of the object such as animacy and definiteness (Bossong 1991; 2003; Aissen 2003, for a counter argument cf. Sinnemäki 2014), and other information structural properties that can only be interpreted at higher constituent levels such as topicality (Leonetti 2008), and telicity (Torrego 1998).

In this study, we investigate emerging patterns of DOM in Clitic Doubling (CLD) and the two related information structures Clitic Left Dislocation (CLLD) and Clitic Right Dislocation (CLRD). We focus on two referential factors: animacy and definiteness as well as on the verbal thematic roles *PATIENT* and *THEME* of the direct object noun phrase which have been less explored as a factor involved in DOM. Both thematic roles are subject to different degrees of affectedness, while *PATIENT* objects undergo a change, *THEME* objects remain unchanged (Naess 2004). In order to understand how some of these factors interact with each other in a continuum of language contact, we conducted a study among three groups of bilingual speakers of one of three indigenous languages spoken in Peru: Asháninka, Quechua, Shipibo and Spanish as well as of three monolingual speakers of Spanish exposed to contact varieties living in a continuum of language contact situations.

It has been shown that DOM in clitic doubling and dislocated structures<sup>1</sup> exhibits high levels of variability across Spanish dialects more generally and particularly in contact varieties (Mayer/Sánchez 2016; 2017). This variability may be due to the specific properties of the languages with which it is in contact, spe-

---

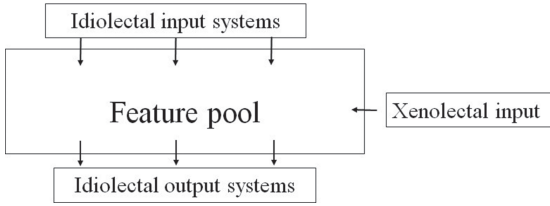
<sup>1</sup> We focus on DOM in these structures as DOM exclusively occurs in them in the indigenous languages.

cifically in languages that lack DOM as is the case of Asháninka, Quechua and Shipibo. The study of the development of DOM in these contexts is of relevance in the understanding of how DOM emerges and evolves and in the understanding of the main semantic and syntactic features that trigger it. In the next Sections, we present a brief overview of the framework we assume for language contact situations, a brief description of case marking in each of the indigenous languages and Spanish, the study design, its results and conclusions.

## 2 Contact and SLA continuum

It is well known that in language contact situations, divergence at the interfaces and typological differences in morphological patterns may result in new systems (Matras 2010) due to a range of ecological factors (Mufwene 2002). These new systems reflect different grammaticalization stages and different choices in the selection from a feature pool (Mufwene 2001; 2002; Matras 2010; Mayer/Sánchez in press) depending on factors such as levels of proficiency and frequency of exposure to bilingual varieties as well as frequency of interaction and activation of features between bilinguals and monolinguals (Putnam/Sánchez 2013). Following Mufwene (2001; 2002) we assume that idiolectal output systems are the result of the selection of the best fitting feature available for communicative purposes from the competing features in any given pool, see Figure 1 below. In the case of the bilingual groups under investigation, contact with xenolectal input contributes different alignment systems to the feature pool in a continuum that goes from very different to closer to Spanish. The feature pool is also impacted by differences in input to Spanish due to varying access to educational opportunities. Bilingual as well as monolingual speakers find themselves confronted with a complex relationship between idiolectal input, xenolectal input in the form of language or dialect contact, as in the case of the monolinguals – and access to educational opportunities. Idiolectal output systems are the result of the selection of a subset of features in individuals and may be subject to crosslinguistic convergence (Sánchez 2003).

Furthermore, differences in the mapping of syntactic features onto morphology are a main factor in differential bilingual development. In second language acquisition studies this has been identified as difficulties in assembling or reassembling features in the L2 that are not activated in the L1. The Feature Reassembly Hypothesis states that these difficulties stem from the need to reassemble features and associate them with new morphological forms in the L2 (Lardiere 1998; 2005). More recently, the Bottleneck hypothesis in SLA (Slabakova 2008; Jensen et al. 2017) has identified functional morphology as the locus of greater levels of difficulty in



**Figure 1:** Feature pool: input varieties and output varieties.

second language acquisition. Monolingual child acquisition of DOM in monolingual contexts has been described as straightforward (Rodríguez-Mondoñedo 2008) and somewhat variable among L2 speakers but as variable among heritage speakers (Montrul 2004). One source of this variability lies in differences between the acquisition of DOM as triggered by object inherent semantic features like animacy and definiteness versus structural features like case (Blake 2001). The former has been identified as more difficult to learn than the latter. Even within semantic features, animacy has been identified as more learnable as a DOM trigger than the discourse-related features definiteness and specificity (Guijarro-Fuentes 2011; 2012).

In this study, we take into account the inherent variability of language acquisition when analysing bilingual and monolingual DOM clitic doubling and information structure systems in contact situations, as these involve individual differences that result from variability in the feature pool and sensitivity to different semantic features.

### 3 The study

In previous work (Mayer/Sánchez in press), we examined the distribution of clitics and DOM in clitic doubled constructions in Amazonian Spanish in contact with Shipibo and Ashéninka Perené – a different data set from the Asháninka data set used here. In this chapter, we focus on crosslinguistic effects and DOM in oral production data from bilingual Spanish of speakers of three typologically different indigenous languages in Peru across urban and rural settings. As a point of comparison, we also include data from a longitudinal study of monolingual Spanish speakers in Lima (LSCV-Lima Spanish contact varieties).<sup>2</sup> We explore how DOM patterns are expressed in bilingual and monolingual clitic doubling

<sup>2</sup> The speakers of LSCV whose data we analyzed are part of a continuum of Quechua and Spanish contact that ranges from passive knowledge of Quechua to monolingualism in Spanish.



and information structures. In order to do so, we investigate the effect of contact on case marking based on the conditioning features animacy and definiteness in definite and indefinite object Determiner Phrases (DP) as well as transitivity expressed through the thematic roles PATIENT and THEME in marked and unmarked direct object arguments in clitic doubled and dislocated information structures. The selection of these information structures stems from the fact that, in our contact data, DOM preferably occurs in those structures and our data sets show practically no evidence of DOM in simple transitive clauses.

The following examples reflect the preferred DOM strategies of each of the three bilingual data sets and the monolingual data set. The bilingual datasets consist of data from two Amazonian bilingual groups and one Andean bilingual group. The monolingual data comes from a longitudinal study of three individuals in Lima representing Limeño Spanish Contact Varieties (LSCV).

Expression of DOM in Quechua-Spanish from a rural Andean area in (1) is strikingly different from the other data sets by marking close to 100% of all direct objects, based on the features animacy and definiteness with 90% each. In terms of thematic roles, the preference for marking THEME over PATIENT is significant. The high percentage for both features of the object DP can be analyzed as the result of a shared nominative/accusative alignment system.

- (1) *Le acarici-a a su perr-o*  
 CL.3SG caress-PRES.3SG DOM POSS dog-M.SG  
 ‘(S)/He caresses her/his dog.’

Huánuco Quechua-Spanish (Sánchez Dataset 2005)<sup>3</sup>

Data from rural Asháninka-Spanish show a different scenario being the only group with a preference for unmarked DPs (55%) over marked DPs (45%). DOM

---

<sup>3</sup> The following abbreviations are used in this chapter: A = agent, ABS = absolutive, ACC = accusative, AUX = auxiliary, CAUS = causative, LEJ/CIS = cislocative, CL = clitic, CMPL = completive, DEM = demonstrative, DET = determiner, DIM = diminutive, DOM = Differential Object Marking, DP = determiner phrase, ERG = ergative, EV/EVID = evidential, F = feminine, GER = gerund, HSY = hearsay, IMPF = imperfective, INC = incorporation, INDEF = indefinite determiner, INF = infinitive, IRR = irrealis, LOC = locative, M = masculine, O/OBJ = object of transitive verb, P/pat = patient, 1 = first person, 2 = second person, 3 = third person, PART = participle, PAS/PAST = past tense, PFV/PRF/PERF = perfective, PL = plural, P/POSS = possessive, PP2 = past completive particle, PP = positive polarity, PR = prospective, PRES = present tense, PRO = pronoun, PRT = preterit, REAL = realis, S = subject of intransitive verb, SG = singular, SUB = subordinate, SS = same subject, TOP = topic, T/TR = transitive, TH = theme. For the examples from South American indigenous languages we keep the original glosses as used by the authors in the publications. Differences with the abbreviations used in Spanish and additional abbreviations are noted in bold.

as in (2) occurs preferably with definite and animate DPs, PATIENTS are strongly preferred over THEMES. Given the fact that Asháninka is a nominative/accusative language with fluid/split transitivity and the availability of gender specific bound morphemes as argument markers it is possible that these factors support acquisition of DOM in clitics and the coding of information structure.

- (2) *Lo bot-aron a la rana*  
 CL.3.M.SG kick-PERF.3PL DOM DET.F.SG frog.F.SG  
 ‘They kicked out the frog.’

Asháninka-Spanish (Mayer Dataset 2016)

Shipibo-Spanish shows exactly the opposite distribution from Asháninka-Spanish for marked (55%) and unmarked objects (45%). The distribution of DOM follows a clear preference of animacy closely followed by the thematic role PATIENT with little attention to the feature definiteness and thematic role. Ergative alignment in Shipibo seems to be a possible cause for the lack of DOM with a human and definite object in (3).

- (3) *Le molest-a Ø el niño*  
 CL.3SG bother-PRES.3SG Ø DET.M.SG child-M.SG  
 ‘He bothers the child.’

Shipibo-Spanish (Sánchez Dataset 2002)

Finally, monolingual Spanish (LSCV) shares a similar distribution as Shipibo-Spanish in terms of marked and unmarked objects however with a slightly bigger margin between marked (60%) and unmarked objects (40%). LSCV shares the hierarchical distribution of features and thematic roles with Shipibo-Spanish however with two important differences. For one, the features definiteness and animacy (61.5% and 56.1% respectively) play a significantly stronger role in marking than the thematic roles PATIENT and THEME (26.2% and 23.8% respectively). Extension of DOM to inanimate, definite, patient direct object in clitic doubled DPs as in (4) and dislocated constructions is common in LSCV. These kinds of extensions of DOM can be linked to LSCV as an acquisitional and mainly oral variety, where speakers are located on a continuum dependent on variable access to formal instruction.

- (4) *Lo licu-o a-l ajo*  
 CL.3.M.SG blend-PRES.1SG DOM-DET.M.SG garlic.M.SG  
 ‘I blend the garlic.’

LSCV (Mayer 2017, 65)

A unifying factor of our different data sets is a complex interaction of direct or indirect contact with typologically different languages, language ecological factors (Mufwene 2002) and stages of language acquisition (Lardiere 2005). We find common syntactic properties across a continuum of speakers and variability in morphological marking, lack of DOM with human objects and general extension to definite inanimate objects in combination with non-agreeing clitics. Overall, definiteness is found to work as a semantic restriction for topic-worthy objects overriding animacy constraints. Due to these facts and as DOM preferably occurs in clitic doubled and dislocated structures, we will analyze the interaction of the above factors as a strong tendency to mark information-structure (Mayer/Sánchez 2016; Mayer 2017).

Given the typological differences between the languages in contact with Spanish and the complex array of syntactic conditions and semantic features that yield great variability of DOM and its relationship to clitic doubling and dislocated structures in Spanish, we formulate the following two research questions:

1. Are there differences in the Spanish DOM patterns found in clitic doubling and clitic dislocated structures in contact bilingual varieties depending on the typological characteristics of the contact language? Do monolingual contact varieties differ too?
2. What is the effect of animacy, definiteness and thematic role on the presence or absence of DOM in bilingual Spanish clitic doubling and dislocated structures in contact bilingual and monolingual varieties?

We formulate the following hypotheses:

1. DOM patterns in clitic doubling and clitic dislocated structures in contact varieties are affected by differences in the configuration of the morphology of the contact language. We expect the Quechua group to exhibit more DOM than the other groups as a result of similarities with Spanish in nominative/accusative alignment. We expect the Shipibo and the Asháninka groups to exhibit lower frequencies of DOM due to differences in alignment between these languages and Spanish.
2. We expect animacy to have an important role in the acquisition of Spanish due to its relevance in the input. Given that none of the indigenous languages has overt definite determiners we expect animacy to play a greater role than definiteness in all bilingual groups but possibly less so in the contact monolingual individuals given the availability of definite determiners. Finally, we expect thematic role to play a more salient role in Asháninka Spanish given the sensitivity to thematic role in the morphology of Asháninka.

In the next Section, we describe the typological differences in the argument marking systems of Shipibo, Huánuco Quechua, Asháninka and Spanish and address the difficulties arising for bilingual speakers from typologically different systems.

## 4 Typological differences in argument marking systems

### 4.1 Argument-marking in Shipibo

In this Section we present the different properties of case marking in each of the indigenous languages and in Spanish. As shown below, each of the three languages differs from Spanish with respect to case marking in significant ways although of the three, Quechua is closer in terms of case alignment.

Shipibo is part of the Panoan language family, it is an agglutinative language with ergative/absolutive alignment in its nominal and pronominal systems. It has null objects and SOV word order (Valenzuela 2010). In an ergative/absolutive system, pronominal subjects of most intransitive verbs as well as overt objects of most transitive verbs are marked as ergative. In (5a), the pronominal form *jato* is the subject (S) of an intransitive verb and in (5b) the object (O) of a transitive verb.

- (5) a. *Moa-ra jato bo-kan-ai*  
 already-EV 3P:ABS go.NON.SG-PL-INC  
 ‘They are leaving already.’

(Valenzuela 2010, 71)

- b. *Ja-n-ra jato keyyo-ke*  
 3-ERG-EV 3P:ABS finish-CMPL  
 ‘S/he exterminated them.’

(Valenzuela 2010, 71)

As shown in (6), subjects of transitive verbs have a different pronominal form *jabaon*.

- (6) *Jaská-a-xon-ki ja-baon no-a onan-ma-a iki*  
 SO-DO.T-A-HSY2 3-PL:ERG 1P-ABS know-CAUS-PP2. AUX  
 ‘And then they (our grandmothers) taught us (the activities that are proper of women).’

(Valenzuela 2010, 71)

When a third person subject co-occurs with a third person and object, verbal subject and object agreement may be null altogether as shown in (7).

- (7) *Nima<sub>i</sub> oin-xon-ra, Jose-kan kena-Ø-<sub>i</sub>ke*  
 Nima see-SS.TR.PRT Jose-ERG call-Ø-PERF  
 ‘When he saw Nima, José called him.’

(Loriot/Lauriault/Day 1993, 56)

Given the ergative/absolutive alignment typically found in Shipibo, there is to the best of our knowledge no evidence of Differential Object Marking.

## 4.2 Argument-marking in Asháninka

Asháninka is an Arawakan language with VS, VO basic constituent order, nominative-accusative alignment in transitive clauses and fluid or split transitivity in intransitive clauses (Payne/Payne 2005; Mihás 2015, 5). Fluid or split transitivity refers to the lexical meaning of the verb root being open to an increase or decrease in valency that determines transitivity. Bound morphemes mark obligatorily A/S (agent/subject) and P/O (patient/object) arguments on the verb. A/S arguments can occur pre- or postverbally and are marked according to semantic role. Of the three indigenous languages discussed here, Asháninka is the only one to mark gender with non-masculine gender as the default.

In transitive clauses in nominative/accusative alignment as in (8), the basic constituent order is VO, with compulsory marking of the subject A with a verbal prefix and the object O with a verbal gender specific suffix (*-ri* masculine, *-ro* non-masculine). Note that these gender-specific forms also mark the semantic role of theme in transitive constructions (Mihás 2015, 200).

- (8) *No-pos-ak-i-ri*  
 1SG.A-hit-PFV-REAL-3M.O  
 ‘I hit him.’

(Mihás 2015, 441)

In transitive clauses co-referential overt pronouns may co-occur as topics additionally to the postverbal object marker as shown in (9) but they are not marked for case.

- (9) *Iri no-pos-ak-e-ri ir-ako-ki*  
 3M.TOP 1SG.A-hit-PFV-REAL-3M.O 3M.POSS-arm-LOC  
 ‘I hit him on his hand.’

(Mihás 2015, 193)

Finally, Asháninka also exhibits ambitransitive transitive clauses which lack fixed transitivity thus allowing for increase or decrease of their valency even without object morphological marking on the verb as in (10). Such clauses exhibit fluid transitivity as explained above.

- (10) *n-a-ak-i kaniri*  
 1SG.S-take-PFV-REAL manioc  
 ‘I obtained manioc roots.’

(Mihás 2015, 194)

Once again, crucial for our present study is that the dislocated object is not marked for accusative case. This is also the case in (11) where the left dislocated object is unmarked too. Gender-specific suffixes appear on both verbs but not on the object.

- (11) *pakitsa ari a-ñ-ak-e-ri apaata a-shiyakant-ak-e-ri*  
 eagle PP 1PL.A-see-PFV-IRR-3M.O later.on 1PL.A-take.picture-PFV-IRR-3M.O  
 ‘The eagle, there we’ll see it, when we later on take pictures of it.’

(Mihás 2015, 616)

As in the case of Shipibo, to the best of our knowledge, there is no evidence of DOM in Asháninka.

### 4.3 Argument marking in Huánuco Quechua

Huánuco Quechua is a nominative/accusative, agglutinative language with SOV word order as shown in (12):

- (12) *Juan Tumas-ta maga-n*  
 Juan Tomas-OBJ hit-3  
 ‘Juan hits Tomas.’

(Weber 1996, 51)

In this sentence, the direct object is marked with the accusative suffix *-ta* and there is no overt nominative marking. Like Shipibo, Asháninka and Spanish, Huánuco Quechua has null subjects as shown in:

- (13) *Maga-ma-ra-n*  
 hit-1-PAS-3  
 ‘S/He hit me.’

(Weber 1996, 249)

When first or third person subjects co-occur with a third person object, the latter is not marked on the verb as shown in (14a) and (14b):

- (14) a. *Maga-Ø-n*  
 hit-Ø-3  
 ‘S/He hits (him, her, it).’  
 b. *Apa-mu-na-:-paq ka-yka-sah-:*  
 bring-LEJ-SUB-1P-PR be-IMP-PRF-1  
 ‘I was about to bring (it).’

(Weber 1996, 168)

While there is no evidence of DOM in Huánuco Quechua of the type that is sensitive to semantic features such as definiteness, animacy or thematic role, some objects inside subordinate clauses may remain unmarked for accusative case in Huánuco Quechua as in:

- (15) *Cristobal-Ø asi-q aywa-ska-:*  
 Cristobal search-SUB go-PRF-1  
 ‘(I) went to look for Cristobal.’

(Weber 1996, 250)

This indicates the possibility of some alternation between direct object marking and lack of direct object marking on the noun. Finally, there is evidence in Huánuco Quechua of left and right topic dislocated structures as shown in (16). The dislocated element is marked with the topic suffix *-ga*.

- (16) a. *Hatun wasi-ta-ga muna:*  
 big house-OBJ-TOP want-1  
 ‘I want a big house.’

- b. *Wasita muna: hatunta-qa*  
 House-OBJ want-1 big-OBJ-TOP  
 ‘I want a big house.’

(Weber 1996, 515)

The existence of these dislocated structures may facilitate the acquisition of clitic doubled and clitic doubling structures.

#### 4.4 Spanish argument marking

Spanish is a nominative/accusative language with minimal case marking on personal pronouns, some relative pronouns, clitics and a range of direct objects. In (17), the subject bears unmarked nominative case, the definite and human direct object receives accusative case through DOM.

- (17) *La niñ-a am-a a-l niñ-o*  
 DET.F.SG girl-F.SG love-PRES.3SG DOM-DET.M.SG boy-M.SG  
 ‘The girl loves the boy.’

(Mayer 2017, 58)

Direct object arguments can be marked by feature-agreeing verbal clitics as anaphors in the examples in (18). Clitic features include person, number, gender (in accusative/direct object clitics) or case (in dative/indirect object clitics). They are phonologically bound to their verbal host and can appear as proclitics with finite verbs (18a) and enclitics in non-finite contexts (18b). As subject information is encoded in verbal inflection, the overt expression of pronominal subjects depends on their informational status (18a).

- (18) a. *(Ella) lo am-a*  
 PRO.3.F.SG CL.3.M.SG love-PRES.3SG  
 ‘She loves him (the boy).’  
 b. *Quier-e am-ar-lo*  
 want-PRES.3SG love-INF-CL.3.M.SG  
 ‘She wants to love him.’

As mentioned in the Introduction, DOM in Spanish marks a range of objects ranked for prominence on a two-dimensional scale based on the interaction of animacy and definiteness and also semantic role (Bossong 1991; 2003; Aissen 2003). According to Leonetti (2008), in most varieties of General Spanish, human



and animate patient arguments as in (19a) are generally marked, they are considered highly topical; indefinite animate objects may be optionally marked because of their specificity/identifiability status as in (18b), and inanimate non-specific core object arguments are excluded from marking (19c).

- (19) a. *Amalia abraz-ó a su tí-a*  
 Amalia hug-PERF.3SG DOMDOM POSS aunt-F.SG  
 ‘Amalia hugged her aunt.’
- b. *El ladrón mat-ó (a) un perr-it-o*  
 DEF.M.SG thief.M.SG kill-PERF.3SG DOM INDEF.M.SG dog-DIM-M.SG  
 ‘The thief killed a little dog.’
- c. *María compr-ó \*a un pian-o nuev-o*  
 María buy-PERF.3SG DOM INDEF.M.SG piano-M.SG new-M.SG  
 ‘Maria bought a new piano.’

In clitic doubling and related configurations expressing information structure such as CLLD and CLRD clitics play an important role in Spanish syntax in object-verb agreement as head markers in conjunction with dependent marking by DOM (Nichols 1986).

This marking strategy is characterized by a continuum of DOM and CLD with both marking strategies exhibiting great diachronic and synchronic variability across time and space. In some varieties of Spanish, DOM in CLD is restricted to pronominal direct objects. Historically, the extension of optional marking to inanimate proper nouns has been documented in the Old Spanish text *Cantar de mio Cid* and it has been related to a specific relationship between the subject and the object which allows for referential identification (Melis 1995; Bresnan/Aissen 2002, 91). Contemporary liberal clitic doubling varieties such as Buenos Aires Spanish and Lima Spanish extend DOM to animate specific (20a) (Mayer 2008; 2017; Zdrojewski/Sánchez 2014) and in the case of Buenos Aires Spanish to inanimate specific direct object arguments (20b).

- (20) a. *La bes-o /ve-o a la maestr-a*  
 CL.3.F.SG kiss-PERF.1SG /see-PERF.1SG DOM DET.F.SG teacher-F.SG  
 ‘I kiss/see the teacher.’
- b. *Lo quier-o mucho a este arbol-ito*  
 CL.3.M.SG like-PRES.1SG much DOM DEM.M.SG tree-DIM.M.SG  
 ‘I like this little tree very much.’

(Suñer 1989, 379)

Extension of optional DOM in contemporary varieties of Spanish in (19b) points to information structural marking (Dalrymple/Nikolaeva 2011; Mayer/Sánchez 2016; Mayer 2017). Clitic configurations such as left dislocated structures (CLLD) as in (21a) and right dislocated structures (CLRD) as in (21b) also known as topicalized structures show the same patterns of head and dependent marking.

- (21) a. *A la maestr-a la salud-é ayer*  
 DOM DET.F.SG teacher-F.SG CL.3.F.SG greet-PERF.3SG yesterday  
 ‘The teacher, I greeted her yesterday.’
- b. *La salud-é ayer a la maestr-a*  
 CL.3.F.SG great-PERF.3.SG yesterday DOM DET.F.SG teacher-F.SG  
 ‘I greeted her, the teacher yesterday.’

The syntactic and semantic factors that determine marking of direct objects including clitic agreement in most varieties of Spanish are summarized in Table 1. They show that DOM depends on the interaction of the referential features animacy and definiteness allowing for variability in animacy with no significant role for thematic roles.

**Table 1:** Spanish direct object marking, case, clitic agreement (CL) and thematic roles (adapted from Mayer 2017, 61).

	Direct Object	
	marked direct object	unmarked direct object
<b>Syntax</b>		
<b>Case</b>	accusative <i>a</i>	∅
<b>CL Agreement</b>	yes	no
<b>Semantics</b>		
<b>Object type</b>	PAT/THEME	PAT/THEME
	[±anim/+def]	[-anim/-def/-spec]
	[+spec/+top]	

Given the complex array of syntactic conditions and semantic features yielding variability of DOM and its relationship to clitic doubling in Spanish, two questions arise immediately. These are a) how do bilinguals in language contact situations map functional features onto morphology, and b) how do monolingual contact speakers navigate the variability arising from the complex settings for DOM. As we will discuss in the rest of the chapter, most varieties of contact Spanish show even greater variability than non-contact varieties as they may lack DOM or extend case marking to new contexts.

## 4.5 The bilingual puzzle

In terms of differences and similarities of the bilingual systems, differently from Spanish, Shipibo exhibits case marking in nouns but the verb is unmarked for subject and object agreement. Important for clitic doubling structures is the fact that while Spanish clitics have person, number and case/gender features, the Shipibo pronominal system lacks gender and definiteness. In order to acquire clitic structures, Shipibo-Spanish bilinguals need to master the assembly of pronominal agreement with multiple features into one single morpheme (Lardiere 2005). Furthermore, they need to acquire sensitivity to semantic features such as definiteness and animacy as triggers of DOM.

Quechua argument marking is closest to Spanish argument marking in relation to case marking (nominative/accusative) as expressed by the accusative suffix *-ta* in Quechua and the DOM marker *a* in Spanish. As our data show, common nominative/accusative alignment with Spanish facilitates the acquisition of DOM for Quechua-Spanish bilinguals despite the morphological differences between the Spanish preposition *a* and the Quechua case marking suffix *-ta*. Like Shipibo, Quechua lacks marking of (in)definiteness and gender when the subject and the object are third person.

Asháninka, unlike Shipibo, shares two important similarities with Spanish, which could contribute significantly to the acquisition of Spanish argument marking in terms of DOM and clitic structures by Asháninka-Spanish bilinguals. These are a set of feature specific bound morphemes that mark objects as verbal affixes and a set of free personal pronouns to mark information structure in preverbal or postverbal position. Given the fact that the first one of these sets is specified for person, number and gender, it could play an important role in the acquisition of Spanish clitics. Also, the fact that these two sets of bound and free forms can co-occur and allow thus for the possibility to restructure word order pragmatically, could potentially facilitate the acquisition of clitic doubled and dislocated structures.

In the next Section, we address the ethnographic description, methodology, and data collection procedures.

## 5 Ethnography and methodology

In order to address this question, we analyzed the data sets of 3 bilingual L2 Spanish speaking groups in contact with typologically different languages living in their communities in rural and urban areas. By way of comparison, the analysis

was extended to a monolingual Spanish acquisitional variety in Lima. The focus is on the interaction of animacy and definiteness in conjunction with the thematic roles, PATIENT and THEME, in Differential Object Marking in clitic doubled and dislocated arguments structures.

## 5.1 Participants

For the bilingual argument structures, the data sets are based on fieldwork with L2 learners of Spanish with a Shipibo community in Lima, with a Quechua community in Huánuco, a rural Andean area, and with two Asháninka communities in the central Peruvian Amazon in the Junín province. The monolingual LSCV data set is a longitudinal study of three months with two sisters and the daughter of one of them living in two socioeconomically different neighbourhoods in Lima. Participant details in terms of numbers, gender, age range, education and place of data collection are specified in Table 2 below.<sup>4</sup>

**Table 2:** Ethnographic details of participants.

<i>Communities</i>	$\Sigma$	<i>Gender</i>	<i>Age</i>	<i>Education</i>	<i>Place</i>
<i>Shipibo</i>	19	10F, 9M	29–56	P:9F,3M S:1F,6M	Cantagallo, Lima
<i>Huánuco Quechua</i>	14	8F, 6M	22–45	P:6F,4M S:3F,2M	Chaglla, Huánuco
<i>Asháninka</i>	21	13F, 8M	18–88	P:5F,1M S:4F,3M PS:4F,4M	Arizona, Puerto Ocopa, Junín
<i>Monolingual LSCV</i>	3	3F	16–45	S:3F	Ventanilla, Surco, Lima

## 5.2 Methodology

For the Quechua and Asháninka groups a picture-based elicitation task was used to elicit a narration. The task was based on Sanchez's (2003) adaptation of Mayer and Mayer's (1992) frog story. A very similar story was used to collect the Shipibo

<sup>4</sup> For the abbreviations in education, P refers to primary schooling, S to secondary schooling and for the Asháninka participants PS to postsecondary which in this case refers to an Agricultural Technical Institute located in the Puerto Ocopa community.

Spanish data using a figure-based narrative elicitation task. In order to investigate the research questions, firstly an ethno-biographical survey about language history, preferences for language use including language attitudes towards their indigenous languages and Spanish was conducted, followed by the oral elicitation tasks. Both were orally administered and digitally recorded. In the case of the Asháninka group these were followed by a short written Spanish language proficiency test adapted from the cloze section of the DELE used in Cuza et al. (2013). Older and illiterate participants were read the proficiency test and they completed it orally. An important point to make here is that specifically in Andean and Amazonian bilingual Spanish rurality and orality are predominant.

The Shipibo narratives were collected in 2002 in Lima, Peru, the Huánuco Quechua narratives in 2005 in Chaglla, Huánuco, Peru and the narratives of two Asháninka bilingual groups in Arizona Portillo and Puerto Ocopa, Junín, Peru in 2016. In the case of the Asháninka groups, argument structures were elicited in order to test the effect of similarities in case marking, gender marking and affixal verbal subject and object marking.

Data collection with the three monolinguals took place in early 2006 in Lima over the course of three months in the form of recorded elicited narratives about their lives, their regular activities and general life stories. Two participants are siblings, born to Quechua-speaking parents in Iquitos who migrated to Lima at ages 14 and 16. Their inclusion is warranted by the fact that the group exhibits gender and DOM and presents as such an invaluable opportunity for reasons of comparison. Data for all groups was transcribed using the CHILDES system. The Asháninka group data was transcribed using ELAN and PRAAT.

## 6 Data sets and results

In this Section, we present the comparative results for the features under study of the three bilingual groups and the three monolingual individuals followed by a discussion of these results and brief conclusion.

### 6.1 Data sets and coding

The size of the data set in terms of individual tokens in accordance with the distribution of clitic doubling and dislocated structures across all four groups as well as the overall tokens is shown in Table 3 below.

**Table 3:** Distribution of clitic doubling and dislocated structures across all groups.

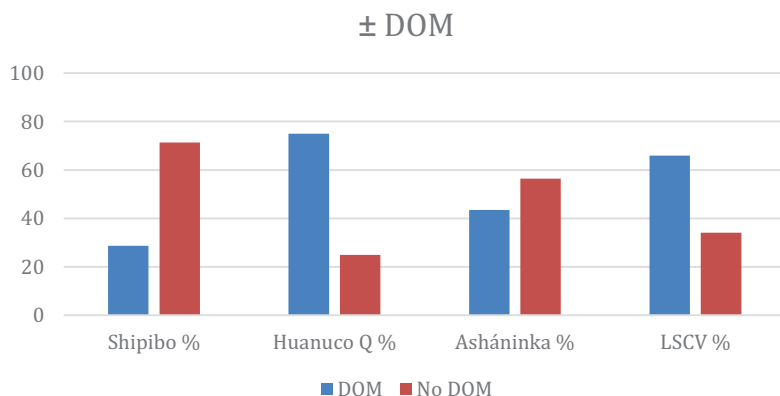
	CLD (%)	CLLD (%)	CLRD (%)	Total
<b>Shipibo</b>	91% (62t)	3% (2t)	6% (4t)	100% (68t)
<b>Asháninka</b>	72% (73t)	23% (23t)	5% (5t)	100% (101t)
<b>Huánuco</b>	77% (40t)	13% (7t)	10% (5t)	100% (52t)
<b>LSCV</b>	54% (76t)	26% (38t)	20% (28t)	100% (142t)

For all four groups all transitive verbs were coded according to each one of the CLD, CLLD and CLRD structures for i) clitics and absence and presence of DOM, and ii) animacy, definiteness and thematic role of the object. The individual occurrences for each structure and the overall occurrences are shown in Table 3. In terms of animacy we counted humans and animals as animate and everything else as inanimate.<sup>5</sup> Objects with a definite determiner were coded as definite and objects with an indefinite determiner or bare nouns were coded as indefinite. Thematic roles were coded depending on transitivity of the verbs, e.g. PATIENTS undergo visible changes and THEMES remain unchanged. Doubtful cases that could not be resolved by double checking audio files and PRAAT were excluded from the analysis all together.

## 6.2 Overall production of DOM

As shown in Figure 2, all bilingual groups and the monolingual individuals produce DOM to varying degrees and pattern in interesting ways. Expression of DOM in the Huánuco Quechua-Spanish dataset is strikingly different from the other data sets by marking 75% of all direct objects produced within clitic structures and leaving 25% unmarked. LSCV exhibits a similar distribution with 66% marked and 34% unmarked direct objects. Data from rural Shipibo-Spanish and rural Asháninka-Spanish show a different scenario. The Shipibo group exhibits the highest percentage of unmarked DPs (71%) over DOM marked DPs (29%), closely followed by the Asháninka Spanish group exhibiting a preference for unmarked DPs although less pronounced (56% unmarked vs. 44% marked). Overall, the groups show a significant difference in their distribution ( $\chi^2(3, N = 400) = 7.8, p < .005$ ).

<sup>5</sup> We acknowledge that animacy in indigenous languages may have other scales.



**Figure 2:** Expression of DOM across groups in CLD, CLLD and CLRD structures.

When comparing the production of DOM in clitic doubling constructions in Asháninka-Spanish bilinguals and Quechua-Spanish bilinguals, a general linear model showed a difference ( $\beta = 3.05$ ,  $SE = .43$ ,  $z = 7.05$ ,  $p < .001$ ). That difference was not found when comparing the Shipibo-Spanish and the Asháninka-Spanish bilinguals ( $\beta = .34$ ,  $SE = .41$ ,  $z = .81$ ,  $p = .42$ ).

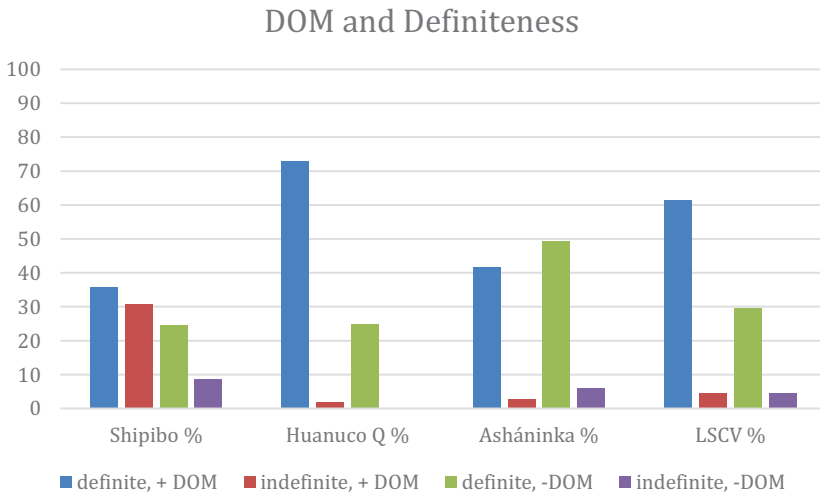
### 6.3 DOM and conditioning factors

Looking at three of the data sets (Huánuco Quechua, Asháninka, and Shipibo)<sup>6</sup> a general lineal model was fitted for definiteness, human, animacy and thematic role as fixed factors, in order to determine their effect on DOM production. There was a main effect of definiteness ( $\beta = 2.24$ ,  $SE = .49$ ,  $z = 4.55$ ,  $p < .001$ ), human ( $\beta = 2.25$ ,  $SE = .62$ ,  $z = 3.66$ ,  $p < .001$ ), animacy ( $\beta = 3.01$ ,  $SE = .51$ ,  $z = 5.90$ ,  $p < .001$ ), and thematic role ( $\beta = -0.59$ ,  $SE = .27$ ,  $z = -2.19$ ,  $p = .028$ ). Given these main effects found for each conditioning factor, in this Section, we present the distribution of DOM according to these factors to the exclusion of human, which we will treat as subsumed under animate.

<sup>6</sup> The LSCV individuals data were not included in the model.

### 6.3.1 DOM and definiteness

The distribution of DOM with definite and indefinite DPs according to group is provided in Figure 3 below. The Huánuco Quechua group exhibits the highest percentage of DOM with definite DPs (73%), followed by the LSCV monolinguals (61.4%), the Asháninka group (41.6%), and the Shipibo group (35.8%).



**Figure 3:** DOM with (in)definiteness across groups.

The Asháninka group shows the highest percentage of unmarked definite DPs (49.5%), followed by the LSCV (29.5%) and the Shipibo group (24.7%). The Huánuco Quechua group and the Shipibo group's percentage of unmarked definite DPs are practically equal and the lowest of all groups (25% and 24.7% respectively). Extension of DOM to indefinites is highest in the Shipibo group (30.9%) followed by very low numbers from LSCV (4.5%), the Asháninka group (3%) and the Huánuco Quechua group (1.9%). Again, independence of preference is proven by a chi-square test  $\chi^2(6, N = 300) = 12.5, p <<.0003$ .

A general linear model (excluding LSCV data) found interactions between definiteness and language in the case of Huánuco Quechua ( $\beta = 5.02, SE = 0.87, z = 5.7, p = 0.00$ ) and Shipibo ( $\beta = 2.39, SE = 1.02, z = 2.34, p > 0.01$ ). When compared to the Asháninka data, the probability that a speaker of Huánuco Quechua or Shipibo would produce DOM with a definite DP is higher, especially among the Huánuco Quechua speakers.

The following examples illustrate the use, extension or lack of DOM with definite and indefinite DP's representative of each of the language varieties under study.



## DOM with definite DP in CLD

- (22) *Le ve-mos también a-l sapo*  
 CL.3SG see-PRES.1PL also DOM-DET.M.SG toad.M.SG

‘We also see the toad.’ (H15, 82)

Huánuco Quechua-Spanish (Sánchez Dataset 2005)

## DOM with indefinite DP in CLD

- (23) *Sí, le he ayud-ado a una señora.*  
 Yes CL.3SG have.PERF.1SG help-PART DOM INDEF.F.SG lady.F.SG

‘Yes, I have helped a Lady.’ (LSCV1)

LSCV (Mayer Dataset 2006)

## DOM with definite DP in CLRD

- (24) *Pued-e morder-le el perr-o a-l*  
 can-PRES.3SG bite.INF-CL.3SG DET.M.SG dog-M.SG DOM-DET.M.SG

*niñ-o*

boy-M.SG

‘The dog can bite the boy.’ (H5,13)

Huánuco Quechua-Spanish (Sánchez Dataset 2005)

## DOM with definite DP in CLLD

- (25) *Yo a mi sobrin-ita la recib-ía acá*  
 PP.1SG DOM POSS.1SG niece-DIM.F.SG CL.3.F.SG receive-IMP.F.3SG here

‘I received my niece here with me.’ (LSCV1)

LSCV (Mayer Dataset 2006)

## Lack of DOM with definite DP in CLD

- (26) *El niño lo agarr-ó Ø el sap-ito*  
 the boy CL.3.M.SG grab-PERF.3SG DOM DET.M.SG toad-DIM.M.SG

‘The boy grabbed the little toad.’ (A 15)

Asháninka-Spanish (Mayer Dataset 2016)

## Lack of DOM with indefinite DP in CLD

- (27) *Le sac-ó un lor-o*  
 CL.3SG took out-PERF.3SG INDEF.M.SG parrot-M.SG

‘(He) took out a parrot.’ (SH2)

Shipibo-Spanish (Sánchez Dataset 2002)

Lack of DOM with definite DP in CLRD (the only occurrence)

- (28) *Un gente le sac-amos la caja*  
 INDEF.SG person CL.3SG take.out-PRES.1PL DET.F.SG box.F.SG  
*Ø este su lor-o*  
 DOM DEM.M.SG POSS parrot-M.SG  
 ‘Somebody took this parrot out of the box.’

Shipibo-Spanish (Sánchez Dataset 2016)

Lack of DOM with definite DP in CLLD

- (29) *Y el sap-ito lo llev-ó*  
 and DET.M.SG toad-DIM.M.SG CL.3.M.SG carry-PERF.3SG  
*Ø la tortuga*  
 DOM DET.F.SG turtle.F.SG  
 ‘And the little toad carried the turtle.’ (PO41)

Asháninka-Spanish (Mayer Dataset 2016)

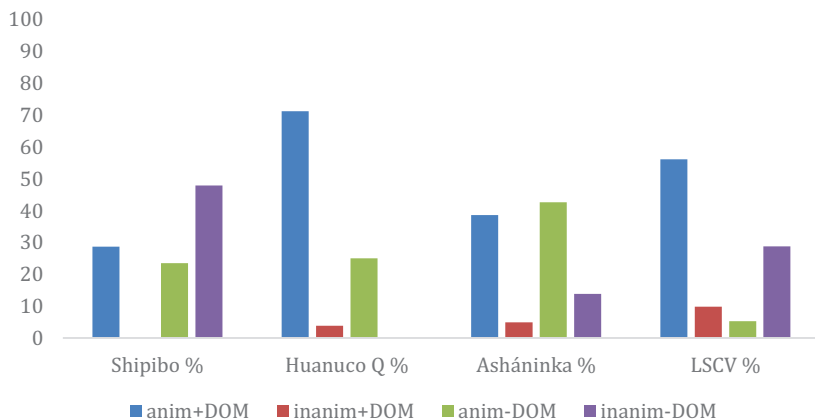
### 6.3.2 DOM and animacy

In the overall results for animacy patterns shown in Figure 4, the Asháninka group leads with the highest percentage of animates (84.3%) and lowest for inanimates (15.7%) followed by the same pattern by the Shipibo group with higher frequency of animates (61%) than inanimates (39%). The Huánuco Quechua group presents a similar pattern to the LSCV monolinguals with a higher production of inanimates (57%) than animates (43%). The monolingual individuals exhibit a reversal from the definiteness results with a higher (59.6%) percentage of inanimate than animate DPs (40.4%). Again, independence of preference is corroborated by  $\chi^2(3, N = 400) = 7.8, p <<.0067$ .

The distribution of DOM and animacy for the four groups shows that in all four groups, DOM occurs with animate DPs led by the Huánuco Quechua group with a very high result (71.2%), followed by the LSCV group (56.1%) and the Asháninka group (38.6%), with the Shipibo group presenting the lowest percentage (28.7%).

The Shipibo group stands out from the other three groups with close percentages for DOM with animate DPs (28.7%) and lack of DOM with animate DPs (23.5%). There is no overextension of DOM to inanimates in the Shipibo group and inanimates without DOM reach 47.8%. All other groups extend DOM to a very low extent to inanimates led by LSCV (9.8%), followed by Asháninka (5%) and Huánuco Quechua (3.8%) and again, a chi-square test corroborates independence of preference  $\chi^2(6, N = 300) = 12.5, p <<.0029$ .

## Animacy and DOM



**Figure 4:** Comparative data for DOM and animacy across groups.

The following examples illustrate the unexpected patterns of lack of DOM with animate DPs (30), overextension of DOM to inanimate DPs (31) and lack of DOM with animate DP in CLLD (32).

Lack of DOM with animate DP in CLLD

- (30) *El niñ-o le riñ-e Ø ese perr-o*  
 The boy-M.SG CL.3SG tell.off-PRES.3SG DOM DEM.M.SG dog-M.SG  
 ‘The boy tells off that dog.’

Shipibo-Spanish (Sánchez Dataset 2002)

DOM with inanimate DP in CLLD

- (31) *A-l aji la hag-o hervir*  
 DOM-DET.M.SG chili.M.SG CL.3.F.SG make-PRES.1SG boil-INF  
 ‘I’ll boil the chili.’(LSCV2)

LSCV (Mayer Dataset 2006)

Lack of DOM with animate DP in CLLD

- (32) *El otr-o le han dej-ado cre-o*  
 DET.M.SG other-M.SG CL.3SG PERF.3PL left-PART think-PRES.1SG  
 ‘The other one, I think they left him.’ (PO50)

Asháninka (Mayer Dataset 2016)

The only instances of DOM marking inanimate DPs in CLRD and CLLD dislocated structures were found in the monolingual LSCV data in LSCV1. The other three bilingual groups did not show extension of DOM to those information structures.

DOM with inanimate DP in CLLD

- (33) *Por ejemplo, yo a-l fresco*  
 for example PRO.1SG DOM-DET.M.SG juice.M.SG  
*le he licu-ado*  
 CL.3SG have-PAST.1SG blend-PART

‘For example, I blended the fresh juice.’ (LSCV1)

LSCV (Mayer Dataset 2006)

Within unmarked animate DPs, we looked at human DPs, usually ranked highest in the animacy hierarchy, and we also found evidence of lack of DOM with human DPs as illustrated in (34) and (35) for CLD constructions and (36) and (37) for CLLD and CLRD respectively.

Lack of DOM with human DP in CLD

- (34) *Y lo agarr-a Ø el niño*  
 and CL.3.M.SG grab-PRES.3SG DOM DET.M.SG boy-M.SG  
 ‘And he grabbed the boy.’ (H16, 98)

Huánuco Quechua-Spanish (Sánchez Dataset 2005)

- (35) *El tortug-a le agarr-a Ø el niño*  
 DET.M.SG turtle-F.SG CL.3SG grab-PRES.3SG DOM DET.M.SG boy-M.SG  
*en su pantalón*  
 PREP POSS trouser

‘The turtle grabs the boy by his trousers.’ (PO, 43)

Asháninka-Spanish (Mayer Dataset 2016)

Lack of DOM with human DP in CLLD

- (36) *Ella yo le llev-é a Iquitos*  
 PP.3.F.SG PP.1SG CL.3SG take-PERF.1SG LOC Iquitos  
 ‘The other one, I took her to Iquitos.’ (LSCV1)

LSCV (Mayer Dataset 2006)

Lack of DOM with human DP in CLRD

(37) *Cuando ella la cort-aban a la barriga*  
 when PP.F.SG CL.F.SG cut-IMP.F.3PL DOM DET.F.SG tummy.F.SG  
 ‘When they cut her tummy.’ (LSCV2)

LSCV (Mayer Dataset 2006)

In terms of DOM as a function of animacy and language, the linear model found animacy to be a predictor of DOM ( $\beta = 4.05$ ,  $SE = 0.39$ ,  $z = 10.26$ ,  $p < 0.01$ ) and Huánuco Quechua was also found to be a predictor of DOM ( $\beta = 2.43$ ,  $SE = 0.36$ ,  $z = 6.62$ ,  $p < 0.01$ ).

### 6.3.3 DOM and thematic roles

The data in Figure 5 show a preference for PATIENTS (64%) over THEMES (36%) overall. However, the Huánuco Quechua group defies this trend by a clear preference for THEMES (66%) over PATIENTS (34%). The Shipibo, LSCV and Asháninka groups show the opposite pattern with the Shipibo group leading the pattern with a preference for PATIENTS (76.7%) over THEMES (23.2%), followed by the Asháninka groups strongly preferring PATIENTS (71%) over THEMES (29%). The LSCV group follows this distribution closely with similar percentages for PATIENTS (61.4%) over THEMES (38.6%). Independence of preference was clearly established by a chi-square test  $\chi^2(6, N = 300) = 12.5$ ,  $p < .005$ .

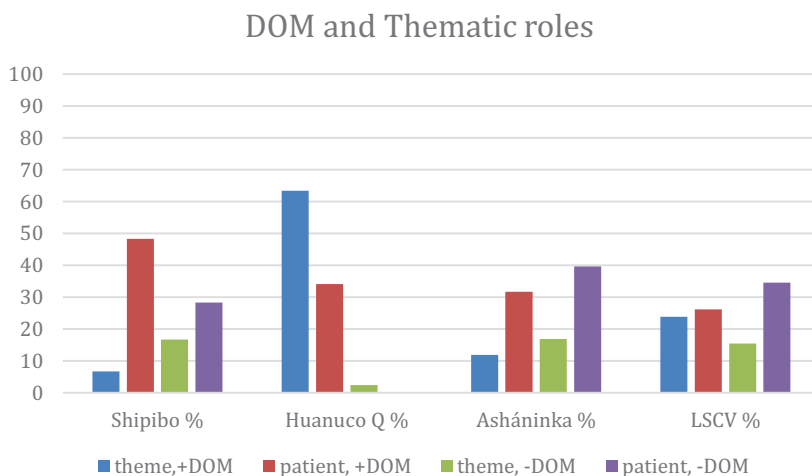


Figure 5: Distribution of thematic roles across groups.

When looking at the distribution of DOM marking according to thematic role, we find that the Huánuco Quechua group has a unique pattern in which the frequency of direct objects with theme roles marked with DOM is the highest (63.4%) followed by those with patient roles with DOM (34.1%) and by unmarked direct objects with a theme role (2.4%). Unlike the Huánuco Quechua group, the Shipibo group shows a higher frequency of DOM marking with PATIENTS (48.3%), than THEMES (6.7%). The Asháninka group shows a similar pattern of preference for DOM with PATIENTS (31.7%) over THEMES (11.9%). Unlike the other groups, the LSCV individuals do not exhibit a clear preference for either and they only show a slightly higher frequency of DOM with PATIENTS (26.2%) than THEMES (23.8 %). The following examples illustrate the most frequent cases of DOM with THEME and PATIENT arguments in each group.

DOM with theme argument in CLD

- (38) *Y el tortuga también le*  
 and DET.M.SG turtle.F.SG also CL.3SG  
*mir-a a-l niñ-o*  
 look-PRES.3SG DOM-DET.M.SG boy-M.SG  
 ‘And the turtle also looked at the boy.’ (H16)  
 Huánuco Quechua-Spanish (Sánchez Dataset 2005)

DOM with patient argument in CLD

- (39) *Le muerde a-l lor-ito*  
 CL.3SG bite-PRES.3SG DOM-DET.M.SG parrot-DIM.M.SG  
 ‘(He) bites the little parrot.’  
 Shipibo-Spanish (Sánchez Dataset 2002)

DOM with theme argument in CLD

- (40) *Est-á mir-ándo-le a-l turtuga*  
 AUX-PRES.3SG look-GER-CL.3SG DOM-DET.M.SG turtle.F.SG  
 ‘He is looking at the turtle.’ (PO43)  
 Asháninka-Spanish (Mayer Dataset 2016)

DOM with theme argument in CLD

- (41) *Tien-e que est-ar viéndo-la a su*  
 has-PRES.3SG that be-INF look.after-CL.3.F.SG DOM POSS  
*hij-it-a*  
 daughter-DIM-F.SG  
 ‘She has to keep looking after her little daughter.’ (LSCV1)  
 LSCV (Mayer Dataset 2006)

For the coding of information structure, the Huánuco Quechua group shows no preference for either of the roles (for 7 CLLD 3 THEMES and 4 PATIENTS and 5 CLRD 2 THEMES and 3 PATIENTS) while the Shipibo group only marks PATIENTS in 2 instances of CLLD and even distribution of both thematic roles for 4 instances of CLRD. The last bilingual group, Asháninka-Spanish differs from both previous groups in terms of high numbers for CLLD and a pronounced preference for PATIENTS (21/23) to mark those with no clear preference for CLRD (3 THEMES, 2 PATIENTS) in (44).

DOM with theme argument in CLRD

- (42) *Y aquí el niño-to le está mir-ando*  
 and here DET.M.SG boy-DIM.M.SG CL.3SG is.PRES.3SG look-PART  
*molesto al sap-ito*  
 annoyed DOM-DET.M.SG toad-DIM.M.SG  
 ‘And here the boy is looking annoyed at the little toad.’ (H11)  
 Huánuco Quechua (Sánchez Dataset 2005)

DOM with patient argument in CLLD

- (43) *Al, el perro eh le pate-a*  
 DOM-DET.M.SG DET.M.SG dog-M.SG EH CL.3SG kick-PRES.3SG  
 ‘The, the dog, ehem, he kicked him.’  
 Shipibo (Sánchez Dataset 2002)

DOM with patient argument in CLLD

- (44) *Al otro sap-ito le hab-ían dej-ado*  
 DOM-DET.M.SG other toad-DIM CL.3SG PERF-3PL leave-PART  
*porque era muy gruñón.*  
 because is.IMPF.3SG very grumpy  
 ‘The other toad, they left him behind he was very grumpy.’ (PO 30)  
 Asháninka Spanish (Mayer Dataset 2016)

The monolingual group shows a clear preference for PATIENTS in both information structures with 26:10 in CLLD (45) and less so in CLRD with 17:11 in (46).

DOM with patient argument in CLLD

- (45) *A la yuca sí Ø pel-o*  
 DOM DET.F.SG yuca.F.SG yes peel-PRES.1SG  
 ‘The yuca, yes, I peel it.’ (LSCV1)

Mayer (Dataset 2016)

DOM with patient argument in CLRD

- (46) *Est-á loc-o la abuel-a porque no lo*  
 is-PRES.3SG crazy-M.SG DET.F.SG grandmother-F.SG because not CL.3.M.SG  
*oblig-ó así a su hij-o cuando Eva nac-ió*  
 oblige-PERF.3SG thus DOM POSS son-M.SG when Eva born-PERF.3SG  
 ‘She is crazy, the grandmother, because she did not oblige her son this  
 way when Eva was born.’ (LSCV2)

Mayer (Dataset 2016)

Huánuco Quechua was also a predictor for DOM, as a function of thematic role and language ( $\beta = 0.94$ ,  $SE = 0.33$ ,  $z = 2.86$ ,  $p = 0.00$ ), but no interactions were found. The monolingual group shows a more extended spread of features in terms of frequency.

The group shows a common strategy with DOM occurring most preferably with definite and animate objects. This is followed by a pronounced drop in frequency for unmarked patient, definite, and inanimate objects. The following bars show an array of combinations including DOM with patient and theme objects, as well as unmarked themes. Although with very low frequencies, these data also show animates without DOM and inanimates with DOM which reflect the differences between the three LSCV speakers. While LSCV3, the Lima-born youngest speaker only marks animate and definite objects, LSCV2 exhibits DOM with inanimate objects to a lesser degree than LSCV1. Lack of DOM with animates occurs in both to the same degree.

## 7 Discussion

The results we have presented above show that all factors under study play a role in determining DOM in the bilingual contact varieties under study and to some extent in the monolingual contact data. However, their effect differs across groups. Typological differences between the languages with which Spanish is in contact also play an important role as they do result in different frequency patterns. We interpret these results as the outcome of a complex interaction between contact with typologically distinct languages for the bilingual groups and language-specific ecological factors for all groups but especially for the monolingual individual data that exhibit the greater level of variability.

In response to our first research question, where we asked a) whether differences found in Spanish DOM patterns in information structures depended on typological characteristics of the contact language, and b) whether monolingual



varieties showed differences as well, our hypothesis was born out. We find that DOM patterns in clitic doubling and clitic dislocated structures in contact varieties are affected by differences in the configuration of the morphology of the contact language, as shown by the emerging Differential Object Marking systems across the bilingual and the monolingual group in Table 4.

**Table 4:** Emerging Differential Object Marking systems.

	<b>+DOM</b>	<b>-DOM</b>
<b>Huánuco Quechua-Spanish</b>	DEF > ANIM > THEME > PATIENT 73.1% > 71.2% > 63.4% > 34.1%	DEF = ANIM > THEME > PATIENT 25% = 25% > 2.4% > 0%
<b>Shipibo-Spanish</b>	PATIENT > DEF > ANIM > THEME 48.3% > 35.8% > 28.7% > 6.7%	PATIENT > DEF > ANIM > THEME 28.3% > 24.7% > 23.5% > 16.7%
<b>Asháninka-Spanish</b>	DEF > ANIM > PATIENT > THEME 41.6% > 38.6% > 31.7% > 11.9%	DEF > ANIM > PATIENT > THEME > 49.5% > 42.6% > 39.6% > 16.8%
<b>LSCV-monolinguals</b>	DEF > ANIM > PATIENT > THEME 61.4% > 56.1% > 26.2% > 23.8%	PATIENT > DEF > THEME > ANIM 34.5% > 29.5% > 15.5% > 5.3%

As hypothesized, the Quechua group data exhibits not only a higher frequency of DOM when compared to the other groups, but also a very different scale with very low frequency of lack of DOM with the semantic features definite and animate, and the thematic role *THEME*. DOM in Huánuco Quechua is linked to the semantic features definite and animate followed by the thematic roles with a prominent result for *THEME*. This result can be attributed to the similarities Huánuco Quechua and Spanish share through a common nominative/accusative alignment system and the availability of an accusative case marker. By not having to focus on differences in case marking Huánuco Quechua speakers can focus on definiteness and animacy. We take this to indicate that morphological differences between the preposition *-a* in Spanish and the Quechua case suffix *-ta* have hardly any effect on the production of DOM. Huánuco Quechua-Spanish bilinguals seem to be able to successfully map the relevant features onto a DOM marker. These results question the assumption that morphology is the locus of second language acquisition difficulties as proposed by the Bottleneck Hypothesis. The data show that differences in morphological patterns across languages do not seem to be a barrier for the acquisition of DOM marking.

Our expectation that the Shipibo and the Asháninka groups will exhibit lower frequencies of DOM due to differences in alignment between these languages and Spanish are born out as well. Of the two Amazonian languages, we see a higher frequency of DOM with definite and animate features in the Asháninka group followed closely by the thematic role *PATIENT* over *THEME*. The fact that

Asháninka is a nominative/accusative language albeit with fluid/split transitivity and the availability of gender specific bound morphemes as argument markers may be factors that support the acquisition of DOM in the relevant clitic structures. However, unlike the Huánuco Quechua group, the Asháninka group also exhibits lack of DOM with definite, animate and patient DPs (49.5%, 42.6%, and 39.6% respectively) and to a lesser frequency with theme DPs (16.8%). We take this to suggest that the fact that it is a language with fluid/split transitivity may also result in greater difficulty in the acquisition of DOM despite some sensitivity to the semantic features we have studied.

Shipibo-Spanish shows a different distribution from Asháninka-Spanish for marked and unmarked objects. The distribution of DOM follows a clear preference of the thematic role *PATIENT* followed by the definite semantic feature over animate and *THEME*, the latter only to a small extent. Interestingly, the same scale is found when looking at lack of DOM but with lower frequencies. A possible cause for a marking hierarchy that ranks the thematic role *PATIENT* higher than all other features in Shipibo could be the fact that Shipibo speakers are beginning to pay attention to *PATIENTS* as objects as a way to overcome the fact that in Shipibo, an ergative language, the subject of a transitive sentence receives case marking while the object remains unmarked. Unlike the Huánuco Quechua and the Asháninka groups who can focus their attention directly on definiteness and animacy, Shipibo speakers must first focus on the thematic role *PATIENT* as a way of identifying the direct object as the recipient of case marking.

Monolingual Spanish (LSCV) shares a similar distribution to the Asháninka-Spanish group in terms of marked objects with higher percentages in each category for LSCV and it shares a similar distribution with the Shipibo Spanish group in terms of unmarked objects although with a reversal in the ranking of *THEME* and animacy at the end of the scale. In LSCV the features definite and animate (61.5% and 56.1% respectively) play a stronger role in marking over the thematic roles (26.2% and 23.8% respectively).

For the second research question asking for the effect of animacy, definiteness and thematic role on the presence or absence of DOM in bilingual Spanish clitic doubling and dislocated structures in contact bilingual and monolingual varieties, our results show that all factors have an effect although their frequency varies in each of the groups.

Our hypothesis that animacy has an important role in the acquisition of Spanish was partly confirmed as there was a main effect for animacy and it is a predictor of DOM, but its relevance in terms of frequency is mostly found in the Huánuco Quechua group. Despite its presence in the input, animacy is the third most frequent feature in DPs with DOM among the Shipibo group and the second ranked feature among the Asháninka and the LSCV individuals. The defi-

nite feature plays a role in DOM as an effect for definiteness was found. It was also ranked highest in the Asháninka and the Huánuco Quechua groups as well as among the LSCV individuals and second in the Shipibo group. We would like to propose that the results for two of the bilingual groups with higher and similar frequencies for definiteness and animacy (Huánuco Quechua and Asháninka) point to an increase marking information structure in conjunction with a two-dimensional scale of animacy and definiteness (Aissen 2003) in a given discourse situation. This does not seem to be the case for Shipibo Spanish speakers given the higher frequency of DOM with patient DPs.

Our expectation for thematic role to play a more salient role in Asháninka Spanish given the sensitivity to thematic role in the morphology of Asháninka was met in terms of frequency. The Asháninka and Shipibo groups lead the preferences for PATIENT over THEME (71.2% and 63.6% respectively) followed by lower frequencies of preferences by the LSCV and Huánuco Quechua groups (38.3% and 23.6% respectively). The higher frequency of patient roles over themes (Dowty 1991) in the former groups can be related to sensitivity to thematic roles in the case of Asháninka and to the higher frequency of production of patient direct objects as well as the need to identify the direct object as a bearer of marking in the Shipibo group.

Finally, the extension of DOM to indefinites among Shipibo speakers seems to reflect the fact that there are no definite determiners in this language and this group seems to be more inclined to focus on the thematic role PATIENT as the semantic feature driving DOM rather than on definiteness. The extension of DOM to inanimates among the Asháninka speakers could be an effect of the low frequency of inanimates in their data. There is some extension of DOM to indefinites in the data of individuals of the LSCV group. These kinds of extensions of DOM can be linked to the fact that LSCV speakers are located on a continuum dependent on variable access to formal instruction.

In relation to the ecological factors, we propose that the scalar DOM systems (Table 4) found across the bilingual groups and the monolingual individuals can be linked to a feature pool with a reduced subset of features shaped by each group individually in accordance with their communicative needs (Mufwene 2001; 2002). In the case of the bilingual groups, the typology of the first languages contributes significantly to the ability to determine the relevance of semantic features as triggers of DOM in clitic doubling constructions.

The monolingual group exhibits clear intragroup differences with a scalar spread from LSCV1 most removed from the prestigious Lima norm, LSCV2 closer to the norm and LSCV3 – born and raised in Lima representing the prestigious norm. The variability found among the three monolingual Spanish speakers demonstrates clearly the choices individual speakers make dependent on the

availability and mix of features they are exposed to in the pool. It also shows that variability naturally exists within communities and reflects individual and group-specific communicative needs (Matras 2010).

For the bilingual group, the feature pool would be shaped by language ecological factors such as contact with language particular alignment systems, input and stage of language acquisition, and social networks. In bilingual systems, functional features from typologically different languages compete against each other and features shared by both languages may converge towards sets of features combined from both languages into a single matrix (Sánchez 2003). Our results as described above support our hypothesis that contact between typologically different languages influences the outcome for DOM in bilingual Spanish. While contact with alignment systems closer to Spanish, as in the case of Huánuco Quechua, strongly support and facilitate the production of DOM based on a two-dimensional scale of the semantic features definiteness and animacy, contact with ergative/absolutive alignment in Shipibo yields a very different system in terms of ranking of features. Asháninka, being closer to Spanish in terms of nominative-accusative alignment but exhibiting differences in marking transitivity, shows less sensitivity to the semantic features and more to transitivity, basing DOM primarily on the thematic role *PATIENT*.

Apart from these typological differences, the feature pool is particularly susceptible, and constrained by access to formal instruction in either or both languages as well as input from their social networks. As Mufwene (2001, 52) argues, in bilingual community languages, the selection from the feature pool works in the same ways as for idiolects. The feature pool represents a space where languages or dialects – varieties of Spanish in the case of the monolingual group – coexist, are acquired and used individually by speakers with differences across individuals that may not necessarily be reflected in the community language.

## 8 Concluding remarks

Our study provides preliminary evidence of typological effects in the development of DOM in clitic doubling structures. It shows that similarity in nominative-accusative alignment in Huánuco Quechua and Spanish favours a higher frequency of production of DOM in Spanish among the Huánuco Quechua speakers than among the Shipibo group and the Asháninka group. We advance the idea that this different distribution could be the result of Shipibo being a language with predominance of ergative-absolutive alignment and Asháninka one with nominative-accusative alignments but also split or fluid transitivity. In addition to case alignment, seman-

tic features such as definiteness and animacy as well as thematic role are present in all groups as determinants of DOM. Definiteness and animacy are more frequent in the Huánuco Quechua and Asháninka groups. We attribute their lower frequency in the Shipibo group to the fact that Shipibo speakers face the challenge of first identifying the constituent with a thematic role PATIENT as potentially subject to special morphological marking. Our data point in the direction that in contact situations typological similarities play a large role but the development of sensitivity to definiteness, animacy and thematic roles is not blocked even when there are typological differences. In the case of the LSCV individuals, we found evidence of higher levels of lack of DOM than among the Huánuco Quechua group, a distribution that seems to be better accounted for by the differences in the ecological factors that determine difference at the idiolectal level.

## Bibliography

- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21:3 (2003), 435–483.
- Alsina, Alex, *The role of argument structure in grammar. Evidence from Romance*, Stanford, Center for the Study of Language and Information, 1996.
- Blake, Barry J., *Case*, Cambridge, Cambridge University Press, 2001.
- Bossong, Georg, *Differential Object Marking in Romance and beyond*, in: Wanner, Dieter/ Kibbee, Douglas A. (edd.), *New analyses in Romance linguistics. Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, April 7–9, 1988*, Amsterdam/Philadelphia, John Benjamins, 1991, 143–170.
- Bossong, Georg, *Nominal and/or verbal marking of central actants*, in: Fiorentino, Giuliana (ed.), *Romance Objects. Transitivity in Romance languages*, Berlin/New York, de Gruyter, 2003, 17–48.
- Bresnan, Joan/Aissen, Judith, *Optimality and functionality. Objections and refutations*, *Natural Language and Linguistic Theory* 20 (2002), 81–95.
- Company Company, Concepción, *Multiple dative-marking grammaticalization. Spanish as a special kind of primary object language*, *Studies in Language* 25:1 (2001), 1–47.
- Company Company, Concepción, *Transitivity and grammaticalization of object. The diachronic struggle of direct and indirect object in Spanish*, in: Fiorentino, Giuliana (ed.), *Romance objects. Transitivity in Romance languages*, Berlin/New York, de Gruyter, 2003, 217–260.
- Cuza, Alejandro/Pérez-Leroux, Ana Teresa /Sánchez, Liliana, *The role of semantic transfer in clitic drop among simultaneous and sequential Chinese-Spanish bilinguals*, *Studies in Second Language Acquisition* 35:1 (2013), 93–125.
- Dalrymple, Mary/Nikolaeva, Irina, *Objects and information structure*, Cambridge, Cambridge University Press, 2011.
- Dowty, David, *Thematic proto-roles and argument selection*, *Language* 67:3 (1991), 547–619.
- Givón, Talmy, *Topic, pronoun, and grammatical agreement*, in: Li, Charles N. (ed.), *Subject and topic*, New York, Academic Press, 1976, 149–188.

- Givón, Talmy (ed.), *Grammatical relations. A functionalist perspective*, Amsterdam/Philadelphia, John Benjamins, 1997.
- Guijarro-Fuentes, Pedro, *Feature composition in Differential Object Marking*, *EUROSLA Yearbook* 11:1 (2011), 138–164.
- Guijarro-Fuentes, Pedro, *The acquisition of interpretable features in L2 Spanish. Personal “a”*, *Bilingualism* 15:4 (2012), 701–720.
- Iemmolo, Giorgio. *Towards a typological study of Differential Object Marking and Differential Object Indexation*, Pavia, University of Pavia dissertation, 2011.
- Jensen, Isabel N./Slabakova, Roumanya/Westergaard, Marit, *The Bottleneck Hypothesis in second language acquisition. A study of L1 Norwegian speakers’ knowledge of syntax and morphology in L2 English*, in: LaMendola, Maria/Scott, Jennifer (edd.), *Proceedings of the 41<sup>st</sup> annual Boston University Conference on Language Development*, Somerville, Cascadilla Press, 2017, 333–346.
- Laca, Brenda, *El objeto directo. La marcación preposicional*, in: Company Company, Concepción (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, vol. 1, México DF, UNAM/Fondo de Cultura Económica, 2006, 423–475.
- Lapesa, Rafael, *Estudios de morfosintaxis histórica del español*, Madrid, Gredos, 2000.
- Lardiere, Donna, *Dissociating syntax from morphology in a divergent L2 end-state grammar*, *Second Language Research* 14:4 (1998), 359–375.
- Lardiere, Donna, *On morphological competence*, in: Dekydtspotter, Laurent/Sprouse, Rex A./Liljestrand, Audrey (edd.), *Proceedings of the 7<sup>th</sup> Generative Approaches to Second Language Acquisition Conference*, 2005, 78–192.
- Leonetti, Manuel, *Specificity in clitic doubling and in Differential Object Marking*, *Probus* 20:1 (2008), 33–66.
- Loriot, James/Lauriault, Erwin/Day, Dwight, *Diccionario shipibo-castellano*, Lima, Instituto Lingüístico de Verano, 1993.
- Matras, Yaron, *Contact, convergence, and typology*, in: Hickey, Raymond (ed.), *The handbook of language contact*, Oxford, Wiley-Blackwell, 2010, 66–85.
- Mayer, Elisabeth, *Clitics on the move. From dependent marking to split marking*, in: Butt, Miriam/Holloway King, Tracy (edd.), *Proceedings of LFG ’08 Conference*, Stanford, CSLI Publications, 2008, 352–437, <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/13/lfg08.html> [last access: 13.09.2021].
- Mayer, Elisabeth, *Spanish clitics on the move. Variation in time and space*, 14 vol., Berlin/Boston, de Gruyter, 2017.
- Mayer, Elisabeth/Sánchez, Liliana, *Object agreement marking and information structure along the Quechua-Spanish contact continuum*, *Revista Española de Lingüística Aplicada* 29:2 (2016), 544–581.
- Mayer, Elisabeth/Sánchez, Liliana, *Feature variability in the bilingual-monolingual continuum. Clitics in bilingual Quechua-Spanish, bilingual Shipibo-Spanish and in monolingual Limeño Spanish contact varieties*, *International Journal of Bilingual Education and Bilingualism* 22:7 (2017), 883–896.
- Mayer, Elisabeth/Sánchez, Liliana, *Feature selection in clitic expression in two bilingual Amazonian Spanish varieties*, in: Zariquiey, Roberto, et al. (edd.), *Spanish diversity in the Amazon. Dialect and language contact perspectives*, Leiden/Boston, Brill, in press.
- Mayer, Mercer/Mayer, Marianna, *One frog too many*, New York, Puffin, 1992.

- Melis, Chantal, *El objeto directo personal en “El Cantar de Mio Cid”*. *Estudio sintáctico-pragmático*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 133–164.
- Mihas, Elena, *A grammar of Alto Perené (Arawak)*, Berlin/Boston, de Gruyter, 2015.
- Montrul, Silvina, *The acquisition of Spanish. Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*, Amsterdam/Philadelphia, John Benjamins, 2004.
- Mufwene, Salikoko, *The ecology of language evolution*, Cambridge, Cambridge University Press, 2001.
- Mufwene, Salikoko, *Competition and selection in language evolution*, *Selection* 3:1 (2002), 45–56.
- Næss, Åshild, *What markedness marks. The markedness problem with direct objects*, *Lingua* 114:9 (2004), 1186–1212.
- Nichols, Johanna, *Head-marking and dependent-marking grammar*, *Language* 62:1 (1986), 56–119.
- Payne, Judith K./Payne, David L., *The pragmatics of split intransitivity in Ashéninka*, *Revista Latinoamericana de Estudios Etnolingüísticos* 10 (2005), 37–56.
- Putnam, Michael T./Sánchez, Liliana, *What’s so incomplete about incomplete acquisition? A prolegomenon to modeling heritage language grammars*, *Linguistic Approaches to Bilingualism* 3:4 (2013), 478–508.
- Rodríguez-Mondoñedo, Miguel, *The acquisition of Differential Object Marking in Spanish*, *Probus* 20:1 (2008), 111–145.
- Sánchez, Liliana, *Quechua-Spanish bilingualism. Interference and convergence in functional categories*, Amsterdam/Philadelphia, John Benjamins, 2003.
- Sánchez, Liliana/Mayer, Elisabeth, *Clitics and argument marking in Shipibo Spanish and Asháninka-Spanish bilingual speech*, in: Fafulas, Stephen (ed.), *Amazonian Spanish. Language contact and evolution*, Amsterdam/Philadelphia, John Benjamins, 2020, 127–154.
- Sinnemäki, Kaius, *A typological perspective on Differential Object Marking*, *Linguistics* 52:2 (2014), 281–313.
- Slabakova, Roumyana, *Meaning in the second language*, 34 vol., Berlin/New York, de Gruyter, 2008.
- Suñer, Margarita, *Dialectal variation and clitic-doubled direct object*, in: Kirschner, Carl/DeCesaris, Janet A. (edd.), *Studies in Romance linguistics. Selected proceedings from the XVII Linguistic Symposium on Romance Languages*, Amsterdam/Philadelphia, John Benjamins, 1989, 377–395.
- Torrego, Esther, *The dependencies of objects*, Cambridge, Mass., MIT Press, 1998.
- Valenzuela, Pilar M., *Ergativity in Shipibo-Konibo, a Panoan language of the Ucayali*, in: Gildea, Spike/Queixalós, Francesc (edd.), *Ergativity in Amazonia*, Amsterdam/Philadelphia, John Benjamins, 2010, 65–96.
- Weber, David J., *Una gramática del quechua del Huallaga (Huánuco)*, Lima, Instituto Lingüístico de Verano, 1996.
- Witzlack-Makarevich, Alena/Seržant, Ilja A., *Differential Argument Marking. Patterns of variation*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 1–40.
- Zdrojewski, Pablo/Sánchez, Liliana, *Variation in accusative clitic doubling across three Spanish dialects. Morphological variation in Spanish*, *Lingua* 151 (2014), 162–176.





Albert Wall and Philipp Obrist

# Multilingualism effects in an elicitation study on Differential Object Marking in Cusco (Peru) and Misiones (Argentina)

**Abstract:** Although Differential Object Marking in Spanish (*a*-marking of direct objects) has been extensively studied from different perspectives and with different methods, its status and functioning in multilingual and language contact settings has thus far received little attention. This paper presents and compares data from monolingual and bilingual speakers of Spanish from two regions in Latin America, namely Argentina/The River Plate and Peru. An experimental elicitation study reveals that there are considerable differences in the DOM systems of Spanish monolinguals vs. bilinguals and between the bilingual groups, with the latter showing more individual variability and lower rates of *a*-marking in general. Our findings also suggest that within monolingual groups, the variation of *a*-marking is strongest for semantics-driven factors rather than syntax-driven ones. From a methodological perspective, we introduce an effective tool for collecting oral production data for a wide range of different DOM-sensitive syntactic configurations.

**Keywords:** language contact, language variation, experimental linguistics, animacy, specificity, reversible predicates

---

**Acknowledgements:** We would like to thank the audience of the Alpes 3 Winter School 2020 in Kandersteg and two anonymous reviewers for their valuable observations on previous versions of this paper. Furthermore, we thank Vania Escalante Orellana and Mike Chacmana from Cusco for establishing contact with the bilingual speakers, all our participants for sharing their ideas and time with us and Patricia de Ramos for finding participants in Misiones and for her assistance in conducting the experiments. We are also indebted to Larissa Binder for her help in the preparation of the experimental materials and during fieldwork in Cusco and to the Swiss National Science Foundation for financial support.

---

**Albert Wall**, University of Vienna, e-mail: [albert.wall@univie.ac.at](mailto:albert.wall@univie.ac.at)  
**Philipp Obrist**, University of Zurich, e-mail: [philipp.obrist@uzh.ch](mailto:philipp.obrist@uzh.ch)

# 1 Introduction

## 1.1 An open question in the study of Differential Object Marking

Since Bossong (1982), the split in Spanish direct object marking between zero marking and the presence of the marker *a* became to be known as Differential Object Marking (DOM). As the examples in (1) show, some objects obligatorily receive the marker, while others cannot receive it.

- (1) a. *Veo*                *\*(a) María.*  
           see.PRS.1SG    DOM    M.  
           ‘I see Maria.’
- b. *Veo*                *(\*a) la bicicleta.*  
           see.PRS.1SG    DOM    the    bicycle  
           ‘I see the bicycle.’

The topic in itself has always been a controversial one in the study of Spanish grammar and has received broad attention from different perspectives: synchronic, diachronic and variational (for a detailed overview cf. Fábregas 2013). It has also been claimed that the extension of DOM, especially with respect to inanimate objects, may differ regionally. Company Company (2002) discusses data from Mexico, suggesting an important increase of *a*-marking in indefinite and inanimate objects. Similar claims have been made for the Rio de la Plata region (Dumitrescu 1997; Montrul 2013; Hoff 2018) and corpus studies have found that factors of different relative strength favour *a*-marking in different varieties (Barraza 2003; Alfaraz 2011; Balasch 2011; Tippets 2011). Many regions, however, lack detailed empirical studies, this being the case for most of the countries along the Pacific coast in South America, including Peru. The same holds for contact scenarios. While there is work on heritage speakers and acquisition of Spanish as a foreign language (cf. Section 2), empirical work on DOM in contact scenarios of predominantly Spanish-speaking countries is rare (cf., however, Mayer/Sánchez, this volume).<sup>1</sup>

Furthermore, many relevant configurations which, according to the literature, show a strong affinity to DOM have not received special attention in the

---

<sup>1</sup> Sporadically, potentially relevant observations on local uses can be found in works with a more general scope, such as Pfänder (2009) on the variety of Cochabamba/Bolivia, where Spanish and Quechua are in contact. The author claims that inanimate objects can be *a*-marked in this variety (Pfänder 2009, 107). However, only one isolated example of this use is given, a sentence with a left-dislocated object under contrastive focus.

above-mentioned studies, most probably because they are not frequent enough in spontaneous spoken data. In many cases, it is not clear whether constructions with particular DOM-relevant features (e.g. reversible predicates based on verbs such as *reemplazar* ‘replace’ or *seguir* ‘follow’) have been included in the analysis of what we will call canonical transitive sentences, such as (1), or whether they have been excluded from the analysis in those studies. The same applies to the other configurations introduced in 1.2.

The present paper addresses this gap both methodologically and empirically. It introduces an approach based on sentence elicitation, performed by speakers from contact regions and control groups with a predominantly monolingual background. The sentence elicitation procedure was set up with an experimental design in order to assess the use of a wider range of relevant structural configurations in a given variety while controlling for a number of factors. Exact replication also allows for the collection of highly comparable data. Furthermore, it allows us to test marginal or infrequent configurations in a more reliable way. In the ideal case, such experimental studies can be backed up with spontaneous spoken language data, acceptability rating tasks, and metalinguistic interviews in a combined approach for a final assessment.

In the remainder of this introductory section, we present a series of configurations characteristic of Spanish DOM, which we will use to check for putative variation patterns in the empirical study. Section 2 addresses the current issues in the investigation of DOM in multilingual settings, with special focus on the two contact regions under discussion. Section 3 presents a description of the elicitation study and its results. Section 4 discusses the findings, and Section 5 presents some conclusions.

## 1.2 DOM as a multifactorial phenomenon

Given that Spanish DOM is a multifactorial phenomenon, different properties should be considered when characterizing the DOM system of a given variety of Spanish. In this Section, we introduce some nominal and verbal semantic properties, as well as configurations in structure and discourse well known to be relevant to DOM.

Animacy, definiteness and specificity:

This is the “core contrast” associated with DOM. NPs with reference to humans and definite interpretations always receive the marker (1a). Other animate NPs (e.g. with reference to animals) are also *a*-marked or show some degree of variation, but they certainly do not reject *a*-marking. For inanimates in canonical transitive struc-

tures as in (1b), the natural intuition in most contexts is strong rejection, hence the claims of ungrammaticality in the literature. However, it has been observed that *a*-marking occurs sporadically with inanimates both in spontaneous spoken and written language (cf. García García 2014 for a monograph-length discussion).

The specificity contrast in inanimate objects is exemplified in (2), an ambiguous sentence, which can mean that María is just looking for someone who fulfils the requirements of translating from or to German (unmarked), or, that there is a previously identified German translator she is trying to find (*a*-marked).

- (2) *María busca*                      (a) *un traductor alemán.*  
 M.    search.PRS.3SG DOM a translator German  
 ‘María is looking for a German translator.’

(López 2012, 10)

According to López (2012, 10), “[t]he object in this sentence can be prefixed by accusative A. With accusative A, it can have a specific reading. Without accusative A, it can only be nonspecific.” Similar contrasts can be observed by modifying the object with ‘a certain’ or ‘no matter who’ (*cierto/cualquiera*), the latter blocking *a*-marking according to López (2012, 17), or with subjunctive/indicative alternations. For further DOM patterns, the properties of the entire construction have to be taken into consideration.

Verb semantics:

After the inherent properties of the object noun and its discourse status, the properties of the verb as the main predicate of the sentence also constitute a crucial factor of DOM (von Heusinger/Kaiser 2011; García García 2014). Among these properties of the verb, a primary focus of attention has been on affectedness, in that it refers to the “persistent change in an event participant” (von Heusinger/Kaiser 2011, 594) and therefore is a crucial ingredient in the definition of transitivity. Von Heusinger/Kaiser (2011) use a scaled notion of affectedness in their empirical study and rank the verbs according to the degree to which the participant is transformed or involved according to the meaning of the verb. Figure 1 shows a simplified form of their scale and gives examples of Spanish verbs for the different categories.

1 DIRECT EFFECT ON PATIENT	2 PERCEPTION	3 PURSUIT	4 KNOWLEDGE	5 FEELING
<i>matar, herir, golpear, tirar</i> (‘kill, injure, beat, shoot’)	<i>ver, oír, escuchar, mirar</i> (‘see, hear, listen, watch’)	<i>buscar, esperar</i> (‘look out for, wait’)	<i>conocer, entender</i> (‘know, understand’)	<i>querer, temer</i> (‘like, fear’)

Figure 1: Affectedness scale from von Heusinger/Kaiser (2011), simplified.

The generalization expressed by the affectedness scale is that there is a decrease of *a*-marking from left to right, as verified by von Heusinger/Kaiser (2011) in their corpus study.

Another important factor involves the semantic roles defined by the argument structure of the verb. García García (2014, 22) proposes a relational notion of semantic roles, namely a ‘decline of agentivity’ between agent and patient. Agentivity is of special importance for the *a*-marking of inanimate objects, as García García (2014) exemplifies with a class of verbs which he labels “reversible predicates” (García García 2014, 147). This class comprises positioning and substitution verbs, such as *preceder* and *sustituir*. In the appropriate readings, such predicates do not express an ‘incline’ in agentivity between their arguments. Thus, *a*-marking is a possible and perhaps even necessary strategy to differentiate subject and direct object.

- (3) *El artículo acompaña al \*el sustantivo.*  
 the article accompany.PRES.3SG DOM+the/ the noun  
 ‘The article accompanies the noun.’

(García García 2014, 144)

Doubled structures:

A further generalization that has emerged from the study of *a*-marking on inanimates is that certain more complex structures involving secondary predication allow for the *a*-marking of objects that would be incompatible with it according to the animacy criterion. One case in point is that of verbs which allow for double accusative constructions, such as *considerar* (‘consider’), *llamar* (‘call’) and *caracterizar* (‘characterize’).

- (4) *Algunos gramáticos [...] no consideran oración a*  
 some grammarians NEG consider.PRS.3PL sentence DOM  
*la secuencia con verbo.*  
 the sequence with verb  
 ‘Some grammarians do not consider the sequence with a verb (to be) a sentence.’

- (5) *Consideren estos datos.*  
 consider.IMP.3PL these data  
 ‘Consider these data.’

(García García 2014, 49)

In contrast to (5), where *a*-marking is excluded, there are two elements qualified to fill the direct object slot in (4) – the predicative complement *oración* and the complex NP complement *la secuencia con verbo*. Such configurations are reported to show high frequencies of *a*-marking (Weissenrieder 1991, 150). Interestingly, López (2012, 10) claims that in such configurations – “small clause complements” in his terminology – an animate argument is also obligatorily marked if it is indefinite and non-specific (6).

- (6) *Considero*            *\*(a) un estudiante inteligente.*  
 consider.PRS.1SG DOM a student intelligent  
 ‘I consider a student to be intelligent.’

García García (2014, 103) presents a more differentiated picture of such constructions. One of the results of his corpus study suggests that the adjacency of the two “objects” is a decisive factor. Sentences where the direct object and the predicative are not adjacent only showed *a*-marking in 21% of cases, whereas adjacent constructions confirm López’ intuition and exhibit *a*-marking in 100% of cases.

Ditransitive sentences, in which the indirect object typically is an animate NP, represent another case of doubled structures, in the sense that if the direct object of such a sentence is animate and specific, both objects look the same overtly. It has been reported that the *a*-marking of the direct object is highly disfavoured in such structures.

- (7) *Pedro presentó*            *\*(a) su mujer a sus amigos.*  
 P. present.PRF.3SG DOM his woman to his friends  
 ‘Pedro introduced his wife to his friends.’

(García García 2014, 53)

#### Complex objects:

AcI structures (8) are similar to the double accusative structures presented above in that they also have an object-related secondary predication (García García 2014, 51). These constructions also allow for the *a*-marking of inanimate objects, especially if the object receives a more agentive description (Torrego 1999, 1792). Thus, in the adapted examples from García García (2014), *a*-marking would be more likely on (8b) than on (8a).

- (8) a. *Veo*            *el/ al*            *agua caer.*  
 see.PRS.1SG the DOM+the water fall.INF  
 ‘I see the water falling.’

- b. *Veo al/ el agua caer muy rápidamente.*  
 see.PRS.1SG DOM +the the water fall.INF very fast  
 ‘I see the water falling swiftly.’

(García García 2014, 51–52)

López (2012, 23–25) also discusses such constructions (“clause union”) and observes that for perception and causation verbs, *a*-marking for animate indefinites is obligatory regardless of specificity. García García (2014, 106) reports that such causative structures also allow for the *a*-marking of inanimate objects.

- (9) *María vio caer \*(a) un niño.*  
 M. see.PRF.3SG fall.INF DOM a child  
 ‘María saw a child falling down.’

- (10) *María hizo llegar tarde \*(a) un niño.*  
 M. make.PRF.3SG arrive.INF late DOM a child  
 ‘María made a child arrive late.’

(López 2012, 24)

Secondary predicates also play a role in the context of the transitive verb *tener* which is notorious for rejecting *a*-marking in most contexts: “A marked object is ungrammatical as the complement of *haber* ‘have’ (existential) and *tener* ‘have’ (possessor or relator). [...] The data surrounding *tener* are extremely intricate. *Tener* can mean something close to ‘hold’ or ‘get’, in which case a marked object is possible. The VP headed by *tener* can include a secondary predicate, in which case a marked object is again possible” (López 2012, 20).

- (11) *María tiene \*(a) tres hijos.*  
 M. have.PRS.3SG DOM three children  
 ‘María has three children.’

- (12) *María tiene (a) un hijo en el ejército.*  
 M. have.PRS.3SG DOM a son in the army  
 ‘María has a son in the army.’

García García (2014, 50) presents similar data and adds that inanimates may also be *a*-marked in such constructions. He also claims that *a*-marking is the preferred option with animates:

- (13) *Ana tiene*                    \*(a) *una hija*            *estudiando derecho.*  
 A. have.PRS.3SG DOM a daughter study.GER law  
 ‘Ana has a daughter who is studying law.’
- (14) *Ana tiene*                    (a) *un ordenador* *calculando* *el problema.*  
 A. have.PRS.3SG DOM a computer calculate.GER the problem  
 ‘Ana has a computer calculating the problem.’

One possible explanation for these findings is that the marked direct object in such configurations can be interpreted as the subject of the secondary predicate and hence as having more agentive properties. It is not the goal of this study to explore the patterns introduced in this Section in greater detail or from a theoretical perspective. Rather, they are listed and explained in order to show that determining the status of DOM in a given variety of Spanish involves taking into consideration very different configurations, and also introducing the types of structures that have been included in the elicitation experiment, where all the distinctions mentioned above are taken into account.

### 1.3 Methodological and grammatical considerations with respect to the elicitation experiment

As outlined above, the goal of this study was to collect equivalent language production data on the variational properties of DOM in Spanish and to compare different contact scenarios. Previous empirical studies using spontaneous spoken data reported considerable variation in the use of Spanish DOM in certain configurations. However, it is not clear to what degree different corpora of spontaneous spoken language from different varieties are actually comparable. Furthermore, many of the configurations introduced above do not occur frequently enough in common corpora to allow for a solid understanding of their behaviour. Both of these issues can be addressed by controlled elicitation. We decided to collect production data rather than acceptability judgments as a first step, because our primary interest was to know what structures speakers actually produce. In Likert-scale acceptability ratings, strictly speaking, only contrasts between sentences can be interpreted. Hence, if there is a correlation between acceptability and use, it is only an indirect one. As mentioned in the introductory section, ideally we should work towards a combination of methods, and this study adds a hitherto missing type of data to the overall picture.



Apart from the methodological considerations, the choice of grammatical configurations that have been included in the study, as well as the number of configurations tested, need further explanation. Obviously, not all DOM-sensitive phenomena described above can be tested thoroughly in one experiment. Therefore, we decided to focus on two configurations, which represent more than half of the experimental items, in order to achieve robust results for them. Four additional configurations are included in the remaining part of experimental items. The data on these four constructions individually are not as robust as for the first two, but they can still provide an exploratory impression of what is possible in these configurations. Section 3.1 provides a detailed description of these six configurations and how they were implemented in the experiment. Since one of the main goals of the experiment was to collect variational data, all these configurations include indefinite and/or inanimate objects in at least one manipulation, often contrasting them with animate and/or definite objects. Therefore, the experiment does not test configurations where *a*-marking has been shown to be obligatory in previous research, such as strong pronouns and proper names (for the latter cf. example 1a above). Often, Spanish DOM is explained or characterized by making use of the animacy/definiteness hierarchy, which is given in a simplified form here in (15).

- (15) pronouns > proper names > definite, animate > indefinite, animate > inanimate

There is wide consensus in the literature that, starting from the left side of the scale, *a*-marking is categorical with pronouns, proper names and definite animates. This is why our study focuses on the “right side” of the scale, where things are less clear and where variation is expected. For the sake of concision, other DOM-related phenomena, such as clitic doubling and *leísmo*, among others, will also not be discussed in this study.

## 2 Differential Object Marking in multilingual environments

### 2.1 Multilingual acquisition, language attrition and contact

In contrast to the detailed accounts of DOM in Spanish or other individual languages, there are only few studies dedicated to DOM in multilingual settings. The Spanish-English contact scenario is among the best explored of such

constellations: Ticio (2015) investigated the early acquisition of the Spanish DOM system of children growing up in a simultaneous acquisition scenario, while Montrul/Bowles (2009) is a study of DOM in heritage speakers of Spanish in the United States. Accounts for other constellations are Döhla (2011), who discusses different contact scenarios with American Indian languages, and Montrul/Gürel (2015) and Montrul (2019) presenting experimental data of learners of Spanish in Turkey and Romania, respectively.

Montrul/Bowles (2009) consider heritage speakers in two experiments which include a general proficiency test, an oral production task, and different acceptability judgment tasks. They show that lower proficiency tends to correlate with a decrease in the production of the *a*-marking of objects that should be marked, and with increasing insecurity in the acceptability judgment tasks. Ticio (2015) finds that, in contrast to monolinguals, bilingual children did not acquire the DOM system in the period under study (until the age of 3;6) and that bilingual acquisition differs from monolingual acquisition in a fundamental way: “[...] DOM seems to be difficult or almost impossible to acquire for L2 learners, and it results in a range of error productions among HS and adult or school-age bilinguals” (Ticio 2015, 70). Similar findings had already been reported by Montrul/Sánchez-Walker (2013) for school-age Spanish-English bilingual children, with the omission of expected marking of over 65% in some cases. By contrast to Ticio’s claims, Döhla (2011, 27) speculates that “[s]ince DOM is very common [cross-linguistically], we suppose that, in case of language contact, and first and foremost bilingualism, a language with DOM can easily transfer the morphosyntactic feature to another language without DOM or exert influence on another language that exhibits DOM.” The same author discusses examples of American Indian languages that presumably already had a DOM system prior to contact with Spanish, and he suggests that in these cases contact does not play a role. On the other hand, for indigenous languages with more recent traits of DOM, Döhla suggests that Spanish might very well have triggered or potentiated its evolution. As a prime example, he cites Paraguayan Guaraní, exhibiting a DOM system similar to that of Spanish. The author concludes that more empirical data is necessary in order to assess the role of contact in all discussed scenarios.

The basic idea in both Montrul/Gürel (2015) and Montrul (2019) is that the existence of a DOM system in the L1, as in Turkish or Romanian, might enhance the acquisition of DOM in another language, such as Spanish, despite some structural differences. They derive the predictions of their study from the so-called Feature Reassembly Hypothesis, according to which grammatical features of lexical and functional items are bundled differently from one language to the other. Consequently, L2 learners would need to work out how the features are bundled in the target language. In this process, reconfiguration of the feature

bundles of L1 comes into play (Montrul/Gürel 2015, 290). The results of these studies confirm this basic assumption: From the Turkish participants, even L3 learners with lower proficiency perform quite well, while learners with higher proficiency significantly outperform the Spanish-English bilinguals and heritage speakers from the previously reported studies. For Romanian, a language genetically and structurally closer to Spanish, the enhancement effect is even stronger than for Turkish.

This is not the place to discuss the different theories of acquisition on which these works are based. The goal of the present study is not to argue in favour or against a certain model of language contact or acquisition, but rather to begin filling an empirical gap in the literature: There is hardly any empirical work on Spanish DOM in scenarios of contact involving bi- or multilingual territories, such as certain regions of the Andes (cf. also Mayer/Sánchez, this volume) or the Misiones Province in Argentina. The following Section summarizes the most important facts about this linguistic space for the purpose of this paper.

## 2.2 The contact scenarios: Andean Spanish and multilingualism in Misiones

Andean Spanish has been identified as a supranational macrovariety of Spanish showing a series of features at all structural levels that diverge from normative standards. Many of these features have been described and studied in some detail. Escobar (2011) provides a detailed overview of the literature here, with a special focus on the Spanish-Quechua contact scenario, which plays a crucial role in the development of this variety. It is well known for a tendency towards OV word order in contrast to other varieties of Spanish, it has some morphological and many lexical borrowings from Quechua, as well as some phonological peculiarities, such as the distinction between /j/ and /ʎ/ (otherwise uncommon in American varieties), strengthening and preservation of consonants and reduction of unstressed vowels, among many other features. Interestingly, however, there is no mention of DOM in the literature on this contact scenario. Mayer/Sánchez (this volume) discuss Spanish-Quechua contact data from Huánuco (central Peru) among other contact scenarios in Peru. According to their data, *a*-marking is quite frequent in Huánuco Spanish, unlike in contact scenarios with languages such as Asháninka or Shipibo. On a more anecdotal note, one could also mention the possible emergence of a new DOM marker in the variety of Cajamarca (northern Peru). In this variety, the substitution of the DOM marker *a* with *onde* has been documented in the writings of Ciro Alegría, whose rural characters from that region use this form (Bossong 2008, 93). However, the precise status of

this form is unclear. For the southern regions of Peru, including Cusco, even less is known. The data and analysis presented below are therefore first steps towards filling this gap.

The northeastern Argentinian region of Misiones has only recently come to be known as a crossroads of language contact. Originally colonized by Jesuits, it was subject to a territorial dispute between Brazil, Paraguay and Argentina until the end of the 19<sup>th</sup> century, this related to the Paraguayan Wars, and is now the youngest province of northern Argentina. At the turn of the 20<sup>th</sup> century, it was almost entirely repopulated by foreign settlers, many of which were from Central Europe (e.g. Ukraine, Poland, Germany). While the remnants of Slavic and Germanic linguistic heritage are still detectable, the majority language of the Province today is Spanish. Current contact languages are Portuguese, particularly in the villages on the banks of the river Uruguay (bordering Brazil), and Guaraní, which is still spoken within the indigenous population. While the population is conscious of their plurilinguism and their particular linguistic identity, a comprehensive description of the provinces' linguistic situation remains a desideratum: whereas the rather impressionistic description of the *habla misionera* by Amable (1975) focusses on lexicon and phraseology, an unpublished dissertation by Sanicky (1981) concentrates on phonology. Recent work by de Ramos (2017) confirms the existence of widespread “leísmo”, already briefly mentioned in cross-variational studies (Fernández-Ordóñez 1999, 1347–1349), attributed to language contact with Guaraní and possibly related to DOM.

### 3 The elicitation study

As noted in the introductory section, the elicitation tasks combine different DOM-sensitive contexts under one general setting. The tasks consist of spontaneously producing a sentence in which input material presented on a display has to be used and some very general instructions followed. Sentences are recorded by means of the SpeechRecorder software (Draxler/Jänsch 2004) for subsequent analysis, in this case, checking for the presence of *a*-marking on direct objects.

#### 3.1 Design and materials

While following an experimental design and striving to avoid confounding factors, we were also interested in receiving the most natural output possible. For some sentences, it was important to ensure specific reference, hence a context

sentence had to be included. For others, the relative order of the direct object with other elements had to be established. This implied a delicate balance between nudging towards the intended structures and favouring spontaneous speech (cf. Bautista-Maldonado/Montrul 2019 for a similar technique). Four different versions of the production task were implemented: In Task1, a context sentence was presented (in black letters) together with some additional unconnected words (in red letters), which had to be used in the production task. The additional unconnected words consisted of two NPs and an inflected verb. Participants were asked to create a sentence with the two NPs and the verb, taking the context sentence into consideration. They were explicitly allowed to add more words to the sentence and to arrange the presented material and the additionally included words as they liked. In Task2, a sentence (in black letters) was presented together with two NPs and an inflected verb (in red letters) and participants were asked to paraphrase the presented sentence with the two NPs and the inflected verb. They were explicitly allowed to add more words to the sentence and to arrange the presented material and the additionally included words as they liked. In Task3 and Task4, unconnected words and phrases were presented (in red letters) and participants were asked to build a sentence with this material. They were explicitly allowed to add more words to the sentence. While Task3 required maintaining the relative order of the presented words, Task4 permitted the rearrangement of the presented material (and any additionally included words) at the participant's discretion. Figure 2 gives one example of how Task1 was prompted by written stimuli and displayed on the participants' screen:

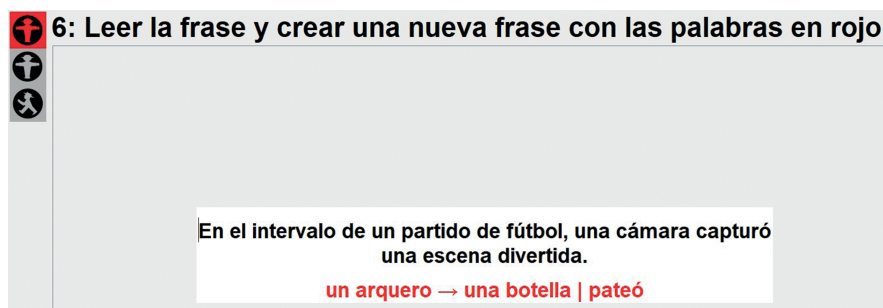


FIGURE 2: Screenshot of the participants' screen.

As can be seen in Figure 2, context sentences or sentences to be paraphrased were presented in black on a white screen. The unconnected words for sentence production were presented in a separate line below in red letters. These chunks of material for sentence construction were graphically separated from each other in

the presentation of stimuli. Usually, a vertical bar separated words or phrases. If two nominal arguments were presented, sometimes an arrow was used between them. This arrow was included as a non-verbal strategy to induce transitivity. Thus, the arrow always pointed from the potential subject to the potential object or, in the double accusative set (see below), from the potential object NP to the predicative complement. Participants were not explicitly told about the function of the arrow. Upon request, it was explained that it represented a connection in meaning between the two elements which had to be involved with one another in the sentence. This separation by bars and arrows is included in the reproduction of the example material below.

Six different types of experimental items were created in order to cover the grammatical configurations introduced in Section 1.2, resulting in a total of 40 experimental items.<sup>2</sup> All six sets of these items crossed two conditions (2x2 design), and the items were distributed over four lists, with each participant assigned one list. In this way, each participant saw each item in only one of the four conditions and all participants saw the same number of conditions of each set. The first dataset contained four repeated measures per condition and list and the second dataset two repeated measures per condition and list. These lists were randomized by the experimental software for each participant.

The *first and largest set* of items were aimed at comparing specific and non-specific indefinite objects (animate and inanimate) in combination with verbs presenting different degrees of affectedness. In order to ensure specific readings, three context sentences were created for every item. While the first described a scenario without introducing any referent, the second introduced two referents as potential subjects related to two potential animate objects, and the third used the same potential subjects but combined with two inanimate potential objects. The material for sentence construction presented along with the context sentences contained two indefinite NPs that matched with the introduced referents of the second context sentence. (16) shows such an item with the three context sentences and the two sets of unconnected linguistic material that was presented for sentence construction.

---

<sup>2</sup> The complete materials and the SpeechRecorder script used for the elicitation is available upon request.

(16) Complete materials of one experimental item of the specificity/animacy set:

Context sentences with and without introducing the animated object referents:

a. *En el puerto se encontraron muchas personas.*

‘At the harbor, many people met.’

b. *En un crucero, dos pasajero y dos tripulantes se encontraron en el mismo bar.*

‘On a cruise ship, two passenger and two crew members met in the same bar.’

Stimuli presented with each of the context sentences (a) and (b):

c. *un pasajero* → *un tripulante* | *vio*  
 a passenger a crew member see.PRF.3SG

Context sentences with and without introducing the animated object referents:

d. *En un viaje en crucero hubo mucho que descubrir.*

‘On a cruise trip there was a lot to be discovered.’

e. *Dos pasajeros de un crucero se aproximaron a dos islas.*

‘Two passengers of a cruise ship approached two islands.’

Stimuli presented with each of the context sentences (d) and (e):

f. *un pasajero* → *una isla* | *vio*  
 a passenger a island see.PRF.3SG

This design allows us to observe putative interactions between specificity and animacy. Since it has been claimed that a higher degree of affectedness enhances *a*-marking, we controlled for this by including four verbs of each of the affectedness groups defined by von Heusinger/Kaiser (2011), cf. Figure 1. In total, the specificity/animacy set had 16 items and the corresponding task was Task1.

*The second set* was created in order to test reversible predicates. The predictions from the literature are that inanimates are always *a*-marked in symmetric configurations (both arguments inanimate) and that inanimate objects are also *a*-marked with animates in the so-called “reversible” interpretations. Eight verbs that allow for reversible structures were selected and combined with animate as well as inanimate subjects and objects in all four possible configurations, resulting in four target sentences per verb/item. For each sentence, a paraphrase was created using the two arguments but not the same verb. This paraphrase was presented in Task2 as a point of departure. The participants were asked to paraphrase the presented sentence using the reversible verb and the two arguments. Such as experimental item is given in (17).

(17) Complete materials of one experimental item of the reversible verbs group:

a. animate subject – animate object

*El alumno tomó el lugar del instructor.*

‘The pupil took the place of the instructor.’

*el alumno* → *el instructor* | *sustituyó*  
 the pupil            the instructor    substitute.PRF.3SG

b. animate subject – inanimate object

*El alumno se hizo cargo del trabajo de la máquina.*

‘The pupil took on the machine’s work.’

*el alumno* → *la máquina* | *sustituyó*  
 the pupil            the machine    substitute.PRF.3SG

c. inanimate subject – animate object

*La máquina continuó con el trabajo del alumno.*

‘The machine continued with the pupil’s work.’

*la máquina* → *el alumno* | *sustituyó*  
 the machine        the pupil        substitute.PRF.3SG

d. inanimate subject – inanimate object

*Hoy en día se usa más el bolígrafo en vez del lápiz.*

‘Nowadays, the pen is used more often instead of the pencil.’

*el bolígrafo* → *el lápiz* | *sustituyó*  
 the pen                the pencil        substitute.PRF.3SG

Given that all four possible configurations of animate vs. inanimate in subject and object positions were included, the results can be analyzed with respect to relative agentivity as well as for the claim of obligatory *a*-marking in reversible readings.

*Set 3* tested double accusative structures. Four verbs licensing such structures were chosen and combined with two sets of two possible objects each. The first set contained an animate NP as a potentially referential expression and a noun that was more apt to serve as a predicative complement. The second set contained an inanimate NP as a potentially referential expression and also a noun that was more apt to serve as a predicative complement. In this way, the claim about high rates of *a*-marking of inanimates can be verified and *a*-marking of inanimates and animates can be directly compared. Since it is difficult to control for (non-)adjacency of the two objects within the general design of the study (cf. the findings of García García 2014, 103, presented in Section 1.2), we decided to test for a related



factor, namely canonical syntactic configuration vs. structures that included displacement. “Displacement” was created by manipulating the relative order of the two objects. The items were presented in the following way: first, the verb in the third person plural of the past tense (indefinido), and second, the two potential direct objects, either the set with the animate or the set with the inanimate NP. By also manipulating the order of the two potential objects, four conditions were created, as shown in Example (18). Only the material for the construction of the sentence was presented (no context etc.) and participants were asked not to change the given order of the words (Task3).

- (18) Complete materials of one experimental item of the double accusative group
- a. *Consideraron* | *un informático* → *testigo experto* [in situ]  
 consider.PRF.3PL a computer scientist witness expert
  - b. *Consideraron* | *experto testigo* ← *un informático* [displaced]  
 consider.PRF.3PL expert witness a computer scientist
  - c. *Consideraron* | *una fábrica* → *empresa pionera* [in situ]  
 consider.PRF.3PL a factory company pioneer
  - d. *Consideraron* | *empresa pionera* ← *una fábrica* [displaced]  
 consider.PRF.3PL company pioneer a factory

Set 4 was intended to look at ditransitive structures. Four verbs with a “transferential” argument structure were chosen and combined with a person name, a second NP containing a kinship noun and a third NP denoting either an animal or an inanimate. The second and third NP were presented either with an indefinite article or a possessive. Example (19) shows the four manipulations of one item in this group. In this way, the effects of ditransitive structures can be tested for interactions between animacy and definiteness.

- (19) Complete experimental materials of one item of the ditransitive group
- a. *Cristina* | *un loro* | *una hermana* | *ayer* | *entregó* [-DEF][+ANIM]  
 C. a parrot a sister yesterday give.PRF.3SG
  - b. *Cristina* | *una bicicleta* | *una hermana* | *ayer* | *entregó* [-DEF][-ANIM]  
 C. a bicycle a sister yesterday give.PRF.3SG
  - c. *Cristina* | *su loro* | *su hermana* | *ayer* | *entregó* [-DEF][-ANIM]  
 C. her parrot her sister yesterday give.PRF.3SG
  - d. *Cristina* | *su bicicleta* | *su hermana* | *ayer* | *entregó* [-DEF][-ANIM]  
 C. her bicycle her sister yesterday give.PRF.3SG

*Set 5* was created to assess DOM in AcI structures. More precisely, it tests whether the potential degree of agentivity of the object would increase the rate of *a*-marking for human indefinites. Four animate nouns were chosen as possible objects and combined with appropriate verbs in the infinitive. Each of these four nouns was then presented either with a causative verb or a perception verb in the third person plural of the perfective past tense (*indefinido*) and with an intensification adverb modifying the infinitive or no further modification, again resulting in four manipulations, as presented in (20). The causation/perception manipulation provides a contrast of agentivity from verbal semantics, and the intensification adverb, according to the literature, could further increase the agentivity of the object.

- (20) Complete materials of one experimental item of the AcI group
- a. *Hicieron* | *correr* | *un mensajero* [CAUSATIVE][-INTENSE]  
make.PRF.3PL run.INF a messenger
  - b. *Hicieron* | *correr rápidamente* | *un mensajero*  
make.PRF.3PL run.INF fast a messenger  
[CAUSATIVE][+INTENSE]
  - c. *Vieron* | *correr* | *un mensajero* [PERCEPTION][-INTENSE]  
see.PRF.3PL run.INF a messenger
  - d. *Vieron* | *correr rápidamente* | *un mensajero*  
see.PRF.3PL run.INF fast a messenger  
[PERCEPTION][+INTENSE]

Finally, *Set 6* tested secondary predicates of the verb *tener* in combination with agentivity contrasts. The items were created in the following way: KINSHIP TERM / *muy* + ADJECTIVE / GERUND (secondary predicate verb) + LOCATION / *tener* in the first person singular of the present tense. Four kinship terms were combined with four manipulations: the verb either denoted a more or less strenuous activity and the adjective denoted a low or a high degree of involvement. The location was chosen to fit the activity expressed by the verb (cf. examples in 21). This allowed us to compare different degrees and sources of agentivity. We considered adjectives overtly expressing strong emotions like *happy*, *excited* and *agitated* to convey more involvement than adjectives not overtly expressing emotions but other transitory states, such as *relaxed*, *exhausted* or *ill*.

- (21) Complete materials of one experimental item of the *tener* group
- a. *padre* | *muy contento* | *descansando en el*  
father very happy rest.GER on the  
*sofá* | *tengo*  
sofa have.PRS.1SG

- b. *padre* | *muy exhausto* | *en descansando el*  
 father very exhausted on rest.GER the  
*sofá* | *tengo*  
 sofa have.PRS.1SG
- c. *padre* | *muy contento* | *trabajando en la*  
 father very happy work.GER in the  
*oficina* | *tengo*  
 office have.PRS.1SG
- d. *padre* | *muy exhausto* | *trabajando en la*  
 father very exhausted work.GER in the  
*oficina* | *tengo*  
 office have.PRS.1SG

The elicitation was conducted in the following way: The stimulus material was presented on a separate screen for participants, this connected to the laptop of the experimenter. The audio data was recorded by a directional microphone (RØDE NGT4), also connected to the same laptop via an audio interface (ZOOM U-22). Participants only saw their own screen. The experimenter was seated across the table and had an overview of the experiment from the laptop. The instructions were presented on the screen and commented on by the instructor. First, the four tasks were introduced. The experimenter told the participants that they were free to add words to the ones presented to them in order to create the sentence, but that they had to use all those that were presented without modifying their form, and also that they were allowed to arrange the words as they liked, except in Task3. Participants were then informed that the tasks would be randomized and that they would always receive a brief instruction for whichever of the four tasks they had to perform. They were also given the total number of sentences to be created. After these general instructions, the recording procedure was explained: Participants would be given time to read the complete instructions and stimulus material on the screen and to think. Once they had the sentence in mind and gave the experimenter a signal, the microphone was activated and their answer was recorded. Participants were able to ask questions between the recordings, but the experimenter would not comment on possible sentences they created. The experimenter would only interfere in the following cases: (i) if a participant distributed the stimulus words in more than one sentence (including coordinated structures); (ii) if a participant modified the stimulus material (verb form, determiner, etc...); (iii) if a participant changed the predefined word order in Task3. The experiment was implemented in closed rooms whenever possible, although with some Quechua L1 speakers in Cusco, recordings had to be conducted outside.

### 3.2 Participants

Participants were recruited randomly through a variety of strategies, such as social media, personal contacts and by spontaneously inviting people in public spaces. In the urban areas, we restricted the pool of participants to university students, excluding students of disciplines with an analytical focus on language (philology, linguistics, literature, etc.). For the rural regions, it was not possible to maintain this restriction. Participants received a monetary reward for their participation. All participants understood that participation was entirely voluntary and that they could interrupt or stop their participation at any time.

In this study, we report the results of 32 participants, 16 from Argentina and Montevideo and 16 from Peru. Half of the participants of each region came from a bilingual location. For Cusco and Misiones, this is the total number of participants in our sample. The same number of participants was randomly selected from larger samples taken in Lima and Montevideo for comparison. Table 1 provides more information about the participants, the total number of recorded sentences and the number of “collaborative” sentences, i.e. sentences that could be included in the results and used for the subsequent analysis. The opening lines of Section 3.3 explain the annotation of the recordings and the classification of the elicited sentences in detail.

**Table 1:** Information on the participants of the elicitation study.

Place	N	Age			Gender		Sentences collaborative/total
		range	mean	SD	F	M	
Cusco	8	30–54	41.1	10.5	1	7	272/300
Lima	8	19–21	19.5	1.22	5	3	252/320
Misiones	8	20–63	40.4	14.88	6	2	309/320
Montevideo	8	21–36	24.9	5.75	7	1	271/320
Total	32				19	13	1,104/1,260

The participants from Lima and Montevideo all had Spanish as their L1, and they lived and studied in their respective cities. As for knowledge of further languages, the students from Lima all had some command of English, with one student additionally mentioning French and one Portuguese. The students from Montevideo had also learned English to different degrees, with some additionally mentioning Portuguese or Italian.

Participants from Misiones all had Spanish as their L1 but reported regular contact to Portuguese in their daily lives. They were accustomed to watching Portu-

guese television, to seeing and hearing the language in other media and to using it with Portuguese-speaking people. Most also mentioned Guaraní, German, Polish or Russian as heritage languages still spoken by their elders, but they claimed not to be able to speak such languages themselves (except for one participant, who had an active knowledge of German). All participants had finished secondary school, one was currently attending college, but had not yet graduated. Among the other seven, five worked in agriculture and commerce and two had retired.

The participants from Cusco were more heterogeneous in their linguistic profiles. Four lived and worked in Cusco City and were only fluent in Spanish. Some had some basic knowledge of English and some reported understanding isolated Quechua words but not to be able to speak the language. They all held university degrees and worked in the local university administration or in the tourist sector. The other four participants came from the surrounding areas of Cusco, were native speakers of Quechua and acquired Spanish at school. They were all bilingual in Spanish and Quechua and report to switching freely between the two languages, although most of them preferred to speak Quechua whenever possible. One of them had only finished primary school, two were secondary school graduates and one was a student at the local university. They all worked in agriculture and transportation.

### 3.3 Results

The experimental design had four lists, and each participant of each group received a different list, yielding a complete set of responses for the entire elicitation experiment. Unfortunately, 20 responses to one of the bilingual lists in Cusco were not recorded due to a technical failure. Recordings were transcribed and annotated for further analysis. In a first step, the transcriptions of each item was compared with the expected output in order to determine whether the participant had been cooperative, partially cooperative, or uncooperative. A trial was considered cooperative if the participant used the direct object and, where applicable, the subject according to the outlined transitive structure. If the participant uttered a sentence with a transitive structure but did not use the arguments as expected, this was considered as partly cooperative. Other utterances were discarded as uncooperative. For the analysis, we considered cooperative trials as well as those partly cooperative trials where the grammatical configuration established in the condition was not violated. Thus, if animate subject and animate object exchanged places within a transitive sentence, this would still be taken into consideration for analysis.

For the purpose of comparison, we established idealized predictions about each individual condition, based on findings in the literature. For configurations claimed to have obligatory *a*-marking, we set the expectation to 100% of *a*-marking responses, and for configurations that reported to reject *a*-marking the expectation was 0%. When variation was expected, we set the prediction to 50%. For instance, according to previous studies, *a*-marking is considered to be obligatory in all four manipulations of the reversible predicates dataset (set 2, cf. example 17 above). Hence, the general expectation would be 100% *a*-marking in this subset of responses. In the case of the animacy/specificity dataset, only 37.5% of *a*-marking is predicted, more specifically 100% for specific animates, 50% for unspecific animates and 0% for both inanimate conditions, cf. example (16) above. This allows us to calculate an overall expectation of *a*-marking for the whole experiment. Table 2 reports the expected overall performance together with that of the different regions and localities.

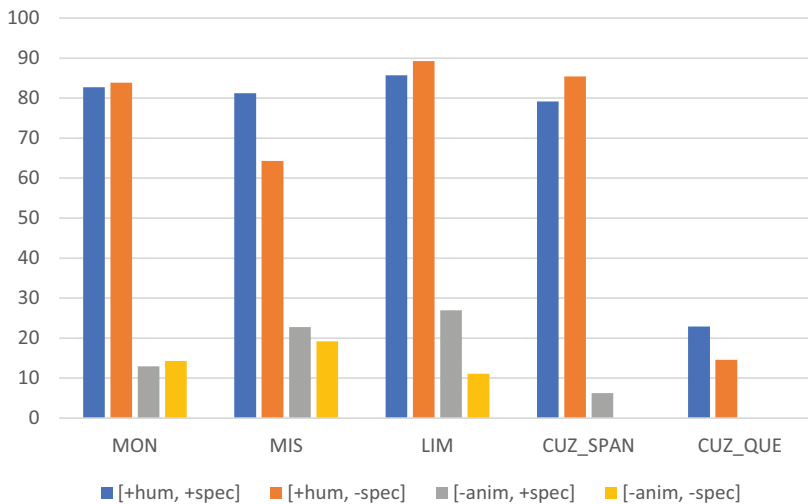
**Table 2:** “Predicted” and found overall rate of *a*-marking.

<b>“Predicted” total 56.25%</b>			
<b>Region</b>		<b>Locality</b>	
Arg/Mon	54.95%	Montevideo	62.12%
		Misiones	47.79%
Peru	42.73%	Lima	59.88%
		Cusco	34.16%

Table 2 suggests that overall, the Argentina/Montevideo group performed almost exactly as expected, whereas Peru showed a considerably lower overall percentage of *a*-marking. However, looking more closely at the four localities under investigation, it transpires that the monolingual groups for both regions actually outperform the predictions, while the multilingual groups, especially Cusco, show a drop in the overall rate of *a*-marking. In what follows we will first take a closer look at the variation found in dataset 1 (specificity and animacy) and 2 (reversible predicates), since the robustness of these datasets allows us to identify interesting patterns of variation; we will then consider trends in individual performances, where the remaining four configurations of the experiment will also be taken into consideration.

### 3.3.1 Animacy and specificity

Before considering the variation between the different groups outlined above, a differentiation has to be made with respect to Cusco. While the participants from this area obviously all had some connection to and experience with Quechua, four of them had Spanish as their L1, while the other four were Quechua natives. Our results show that this leads to a considerable contrast in performance, and therefore we will report the two groups from Cusco separately. Figure 3 shows the percentages of *a*-marking in the four manipulations of the first set of stimuli.



**Figure 3:** Results for animacy and specificity across groups.

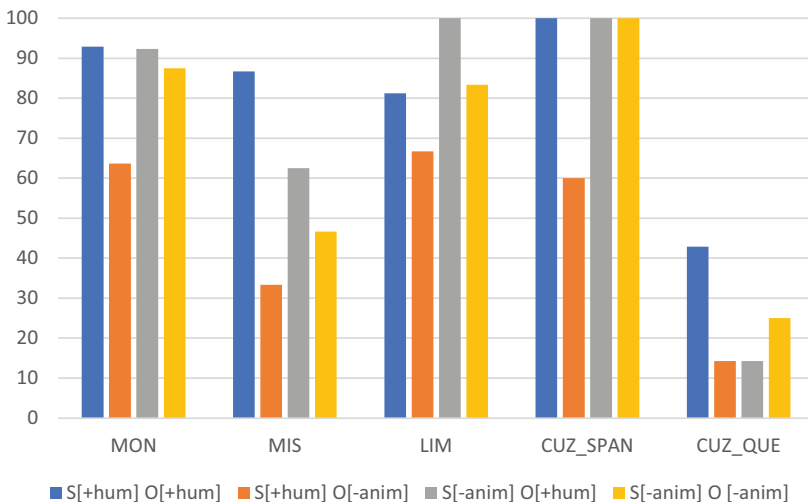
As expected, *a*-marking with human nouns is mostly very high, although never reaching categorical marking. Cross-regional variation in the data is limited, with the exception of the only group that consists of non-native speakers of Spanish, namely the Quechua-Spanish bilinguals from Cusco.<sup>3</sup> There, the rate of *a*-marking with human-reference nouns is roughly comparable to that of *a*-marking with inanimates in the other regions. Specificity plays a less prominent role than animacy in this dataset. Only in Misiones do specific human nouns trigger *a*-marking considerably more often than non-specific nouns. For inanimate objects, only Lima shows a notable contrast with respect to specificity. However,

<sup>3</sup> It should be noted that most cases of *a*-marking from the Quechua L1 group come from only one of the four participants. Cf. Section 3.3.4 for details.

a clear tendency or interaction across groups cannot be identified, neither for animate or for inanimate-reference nouns. Furthermore, note that inanimates are *a*-marked in more than 10% of cases in Montevideo and Lima, at around 20% in Misiones, but almost never in Cusco.

### 3.3.2 Reversible predicates

For reversible predicates, the literature predicts categorical marking. While the overall rates of *a*-marking actually turn out to be very high, even for inanimate objects, the supposed generalization cannot be confirmed, as shown in Figure 4. Instead, we find an interesting pattern of variation across groups.



**Figure 4:** Results for reversible predicates across groups.

Again, the L1 Quechua speakers hardly ever employ *a*-marking. The Spanish L1 group from Cusco, on the other hand, comes closest to categorical marking. Only in the “prototypical” pattern with animate subject and inanimate object do we find that the rate of *a*-marking is not 100%. Lima and Montevideo also have very high ratings for the same three conditions, while in Misiones only the symmetric configuration with human nouns achieves very high percentages of *a*-marking, showing lower rates of *a*-marking in general compared to the other Spanish L1 groups. In all varieties, the “prototypical” pattern produces the lowest rate of *a*-marking.



### 3.3.3 The remaining four configurations

The remaining four configurations will not be discussed individually because, due to the rather low number of observations in each group of speakers, the variation for particular conditions across groups could not be interpreted with a high degree of certainty. Nevertheless, the global pattern found for the previous two sets of stimuli is confirmed. Montevideo, Lima and Spanish L1 speakers from Cusco show the highest rates of *a*-marking in the expected conditions, while it remains close to zero in the L1 Quechua group from Cusco. Speakers from Misiones perform somewhere between these two extremes.

On closer inspection, Quechua L1 speakers only show some marginal *a*-marking in the AcI dataset, while it is zero for all other configurations. The AcI dataset has *a*-marking at very high or categorical levels for all manipulations in the other groups. Example (22) repeats the manipulations from Example (20) in a combined presentation.

- (22) Hicieron / vieron correr (rápidamente) a un mensajero.  
 make.PRF.3PL see.PRF.3PL run.INF fast DOM a messenger  
 ‘They {made a messenger run/saw a messenger running} (fast).’

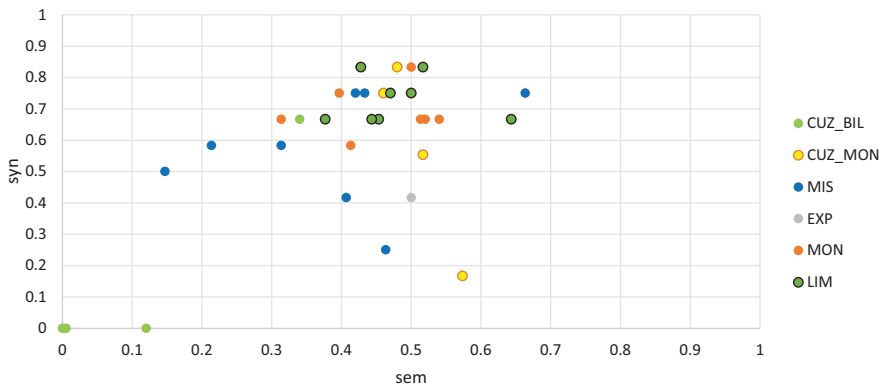
The double accusative structures (*They considered a computer scientist an expert*, cf. example 18 for details) are second highest in eliciting *a*-marking, while secondary predication with *tener* elicited the lowest numbers of all configurations (cf. example 21). The blocking effect expected for the ditransitive structures did not affect definite animate objects in Montevideo and Cusco (with Spanish L1) at all, while the data from Lima and Misiones show a drop in marking for this highly DOM-favouring context (*Cristina gave a parrot to her sister yesterday*). In the remaining three conditions, where we have either indefinite or inanimate features on objects (or both), there is only one single observation of *a*-marking in the whole dataset. This could mean that the blocking effect is stronger for configurations with “optional” marking, but further research is needed to confirm this possibility.

The shared characteristic of these four configurations (AcI, ditransitives, complex objects featuring predicatives or secondary predication) is that they are defined by some structural or constructional property. Thus, their specific syntax plays a more prominent role. The canonical transitive sentences of the first set of stimuli and the reversible predicates of set 2, on the other hand, are straightforward SVO sentences without further structural complexity, and the semantic or discourse properties of the object, such as animacy or definiteness, can be considered as more decisive for the use of *a*-marking than the specific structure of the

sentence. This contrast will be used in the next Section for generating profiles of the performance of individual speakers.

### 3.3.4 Individual profiles

Comparing individual performances yields two further insights: First, we can observe the variability and dispersion within each group, something that has not been shown in the figures above; second, it allows us to see how much overlap there is between the different groups. For this analysis, we calculated two indices for each participant. The first index is the mean rate of *a*-marking in all conditions of the four predominantly syntax-driven configurations, and the second is the mean rate of *a*-marking in all conditions of the merely semantics-driven configurations. When plotted against each other, the picture in Figure 5 emerges.



**Figure 5:** Individual rates of *a*-marking in the elicitation study.

The EXP-point in Figure 5 represents the indices of the predicted results and is located roughly at the midpoint of both axes. Looking at the general pattern of the five groups, Montevideo and Lima show a denser clustering in the upper half of the plot with very few outliers and a high degree of overlap. In terms of the two axes, there seems to be hardly any difference with regard to syntax (y-axis) while it could be argued that Montevideo has somewhat higher *a*-marking rates on the semantic dimension (x-axis), since most of the respective dots are further to the right than those representing Lima. The three remaining groups show much more dispersion, also partly occupying the lower half of the plot. The eight speakers from Misiones have a wide range of dispersion on both axes. They are clearly the

most dispersed group, showing no clear cluster. The Spanish L1 group from Cusco also shows more dispersion, but essentially on the syntactic dimension, while they cluster around the midpoint of the scale as far as semantics is concerned. Again, the Quechua L1 speakers from Cusco requires special comment. As can be seen from the plot, and as already mentioned in footnote 3, a great deal of *a*-marking in this group is due to just one participant, while the other speakers show almost no marking at all. For two speakers, the overall rate of *a*-marking is zero, while one participant marked one object from the canonical transitive set and one from the reversible predicates set. The performance of this one exceptional Quechua L1 speaker is closer to the Spanish L1 groups than to the rest of the speakers of Quechua, but even as an outlier, his profile is still located in the transitional area between his fellow Quechua natives and the core of the Spanish L1 speakers.

## 4 Discussion

In this Section, we would like to focus on the following three issues with respect to the results presented above: (i) the reliability of the data, (ii) the variation found in the data and its import on claims in the literature, and (iii) the value of the results regarding the status of DOM in the examined varieties.

While the data collection followed a strict experimental design and the same protocol in all locations, and thus allows for a high degree of comparability of the linguistic material under investigation, one question that arises is the robustness of the findings, since the sample of participants for each location is relatively small. Another issue that could be raised is that the elicitation tasks are somewhat artificial and hence might not represent normal language use. For both caveats, it is important to point out that the results presented are part of a larger research project on the variation of DOM in different locations of the Spanish-speaking world. Wall et al. (2020) present a more robust dataset of more than 40 participants from the same experiment in Lima and Montevideo. Expanding the dataset for these two varieties does not change the general tendencies drastically. In fact, for Lima all participants cluster around the same region indicated in Figure 5 above. While the expanded dataset for Montevideo shows more dispersion in this respect, it does not reach the amount found for Misiones. Unfortunately, there are no larger datasets for the contact zones. Nevertheless, there are more than 270 individual recorded sentences for each location in the dataset presented above, of which – still per region – more than 100 correspond to the canonical transitive set and more than 50 to the reversible predicates set. Thus,

at least for these two datasets we have a considerable number of data points per speaker. It goes without saying that these results should by no means be considered as final and representative for the respective regions. However, this is true for any isolated experiment, for which replication is crucial. Regarding the two contact zones, it is furthermore unclear whether we should assume stable varieties in these contact scenarios in the first place, and it is even less clear what representativity would mean even for larger groups of speakers in those areas. What the findings of this study can provide is a first indication of putative differences in the two contact regions with respect to predominantly monolingual speakers. They also can give us a first impression of some general tendencies for canonical transitive sentences and reversible predicate constructions.

As for the artificiality of the elicitation process, participants were asked about their experiences with the tasks and as to the possible purpose of the experiment. Almost no one was able to guess the research subject; only one participant noted that he had been adding the preposition *a* to his sentences multiple times, but he was not able to identify the part of speech of interest to us or comment on the argument structure of the sentences. Some participants reported that they needed time to get used to the task, which was not a problem since there were no time constraints in the experiment. While the form of presentation of the stimuli requires a certain degree of literacy, which was checked for beforehand, none of the recruited participants found it impossible to construct sentences out of the presented material. Most found the experience interesting or challenging in a positive way, and none aborted the experiment. It is of course impossible to say whether participants would produce exactly the same amount of variation in more spontaneous conversations, yet the elicited sentences do reflect some of the general tendencies reported in the literature. Also, the results do not show inconsistencies or contradictory behaviour. In our experience, *a*-marking of direct objects is also relatively unsusceptible to the drawbacks of a more artificial elicitation task. The form does not carry any prominent social or expressive meaning and its use is highly unconscious and automatized. Neither in the literature nor in our fieldwork experience have we come across evidence suggesting that *a*-marking of direct objects might require a notable amount of preparatory processing or that performance constraints would have a strong impact on it. Therefore, we argue that our results are quite a good approximation to normal language use.

While our results reproduce tendencies that have been described in the literature, they do not fully match the predictions we derived from prior studies. As has been pointed out above, these predictions are idealized and should be taken with a grain of salt. However, we would also like to answer the question as to why the results diverge from the predictions in the way they do, at least for the two more robust datasets. One first case in point would be that animate specific indef-

indefinites should be categorically marked, and they are not. As Figure 3 shows, they are at around 80% for most groups, only exceptionally reaching 90%. Of course, reaching 100% in performance could be considered unrealistic in general, and even more so since the use of a marker has a probabilistic component. In our view, however, a score as low as 80% can probably not be attributed solely to confounding factors. Rather, we suspect that in addition, the context sentences did not always work as expected and the priming context intended to implicate specificity might not have been strong enough. This seems to be corroborated by the fact that the (non-)specificity manipulation of our context sentences did not produce contrasts in most groups and that in most of them, *a*-marking rates are indeed slightly higher when the referent was not introduced in the context sentence. Thus, the lower numbers of *a*-marking on animate indefinites (specific and non-specific) could be due to speakers constructing the (non-)specificity of those referents based on factors other than the cues from the context sentences. The role of the context sentence in this kind of elicitation study clearly needs further investigation.

The sentences with reversible predicates produced very high rates of *a*-marking in general but clearly did not lead to categorical marking, with the exception of the Spanish L1 group from Cusco, where we have categorical marking in three out of four conditions. It is important to recall that the arguments in this set of stimuli were given as definite NPs. Unlike the set of canonical transitive sentences, where the arguments were formally indefinite, these stimuli produce the expected high rate of *a*-marking, averaging around 90% in most groups. Here, context cannot be invoked to explain the lack of *a*-marking. Further investigation is needed in order to determine whether other factors might be involved here, or whether this is the range of the probabilistic component in DOM for definite animates in language use or whether it is a consequence of the given task after all. While this issue cannot be resolved here, what we can learn from our results is that the claims in the literature have been oversimplified, since they do not differentiate between the four possible combinations of (in)animate subjects and objects. Our results, however, show that for the “prototypical” alignment of animate subject and inanimate object, the rate of *a*-marking is considerably lower than for the other three conditions, although remaining above 50% in most groups. Another interesting observation is that the symmetrical alignment with inanimate arguments produces considerably higher rates of *a*-marking than the “prototypical” alignment. This is not only a new observation; it is also strong evidence for theories that argue for a “global” explanation of *a*-marking where not only local factors (i.e. the object domain and how the object is related to the verb) are considered as relevant, but also the configuration of other parts of the sentence (for instance, the type of subject).

Turning to the third and final open issue, as was noted in Section 1.3, this study focuses on the “righthand” side of the animacy and definiteness scale, as provided in (15). This is the area where variation is expected to be present, namely in the transition from definite animates to indefinite animates. Thus, pronouns and proper names are not included in the study. Nevertheless, a number of conclusions can still be drawn with respect to the other three categories, which constitute an important part of the scale. First, compared to the other four groups, the Quechua L1 group clearly stands out by producing almost no *a*-marking.<sup>4</sup> It is of course not possible to determine whether the Quechua L1 speakers feature a DOM system as part of their Spanish grammar at all, given the previously mentioned restrictions of the study. In any case, that DOM system would not involve the marking of animate definites, which are well represented in the first two sets of stimuli. This finding is unexpected on the view that DOM should be easily transferable for speakers of languages that have nominative-accusative alignment in their L1 (Döhla 2011). If it is easily transferable to their L1, it should arguably also be easily acquirable in the L2 before. Quechua is a language with nominative-accusative alignment, but while our participants freely switch between Spanish and Quechua in their daily lives, they have not acquired DOM as expected. Compared to learners of Spanish that have DOM in their L1 (such as Romanian or Turkish), who acquired a very good command of Spanish DOM after a few years (Montrul/Gürel 2015; Montrul 2019), this is remarkable. Thus, nominative-accusative alignment alone might not be sufficient for an “easy transfer”. It should be recalled that for a nominative-accusative language lacking DOM, such as English, this property of Spanish is among the most difficult to master, and that in heritage speakers, the DOM system is among the first features to be lost by interference (Montrul/Bowles 2009; Montrul/Walcker-Mayer 2013). We cannot exclude the possibility that our Quechua speakers show *a*-marking on stressed pronouns, but even if they do, this system would be rather reminiscent of such rudimentary systems as those found, for instance, in Portuguese, but not the highly grammaticalized ones common in Spanish. Interestingly, the DOM system of the Spanish L1 speakers from Cusco comes closest to the predictions in the literature and shows little variation: for reversible predicates, categorical *a*-marking in three out of four conditions; for canonical transitive sentences, high rates of *a*-marking on animates and practically none on inanimates. The issue of whether this pattern is generalized among Spanish L1 speakers in that region needs further investigation.

---

<sup>4</sup> Remember that almost all the (few) occurrences of the marker are from one individual, whose profile is somewhat different from those of the other three.

Finally, for inanimate objects we have been able to confirm *a*-marking in the Rio de la Plata region and show that in Misiones and Lima similar rates can be expected in canonical transitive sentences. This is the first dataset that allows for such direct comparison. Tippets (2011) reports 8% of *a*-marking on inanimates for Buenos Aires, a slightly lower rate than ours for Montevideo, and considerably lower than the results from Misiones. However, since it is not clear whether Tippets only considered what we call canonical transitive sentences (we suspect that this was not the case), it is difficult to relate our findings to those. Already the direct comparison between canonical transitives and reversible predicates shows that although both sentences have a simple SVO structure, *a*-marking rates are very different. We can expect this contrast to become stronger for different and more complex structures.

## 5 Summary and conclusion

We have presented a new elicitation tool for collecting highly comparable datasets on the variational range of DOM in Spanish, and the first results for four varieties from two regions in South America. The focus of the study was on two varieties from zones where Spanish is in contact with other languages, namely Cusco and Misiones. For both contact zones we included reference groups from predominantly monolingual surroundings: Lima as a point of comparison for Cusco and Montevideo as a representative of River Plate Spanish and reference point for Misiones. The stimuli included in the elicitation task provided data on six DOM-sensitive constructions, two of which were explored in this study in more detail.

As for the contact regions, the general findings were that Quechua L1 speakers produced almost no *a*-marking in any set of stimuli. We have not investigated the use of Quechua object syntax of these speakers, but given that most of them do not show signs of a developed DOM system in their Spanish, it seems doubtful that they have transferred DOM from Spanish into their variety of Quechua. Thus, although we did not investigate the contact language, a highly plausible interpretation is that no support has been found for the speculation in Döhla (2011) that DOM systems being common in many languages makes them easily transferable from one language to another. The findings of our study are in line with other recent empirical findings on multilingual settings and on the acquisition of Spanish DOM, which is not easily acquired when the L1 does not have a similar DOM system. The speakers from Misiones do have an articulated DOM system, yet it differs in some aspects from that of the predominantly monolingual speak-

ers. For that region, we also found stronger contrasts between individual speakers than in the predominantly monolingual zones, where we observed stronger clustering.

While other experimental studies have focused on less phenomena in favour of more statistical power in the results, our findings are, for the time being, limited to a preliminary overview. However, the method used in the present study can also yield more robust quantitative results if the database is expanded. With our method, we were able to replicate several general tendencies described in the literature, such as *a*-marking on inanimates and high rates of *a*-marking in sentences featuring reversible predicates, and we showed that these cases have to be treated separately in empirical terms. We also provided, for the first time, a highly comparable dataset that allows for direct comparison of the variation in the varieties under investigation and that promises even more interesting findings once applied to further regions of the Spanish-speaking world.

## Bibliography

- Alfaraz, Gabriela G., *Accusative object marking. A change in progress in Cuban Spanish?*, Spanish in Context 8:2 (2011), 213–234.
- Amable, Hugo W., *Figuras del habla misionera*, Santa Fe, Colmegna, 1975.
- Balasz, Sonia, *Factors determining Spanish Differential Object Marking within its domain of variation*, in: Michnowicz, Jim/Dodsworth, Robin (edd.), *Selected proceedings of the 5<sup>th</sup> workshop on Spanish sociolinguistics*, Somerville, Mass., Cascadilla, 2011, 113–124.
- Bautista-Maldonado/Montrul, Silvina, *An experimental investigation of Differential Object Marking in Mexican Spanish*, Spanish in Context 16:1 (2019), 22–50.
- Bossong, Georg, *Historische Sprachwissenschaft und empirische Universalienforschung*, Romanistisches Jahrbuch 33 (1982), 17–51.
- Bossong, Georg, *Die romanischen Sprachen. Eine vergleichende Einführung*, Hamburg, Buske, 2008.
- Company Company, Concepción, *El avance diacrónico de la marcación prepositiva en objetos directos inanimados*, in: Bernabé Pajares, Alberto, et al. (edd.), *Presente y futuro de la lingüística en España. La Sociedad Lingüística, 30 años después. Actas del II Congreso de la Sociedad Española de Lingüística*, vol. 2, Madrid, Consejo Superior de Investigaciones Científicas, 2002, 146–154.
- De Ramos, Patricia, *O quadro de estratégias de realização do objeto direto anafórico de terceira pessoa no espanhol de Oberá-Misiones (Argentina). Resultados*, in: *Anais do XVII Colóquio de Pós-graduação e Pesquisa em Letras Neolatinas*, Rio de Janeiro, UFRJ, 2017, 220–228.
- Döhla, Hans Jörg, *Differential Object Marking (DOM) in some American Indian languages. Contact induced replication and convergence or internal development?*, in: Mendoza, Imke/Pöll, Bernhard/Behensky, Susanne (edd.), *Sprachkontakt und Mehrsprachigkeit als*



- Herausforderung für Soziolinguistik und Systemlinguistik*, München, Lincom Europa, 2011, 27–45.
- Draxler, Christoph/Jänsch, Klaus, *SpeechRecorder. A Universal platform independent multi-channel audio recording software*, in: *Proceedings of the IV. International Conference on Language Resources and Evaluation*, Lisbon, European Language Resources Association, 2004, 559–562.
- Dumistrescu, Domnita, *El parámetro discursivo en la expresión del objeto directo lexical. Español madrileño vs. español porteño*, *Signo y Seña* 7 (1997), 303–354.
- Escobar, Anna María, *Spanish in contact with Quechua*, in: Díaz-Campos, Manuel (ed.), *Handbook of Hispanic sociolinguistics*, Oxford, Blackwell, 2011, 323–352.
- Fábregas, Antonio, *Differential Object Marking in Spanish. The state of the art*, *Borealis* 2:2 (2013), 1–80.
- Fernández-Ordóñez, Inés, *Leísmo, laísmo y loísmo*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española*, vol. 1, Madrid, Espasa-Calpe, 1999, 1317–1397.
- Flores, Marcela/Melis, Chantal, *El leísmo desde la perspectiva del “marcado diferencial de objeto”*, *Revista de Historia de la Lengua Española* 2 (2007), 83–107.
- García García, Marco, *Differentielle Objektmarkierung bei unbelehten Objekten im Spanischen*, Berlin/Boston, de Gruyter, 2014.
- Hoff, Mark, *Are Argentines “a”-blind? Acceptability of “a”-marked inanimate direct objects*, in: MacDonald, Jonathan E. (ed.), *Contemporary trends in Hispanic and Lusophone linguistics. Selected papers from the Hispanic Linguistic Symposium 2015*, Amsterdam, John Benjamins, 2018, 121–142.
- López, Luis, *Indefinite objects. Scrambling, choice functions and Differential Marking*, Cambridge, Mass., MIT Press, 2012.
- Montrul, Silvina/Bowles, Melissa, *Back to basics. Incomplete knowledge of Differential Object Marking in Spanish heritage speakers*, *Bilingualism: Language and Cognition* 12:3 (2009), 363–383.
- Montrul, Silvina/Gürel, Ayşe, *The acquisition of Differential Object Marking in Spanish by Turkish speakers*, in: Judy, Tiffany/Perpiñán, Silvia (edd.), *The acquisition of Spanish in understudied language pairings*, Amsterdam, John Benjamins, 2015, 281–308.
- Montrul, Silvina/Walcker-Mayer, Noélia, *Differential Object Marking in child and adult Spanish heritage speakers*, *Language Acquisition* 20:2 (2013), 109–132.
- Montrul, Silvina, *La marcación diferencial del objeto directo en el español de Argentina. Un estudio experimental*, in: Colantoni, Laura/Rodríguez-Louro, Celeste (edd.), *Perspectivas teóricas y experimentales sobre el español de la Argentina*, Madrid, Iberoamericana, 2013, 207–228.
- Montrul, Silvina, *The acquisition of Differential Object Marking in Spanish by Romanian speakers*, *Spanish Journal of Applied Linguistics* 32:1 (2019), 185–219.
- Pensado, Carmen, *La creación del complemento directo preposicional y la flexión de los pronombres en las lenguas románicas*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 179–233.
- Pfänder, Stefan, *Gramática mestiza. Con referencia al castellano de Cochabamba*, La Paz, Instituto Boliviano de Lexicografía y otros Estudios Lingüísticos, 2009.
- Sanicky, Cristina A., *The pronunciation of Spanish in Misiones, Argentina*, Davis/California, PhD Thesis, University of California, 1981.

- Silverstein, Michael, *Hierarchy of features and ergativity*, in: Dixon, Robert M. W. (ed.), *Grammatical categories in Australian languages*, Canberra, Australian Institute of Aboriginal Studies, 1976, 112–171.
- Ticio, M. Emma, *Differential Object Marking in Spanish-English early bilinguals*, *Linguistic Approaches to Bilingualism* 5:1 (2015), 62–90.
- Tippets, Ian, *Differential Object Marking. Quantitative evidence for underlying hierarchical constraints across Spanish dialects*, in: Ortiz-López, Luis A. (ed.), *Selected proceedings of the 13<sup>th</sup> Hispanic Linguistics Symposium*, Somerville, Mass., Cascadia, 2011, 107–117.
- Torrego, Esther, *El complemento directo preposicional*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española*, Madrid, Espasa-Calpe, 1999, 1780–1805.
- von Heusinger, Klaus/Kaiser, Georg A., *Affectedness and Differential Object Marking in Spanish*, *Morphology* 21:3–4 (2011), 593–617.
- Wall, Albert, et al., *The variation of Differential Object Marking in Spanish. Experimental data from four varieties across six constructions*, Talk given at the 53<sup>rd</sup> annual meeting of the Societas Linguistica Europaea, 2020 <<https://osf.io/ekn5t/>> [last access: 23.11.2020].
- Weissenrieder, Maureen, *A functional approach to the accusative “a”*, *Hispania* 74:1 (1991), 146–156.

Alina Tigău

# Differential Object Marking in Romanian and Spanish

A contrastive analysis between differentially marked and unmarked direct objects

**Abstract:** This paper discusses some aspects related to the syntax and semantics of Romanian direct objects (DOs) from a comparative stance with their Spanish counterparts. Spanish and Romanian differentially object marked DOs (DOMed DOs) function as KPs and may have a specific or a wide scope reading, they are disallowed in contexts requiring property denoting nominals and allowed in contexts necessitating nominals with ‘real argumenthood’ i.e., denoting entities or generalized quantifiers. The two languages differ with respect to unmarked DOs (which generally have a DP status): these pattern with DOMed DOs in Romanian. Spanish unmarked DOs, on the other hand, never show a specific or wide scope interpretation, they are not allowed in contexts which require entity denoting nominals but are suitable for contexts where property denoting expressions are required. We posit that the parameter differentiating between Romanian and Spanish concerns the division these languages make regarding DO types: while Spanish draws a line between KPs on the one hand and DPs and NPs/NumPs on the other, the relevant cut-off point in Romanian is that between KPs and DPs on the one hand and NPs/NumPs on the other. An analysis is proposed for Romanian, starting from these observations and building on López (2012)’s analysis for Spanish.

**Keywords:** direct objects, Differential Object Marking, specificity, scope, argumental DPs

---

**Acknowledgments:** I would like to thank the audience of the conference *Differential Object Marking in Spanish – diachronic change and synchronic variation*, University of Zürich, in June 2018, for their critical comments and suggestions and two anonymous reviewers for very helpful comments. The research for this paper has been funded by the Alexander von Humboldt Foundation.

---

**Alina Tigău** University of Bucharest [alina.tigau@lils.unibuc.ro](mailto:alina.tigau@lils.unibuc.ro)

# 1 Introduction

The aim of this paper is twofold: to investigate the behaviour of Romanian direct objects (DOs) from a comparative perspective with their counterparts in Spanish and to propose a tentative analysis of Romanian DOs by parametrizing the account put forth in López (2012) for Spanish DOs. It will be shown that while differentially object marked DOs (DOMed DOs) pattern similarly in both languages, are able to exhibit a specific interpretation, outscope other scope bearing expressions and occur in a number of contexts selecting only DPs possessing real argumenthood,<sup>1</sup> unmarked DOs differ with respect to these properties in the two languages. Thus, Spanish unmarked DOs pattern with bare nominals in that they read unambiguously non-specifically, never outscope other scope bearing expressions and may be only employed in contexts requiring property-denoting nominals. Romanian unmarked DOs, on the other hand, seem to be able to have the same distribution and acquire the same readings as DOMed variants: they may read specifically and have wide scope interpretations and may also occur in contexts restricted to nominals possessing real argumenthood.

As it seems, Spanish draws a clear-cut distinction between marked DOs on the one hand and unmarked DOs and bare nominals on the other, whereas Romanian groups marked and unmarked DOs together and distinguishes between these and bare nominals. Given these observations, DOMed DOs will be analyzed as KPs on a par with their Spanish counterparts along the lines of López (2012): they will be argued to move into an intermediary position between  $v$  and  $V$  for reasons of case. Aspects related to specificity, scope etc. are shown to derive from this position as a consequence of a special mode of semantic composition of the DO DP with the predicate.

Romanian unmarked DOs will be shown to differ from their Spanish counterparts exhibiting a twofold behaviour i.e., as real DPs and as NPs/NumPs: as DPs, they may have argumental status and behave on a par with marked DOs. On the other hand, they may undergo reanalysis to NP/NumP and pattern similarly to other property denoting nominals (e.g., bare plurals). In this latter case, they closely follow in the footsteps of Spanish unmarked DOs.

The article has the following structure: Section 2 brings forth some relevant semantic and syntactic aspects regarding indefinite DOs in Spanish and Roma-

---

<sup>1</sup> In line with López (2012), *real arguments* are nominals of the semantic type  $e$  or  $\langle\langle e, t \rangle t \rangle$ , which never denote properties and never incorporate into the  $V$  for case checking reasons. As will be seen, Spanish marked DOs have real argumenthood in this respect, while unmarked DOs are property denoting nominals which always check their case as a way of incorporation into the  $V$ .

nian; Section 3 discusses the analysis proposed in López (2012) for Spanish DOs; in Section 4 we propose a parametrization of this analysis for Romanian; Section 5 contains the conclusions.

## 2 Comparing Spanish and Romanian indefinite DOs

### 2.1 Spanish indefinite DOs

Spanish draws a clear-cut distinction between DOMed DOs on the one hand and unmarked DOs and bare nominals on the other with respect to a number of syntactic and semantic phenomena: DOMed DOs may exhibit a specific interpretation, outscope other scope bearing expressions, occur in certain contexts which exclude property denoting nominals and shun those contexts requiring property denoting nominals.<sup>2</sup>

Unmarked DOs evince a different behaviour in that they never enable specific or wide scope readings, occur in contexts requiring property denoting nominals and get discarded from contexts eliciting real arguments (i.e., nominals denoting entities or generalized quantifiers). This section will extend upon this clear-cut distinction that Spanish seems to draw between DOMed DOs and their unmarked counterparts, making use of the data presented in López (2012). By drawing on this different behaviour, López (2012) offers an account for Spanish marked and unmarked DOs, which will be presented in section 3 and which will be further adapted for Romanian in the subsequent sections after a comprehensive discussion about the relevant differences holding between the DOs in the two languages.

#### 2.1.1 Specificity

Spanish indefinite DOs may be introduced by the differential object marker *a* which seems to have an important interpretive import: marked DOs may acquire

---

<sup>2</sup> A note of caution is in order here: as pointed out by one of our reviewers, the data on the variation of DOM in Spanish is quite vast and the brief presentation in this article does not do justice to the richness of all the studies in the literature. In this respect, we have chosen López's (2012) work, given that it provides an accurate picture of the basic tenets of DOM marking in Spanish and the lack thereof, providing us with a sturdy point of comparison for the Romanian data.

a specific reading, unmarked DOs may not do so.<sup>3</sup> This is illustrated in (1) below, where the variant *a un traductor de alemán* refers to a specific translator that Mary is looking for, while in the unmarked variant Mary is simply looking for some non-specific individual who has the property of being a translator:

- (1) *María busca a/∅ un traductor de alemán.*  
 Maria seeks DOM/∅ a translator of German  
 ‘Maria is looking for a German translator.’

López (2012, 10)

Expectedly, a modifier such as *cierto* (‘a certain’), which has been shown to foreground an epistemically specific interpretation given that it forces the referent denoted by the indefinite DO to become fixed with respect to the speaker’s epistemic modal base, imposes the use of *a*. An unmarked indefinite is ungrammatical when preceded by *cierto*:

- (2) *Juan buscó a/\*∅ un cierto futbolista.*  
 Juan sought DOM/\*∅ a certain soccer player  
 ‘Juan looked for a certain soccer player.’

López (2012, 17)

On the other hand, the free choice indefinite *cualquiera* (‘any’) drives DOMed DOs to becoming non-specific. In (3) below, *a un futbolista cualquiera* may only be interpreted non-specifically. As expected, the unmarked variant is also fine in this context:

- (3) *Juan buscó a/∅ un futbolista cualquiera.*  
 Juan sought DOM/∅ a soccer player any  
 ‘Juan looked for a soccer player, no matter who.’

López (2012, 17)

---

<sup>3</sup> With respect to the type of specificity that marked indefinite DOs may evince, López argues in favour of *epistemic* and *partitive* specificity (Farkas 1999) disregarding other types of specificity e.g., specificity as referential anchoring (von Heusinger 2011). There is also a split between *wide scope* and *epistemic specificity*, which other studies have subsumed as types of specificity (consider Farkas’ 1994 notion of *scopal specificity*). Furthermore, other studies have shown that partitives need not necessarily be specific (Kornfilt/von Heusinger 2008). In this article, we simply adopt the specificity distinctions endorsed by López in an attempt to capture a parallelism of the Spanish and Romanian data and do not engage in a discussion regarding specificity types.

Mood has also been proposed as a useful tool to tease a specific interpretation apart from a non-specific one (Rivero 1979). In (4) the DO *una gestora* ('a manager') has been modified by a relative clause whose predicate bears the subjunctive mood. As a consequence, the indefinite DP may only be interpreted as non-specific. Both marked and unmarked DOs may be used in this context.

However, if the mood of the predicate in the relative clause modifying the DO is the indicative, the only available type of DO is a DOMed one and the only available interpretation is the specific one, as shown in (5) where the German speaking manager sought for by Mary is necessarily interpreted as specific:

- (4) *María buscó a/∅ una gestora que hablara alemán.*  
 Maria sought DOM/∅ a manager that spoke.SUBJ German  
 'Maria was looking for a manager that spoke.SUBJ German.'  
 López (2012, 1)

- (5) *María buscó a/\*∅ una gestora que hablaba alemán.*  
 Maria sought DOM/\*∅ a manager that spoke.IND German  
 'Maria was looking for a manager that spoke.IND German.'  
 López (2012, 2)

The subjunctive test may be further combined with *un cierto* or *cualquiera*: while *un cierto* needs to be combined with the indicative mood and a DOMed DO (6a), *cualquiera* imposes the use of the subjunctive and an unmarked DO (6b):

- (6) a. *María buscó a/\*∅ una cierta gestora que habla/*  
 Mary searched DOM/\*∅ a certain manager who speaks.IND/  
 \*hable alemán.  
 speak.SUBJ German  
 'Maria looked for a certain manager that speaks.IND/\*SUBJ German.'
- b. *María buscó \*a/ una gestora cualquiera que*  
 Mary searched DOM/∅ a manager any who  
 \*habla/ hable alemán.  
 speaks.IND/ speak.SUBJ German  
 'Maria looked for a manager (no matter what) that speaks. \*IND/SUBJ  
 German.'
- López (2012, 19)

### 2.1.2 Scope

Besides showing a propensity for a specific interpretation, DOMed indefinites also seem to favour a wide scope reading when co-occurring with extensional quantifiers and various sentence operators. In (7a), *a una mujer* ('a woman') may outscope the universal QP *todo hombre* ('every man') enabling an interpretation according to which 'there was (at least) one woman such that every man loved one'. A narrow scope interpretation for the indefinite DO according to which 'each man loved a (possibly) different woman' remains an option. The same is at stake in (7b), with the DOMed DO allowing both a wide as well as a narrow scope reading with respect to the QP subject:

- (7) a. *Todo hombre amó a una mujer.*  
 every man loved DOM a woman  
 'Every man loved a woman.'  
 $\exists > \forall$  and  $\forall > \exists$
- b. *La mayoría de los hombres amó a una mujer.*  
 the majority of the men loved DOM a woman  
 'Most men loved a woman.'  
 $\exists > \text{Most}$  and  $\text{Most} > \exists$

López (2012, 10)

Unlike its DOMed counterpart, the unmarked DO may not outscope the QP subject and may only give rise to a narrow scope interpretation as a consequence: the only available reading in (8) is one according to which for every man/most men there exists (at least) one woman such that the respective man loves her.

- (8) a. *Todo hombre amó una mujer.*  
 every man loved a woman  
 'Every man loved a woman.'  
 $* \exists > \forall$  and  $\forall > \exists$
- b. *La mayoría de los hombres amó una mujer.*  
 the majority of the men loved a woman  
 'Most men loved a woman.'  
 $* \exists > \text{Most}$  and  $\text{Most} > \exists$

López (2012, 10)

Marked indefinite DOs may also outscope negation. The unmarked indefinite only allows for a narrow scope interpretation:



- (9) a. *Juan no amó a una mujer.*  
 Juan not loved DOM a woman  
 ‘There was a woman such that Juan did not love.’  
 \* ‘Juan did not love any woman.’  
 $\exists > \neg$  and  $*\neg > \exists$   
 López (2012, 10)
- b. *Juan no amó una mujer.*  
 Juan not loved a woman  
 \* ‘There was a woman such that Juan did not love’  
 ‘Juan did not love any woman.’  
 \*  $\exists > \neg$  and  $\neg > \exists$   
 López (2012, 10)

Furthermore, Spanish unmarked indefinite DOs may not outscope the conditional operator. DOMed indefinites, on the other hand, may acquire wide scope with respect to the conditional:

- (10) a. *Si Lud invita a un filósofo, Bert se ofenderá.*  
 ‘If Lud invites DOM a philosopher, Bert REFL offend.FUT.’  
 $\exists > \rightarrow$  and  $\rightarrow > \exists$
- b. *Si Lud invita  $\emptyset$  un filósofo, Bert se ofenderá.*  
 ‘If Lud invites  $\emptyset$  a philosopher, Bert REFL offend.FUT.’  
 \*  $\exists > \rightarrow$  and  $\rightarrow > \exists$   
 López (2012, 2)

Thus, just like in the case of specific readings, the split between DOMed DOs and unmarked DOs also holds with respect to scope dependencies: while marked DPs may outscope other scope bearing expressions, unmarked correspondents only exhibit dependent readings.

### 2.1.3 Some contexts which prohibit the use of DOMed indefinite DOs

Bleam (2005) discusses a number of contexts where DOMed DOs are discarded as infelicitous. One such context is that of the existential *haber* (‘have’) and the possessor or relator *tener* (‘have’), which only allow unmarked DOs.

*Haber* always selects unmarked indefinites, which are property denoting: DOMed DPs are disallowed from these contexts. Unmarked variants on the other hand, which are of type  $\langle e, t \rangle$  are felicitous.

- (11) *En el patio hay \*a/∅ un niño.*  
 in the yard EXIST \*DOM/∅ a boy  
 ‘There is a boy in the yard.’

López (2012, 20)

*Tener* allows for both DOMed and unmarked DOs. A difference in interpretation is, however, at stake. Bleam (1999; 2005) distinguishes between an individual level *tener* and a stage-level one. In (12a) *tener* functions as an individual level predicate and is the equivalent of *own* in (13a) in that the possession relationship is not associated with or restricted by a particular spatio-temporal location. In this particular context, the use of DOM is disallowed. (12b) on the other hand prompts the stage level reading and the use of *a* is permitted:

- (12) a. *María tiene \*a/∅ tres hijos.*  
 ‘Maria has \*DOM/∅ three children.’  
 b. *María tiene a/ un hijo en el ejército.*  
 ‘Maria has DOM/ a son in the army.’

López (2012, 20–21)

- (13) *I have a car.*  
 a. *I own a car* (individual-level)  
 b. *I have a car (with me today)* (stage level)

In order to account for these facts, Bleam suggests that *tener* always selects a property-denoting expression. Nevertheless, in its stage-level use, *tener* takes a complement of type <s,t>, denoting a property over events, which syntactically amounts to a small clause containing the subject DP and a spatio-temporal predicate. In its individual-level interpretation, *tener* selects an <e,t> complement, denoting a property of individuals. This also accounts for the fact that in this latter use, the NP complement may never be definite (see Bleam 1999 for discussion): as known, definite descriptions semantically correspond to individual constants and are of type e.

In the stage-level use of *tener*, definite nominals are allowed and this is so due to the fact that the DP is not itself a complement of *tener*, but the subject of the small clause, which *tener* takes as its complement. As such, the DP subject occupies an argumental position.

The difference of behaviour between marked and unmarked DOs in these contexts thus suggests a difference of status: while DOMed DOs are true arguments of the verb, being able to occur in the stage-level use of *tener*, their

unmarked correspondents do not have real argumenthood, being interpreted as property-denoting nominals and only co-occurring with the individual-level of *tener*.

### 2.1.4 Some syntactic phenomena which do not involve scope or specificity

López (2012) identifies three special contexts where the use of DOM seems to be required in the absence of any semantic triggers such as *specificity* or *scope*. The obligatoriness of DOM in these contexts prompts López (2012) to propose that the marking mechanism is actually syntactically triggered.

#### Small clause complements

One context where the use of DOM is compulsory is that of *small clause complements*: in (14) *un estudiante*, the argument within the small clause complement is obligatorily DOMed; the unmarked variant is discarded as ungrammatical and so is the bare plural in (15):

- (14) *Considero a/\*∅ un estudiante inteligente.*  
 consider.I DOM/\*∅ a student intelligent  
 ‘I consider a student intelligent.’

López (2012, 10)

- (15) *El profesor consideró a/\*∅ estudiantes inteligentes.*  
 the professor considered DOM/\*∅ students intelligent  
 ‘The professor considered students intelligent.’

López (2012, 23)

The fact that DOM has not been called for by any semantic trigger in this context is accounted for by examples such as (16) below where the argument of the small clause is an indefinite DP with a non-specific interpretation, given the use of the subjunctive in the modifying relative. The use of DOM is compulsory:

- (16) *Juan no considera honrado a/\*∅ un hombre que*  
 Juan not considers honest DOM/\*∅ a man that  
*acepte sobornos*  
 accepts.SUBJ bribes  
 ‘Juan does not consider honest a man that accepts bribes.’

López (2012, 25)

**Object control predicates**

The object of an *object control predicate* also requires the use of DOM:

- (17) *Juan forzó a/\*∅ un niño a hacer los deberes.*  
 Juan forced DOM/\*∅ a boy to do.INF the homework  
 ‘John forced a child to do his homework.’

López (2012, 25)

Just as in the case of small clause complements, the use of DOM with object control predicates does not seem to be imposed by any semantic considerations: in (18) the DOMed DO has been modified by a relative clause whose predicate bears the subjunctive mood:

- (18) *María forzaría a/\*∅ una empleada que tuviera*  
 Maria force.COND DOM/\*∅ an employee that had.SUBJ  
*depresión a venir al trabajo.*  
 depression to come.INF to work  
 ‘Mary would force an employee who were depressed to come to work.’

López (2012, 25)

**Clause union**

A third syntactic context where DOM is required in Spanish is that provided by accusative affected arguments in *clause union*. In the causative construction presented under (19) below, the *causee* in the accusative is necessarily DOMed. An unmarked indefinite or a bare plural is out.

- (19) a. *María hizo llegar tarde a/\*∅ un niño.*  
 Maria made arrive late DOM/\*∅ a boy  
 ‘Maria made a boy arrive late.’  
 b. *\*María hizo llegar tarde a/ niños.*  
 Maria made arrive late DOM/ boys  
 ‘Maria made boys arrive late.’

López (2012, 24)

Again, DOM is not triggered by specificity as may be seen from (20) where the relative clause modifying the *affectee* contains the subjunctive mood.

- (20) *María hace quedarse en clase a/\*∅ un niño que*  
 Maria does stay.INF in class DOM/\*∅ a boy that  
*no haya terminado los deberes.*  
 no has.SUBJ finished the homework.  
 'Maria makes a boy that has not finished the assignment stay in class.'  
 López (2012, 25)

Such phenomena, where specificity or scope considerations play no role with respect to the requirement for the use of DOM prompt López (2012) to observe that the mechanism comes as a consequence of the suitable environmental conditions in which the DO may find itself and that scrambling represents a prerequisite for these conditions. DOM is thus argued to be the morphological expression of a syntactic configuration (López 2012, 28).

Furthermore, the three contexts above seem to lead to the same conclusion as do the observations regarding DO behaviour with respect to specificity and scope dependencies: Spanish seems to draw a distinction between marked indefinite DOs on the one hand and unmarked indefinites and bare plurals on the other.

The syntactic and the semantic properties of these nominals seem to go hand in hand. Indeed, as will be seen in section 3: DOMed DOs are argued to function as KPs and to always leave their merge position, scrambling to a position outside the VP. By so doing, they check case against a functional projection ( $\alpha$ P in López' terms) and have access to a mode of semantic composition with their predicate which enables the specific/wide scope readings shown to be available for them.

Unmarked DOs and bare nominals, on the other hand, will be argued to stay within the VP and to incorporate into the predicate for case checking. As a consequence, they only get a non-specific/narrow scope interpretation.

## 2.2 Romanian indefinite DOs

Romanian has also been grouped with those languages allowing DOM and the same interpretive effects have been argued to arise as in the case of Spanish marked DOs (Dobrovie-Sorin 1994; Cornilescu 2000). The differential object marker in Romanian is *pe*, a derivative of the locative preposition *p(r)e* ('on'). Romanian DOM exhibits a further complication in that it is more often than not accompanied by Clitic Doubling (CD) and many linguists have argued that the two mechanisms actually pertain to the same phenomenon (Dobrovie-Sorin 1990; Gierling 1997; Tigău 2010; Chiriacescu/von Heusinger 2011a; 2011b).

In the following subsections we focus on the same semantic and syntactic aspects discussed above for Spanish and show that Romanian patterns with

Spanish up to a certain extent: while DOMed DOs behave on a par in the two languages, unmarked DOs exhibit different behaviours with respect to their (im) possibility to acquire specific and wide scope readings etc.

### 2.2.1 Specificity<sup>4</sup>

One of the most widely discussed topics in relation to the behaviour of Romanian marked and unmarked DOs has been their ability to acquire a specific reading. Both DOMed as well as CDed+DOMed indefinite DOs have been argued to have a non-ambiguous specific interpretation and Romanian linguists have different opinions as to which of the two mechanisms i.e., DOM or CD bears responsibility for this interpretation. Thus, Farkas (1987), Dobrovie-Sorin (1990, 388; 1994, 234), Cornilescu (2000, 103) claim that the DOM marked DO induces specificity, while Steriade (1980), Dobrovie-Sorin (1990, 377; 1994, 224), Gierling (1997, 72ss. a.o.) establish a correlation between specificity and clitic doubling. Tigău (2010), Chiriacescu/von Heusinger (2011a; 2011b) a.o. argue on the other hand that specificity is a joint effect of *pe* and CD.

Thus, in an example such as (21) below, Dobrovie-Sorin (1994) argues that *pe* disambiguates the interpretation of the indefinite DO towards a specific one: the indefinite in (21b) is claimed to only allow a non-specific interpretation, while the non-marked variant in (21a) is said to allow for both a non-specific and a specific reading:

- (21) a. *Caut o secretară.*  
 look.I for a secretary  
 ‘I am looking for a secretary.’

(Dobrovie-Sorin 1994, 224)

- b. *Caut pe o secretară.*  
 look.I for DOM a secretary  
 ‘I am looking for a secretary.’

(Dobrovie-Sorin 1994, 234)

Note that Romanian unmarked DOs may also give rise to epistemically specific readings, unlike their Spanish counterparts. Consider (22), where an unmarked DO is allowed even if the only available interpretation is a specific one whereby the referent denoted by the indefinite DP is known to the speaker.

<sup>4</sup> See section 2.1.1 for a comparison with the corresponding phenomenon in Spanish.

- (22) *Ieri am văzut pe/∅ un copil de-al meu în bibliotecă.*  
 yesterday have.I seen DOM/∅ a child of mine in library  
 ‘Yesterday I saw a child of mine in the library.’

Furthermore, Romanian allows both marked and unmarked DOs to co-occur with the equivalent of *a certain*, and to acquire a specific interpretation. In (23) both the DOMed and the unmarked variants of *un fotbalist* (‘a soccer player’) may be preceded by *anumit* (‘a certain’). As discussed in 2.1.1, Spanish examples containing unmarked DOs are infelicitous when these co-occur with *un cierto*:

- (23) *Ion caută pe/∅ un anumit fotbalist.*  
 John seeks DOM/∅ a certain soccer player  
 ‘John is looking for a certain soccer player.’

Just like their Spanish counterparts, DOMed indefinite DOs may combine with the free choice indefinite *oarecare* (‘any’), blocking their specific interpretation. The same holds for unmarked indefinites:

- (24) *Alege pe/∅ un coleg oarecare și roagă-l să te ajute.*  
 select DOM/∅ a colleague any and ask-him.ACC SUBJ you.ACC  
 help.he  
 ‘Pick up any colleague and ask him to help you.’

The mood of the predicate from within the relative clause modifying the indefinite DO has also been argued to be relevant with respect to specificity in the case of Romanian DOs. Thus, Farkas (1982, 109–130) and Dobrovie-Sorin (1994) argue that while the unmarked indefinite DO in (25a) may exhibit both a specific and a non-specific interpretation, the same DP in (25b) may only allow for a non-specific reading on account of being modified by a relative clause whose predicate is in the subjunctive mood. On the other hand, the use of the indicative mood in (25c) forces a specific interpretation on the DO.

- (25) a. *Maria caută o croitoreasă.*  
 Maria looks for a seamstress.  
 ‘Maria is looking for a seamstress.’

- b. *Maria caută o croitoreasă care să îi facă*  
 Maria looks for a seamstress who SUBJ she.CL.DAT make  
*rochia de mireasă.*  
 dress.the of bride  
 ‘Maria is looking for a seamstress who should make her bridal gown.’
- c. *Maria caută o croitoreasă care i-a făcut*  
 Maria looks for a seamstress who she.CL.DAT-has made  
*rochia de mireasă.*  
 dress.the of bride  
 ‘Maria is looking for a seamstress who should make her bridal gown.’

It thus seems that, just like with the Spanish data, the mood in the relative clause modifying an indefinite DO in Romanian has consequences on the specific/non-specific reading of this DP. However, while Spanish DOs clearly split into unmarked variants, which may only be modified by relatives whose predicates are in the subjunctive, and marked ones allowing both moods in the modifying relative, Romanian does not draw such a clear-cut distinction as both marked and unmarked DOs occur within both types of contexts.

This is shown in (26) where a relative clause with the predicate in the subjunctive mood may modify both the unmarked DO *câțiva seminariști* (‘some tutors’) in (26a) and its DOMed variant in (26b) on a non-specific interpretation; the CDed+DOMed variant is also possible as shown in (26c):

- (26) Context: *Examenul de lingvistică a fost foarte greu și majoritatea studenților au picat.*

‘The linguistics exam was very difficult and most of the students failed.’

- a. *Aceștia caută acum câțiva seminariști care să le*  
 these look for now some seminar tutors who SUBJ them.CL.DAT  
*explice din nou materia.*  
 explain again subject matter.the  
 ‘They are now looking for some seminar tutors who would explain the subject matter to them once more.’
- b. *Aceștia caută acum pe câțiva seminariști care să*  
 these look for now DOM some seminar tutors who SUBJ  
*le explice din nou materia.*  
 them.CL.DAT explain again subject matter.the  
 ‘They are now looking for some seminar tutors who would explain the subject matter to them once more.’



- c. *Aceștia îi caută acum pe câțiva seminariști*  
 these them.CL.ACC look for now DOM some seminar tutors  
*care să le explice din nou materia.*  
 who SUBJ them.DAT explain again subject matter.the  
 ‘They are now looking for some seminar tutors who would explain  
 the subject matter to them once more.’

Moreover, both marked and unmarked DOs may be modified by a relative clause in the indicative. The interpretation in this case is a specific one:

- (27) *Studentzii (îi) caută (pe) câțiva seminariști*  
 students.the (them.CL.ACC) search (DOM) some tutors  
*care le explică bine materia.*  
 who them.CL.DAT explain well subject matter.the  
 ‘The students are looking for some tutors who explain the subject matter  
 well to them.’

Also, Romanian DOMed indefinite DOs may co-occur with the free choice indefinite *oarecare* (‘any’) thereby losing their specific interpretation. The same happens with unmarked indefinites:

- (28) *Alege pe/Ø un coleg oarecare și roagă-l să*  
 select DOM/Ø a colleague any and ask-him.ACC SUBJ  
*te ajute.*  
 you.ACC help.he  
 ‘Pick up any colleague and ask him to help you.’

When drawing together the subjunctive test and expressions such as *oarecare* (‘any’) or *un anumit* (‘a certain’), the result is again somewhat different from Spanish with respect to the behaviour of unmarked indefinites: while DOMed DOs behave similarly to their DOMed counterparts in allowing co-occurrence with the indicative and *a certain*, contrary to their Spanish counterparts, Romanian unmarked indefinites allow co-occurrence with the indicative and *a certain* (30b). Both marked and unmarked indefinites are felicitous in the context of *oarecare* and the subjunctive (29a). The combination between the subjunctive with *a certain* seems to be problematic both for marked and unmarked indefinites, as expected (30a). Finally *oarecare* and the indicative seem problematic together for marked and unmarked DOs (29b).

- (29) a. *Caut pe/∅ un student oarecare care să*  
 look.I DOM/∅ a student any who SUBJ  
*vorbească bine englezește la clasă.*  
 speak well English in class  
 ‘I am looking for a (any) student who might speak English well.’
- b. ?*Caut pe/∅ un student oarecare care vorbește*  
 look.I DOM/∅ a student any who speaks  
*bine englezește la clasă.*  
 well English in class  
 ‘I am looking for a (any) student who speaks English well.’
- (30) a. ?*Caut pe/∅ un anumit student care să*  
 look.I DOM/∅ a certain student who SUBJ  
*vorbească bine englezește la clasă.*  
 speak well English in class  
 ‘I am looking for a certain student who might speak English well.’
- b. *Caut pe/∅ un anumit student care vorbește bine*  
 look.I DOM/∅ a certain student who speaks well  
*englezește la clasă.*  
 English in class  
 ‘I am looking for a (any) student who speaks English well in the class.’

By drawing on the data presented in this section, we may conclude the following: unlike their Spanish counterparts, Romanian unmarked indefinite DOs have access to a specific interpretation. As such, these DPs may be used in the context of *anumit* (‘a certain’) or allow modification by relative clauses in the indicative mood. Romanian DOMed DOs on the other hand seem to lead to the same pattern of behaviour as their Spanish correspondents, allowing both a specific and a non-specific interpretation.

### 2.2.2 Scope<sup>5</sup>

Similarly to the Spanish data, the relevant literature on Romanian DOs has claimed that DOMed variants favour a wide scope interpretation (Dobrovie-Sorin 1994; Tigău 2010). In this section we focus on the behaviour of marked and

<sup>5</sup> See section 2.1.2 for a comparison with the corresponding phenomenon in Spanish.

unmarked DOs in the contexts of extensional QPs and intensional operators. It will be shown that the same difference is at stake between Spanish and Romanian DOs as above: Spanish DOs differentiate themselves from their Romanian correspondents. Only marked DOs may acquire a wide scope reading in Spanish, while both marked and unmarked DOs may outscope other scope bearing expressions in the latter Romanian.

In (31) the DOMed indefinite *pe o femeie* may outscope the QP Subject, with the wide scope reading being actually preferred over the narrow scope one:

- (31) a. *Fiecare bărbat a iubit pe o femeie pe când era tânăr.*  
 every man loved DOM a woman when was.he young  
 ‘Every man loved a woman when he was young.’  
 $\exists > \forall$  and  $\forall > \exists$
- b. *Majoritatea bărbaților au iubit pe o femeie în tinerețe.*  
 majority.the men have loved DOM a woman in youth  
 ‘Most men loved a woman in their youth.’  
 $\exists > \text{Most}$  and  $\text{Most} > \exists$

Unlike Spanish correspondents, unmarked indefinite DOs allow both a narrow scope interpretation and a wide scope one:

- (32) a. *Fiecare bărbat a iubit o femeie pe când era tânăr.*  
 every man loved a woman when was.he young  
 ‘Every man loved a woman when he was young.’  
 $\exists > \forall$  and  $\forall > \exists$
- b. *Majoritatea bărbaților au iubit o femeie în tinerețe.*  
 majority.the men have loved a woman in youth  
 ‘Most men loved a woman in their youth.’  
 $\exists > \text{Most}$  and  $\text{Most} > \exists$

Marked DOs may outscope the negation operator. Unmarked DOs may also do so, however, contrary to their Spanish counterparts:

- (33) a. *Ion nu suportă pe o femeie în casa lui.*  
 John not stands DOM a woman in house his  
 ‘There is a woman such that Juan does not stand in his house.’  
 ‘Juan stands no woman in his house.’  
 $\exists > \neg$  and  $\neg > \exists$
- b. *Ion nu suportă o femeie în casa lui.*  
 John not stands a woman in house his  
 ‘There is a woman such that Juan does not stand in his house.’  
 ‘Juan stands no woman in his house.’  
 $\exists > \neg$  and  $\neg > \exists$

Furthermore, DOMed indefinite DOs may acquire wide scope with respect to the conditional operator. Unmarked Romanian indefinites seem to also be able to outscope the conditional, as opposed to their Spanish counterparts.

- (34) a. *Dacă Ion admiră o actriță, Maria îl va certa.*  
 if John admires an actress Mary him.CL.ACC will scold  
 ‘If John admires an actress, Mary will scold him.’  
 $\exists > \rightarrow$  and  $\rightarrow > \exists$
- b. *Dacă Ion (o) admiră pe o actriță, Maria îl va certa.*  
 if John (her.ACC) admires DOM an actress Mary  
 him.CL.ACC will scold  
 ‘If John admires an actress, Mary will scold him.’  
 $\exists > \rightarrow$  and  $\rightarrow > \exists$

As it seems, Romanian data pattern with the Spanish ones only with regard to the marked DO: these DPs may outscope the negation or the conditional operator, the wide scope interpretation being actually favoured. Unmarked DOs may also do so, however, contrary to their Spanish correspondents.

### 2.2.3 The context of *a avea* (‘to have’)<sup>6</sup>

The Romanian counterpart of the Spanish *tener, a avea* also allows for a stage level and an individual level reading, with the same restrictions as in Spanish: when

<sup>6</sup> See section 2.1.3 for a comparison with the corresponding phenomenon in Spanish.

functioning as an individual level predicate, *a avea* disallows the use of DOMed DOs (35a), while the stage level *a avea* allows for DOMed complements (35b):

- (35) a. Maria are (\*pe) o soră.  
 Mary has (\*DOM) a sister.  
 b. Am pe o soră (de-a lui Mihai la mine luna asta).  
 have.I DOM a sister (of Michael at me month this)  
 ‘I have one of Michael’s sisters living with me this month.’

Thus, similarly to its Spanish counterpart *tener*, the individual-level *a avea* may only select a property-denoting complement, not allowing a DOMed DP as its complement, since such DPs are never property denoting. The complement of the stage-level *a avea* may be marked by *pe* because the stage-level *a avea*, denotes a property of events (type  $\langle s, t \rangle$  and its complement is syntactically represented as a small clause containing a subject DP and a spatio-temporal predicate. As such, the subject DP may be of type *e* or  $\langle \langle e, t \rangle, t \rangle$ , which allows the use of *pe*.<sup>7</sup>

Thus, *pe* marked DPs may never be used with individual-level *a avea* as it only selects property-denoting complements, but they may be used as complements of the stage level *a avea* as *e* or  $\langle \langle e, t \rangle, t \rangle$  type DPs are allowed as subject of the small clause selected by this verb. In this respect we may view these data as an argument that DOMed nominals are not property denoting.

#### 2.2.4 Some syntactic phenomena which do not involve scope or specificity<sup>8</sup>

López (2012) includes Romanian among the languages that enforce DOM on small clause subjects, object control and causative-permissive structures, similarly to Spanish.

##### Small clause subjects

- (36) Ion (îl) consideră pe/\*∅ un student tâmpit.  
 John him.ACC considers DOM/\*∅ a student stupid  
 ‘John considers the student stupid.’

López (2012, 143)

<sup>7</sup> See Bleam (1999) for an extensive discussion on the semantic types allowed with the two versions of *have* and its correspondent in Spanish and Cornilescu (2000) for a similar discussion about Romanian.

<sup>8</sup> See section 2.1.4 for a comparison with the corresponding phenomenon in Spanish.

**Object control**

- (37) Ion (l)-a forțat pe/\*∅ un băiat să-i.  
 John him.CL.ACC-has forced DOM/\*∅ a boy SUBJ-him.CL.DAT  
 facă teme.  
 do homeworks.the  
 'John forced a boy to do his homework.'

López (2012, 144)

**Causative – permissive structure**

- (38) Ion l-a lăsat pe/\*∅ un copil să  
 John him.CL.ACC-has let DOM/\*∅ a child SUBJ  
 joace Nintendo.  
 play Nintendo  
 'John let a child play Nintendo.'

López (2012, 144)

The claim is too strong as native speakers of Romanian actually accept the unmarked variant in all the three contexts above.

**Small clause complements**

Romanian patterns with Spanish up to a certain extent when it comes to the obligatoriness of DOM in small clause complements: while it is true that preference is given to the DOMed nominals, unmarked DOs are also possible. Romanian only discards bare nominals from this context:<sup>9</sup>

---

<sup>9</sup> The use of unmarked indefinites within small clauses is broadly attested in the language. Here are some examples that came out as a result of a simple Google search:

- (1) a. *Eu consider unii posesori de R1 și R6 cazuri speciale.*  
 I consider some owners of R1 and R6 cases special  
 'I consider some owners of R1 and R6 to be special cases.'  
 (<http://www.motociclism.ro/forum/index.php?/topic/747673-motocicleta-moarta-de-batranete/page-3>)
- b. *Și eu consider multe fete de aici prietene.*  
 and I consider many girls from here friends  
 'I also consider many girls here as my friends.'  
 (<http://www.culinar.ro/forum/continut/pentru-cei-mici/26350/ingrediente-mai-mult-sau-mai-putin-nocive-ptr-copii/page-2>)

- (39) a. (*Îl*)                    *consider pe un student inteligent.*  
 (him.CL.ACC) consider.I DOM a student intelligent.  
 ‘I consider a student intelligent.’
- b. *Consider un student inteligent.*  
 consider.I a student intelligent
- c. \**Consider studenți inteligenți.*  
 consider.I students intelligent.  
 ‘I consider students intelligent.’

Note further that both marked and unmarked indefinites are allowed as arguments in small clauses when modified by a relative containing a predicate in the subjunctive, with a preference for the unmarked variant:

- (40) *Ion nu ar considera cinstit (pe) un politician care*  
 John not would consider honest (DOM) a politician who  
*să accepte mită de la alegători.*  
 SUBJ accept bribe from electors  
 ‘John would not consider to be honest a politician who would accept bribes from the electors.’

### Object control predicates

Just as in the case of small clause complements, Romanian differs from Spanish in that it allows both marked and unmarked DOs to surface as objects in control predicates. It is only bare nominals that are discarded from this context (42b):<sup>10</sup>

---

<sup>10</sup> Consider also the examples below found on the internet:

- (1) *Angajatorul a forțat un angajat să își dea demisia.*  
 Employer has forced an employee SUBJ REFL. resign  
 ‘The employer forced an employee to resign.’  
 ([https://www.avocatnet.ro/forum/discutie\\_593913/Poti-sa-demisionezi-si-fara-preaviz.html\\*1](https://www.avocatnet.ro/forum/discutie_593913/Poti-sa-demisionezi-si-fara-preaviz.html*1))
- (2) *a forțat mulți părinți să transfere altora îndatoririle parentale.*  
 has forced many parents SUBJ transfer others.DAT duties parental  
 ‘has forced many parents to transfer their parental duties to others.’  
 (<https://books.google.de/books?id=SF7nDQAAQBAJ&pg=PT104&lpg=PT104&dq=%22forțat+mulți%22&source=bl&ots=T39jvyO4V&sig=54nq0igZ7V8TXdjMhscXotpWpBg&hl=ro&sa=X&ved=0ahUKEwjcuprNz5bbAhVNY1AKHbpiDSAQ6AEIKjAB#v=onepage&q=%22forțat%20mulți%22&f=false>)

- (41) *Ion a forțat (pe) un băiat să-i facă*  
 John has forced (DOM) a boy SUBJ-him.CL.DAT do  
*temele.*  
 homework.the  
 ‘John forced a boy to do his homework.’
- (42) a. *Șeful a forțat (pe) mulți angajați să rămână*  
 boss has forced (DOM) many employees SUBJ remain  
*la serviciu peste program.*  
 at work overtime  
 ‘The boss forced many employees to do overtime.’
- b. ? *Șeful a forțat angajați să rămână la*  
 boss.the has forced employees SUBJ remain at  
*serviciu peste program.*<sup>11</sup>  
 work overtime  
 ‘The boss forced employees to do overtime.’

### Clause union

Just as for Spanish, Romanian DOM is argued to also be obligatory in these contexts. This is, nevertheless, inaccurate as unmarked indefinites are actually considered acceptable:

- (43) *Ion a lăsat un copil să joace Nintendo.*  
 John has let a child SUBJ play Nintendo  
 ‘John let a child play Nintendo.’

Note, however, that Romanian causative constructions actually amount to raising to object sentences and as such they are not relevant for the current discussion. In the causative-permissive structure, for instance, the use of DOMed and unmarked affectees is allowed, as well as that of bare plurals:

---

**11** This example might be acceptable under a reading according to which one explains the more recent actions of the boss in questions along the lines of (1):

- (1) So what’s new about your boss?  
 He forced employees to do overtime



- (44) a. *A lăsat (pe) un străin să-i intre în casă, fără să se gândească la consecințe.*  
 has let (DOM) a stranger SUBJ-him.CL.CAT enter in house without SUBJ REFL think of consequences  
 ‘He let a stranger enter his house without thinking about the consequences.’
- b. *A lăsat străini să-i intre în casă, fără să se gândească la consecințe.*  
 has let strangers SUBJ-him.CL.DAT enter in house without SUBJ REFL think of consequences  
 ‘He let strangers enter his house without thinking about the consequences.’

The three contexts discussed in this section point to the same differences holding between Spanish and Romanian DOs: Romanian patterns with Spanish in allowing DOMed indefinite DOs in these configurations (and disallowing bare plurals). Nevertheless, Romanian also allows unmarked indefinites to occur in small clauses, clause union and object control configurations as opposed to Spanish.<sup>12</sup>

## 2.3 Conclusions

This section has compared Spanish and Romanian DOs with respect to a number of contexts. The following conclusions have been reached:

Spanish and Romanian marked DOs pattern alike in allowing for a specific interpretation and a wide scope reading; the non-specific and the narrow scope interpretations remain an option with these DPs however.

The two languages differ with respect to the behaviour of unmarked DOs: while Spanish unmarked DOs always get interpreted non-specifically and may only exhibit a narrow scope reading when co-occurring with other scope bearing expressions, their Romanian counterparts pattern with their DOMed correspondents, allowing both specific as well as wide scope interpretations.

<sup>12</sup> Note again that only two of these contexts are relevant for Romanian: small clause complements and object control predicates.

A further point of similarity between Romanian and Spanish marked DOs has to do with the occurrence of these DPs within the context of *have*: while this verb only enables co-occurrence with unmarked DOs in its individual level reading, marked DOs are allowed with its stage level variant. These restrictions show, in turn, that marked DOs have real argumenthood and never denote properties.

Finally, Spanish DOM seems to be syntactically triggered in three contexts: small clauses, object control and clause union in causative configurations. Again, Romanian differs in this respect: not only DOMed indefinites are allowed in these contexts, unmarked indefinites are also acceptable. Bare plurals, similarly to their counterparts in Spanish, are rejected from these configurations. Thus, while Spanish draws a line between DOMed objects on the one hand and unmarked and bare nominals on the other, Romanian seems to group marked and unmarked objects together when it comes to the aforementioned contexts and to distinguish between these DPs on the one hand and bare plurals on the other. The table below summarizes the data discussed in this section:

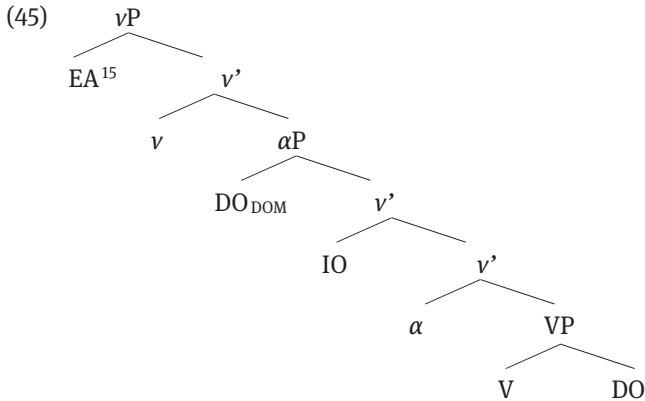
**Table 1:** Marked and unmarked DOs in Spanish and Romanian.

	Spanish		Romanian	
	DOMed DO	unmarked DO	DOMed DO	unmarked DO
Specificity	+/-	-	+/-	+/-
Scope	+/-	-	+/-	+/-
Small clause complements	+	-	+	+
Object control	+	-	+	+

### 3 Analyzing Spanish Data

#### 3.1 DOMed DOs move, unmarked DOs don't

Building on the Spanish data presented and discussed above, López (2012) argues that the interpretive and behavioural differences between Spanish DOMed and unmarked DOs boil down to a syntactic one: DOMed DOs undergo scrambling to a  $vP$  intermediary position, while unmarked DOs stay in-situ and incorporate with  $V$  (45):



López (2012) argues in favour of an indirect mapping between syntax and semantics in the sense that the syntactic position occupied by the DO affects the mode of semantic composition, which in turn affects the interpretation of the sentence. The different modes of semantic composition thus explain the lack of specific readings with unmarked DOs and the availability of a specific interpretation with marked DOs, given that the former get interpreted by means of Restrict (cf. Chung/Ladusaw 2004; 2006),<sup>14</sup> while the latter end up in a position where they get interpreted by means of Choice Functions, in line with Reinhart (1997).<sup>15,16</sup>

<sup>13</sup> EA stands for *external argument*.

<sup>14</sup> Chung/Ladusaw (2004; 2006) propose an operation, *Restrict*, which enables the combination of the DO with its predicate by way of conjunction. As a consequence, the predicate remains unsaturated. López (2012) argues that unmarked indefinite DOs get interpreted by means of this mechanism, which explains why they necessarily have a narrow scope reading.

<sup>15</sup> A choice function variable shifts a DP from a property denotation  $\langle e, t \rangle$  to an entity one  $\langle e \rangle$ . As a consequence, the DP may then be composed by Functional Application. Reinhart (1997) proposes that the choice function variable may be bound by an existential quantifier, which may be merged at different points within the derivational tree. This enables a suitable explanation for the different scopes that indefinites may give rise to e.g., the intermediate scope of the DO *a friend* in (1) below:

(1) Every woman is convinced that if John invites a friend of his to the party, it will be a disaster.  
 $\forall x \exists f CH(f)[\text{woman}(x) \rightarrow \text{convinced}(x, [\text{invite}(\text{John}, f(\text{friend of mine}), \text{party}) \rightarrow \text{disaster}(\text{party})])]$   
 ‘For every woman  $x$  there exists a (choice function that picks out a) friend of John’s such that if John invites the friend picked out by the choice function to the party,  $x$  is convinced it will be a disaster’

López (2012, 7)

<sup>16</sup> In line with Diesing (1992), López (2012) argues that there is a correlation between the syntactic position occupied by the indefinite DO and its ability of acquiring a specific reading. Unlike

According to López (2012), indefinite DOs which remain in their merge position may only be composed with their selecting predicate by means of *Restrict*. On the other hand, DOs which undergo (short) scrambling may only be interpreted by choice functions. The landing site of the (short-)scrambled DO is a position governed by small  $v$ , wherefrom the DP may not c-command the subject or other scope-taking operators within the clause. The fact that the DO in question may acquire a wide scope or a specific interpretation is a consequence of it having been interpreted by means of a choice function. Note, however, that the wide scope/specific reading does not always obtain as this interpretation is merely possible with marked DOs and may even be lost in appropriate contexts (e.g., the co-occurrence with *cualquiera*, or relatives containing a predicate in the subjunctive). This comes as a result of the indirect mapping between syntax and semantics, as argued in López (2012).

One argument in favour of the hypothesis that DOMed DOs move comes from binding dependencies with the IO and the subject DPs: marked DOs may bind the IO but not the Subject; unmarked DOs may bind neither the IO nor the subject: in (46) the DOMed DO *a ningún<sub>i</sub> prisionero* may bind the possessive within the IO *a su<sub>i</sub> hijo*. The same bound interpretation may obtain in (47). Unmarked DOs, on the other hand, stay in situ and will not be able to bind the IO. This is at stake in (48a) where the unmarked DO *un hombre* may not bind the anaphor within the IO *a sí mismo*. Expectedly, marking saves the derivation (48b):

- (46) [Context: What did the enemies do? The enemies delivered X to Y and Z to W, but....]  
*Los enemigos no entregaron a su<sub>i</sub> hijo a/\*∅*  
 The enemies not delivered.they to his son DOM/\*∅  
*ningún<sub>i</sub> prisionero.*  
 no prisoner  
 ‘The enemies did not deliver any prisoner to his son.’

López (2012, 41)

---

Diesing, however, López proposes an indirect mapping between configuration and semantic interpretation in that this mapping is mediated by the kind of operations which may apply in order to construct the compositional meaning of the verb and its complement. López thus pairs the syntactic positions occupied by the DO with different modes of semantic composition. See also Tigău (2010) for the relevance of the Mapping Hypothesis in Diesing (1992) with respect to the Romanian indefinite DOs.

- (47) *Los enemigos no entregaron a su<sub>i</sub> hijo a nadie<sub>i</sub>.*  
 The enemies not delivered.they to his son DOM no.one  
 ‘The enemies delivered no one to his son.’  
 López (2012, 41)

- (48) a. \**María le entregó a sí mismo un hombre.*  
 María him.CL.DAT delivered to himself a man  
 ‘Maria delivered a man to himself.’  
 b. *María le entregó a sí mismo a un*  
 María him.CL.DAT delivered to himself DOM a  
*hombre (cualquiera).*  
 man (no matter who)  
 ‘Maria delivered a man to himself.’  
 López (2012, 41)

On the other hand, a DOMed DO may not bind into the subject: this is so due to the fact that the position that the marked DO reaches by undergoing scrambling does not c-command the EA. Example (49) shows this at stake: a quantifier-variable interpretation is not possible in this case, given that the DOMed DO *ningún niño* may not bind the possessive within the subject.

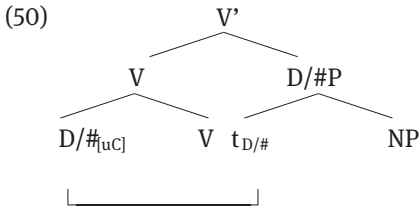
- (49) [Context: So, what happened yesterday?]  
*Ayer no atacó su<sub>i</sub> propio padre a ningún<sub>i</sub> niño*  
 yesterday not attacked his own father DOM no boy  
 ‘His own father attacked no boy yesterday.’  
 López (2012, 43)

The binding data thus suggest that while Spanish unmarked DOs remain inside VP and never leave their in-situ position as complements of the lexical verb, marked DOs move to a position wherefrom they may c-command IOs but in which they are c-commanded by the external argument.

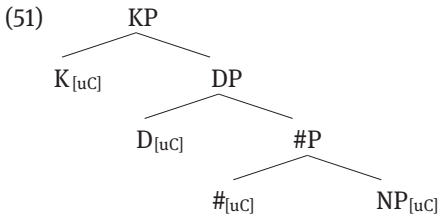
### 3.2 Movement and DP internal structure

As seen above, binding dependencies with the IO and Subject DPs uncover the different behaviour between DOMed DOs and their unmarked correspondents, supporting the hypothesis that the former move into a position wherefrom they may c-command the IO, while the latter stay in-situ. López (2012) argues that

movement is motivated by the internal structure of the respective DP and directly related to the case checking mechanism at stake: unmarked DOs do not project a full structure and as a consequence they check case in-situ by incorporation into the verb (50). DOMed DOs, on the other hand, project as KPs and their K layer blocks incorporation (51). Consequently, these phrases have to move out of VP so as to acquire case. Movement to Spec $\alpha$ P is thus triggered by the case feature [ $u$ C] of the marked DO. This allows  $v$  to probe the marked DO and to assign it [Acc]:



López (2012, 48)



López (2012, 49)

As already pointed out, the syntactic position that the DO occupies affects the mode of semantic composition, and this ultimately affects the interpretation of the sentence: DOs which remain in their merge position may only be composed with their selecting predicate by means of *Restrict* (cf. Chung/Ladusaw 2004; 2006). DOs which undergo (short) scrambling may only be interpreted by choice functions (Reinhart 1997). Short scrambling of the marked DO thus amounts to a precondition for the application of a choice function, which in turn accounts for the (optional) specific and wide scope readings with that DO. Unmarked DOs, which are of  $\langle e, t \rangle$  type never leave their in-situ position and compose with the verb via *Restrict*.<sup>17</sup>

<sup>17</sup> One reviewer draws attention to the existence of unmarked inanimate DOs which may receive a specific interpretation in Spanish. We suggest that such nominals pattern with Romanian unmarked DOs, moving out of the VP and acquiring the specific reading by way of the application of a choice function (see section 4 for an analysis of Romanian unmarked DO). It might be possible that the strict division between marked and unmarked DOs described for Spanish in López

## 4 A tentative analysis for Romanian objects

### 4.1 On the differences between Spanish and Romanian DOs

As seen in section 2, Romanian and Spanish marked DOs exhibit a similar behaviour, while unmarked DOs behave differently in the two languages. In this respect, Romanian unmarked DOs pattern with marked DOs: both DOMed and unmarked DOs may be interpreted as specific and may outscope other scope bearing expressions. Secondly, both marked and unmarked DOs are allowed as arguments in small clauses, and object control contexts.

Finally, both DOMed and unmarked DOs seem to be able to bind into the IO as may be seen in (52)–(54) below:

- (52) *Teroriștii au predat comandantului său<sub>i</sub> (pe) fiecare soldat<sub>i</sub> luat prizonier.*  
 Terrorists.the have handed in commander.DAT his DOM every soldier taken hostage  
 ‘The terrorists handed every soldier taken hostage over to their commander.’

- (53) a. *Teroriștii nu au predat comandantului său<sub>i</sub> (pe) nici un soldat luat prizonier<sub>i</sub>.*  
 Terrorists.the not have handed in commander.DAT his (DOM) no one soldier taken prisoner  
 ‘The terrorists did not hand any soldier taken prisoner over to their commander.’
- b. *Teroriștii nu au predat comandantului său<sub>i</sub> (pe) nimeni<sub>i</sub>.*  
 Terrorists.the not have handed in commander.DAT his (DOM) no one  
 ‘The terrorists handed nobody to his commander.’

---

(2012) is more flexible and patterns more closely with the Romanian data. As the reviewer points out, there is a large body of variational research showing that, despite having complementary core areas, Spanish DOM exhibits a blurry zone between them. We leave this matter for further research.

- (54) *Este vorba despre un roman SF în care, la un moment dat, personajul principal, Bill, ajunge să prezinte chiar ei înseși<sub>i</sub> (pe) o prietenă<sub>i</sub>, dar la o altă vârstă. Autorul vrea să arate astfel că există lumi paralele și că aceiași indivizi pot evolua diferit în contexte diferite.*

‘This is a Science Fiction novel where, at some point, the main character, Bill, ends up introducing a friend<sub>i</sub> to herself<sub>i</sub>, but at a different age. The author wants to show that parallel worlds exist and that the same individuals may evolve differently in different contexts.’

On the other hand, neither the DOMed DO nor its unmarked variant seem to be able to bind into the subject: a quantifier-variable reading is not available (55).<sup>18</sup>

- (55) \**Ieri, la ședința cu părinții, nu a criticat*  
 Yesterday at meeting.the with parents, no has.he criticized  
*tatăl său<sub>i</sub> (pe) nici un copil<sub>i</sub>.*  
 father his DOM no one child

Intended reading: ‘At the parents meeting yesterday his<sub>i</sub> father did not criticize any child<sub>i</sub>.’

Given that Romanian and Spanish pattern similarly when it comes to DOMed DOs, these DPs may be analyzed on a par with their Spanish correspondents along the lines of López (2012) extended upon above. We will thus take DOMed objects to have KP status and to leave their merge position within VP and scramble to Spec<sub>CP</sub>.

Given that Romanian unmarked DOs behave on a par with their marked counterparts with respect to all the relevant phenomena (specificity, wide scope, binding, small clauses, object control), we may posit that these DOs also undergo the same checking mechanism by way of scrambling to Spec<sub>CP</sub> as marked DO do. Since unmarked DOs are DPs, we may posit the following difference between

---

<sup>18</sup> Dobrovie-Sorin and Cornilescu (2008) draw attention to the fact that when the DOMed DO has also been clitic doubled, it is possible for this DP to bind into the subject. Thus the CD and DOMed DO *pe nici un copil* (‘any child’) is said to bind the possessive within the dative DP:

- (1) *Ieri, la ședința cu părinții, nu l-a criticat tatăl său<sub>i</sub>*  
 Yesterday at meeting.the with parents, no him.CL.ACC-has.he criticized father his  
*pe nici un copil<sub>i</sub>.*  
 DOM no one child

Intended reading: ‘At the parents meeting yesterday his<sub>i</sub> father did not criticize any child<sub>i</sub>.’



Romanian and Spanish: Romanian allows both KPs and DPs to scramble, while Spanish only allows scrambling of KPs.

This proposal becomes problematic, however, if we consider some contexts where unmarked DOs seem to pattern with their Spanish correspondents: they never acquire a specific or wide scope readings and they seem to exhibit property status. Section 4.1.1 dwells on this.

#### 4.1.1 Romanian strong and weak DPs

As already pointed out above, not all unmarked indefinites exhibit specific or wide scope readings: in (56) the unmarked indefinite *un șoc* ('a shock') may only allow a narrow scope reading with respect to the QP subject.

- (56) *Orice tânăr care nu învață de acasă să își*  
 Any young man who not learns from home SUBJ REFL.DAT  
*poarte singur de grijă, va suferi un șoc atunci când*  
 carry alone of care will suffer a shock then when  
*va pleca la facultate într-alt oraș.*  
 will.he go at faculty to another city.  
 'Any young man who does not learn from home how to take care of himself,  
 will suffer a shock when he will have to go to the university in a different  
 city.'

Similarly, the unmarked indefinite *o rujeolă* ('measles') may not outscope the QP subject or the conditional in (57).

- (57) *Dacă nu se iau măsuri, toți copiii din școală vor*  
 If not REFL take measures all children from school will  
*face o rujeolă de toată frumusețea.*  
 make a measles of all beauty  
 'If measures are not taken, all the children in school will catch really  
 serious measles.'

Consider also the following contexts, which require the use of nominals with property reading and disallow nominals denoting entities. As can be verified, unmarked indefinites and bare plurals are allowed in these contexts. Marked indefinites are out (for an extensive discussion see Cornilescu 2000, 31ff).

Consider first the case of post-verbal subjects with *se* (reflexive) passive. As shown by Cornilescu (2000), only nominals denoting properties are allowed in

this context. As a consequence, bare plurals are allowed in this context (58b), while proper names, given their (only) <e> type reading are discarded as ungrammatical (58a):

- (58) a. \**S-a adus Maria de la școală.*  
REFL.-has brought Mary from school  
'Mary was brought from school.'
- b. *S-au băgat banane la aprozar.*  
REFL.-have.they brought bananas at store  
'They have supplied the store with bananas.'
- c. *S-a adus de curând o girafă la grădina botanică.*  
REFL.-has brought of soon a giraffe at garden zoological  
'The zoo was recently endowed with a giraffe.'

As also noticed above, unmarked indefinites may function as the DO of *have* in its individual-object reading, a situation in which bare singulars and plurals are also allowed, while marked DOs are discarded. The observation was first made by Cornilescu (2000) building on Bleam (1999):

- (59) a. *Maria are copii deștepți.*  
Mary has children intelligent  
'Mary has intelligent children.'
- b. *Maria are un copil deștept.*  
Mary has a child intelligent  
'Mary has an intelligent child.'
- c. \**Maria are pe un copil deștept.*<sup>19</sup>  
Mary has DOM a child intelligent

Marked indefinites along with proper names or personal pronouns are rejected from these contexts where an individual level *have* occurs. This is because these nominals only have the object-level, <e> type reading. The contribution of K is to secure argumenthood.

---

**19** This example may become acceptable under a reading along the lines of (1) below. Note that the meaning in this case is, however, significantly different:

- (1) *Maria are pe un copil deștept care îi va rezolva problema în câteva minute.*  
Mary has pe a child smart who her.DAT will solve problem.the in few minutes  
'Mary knows a smart child who will solve the problem in a few minutes.'

Together with the previous sections, the examples above show that unmarked DOs have a twofold behaviour: they may pattern with marked DOs in allowing for a specific and a wide scope interpretation, but they may also cause a similar behaviour to bare nouns being able to occur in those contexts which disallow marked DOs and to have a property-denotation. The idea that we would like to put forth is that in these latter contexts unmarked DOs may be reanalyzed as NP/NumP (have the same status as bare nouns) and have a property reading. The following section offers some insights into the internal structure of bare nominals and draws a parallelism between these and unmarked DOs when used a property denoting expressions (see Cornilescu and Tigău ms. for a more extensive discussion).

## 4.1.2 Some insights from bare nominals and weak definites

### 4.1.2.1 Bare nominals

Tănase-Dogaru (2009) analyzes bare objects i.e., determinerless nominals (bare singulars and plurals):

- (60) a. *Maria vrea bărbat bogat.*  
 Mary wants husband rich  
 ‘Mary wants a rich husband.’  
 b. *Maria n-are copii.*  
 Mary not-have children  
 ‘Mary does not have children.’

As it seems, these nominals never allow for specific readings and may not outscope extensional QPs or intensional operators. In fact, as pointed out by Carlson (1977), these nominals have narrowest scope. Indeed, the bare plural *colegi* in (61a) may only exhibit a dependent reading on the QP subject according to which the colleagues who get invited to dinner vary according to the employee. A wide scope reading interpretation according to which there are some colleagues such that any employee would invite them to dinner is completely out. Along the same lines, the bare plural in (61b) cannot outscope the conditional as no particular group of colleagues is targeted such that if they get invited, John gets upset:

- (61) a. *Orice angajat mai invită colegi la masă*  
 Any employee more invites colleagues at dinner  
*din când în când, chiar dacă soția se supără.*  
 sometimes, even if wife.the REFL annoys  
 ‘Any employee invites colleagues to dinner from time to time, even though this might upset the wife.’
- b. *Dacă Maria invită colegi la masă, Ion se supără.*  
 if Mary invites colleagues at dinner, John REFL annoys  
 ‘If Mary invites colleagues to dinner, John gets upset.’

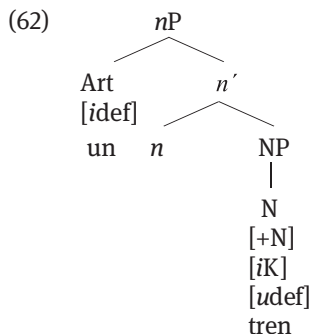
In (61) above, the constituent, ‘inviting colleagues’ designates a particular type of complex activity, where the bare nominal acts as a restrictive modifier on the predicate.

Tănase-Dogaru (2009) shows that bare nominals are ‘small size nominals’ projecting a structure smaller than DP i.e., NP or at most NumPs. As such they can only combine with the verb by way of incorporation – at least when they merge in complement position. One further argument supporting this hypothesis is that bare singulars may only merge as VP complements. Similar observations have been drawn with respect to Spanish bare plurals, which also seem to obey the restriction of only surfacing in post-verbal position (Brugè/Brugger 1996).

The contexts presented in 4.1.1 above show that Romanian unmarked indefinites may pattern with bare nominals. In this particular case, we propose that unmarked DOs are reanalyzed as NP/NumP and remain in-situ and incorporate into the verb (just like bare nominals in Tănase-Dogaru 2009). As a consequence, they get interpreted via Restrict and never acquire specific, or wide scope readings. The examples in the following section building on weak definites further strengthen this hypothesis.

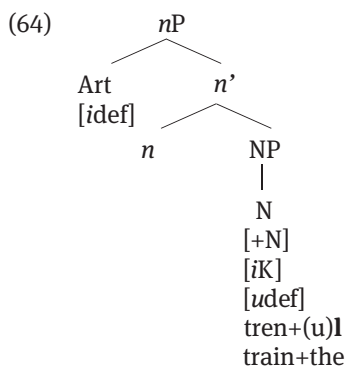
#### 4.1.2.2 Weak definites

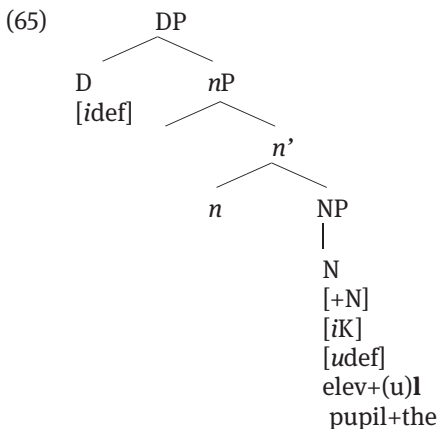
When used in their property reading, unmarked DOs also exhibit similar characteristics to definite descriptions, which have been shown to be able to function as property denoting nominals. Cornilescu and Nicolae (2012 and subsequent publications) argue that in this particular case these nominals have a weak reading and get projected as NPs: more specifically, the D head is not projected as such and the weak indefinite article functions as a quantifier or as an adjective, occupying the Specifier position of the nominal head:



This is in line with the proposals already advanced for definite descriptions, which may also exhibit property readings and which have been analyzed as NPs in this case (Cornilescu/Nicolae 2012). Consider first the examples under (63), where the definite nominal exhibits a property reading. In this case the definite expression seems to be a NP/NumP, where the definite article functions as the specifier (64) along the lines of Emonds (1985) who argues that determiners are in fact former adjectives and as such function as specifiers of nominal heads. Cornilescu/Nicolae (2012) support this hypothesis of the reanalysis of the DP arguing that reanalysis is only necessary for property readings of definites. Strong determiners will follow the lines of analysis in (65) since they are referential:

- (63) *a lua trenul/ a bate drumurile*  
 to take train.the/ to beat roads.the  
 ‘to take the train/ to wander through the roads’





On a par with definite descriptions in their weak readings and bare plurals, unmarked DOs which do not scramble will be analyzed as NPs.

## 4.2 A parametrization

As discussed, marked indefinites are KPs and only have the object-level, <e> type reading. The contribution of K is to secure argumenthood. As a consequence, nominals thus marked may only be interpreted as arguments (never as properties) and need to scramble to a position where they may combine with the verb by means of functional application. In this respect, Romanian marked indefinites pattern with marked DOs in Spanish. Bleam (1999; 2005) shows that DOMed nominals denote individuals <e> or generalized quantifiers <<e,t>> and argues that the DOM marker acts as a filter on the DP denotation, filtering away the property reading.

On the other hand, unmarked indefinites **may** function as DPs and as such may give rise to the same readings as marked indefinites (specificity, wide scope, binding dependencies) and undergo scrambling to Spec $\alpha$ P for case reasons. The fact that they reach this position enables them to get interpreted via Choice Functions, on a par with their marked counterparts. Nevertheless, Romanian unmarked DPs may also undergo reanalysis, receiving a NP status and stay in-situ just like unmarked DOs in Spanish.

A suitable analysis of Romanian data should thus account for the contrasts between Romanian and Spanish unmarked indefinites: the Spanish *a*-stripped nominals may only be property denoting combining with the verb by means of incorporation and never leaving their merge position. Romanian unmarked

indefinite DOs exhibit a twofold behaviour: as DPs they may have argumental status and pattern on a par with marked DOs, or they may undergo reanalysis to NP and behave similarly to other property denoting nominals (e.g., bare plurals). In this latter case, they closely follow in the footsteps of Spanish unmarked DOs. The basic tenets of DO analysis in the two languages may be schematically represented as in table (2). As it seems, while DPs in Spanish incorporate (López 2012), Romanian only allows for the incorporation of nominals with an even smaller internal structure i.e., NP/NumP;<sup>20</sup> nominals with DP status in Romanian do not incorporate:

**Table 2:** DP status in Spanish and Romanian.

Language	DOMed DOs	non-DOMed indefinite DOs		bare plurals
<b>Spanish</b>	argument status	non-argument status		non-argument status
	<e> or <<e,t>t>	<e,t>		<e,t>
	KPs	DPs		NP/NumPs
<b>Romanian</b>	argument status	argument status	non-argument status	non-argument status
	<e> or <<e,t>t>	<e> or <<e,t>t>	<e,t>	<e,t>
	KPs	DPs	NP	NP/NumP

The parameter that differentiates between Spanish and Romanian thus has to do with the types of nominals that must undergo scrambling: while scrambling is restricted to KPs in Spanish, Romanian allows both KPs and DPs to scramble. The parametrization we arrived at:

- a. Spanish only allows KPs to scramble. DPs stay in-situ along with nominals exhibiting a smaller structure e.g., NPs.
- b. Romanian allows the scrambling of both KPs and DPs. Nominals with a smaller structure do not scramble.

Also, while Spanish DPs incorporate, Romanian allows incorporation only with nominal expressions whose internal structure is smaller than DP i.e., NP/NumP (66).

(66) KP > DP > NumP > NP

As argued by Longobardi (1994 and subsequent publications), the locus of argumenthood is the determiner, which is sufficient for reference. Consequently,

<sup>20</sup> One reviewer points out that weak definites may not have a NumP status given that they are number neutral (Poesio 1994; Carlson/Sussman 2005; Carlson et al. 2006 a.o.).

nominal expressions with supplementary marking of referential status also count as arguments and, as such, must undergo scrambling and compose with the predicate by means of Function Application.

## 5 Conclusions

This paper has discussed various aspects regarding the syntax and semantics of Romanian DOs from a comparative perspective with their Spanish correspondents. It has been shown that DOMed DOs behave on a par in the two languages, allowing a specific interpretation, wide scope readings, binding into the IO etc. In terms of their internal structure these nominals are KPs and scramble from their merge position in order to value their case feature. This movement also allows their interpretation via choice functions, which in turn accounts for their possible specific/wide scope interpretation.

Unmarked DOs, on the other hand, constitute a locus of difference between the two languages: Romanian unmarked DOs exhibit the same distribution as DOMed DOs. In contrast, Spanish unmarked DOs, which have DP status, never scramble from their merge position and never acquire wide scope/specific readings.

The similarity between the marked DOs and the differences regarding their unmarked correspondents in the two languages led to a parametrization of the syntactic account put forth in this article: Romanian DOs merge as c-selected complements of V and need to check Accusative case. On a par with Spanish, Romanian resorts to two strategies of case valuation:

- a. DOs which project as NPs check case *in situ* by means of (semantic) incorporation. This amounts to having their [*uCase*] feature valued when the lexical verb moves to *v*. Incorporating objects are interpreted as predicates/ properties (expressions of type  $\langle e, t \rangle$ ) and form complex predicates with V.
- b. DOs projecting as DPs or KPs in Romanian may not incorporate and have to move to a higher position between *v* and V, where they verify their [*uCase*] feature by way of agreement with *v*. The fact that these nominals undergo scrambling finds support in the availability of strong readings i.e., wide scope, specificity a.o. Scrambling DOs are interpreted as arguments and combine with the verb by Function Application.

As seen, marked and unmarked DOs exhibit the same distribution in Romanian. Nevertheless, marked DOs *necessarily* scramble, while this remains an option for unmarked objects, which *may* scramble. This difference follows from the fact



that, while DPs may be re-analyzed as NP/NumPs, KPs may not. Scrambled DOs, whether marked or unmarked are semantically equivalent, in that they have the same scope properties, their reference may be *stable* (“epistemically specific”) or *unstable* (non-specific).

## Bibliography

- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21 (2003), 435–483.
- Bleam, Tonia, *Leísta Spanish and the syntax of clitic doubling*, PhD Dissertation, University of Delaware, 1999.
- Bleam, Tonia, *The role of semantic type in differential object marking*, in: Voegeleer, Svetlana (ed.), *Bare plurals, indefinites, and weak–strong distinction*, *Belgian Journal of Linguistics* 19 (2005), 3–27.
- Brugè, Laura/Brugger, Gerhard, *On the accusative “a” in Spanish*, *Probus* 8 (1996), 1–51.
- Carlson, Gregory, *Reference to kinds in English*, New York, Garland, 1980.
- Carlson, Greg/Sussman, Rachel, *Seemingly indefinite definites*, in: Kepsar, Stephan/Reis, Marga (edd.), *Linguistic evidence. Empirical, theoretical, and computational perspectives*, Berlin/New York, Mouton de Gruyter, 2005, 71–86.
- Carlson, Greg/Sussman, Rachel/Klein, Natalie/Tanenhaus, Michael, *Weak definite noun phrases*, in: Davis, Christopher/Deal, Amy Rose/Zabbal Youri (edd.), *Proceedings of NELS 36*, Amherst, Mass., GLSA, 2006, 179–196.
- Chiriacescu, Sofiana/von Heusinger, Klaus, *The discourse structuring potential of two types of definite noun phrases in Romanian*, in: van Deemter, Kees/Gatt, Albert/van Gompel, Roger P. G./Krahmer, Emiel (edd.), *Proceedings of the Pre-Conference of the Cognitive Science Society*, 2011, 1–7 (=2011a).
- Chiriacescu, Sofiana/von Heusinger, Klaus, *The discourse structuring potential of definite noun phrases in natural discourse*, in: Branco, António H./Hendrickx, Iris/Mitkov, Ruslan/Sobha, Lalitha Devi (edd.), *Proceedings of the 8th Discourse Anaphora and Anaphor Resolution Colloquium*, Lisbon, Edições Colibri, 2011, 151–162 (=2011b).
- Chung, Sandra/Ladusaw, William, *Chamorro evidence for compositional asymmetry*, *Natural Language Semantics* 14 (2006), 325–357.
- Chung, Sandra/Ladusaw, William, *Restriction and saturation*, Cambridge, Mass., MIT Press, 2004.
- Cornilescu, Alexandra, *Notes on the interpretation of the prepositional accusative in Romanian*, *Bucharest Working Papers in Linguistics* 2:1 (2000), 91–106.
- Cornilescu, Alexandra/Dobrovie-Sorin, Carmen, *Clitic doubling, complex heads, and interarboreal operations*, in: Kallulli, Dalina/Tasmowski, Tasmowski (edd.), *Clitic doubling in the Balkan languages*, Amsterdam, John Benjamins, 2008, 289–319.
- Cornilescu, Alexandra/Nicolae, Alexandru, *Nominal ellipsis as definiteness and anaphoricity. The case of Romanian*, *Lingua* 122:10 (2012), 1070–1111.
- Cornilescu, Alexandra/Tigău, Alina, *Romanian direct objects*, manuscript, University of Bucharest.
- Diesing, Molly, *Indefinites*, Cambridge, Mass., MIT Press, 1992.

- Dobrovie-Sorin, Carmen, *Clitic doubling, wh-movement, and quantification in Romanian*, *Linguistic Inquiry* 21 (1990), 351–397.
- Dobrovie-Sorin, Carmen, *The syntax of Romanian*, Berlin/New York, Mouton de Gruyter, 1994.
- Emonds, Joseph, *A unified theory of syntactic categories*, Dordrecht, Foris, 1985.
- Farkas, Donka, *Intensionality and Romance subjunctive relatives*, PhD Dissertation, Pennsylvania State University, 1982.
- Farkas, Donka, *Direct and indirect object reduplication in Romanian*, *Papers of the Chicago Linguistics Society* 14 (1987), 88–97.
- Farkas, Donka, *Specificity and scope*, in: Nash, Léa/Tsoulas, Georges (edd.), *Langues et grammaires* 1 (1994), 119–137.
- Gierling, Diana, *Clitic doubling, specificity and focus in Romanian*, in: Black, James R./Motapanyane, Virginia (edd.), *Clitic, pronouns and movement*, Amsterdam, John Benjamins, 1997, 63–85.
- Kornfilt, Jaklin/von Heusinger, Klaus, *Specificity and partitivity in some Altaic languages*, in: Vermeulen, Reiko (ed.), *Proceedings of the 5<sup>th</sup> Workshop on Formal Altaic Linguistics*, Cambridge, Mass., MIT Press, 2008, 19–40.
- Longobardi, Giuseppe, *Reference and proper names. A theory of N-movement in syntax and Logical Form*, *Linguistic Inquiry* 25 (1994), 609–665.
- López, Luis, *Indefinite objects. Scrambling, choice functions, and differential marking*, Cambridge, Mass., MIT Press, 2012.
- Poesio, Massimo, *Weak Definites*, in: Harvey, Mandy/Santelmann, Lynn (edd.), *Proceedings of SALT IV*, Ithaca, New York, Cornell University, 1994, 282–292.
- Reinhart, Tanya, *Quantifier scope. How labor is divided between QR and choice functions*, *Linguistics and Philosophy* 20 (1997), 335–397.
- Rivero, Maria Luisa, *Mood and presupposition in Spanish*, *Foundations of Language* 7 (1979), 305–336.
- Steriade, Donca, *Clitic doubling in the Romanian wh-constructions and the analysis of topicalization*, *Papers of the Chicago Linguistic Society* 16 (1980), 282–297.
- Tănase-Dogaru, Mihaela, *The category of Number*, Bucharest, Editura Universității din București, 2009.
- Tigău, Alina, *Syntax and interpretation of the direct object in Romance and Germanic languages with an emphasis on Romanian, German, Dutch and English*, PhD Thesis, University of Bucharest, 2010.
- von Heusinger, Klaus, *Specificity*, in: von Heusinger, Klaus/Maienborn, Claudia/Portner, Paul (edd.), *Semantics. An international handbook of natural language meaning*, Berlin/Boston, Mouton De Gruyter, 2011, 1024–1057.

Niklas Wiskandt

# Scale-based object marking in Spanish and Portuguese

*Leísmo*, null objects and DOM

**Abstract:** The prepositional marking of direct objects in Spanish is a well-known instance of Differential Object Marking (DOM), driven mainly by animacy and specificity. But it is not the only domain where Ibero-Romance object marking varies due to semantic factors: Spanish and Portuguese feature several patterns of heterogeneous object marking. This paper provides a comparative overview of prepositional DOM, also known as *a*-marking, *leísmo* and null objects. All three phenomena are found in both languages, but with considerable differences among their varieties in Europe and Latin America. It is argued that where they are present in a certain variety, they are always based on the same scale that ranks all potential direct objects with respect to specificity, animacy, (pro-)nominality and gender. This Ibero-Romance object marking scale is their most central common denominator, which is why I propose “scale-based object marking” as an umbrella term.

**Keywords:** Differential Object Marking (DOM), paradigmatic DOM, null object, semantic scale, Spanish, Portuguese

## 1 Introduction

It is widely accepted that Spanish features Differential Object Marking by means of the preposition *a*, which marks some, but not all, direct objects. The notion of Differential Object Marking was introduced by Bossong (1985), who first described it in Iranian languages and later applied it to Spanish (1991). In (1) we see an example of a direct object marked by *a*.

- (1) *María ve a la amiga.*  
María see.3SG OBJ DEF.F.SG friend  
'María sees the friend.'

(Spanish)

---

Niklas Wiskandt, Heinrich Heine University Düsseldorf, e-mail: wiskandt@phil.hhu.de

As part of the morphosyntactic development from Latin to Romance languages, morphological case marking of nouns has been lost. This gave rise to object marking by prepositions, as in Spanish. In Portuguese, the language most closely related to Spanish, *a*-marking of direct objects is attested, yet is much less frequent. But in fact, prepositional marking is not the only field where different types of direct objects are marked in different ways in Spanish and Portuguese. There are various types of heterogeneous marking of the direct object in both languages, and all of these show some important common properties. For example, the use of dative instead of accusative pronouns for some direct objects in European Spanish, exemplified in (2), has been described under the term *leísmo*.

- (2) *El criado no trabaja bien. Le/?Lo voy a reemplazar.*  
 3SG.DAT/ACC AUX.FUT.1SG FUT replace  
 ‘The servant does not work well. I will replace him.’  
 (European Spanish)

Its relation to other object marking phenomena is still under debate (Flores/Melis 2007; Fernández-Ordóñez 2012, among others). The use of so-called *null objects* plays a crucial role in research on Brazilian Portuguese (Cyrino 1994; 2017b; Schwenter/Silva 2002; 2003 Cyrino/Lopes 2016, among others). Such research understands a null object to be defined as follows: The direct object of a transitive sentence is not expressed overtly, even though it is required by the semantics of the verb; however, the sentence is grammatical and the object is inferred from the context. (3) contains an example of a null object.

- (3) *Comprei um pãozinho e comi Ø a caminho.*  
 buy.PST.PF.1SG INDEF.M.SG bun and eat.PST.PF.1SG on way  
 ‘I bought a bun and ate it on the way.’  
 (Portuguese)

The goal of this paper is to provide an overview of the different patterns of direct object marking found in Spanish and Portuguese, and to define the common denominator of prepositional DOM, *leísmo* and null objects, in order to establish a uniform analysis that includes all three phenomena. It is based on a synchronic analysis; a few remarks on the diachrony of the phenomena will be necessary, however.

The current discussion of the widespread Ibero-Romance languages Spanish and Portuguese will include varieties spoken in Europe and in Latin America.<sup>1</sup> My study cannot, however, address all existing varieties, an impossible task within the limits of a paper like this one. The examples discussed come from various sources. Besides examples cited from previous works, I illustrate my argumentation with new data that have been checked through the grammaticality judgments of native speakers. These speakers were presented with the example sentences and distracting items in written form and were asked to judge them spontaneously as perfect, acceptable, questionable, or impossible. Where speakers did not judge an example to be impossible, they were asked to translate it. However, we cannot guarantee that such judgments would be rendered analogously by all speakers of the respective language or variety.

The paper is structured as follows: After this introductory section, I discuss prepositional object marking in Section 2. Section 3 is concerned with the heterogeneous case marking of direct object pronouns, known as *leísmo* in European Spanish. In Section 4 I explore the contrast between overt and null objects. Section 5 defines a common denominator among the different phenomena of heterogeneous object marking. Some concluding remarks and suggestions for further research are given in Section 6.

## 2 Prepositional object marking in Spanish and beyond

### 2.1 Finding and defining Differential Object Marking

Differential Object Marking was identified in Spanish long ago. In works like Comrie (1979; 1981) the phenomenon itself was described even before Bossong (1985) first coined the now widely adopted term. Since then, a large number of studies have been conducted on Spanish DOM, classic among them being Bossong (1991); more recent studies include Fábregas (2013), García García (2014; 2018), Kabatek (2016), López (2012), Melis (2018) and von Heusinger (2018), to name but a few.

---

<sup>1</sup> I will differentiate mainly between European Spanish (ES), American Spanish (AS), European Portuguese (EP) and Brazilian Portuguese (BP). Whenever one of the languages is mentioned without any specification of the variety, the statement or example is considered valid for all varieties.

When dealing with phenomena of Differential Object Marking, a clear definition of the notion is essential. Adopting the findings of Bossong (1985; 1991; 2008) and Aissen (2003), in the current study I define Differential Object Marking (DOM) as being present in a language if all, or at least a subset of, transitive verbs show the following peculiarity in the marking of their arguments: Some, but not all, direct objects are marked by an overt morphological device, that is, there is a split between marked and unmarked direct objects. As stated by Bossong (2008, 40), the morphological marking of the direct object does not necessarily occur in the form of a case affix; a preposition is also a possible object marker, as well as a marker on the verb or any other morpheme that has the function of marking the grammatical relation of the direct object.

My current object of research here is DOM in a narrow sense, with the following characteristics: it occurs or not within the same syntactic and semantic environment, that is, it is not caused only by word order or by lexical peculiarities of the predicate; and it can be traced back to properties of the respective object referents.

For DOM mechanisms that depend on properties of the object itself, it has been observed that they operate on grammatical scales: All potential direct objects are arranged on a scale that is based on animacy, definiteness/specificity or person hierarchy, or on a combination of such hierarchies. Direct objects that are located above a language-specific cut-off point on that scale bear overt morphological marking; direct objects below that point do not. The scale is implicational in the sense that, if a particular object is overtly marked, all objects that are ranked above it on the scale will also be overtly marked, and vice versa. Typical distinctions relevant for DOM comprise human vs. non-human, animate vs. inanimate, definite vs. indefinite, specific vs. unspecific, 1<sup>st</sup>/2<sup>nd</sup> person vs. 3<sup>rd</sup> person and pronoun vs. noun. Aissen (2003) integrates most of these distinctions in two generalized scales, the animacy scale and the definiteness scale, cited here (4).

(4a) Animacy scale: human > animate > inanimate

(4b) Definiteness scale: personal pronoun > proper name > definite NP > indefinite specific NP > non-specific NP

(Aissen 2003, 437)

The orientation on grammatical scales as described above should not be considered to be a part of the definition of DOM but is a frequently observed characteristic of differential marking phenomena. Assumptions about grammatical scales

has had a significant influence on the description and analysis of Differential Object Marking in Spanish, which will be discussed in the next Subsection.

## 2.2 Prepositional object marking in Spanish

In Spanish, animacy and specificity have been shown to be the dimensions that are most relevant for the distribution of *a*-marking (Comrie 1979; Bossong 1991, among others). Fábregas (2013) gives a detailed report on the state of the art in investigating Differential Object Marking in Spanish and concludes by naming some fields that require further research. One of those is the “potential correlation between *a*-marking and other phenomena where animacy is relevant” (2013, 75), which is what I will be concerned with here. A recent reflection on the historical and current development of Spanish prepositional object marking is Kabatek (2016). He describes the system of prepositional object marking as transparent and predictable, and differentiates between cases where the marking is obligatory, facultative or impossible. For the sake of concision, this paper will largely disregard the facultative cases, although we might note that these merit more extensive study. Concerning the diachrony of Spanish object marking, Kabatek notes that the frequency of objects marked by *a* is increasing. This is an interesting tendency. It deserves further attention in the future, and the synchronic description in the present study will hopefully be of use in any such work.

The great majority of studies on Spanish DOM exclusively address prepositional object marking, or at least they focus mainly on it. In most varieties of Spanish, the preposition *a* marks all indirect objects, except those in the form of pronominal clitics, and some, but not all, direct objects. Since it does not express any distinction between direct and indirect objects, *a* will be glossed by OBJ in this paper. Differential marking of direct objects by the same morpheme that marks indirect objects is not at all unusual in the world’s languages. In Hindi (< Indo-Aryan < Indo-European), for instance, the postposition *ko* marks the indirect object and is thus identified as dative, but also figures as a differential marker for certain direct objects. For a recent study of DOM in Hindi, cf. Montaut (2018).

I will now offer a sketch of DOM, as defined above, in Spanish, focusing on the influence of the two most relevant dimensions, animacy and specificity. Animacy is probably the most obvious factor that influences Spanish DOM and has been shown to be relevant here in a number of works, among them Comrie (1979), Bossong (1991) and Fábregas (2013). For the sake of clarity, the examples used to illustrate the relevance of animacy all feature definite direct objects. The inanimate object in (5a) does not receive any marking, while the animate object in (5b), repeated from (1), is preceded by *a*:

(5a) *María ve la mesa.*  
 María see.3SG DEF.F.SG table  
 ‘María sees the table.’

(5b) *María ve a la amiga.*  
 María see.3SG OBJ DEF.F.SG friend  
 ‘María sees the friend.’

(Spanish)

García García (2018) argues that *a*-marking is obligatory for specific human objects but is optional for animals. He illustrates this claim with example (6), which he states is grammatical both with and without *a*.

(6) *Pepe ve Ø/a la vaca.*  
 Pepe see.3SG OBJ DEF.F.SG cow  
 ‘Pepe sees the cow.’

(Spanish; García García 2018, 211)

Indeed, this seems to be true at least for most speakers of Spanish. Certainly, there are some speakers who only accept the sentence with *a*, while there are probably a few who only accept it without *a*. Yet García García is right, in that one cannot generalize that *a*-marking is obligatory with animal objects. Animals can seemingly be classed either with humans or with inanimate objects; I hypothesize that this may depend on whether the speaker attributes animacy to them. Admittedly, cases of *a*-marked inanimate objects are also attested; for the current study they are regarded as peripheral. For a detailed investigation of DOM on inanimate direct objects of Spanish, cf. García García (2014).

We have seen that animacy is decisive for DOM in Spanish. So far, we have only looked at definite objects. When turning to indefinite objects, we quickly see that animacy is not the only relevant category. In order to concentrate on the categories that Aissen (2003) includes in her definiteness scale, the following examples will contain only animate direct objects. We know that definite NPs are ranked higher on the scale than indefinite NPs, and indeed all animate definite objects are *a*-marked. However, indefinite specific objects are marked, too. This is illustrated in the examples in (7), which are comparable to the examples presented in Comrie (1979, 15; 1981, 127), but have been revised and extended:

(7a) *Buscamos al empleado (que nos atendió ayer).*  
 look\_for.1PL OBJ\_DEF.M.SG employee  
 ‘We look for the employee (that served us yesterday).’



(7b) *Buscamos a un empleado (que lleva gafas).*

look\_for.1PL OBJ INDEF.M.SG employee

‘We look for an employee (that wears glasses).’

(7c) *Buscamos un empleado (cualquiera).*

look\_for.1PL INDEF.M.SG employee

‘We look for an (any) employee.’

(Spanish)

The split of *a*-marking is not between definites (7a) and indefinites (7b–c), but between two types of indefinites, as in (7b) and (7c). In line with many previous studies, I argue that the relevant category is specificity:<sup>2</sup> In (7b) the object looked for is a specific individual who is an employee; in (7c) the object is any member of the set of employees, without further specification as to which individual is being sought. These examples illustrate that, among animate direct objects, the distinction [ $\pm$  specific] dictates whether the object is marked by *a*, as it has been shown by Comrie (1979; 1981), Bossong (1991),<sup>3</sup> and others. Following García García (2018), *a*-marking is possible with non-specific objects in restricted cases, at least when the non-specific reading of the object is enforced by other means, e.g. marked by a relative clause in the subjunctive mood, as in (8). Subjunctive relative clauses indicate non-specificity in Romance languages.

(8) *Pepe busca Ø/a una actriz que hable arameo.*

Pepe look\_for.3SG OBJ INDEF.F.SG actress REL speak.SBJ.3SG Aramaic

‘Pepe is looking for an actress who speaks Aramaic.’

(Spanish; García García 2018, 211)

Yet the default case for non-specific direct objects is still the absence of *a*-marking. The marking is never obligatory with such objects, not even in cases like (8), and not all speakers of Spanish accept it. Although García García’s point

<sup>2</sup> Specificity as the main factor for the marking of indefinite objects is not wholly uncontroversial. For example, López (2012) explains it by movements of functional elements in an underlying syntactic representation. The claim is based on the theoretical framework of minimalist generative grammar, which I do not adopt here. For the purposes of describing the phenomenon, influenced rather by typology than the aim of the theoretical modelling of the syntax, it seems justifiable to stick to the notion of specificity.

<sup>3</sup> Bossong uses the term “referential” instead of “specific”. Although referentiality and specificity are separate concepts in the terminology of some authors, Bossong’s understanding of “referential” is arguably equivalent to what I call “specific” here, and thus I continue to use only the latter term.

is a valuable contribution to the study of the variability of Spanish DOM, it does not pose a problem for the fundamental hypothesis that the distinction [ $\pm$  specific] is decisive for *a*-marking.

So far we have looked at the differential marking of nominal direct objects. But since a large number of Spanish objects appear in the form of clitic pronouns, it is necessary to include pronominal object marking in our analysis. Spanish has distinct clitics for direct and indirect objects. The direct object clitics are specified for gender: *lo* for masculine and *la* for feminine objects. The indirect object clitic *le* is not specified for gender. In contrast to *a*-marking, which does not differentiate between direct and indirect objects, the clitic pronouns distinguish between the two syntactic relations. Thus, I consider it unproblematic to stick to the traditional terminology and call them accusative and dative pronouns, glossed as ACC and DAT respectively. Direct object clitics are used for all direct objects, regardless of their position on the animacy scale, for example, which means that all those objects are overtly marked. The human object in (9a) and the inanimate object in (9b) are both referred to by the same object clitic *lo*. Thus, in the pronominal domain, there is no Differential Object Marking in the narrow sense of the definition, i.e. the presence vs. absence of an overt object marker. A cut-off in marking between personal pronouns and nouns is included as a possible DOM split, e.g. on Aissen's definiteness scale (2003).

(9a) *El empleado? Lo vi.*  
 DEF.M.SG employee 3SG.M.ACC see.PST.PF.1SG  
 'The employee? I saw him.'

(9b) *El coche? Lo vi.*  
 DEF.M.SG car 3SG.M.ACC see.PST.PF.1SG  
 'The car? I saw it.'

(Spanish)

However, there is evidence for something at least closely related to DOM with Spanish pronominal objects. Not all objects can be treated in the same way. The phenomenon of clitic doubling is not possible for all potential objects. Anaphoric direct objects that refer to a human being can be realized in two ways, namely by an object clitic and by a tonic pronoun marked by *a*, these in the same sentence, as in (10a). This is not possible for inanimate referents such as that in (10b), or those without *a*-marking of the tonic pronoun.

- (10a) *Viste a Pedro o a María? – Lo vi*  
 see.PST.PF.2SG OBJ Pedro or OBJ María 3SG.M.ACC see.PST.PF.1SG  
*a él, no a ella.*  
 OBJ 3SG.M NEG OBJ 3SG.F  
 ‘Did you see Pedro or María? – I saw him, not her.’
- (10b) *Viste el pan o la manteca? – \*Lo vi*  
 see.PST.PF.2SG DEF.M.SG bread or DEF.F.SG butter 3SG.M.ACC  
*vi (a) él, no (a) ella.*  
 see.PST.PF.1SG OBJ 3SG.M NEG OBJ 3SG.F  
 ‘Did you see the bread or the butter? – I saw (the bread), not (the butter).’  
 [intended]
- (Spanish)

The pronouns *él* and *ella* seem to be obligatorily *a*-marked when in the direct object position. However, (10b) is still ungrammatical when the tonic pronouns are *a*-marked. Inanimates are regularly not eligible as antecedents for tonic pronominal objects; see Thun (1986) for a description of the scarce cases of tonic pronouns with inanimate antecedents. Clitic doubling is also possible with full nouns or proper names. Just as in the examples above, it is required that the referent of the object is human, as in (11a). Clitic doubling of inanimate nominal objects, for example in (11b) below, is ungrammatical.

- (11a) *Lo vi a Pedro, no a María.*  
 3SG.M.ACC see.PST.PF.1SG OBJ Pedro NEG OBJ María  
 ‘I saw Pedro, not María.’
- (11b) *\*Lo vi el pan, no la manteca.*  
 3SG.M.ACC see.PST.PF.1SG DEF.M.SG bread NEG DEF.F.SG butter  
 ‘I saw the bread, not the butter.’ [intended]
- (Spanish)

What we see here, I argue, is not Differential Object Marking in the narrow sense of the definition, because the clitic pronoun is always an instance of overt object marking. Yet it is evidently the case that not all objects can be marked in the same way; the pattern of pronominal object marking is heterogeneous. Free pronouns can only be used for human, or possibly human-like, specific referents, thus they are always marked by *a* in the object position. In addition to this, nominal objects that are human and specific can be doubled by an object clitic, while inanimate

objects cannot be doubled. Melis (2018) also points out the indexing function of clitic doubling, concluding that the category that is decisive for whether an object can be doubled is its topicworthiness, a notion which is, as she notes, closely linked to both animacy and specificity.

### 2.3 Prepositional object marking beyond Spanish

Thus far I have described the DOM system of Spanish. Since the focus of this paper is on language comparison, I will now turn to the situation of prepositional object marking in Portuguese. Unlike *a*-marking in Spanish, it has not featured widely in synchronic research.<sup>4</sup> Schäfer-Prieß/Schöntag (2012, 146), for instance, note that the prepositional accusative<sup>5</sup> only plays a marginal role. However, it is not wholly absent. It occurs, for example, in cases of clitic doubling, a phenomenon we have just seen in Spanish, where DOM precedes tonic personal pronouns in direct object position:

- (12) *Paulo me vê a mim, não a ti (EP)/você (BP).*  
 Paulo 1SG.OBJ see.3SG OBJ 1SG.TON NEG OBJ 2SG.TON  
 ‘Paulo sees me, not you.’

(Portuguese)

In contrast to Spanish, nominal direct objects in simple transitive sentences cannot be *a*-marked in Portuguese, even when animate and specific, and regardless of gender.

- (13a) *Maria vê o /\*ao homem.*  
 Maria see.3SG DEF.M.SG /OBJ\_DEF.M.SG man  
 ‘Maria sees the man.’

- (13b) *Maria vê a /\*ã mulher.*  
 Maria see.3SG DEF.F.SG /OBJ\_DEF.F.SG woman  
 ‘Maria sees the woman.’

(Portuguese)

<sup>4</sup> There are several historical investigations, an early one being Delille (1970). However, the diachrony is not in the focus of this paper.

<sup>5</sup> “Prepositional accusative” is a term that many older works, as well as contemporary traditionalist ones, use for the function of *a*-marking of direct objects. I will not use this term in the present paper because I consider it imprecise, since Spanish and Portuguese do not have inflectional case on nouns, and since *a* is also the marker of indirect objects.

Yet animacy also plays a role in Portuguese. In the cases of clitic doubling in (12), *a* occurs with personal pronouns that usually refer to humans. But animacy is also decisive for *a*-marking of nominal objects. Cyrino (2017a) shows that the relation between animacy and object marking is relevant for the grammaticality and interpretation of the examples cited in (14).

(14a) *Eu vi o menino e ao professor também.*  
 1SG see.PST.PF.1SG DEF.M.SG boy and OBJ\_DEF.M.SG professor too  
 ‘I saw the boy and (I saw) the professor, too.’

(14b) *Eu vi o livro e o caderno também.*  
 1SG see.PST.PF.1SG DEF.M.SG book and DEF.M.SG notebook too  
 ‘I saw the book and (I saw) the notebook, too.’

(14c) *Eu vi o livro e \*ao caderno também.*  
 1SG see.PST.PF.1SG DEF.M.SG book and OBJ\_DEF.M.SG notebook too  
 ‘I saw the book and (I saw) the notebook, too.’ [intended]

(14d) *Eu vi o menino e o professor também.*  
 1SG see.PST.PF.1SG DEF.M.SG boy and DEF.M.SG professor too  
 ‘I saw the boy and the professor (saw him), too.’

(Portuguese; Cyrino 2017a, 99)

When two human direct objects appear next to each other due to verb ellipsis, the second one is marked by *a*, as in (14a). In a coordination of inanimate direct objects, like in (14b), there is no *a*-marking. The same sentence with *a* preceding the second object (14c) is ungrammatical. The motivation behind the differential marking of the human object is arguably disambiguation. When the *a*-marking of the human direct object is left out, the reading changes: The construction in (14d) is not understood as a coordination of direct objects, but rather as a case of VP ellipsis with a second subject. The ellipsis reading is not available for the inanimate subject in (14b), probably for semantic reasons: The verb *ver* ‘see’ requires an animate subject. Disambiguation as the background for *a*-marking is also recognized by Schäfer-Prieß (2003). In (15) the *a* makes clearly understandable that the subject treats the objects as if the latter were humans, not as if the subject itself were human.

- (15) *Tratava-os*                                *como a homens, como a*  
 treat.PST.IP.3SG-3PL.M.ACC like    OBJ man.PL like    OBJ  
*amigos.*  
 friend(M).PL  
 ‘He/she treated them like men, like friends.’  
 ([European] Portuguese; Schäfer-Prieß 2003, 405)

The disambiguation is necessary because the referents of *homens* and *amigos* are ideal subjects for the verb *tratar*, and thus not typical objects. This is due to their property of being human. Human direct objects are also frequent with psych-verbs of a certain kind, namely those that denote feelings such as love or hate. Consequently, Schäfer-Prieß (2003, 405) notes that direct objects of these “verbos de sentimento” are frequently *a*-marked:

- (16) *Só*                                *não amava*                                *a Jorge como amava*  
 just                                NEG love.PST.IP.3SG    OBJ Jorge like    love.PST.IP.3SG  
*ao*                                *filho.*  
 OBJ\_DEF.M.SG    son  
 ‘He/she just did not love Jorge like he/she loved the son.’  
 (Portuguese; Cunha/Cintra 1992 *apud* Schäfer-Prieß 2003, 405)

For such verbs, the need for overt object marking seems reasonable because the set of possible subjects and the set of possible objects largely overlap. Disambiguation is a frequent motivation for DOM cross-linguistically, which is an effect of the discriminatory function of case: Seržant (2019) argues that the discriminatory function of case influences DOM as a weak universal force, being visible in the diachrony but also “leav[ing] behind traces in the form of local disambiguation” (loc. cit. 167). He explicitly mentions such cases in Spanish (loc. cit. 157). In Portuguese, disambiguation seems to be a decisive kind of motivation for prepositional object marking, while animacy of the object referent is a necessary condition for it.<sup>6</sup>

<sup>6</sup> An anonymous reviewer objects that contemporary BP has stressed direct object pronouns with animate antecedents without *a*-marking, which seems to pose a problem for a scale-based analysis. Indeed, the overall picture would be more consistent if all stressed object pronouns were marked by *a*. However, I do not consider this a major obstacle; it does not provide evidence for the scale, but does not contradict it either: In contexts where the prepositional marking is possible, e.g. for disambiguation, its occurrence is restricted by properties of the object itself. And where the marking is restricted by properties of the object, it follows the predicted scale.

### 3 Accusative vs. dative: *leísmo* in Spanish and Portuguese

Prepositional DOM is surely the most prominent instance of heterogeneous object marking in Spanish and Portuguese, but it is not the only one. In this Section I will discuss what is widely known by the Spanish term *leísmo*. Found mainly in some varieties of European Spanish, such as Castilian and Basque Spanish, among others, the use of dative pronouns referring to direct objects has been described in numerous studies and has also been discussed in normative grammars. I will start with a description of *leísmo* in Spanish, before turning to equivalent findings for Portuguese.

In Spanish, the 3<sup>rd</sup> person dative clitic *le*, not specified for gender, prototypically occurs in the indirect object position. Direct objects are, by default, referred to by the accusative clitics *lo/la*. In ES,<sup>7</sup> *le* can also be used as a direct object.<sup>8</sup> However, for most ES speakers, not all direct objects can be referred to by *le*. The examples in (17) hint at the relevance of animacy for *leísmo*.

(17a) *El criado no trabaja bien. Le/?Lo voy a reemplazar.*  
   3SG.DAT/ACC  AUX.FUT.1SG  FUT  replace  
   ‘The servant does not work well. I will replace him.’

(17b) *El escritorio no me gusta más. \*Le/Lo voy a reemplazar.*  
   3SG.DAT/ACC  AUX.FUT.1SG  FUT  replace  
   replace  
   ‘I do not like the desk anymore. I will replace it.’  
   (European Spanish)

The human direct object in (17a) is referred to by the dative clitic, while this is impossible for the inanimate object in (17b); the inanimate object would require the accusative clitic *lo*. The relevance of animacy for object marking reminds us of the scale adopted for *a*-marking in Section 2, and the neutralization of the distinction between direct and indirect objects, that is, the use of the indirect object marker for some direct objects also links *leísmo* to prepositional DOM. Yet there are some significant differences between the two phenomena. When dis-

<sup>7</sup> This is a generalization about ES. There may be smaller European varieties that behave differently, but this claim captures the behaviour of the majority of ES speakers.

<sup>8</sup> In order to simplify the analysis, I will only consider singular examples in this paper.

cussing *a*-marking in the pronominal domain, I mentioned that all pronominal direct objects are overtly marked. This is also relevant for the classification of *leísmo*: *leísmo* is not DOM in the sense of the presence vs. absence of marking. In the domain where it occurs, all direct objects are marked, but some of them are marked in a different way. This does not fit the definition of DOM given in Section 2. However, it can be classified as a case of paradigmatic Differential Object Marking as understood by de Swart (2014). Paradigmatic DOM, he argues, does not mark the differentiation between subject and object, but rather between different types of objects. Thus, in contrast to DOM in the sense of the definition given in Section 2, which de Swart calls “syntagmatic DOM”, the marking does not happen due to the presence instead of absence of an object marker, but through the choice of one object marker instead of another.

A further difference between *leísmo* and *a*-marking lies in the categories that are available to influence the marking. Both phenomena depend on animacy. For *leísmo*, gender is available as a criterion, demonstrating interesting differences between varieties. In Castilian Spanish, gender conditions the case of the object clitic for human referents. Masculine direct objects are referred to by the dative clitic (18a), while feminine direct objects are referred to by the accusative clitic (18b).

(18a) *Vi*                    *al*                    *chico. Le*            *vi.*  
 see.PST.PF.1SG OBJ\_DEF.M.SG boy 3SG.DAT see.PST.PF.1SG  
 ‘I saw the boy. I saw him.’

(18b) *Vi*                    *a la*                    *chica. La*            *vi.*  
 see.PST.PF.1SG OBJ DEF.F.SG girl 3SG.F.ACC see.PST.PF.1SG  
 ‘I saw the girl. I saw her.’

(Castilian Spanish; Fábregas 2013, 45)

Basque Spanish is different from Castilian in that gender is not relevant for *leísmo*. For human direct objects, the dative clitic is used regardless of their gender. The masculine object in (19a) and the feminine object in (19b) are reflected by the same pronoun.

(19a) *Vi*                    *al*                    *chico. Le*            *vi.*  
 see.PST.PF.1SG OBJ\_DEF.M.SG boy 3SG.DAT see.PST.PF.1SG  
 ‘I saw the boy. I saw him.’



- (19b) *Vi a la chica. Le vi.*  
 see.PST.PF.1SG OBJ DEF.F.SG girl 3SG.DAT see.PST.PF.1SG  
 ‘I saw the girl. I saw her.’

(Basque Spanish; Fábregas 2013, 45)

According to Fernández-Ordóñez (2012, 88), the Basque type of *leísmo* is also found in some small varieties of AS, such as Paraguayan and Ecuadorian Spanish. These all allow for examples such as (20).

- (20) *Le veo a María.*  
 3SG.DAT see.1SG OBJ María  
 ‘I see María.’

(Basque/Paraguayan/Ecuadorian Spanish; Fernández-Ordóñez 2012, 88)

Granda (1982) elaborates on the use of *leísmo* in Paraguayan Spanish. He claims that *le* is used as a direct object pronoun regardless of animacy or gender, and without any diatopic differences within the country (loc. cit., 262–263). He mentions that in formal registers *leísmo* occurs preferably for human direct objects, thus resembling European *leísmo*. This might be due to the fact that normative grammars are oriented towards the European standard. But the default in spoken Paraguayan Spanish is a generalized *leísmo*, which Granda explains through contact with Guaraní. The pronominal system of Guaraní does not distinguish between direct and indirect objects; cf. Palacios (1998, 439–440) for a concise description of the pronominal paradigm. This generalized replacement of direct object pronouns by *le* is a totally different phenomenon from *leísmo* in Europe. As the distinction between accusative and dative pronouns is completely lost, no heterogeneous object marking takes place. This pattern is not of interest for us here; cf. Fernández-Ordóñez (2012, 93ss.) for a variationist description of similar behaviour in other varieties of Spanish. For the purpose of the present study, this brief description of the phenomenon and the relevant categories must suffice.

Moreover, we have to separate yet another phenomenon from our primary object of research: García (1975, 307ss.), who is working primarily with data from the variety of Buenos Aires, describes an alternation between *lo/la* and *le* in which the preference for either the accusative or the dative depends on properties of the subject. The dative is favoured for objects in sentences with an inanimate subject, while the accusative is preferred in the presence of an animate subject. As indicated in section 2.1, this paper concentrates on differential marking phenomena which occur within the same syntactic and semantic environment and can be traced back to properties of the respective object referents. This condition does not apply to the phenomenon described by García. The same holds true for

another type of alternation between *lo/la* and *le* (loc. cit., 317ss.) in which the preference for either the dative or the accusative is determined by the role that the object has in the situation referred to. The role that an individual takes in a situation is not a property of that individual, but emerges from the situation.

The situation in Portuguese is just as diverse as in Spanish. There is no valid evidence for a corresponding phenomenon, such as *lheísmo*,<sup>9</sup> in EP. The 3<sup>rd</sup> person dative clitic *lhe* only occurs in the indirect object position. Direct objects are referred to by the accusative clitics *o/a*. However, the frequently made claim that *leísmo* does not exist in Portuguese is false. Several studies have shown the existence of a *leísmo* phenomenon in Brazilian Portuguese. Following Nascimento (2010), the function of *lhe* is not restricted to denoting the indirect object in BP, as we can conclude from examples (21) and (22) below; it may denote both 2<sup>nd</sup> person and 3<sup>rd</sup> person direct objects:

(21) *Lucas lhe viu ontem.*  
 Lucas 3SG.DAT see.PST.PF.3SG yesterday  
 ‘Lucas saw him yesterday.’

(22) *Eu lhe amo muito.*  
 1SG 3SG.DAT love.1SG much  
 ‘I love you very much.’

(Brazilian Portuguese; Nascimento 2010, 54)

Almeida (2011, 2406) claims that the use of *lhe* for direct objects is found throughout Brazil, but is most common in the north-eastern states. Marroquim (1934, 182ss.) mentions that it is attested in the dialects of the states of Alagoas and Pernambuco, Nascentes (1953, 127ss.) describes it for Rio de Janeiro and Almeida (2009) analyzes it extensively for the dialect spoken in Salvador da Bahia. Just as in Spanish, the appearance of the dative pronoun instead of the accusative seems to be conditioned by animacy. Citing the dissertation of Conceição de Maria Araújo Ramos,<sup>10</sup> Almeida (2011, 2403) points out that the same tendency is evident in African Portuguese varieties and in Galician.

<sup>9</sup> The term *lheísmo* is a Portuguese adaptation of the Spanish term *leísmo*, alluding to the dative pronoun *le*. Henceforth, the widely-used Spanish term will be used in the description of both languages, in order to allow for a generalization of the findings.

<sup>10</sup> Doctoral dissertation at the Federal University of Alagoas, 1999, entitled *O clítico de 3a pessoa: um estudo comparativo português brasileiro/espanhol peninsular*. The work itself is not accessible, thus not cited in the original.

Pereira/da Silva (2014) describe the diachrony of *lhe* as a direct object clitic; they point out that this use has been present for at least two centuries. As a possible motivation of *leísmo* in both Spanish and Portuguese, they assume an analogy with the use of the pronominal object clitics of the 1<sup>st</sup> and 2<sup>nd</sup> person, *me* and *te*, which are used equally for direct and indirect objects. 1<sup>st</sup> and 2<sup>nd</sup> person pronouns usually have human referents, and *leísmo* also occurs with human referents. The proposed motivation, then, is plausible.

In contrast to ES *leísmo*, but just like AS *leísmo*, the corresponding phenomenon in BP is still not accepted by normative grammars, a decision which was criticized as early as Nascentes (1960). This is perhaps due to the fact that it has not been studied as extensively as in Spanish. More empirical research is needed to show how widespread it is in BP and in order to argue for its inclusion in normative grammars. Even if it is found that a large proportion of the speakers of Portuguese do not produce *leísmo* themselves, it should be investigated whether they accept it when asked for grammaticality judgments, and if they do, whether they prefer it in certain contexts over others.

## 4 Total absence of marking: the case of null objects

In the previous Section, we looked at heterogeneous object marking in the domain of clitic pronouns, which consists of the contrast between accusative and dative clitics. Here we turn to the absence of anaphoric pronouns in the direct object position, commonly called null objects. This phenomenon is well-known in Brazilian Portuguese but is also found to a limited extent in EP and in Spanish. The sentence in (23), repeated from (3), which is grammatical in both BP and EP, shows an example of a null object.

- (23) *Comprei*            *um*            *pãozinho*   *e*        *comi*            *a*  
 buy.PST.PF.1SG   INDEF.M.SG   bun            and   eat.PST.PF.1SG   on  
*caminho.*  
 way  
 ‘I bought a bun and ate it on the way.’

(Portuguese)

Both verbs *comprei* ‘bought’ and *comi* ‘ate’ are transitive and thus require a direct object. In the first clause, *pãozinho* ‘bun’ is realized as the direct object, while in the second clause there is no overt object. Yet the second clause is neither ungrammat-

ical nor underspecified in terms of its object; *pãozinho* is understood as the object of *comi* without the use of an anaphoric pronoun, which would be obligatory in English ‘[...] ate it [...]’, for example. In BP, such null objects are very frequent. This is favoured by the condition that spoken BP has almost completely lost the direct object clitics in the 3<sup>rd</sup> person. However, not all anaphoric direct objects are equally eligible to be realized as a null object. 1<sup>st</sup> and 2<sup>nd</sup> person objects are inherently animate and specific, thus always require pronominal realization, and there are also restrictions among 3<sup>rd</sup> person objects. In example (24), the object introduced in the first sentence will usually not be referred to by a null object in the second sentence.

- (24) *Você conhece o Paulo? Eu vi ele/∅*  
 2SG know.2SG DEF.M.SG Paulo 1SG see.PST.1SG 3SG.M  
*ontem.*  
 yesterday  
 ‘Do you know Paulo? I saw him yesterday.’

(Brazilian Portuguese)

Why is it that the null object is possible in (23), but questionable in (24)? The most striking difference between the two objects lies in the dimension of animacy: The direct object in (24), *Paulo*, is human, while *pãozinho* in (23) is inanimate. If a morphological device, a pronoun in this case, is required to mark human direct objects but is unnecessary for inanimate direct objects, then this hints at a relation to DOM: It is an asymmetry between different types of direct objects that has an impact on whether they receive overt marking, and this asymmetry is conditioned by a parameter that has been shown to be relevant for DOM in several languages.

Null objects in BP have been studied extensively (cf. Cyrino 1994; 2016; 2017b; Cyrino/Lopes 2016, among others). Among the interesting illustrations provided by the former author, we find the following example, which is grammatical with either a null object or a pronominal object. The difference in object marking leads to a difference in meaning:

- (25) *Eu nunca vejo o meu pai. Nem me*  
 1SG never see.1SG DEF.M.SG POSS.1SG father not\_even 1SG.REFL  
*lembro do rosto dele. Acho que*  
 remember.1SG of\_ DEF.M.SG face POSS.3SG.M think.1SG that  
*já esqueci ∅ /ele.*  
 already forget.PST.1SG 3SG.M  
 ‘I never see my father. I do not even remember his face. I think I already forgot it/him.’

(Brazilian Portuguese; cf. Cyrino 2016, 181)

In the case of the null object, the father's face is understood as the object of *esqueci* 'I forgot'. If the pronoun *ele* is used as the object, the situation is different: Cyrino (2016, 182) states that the pronoun can be understood as referring either to the face, or to the father as a whole. However, speakers prefer the interpretation of *ele* as being coreferent with *o meu pai* 'my father', i.e. referring to the person, not only to a part of his body. This judgment is also argued for by Schwenter/Silva (2002, 582), who analyze a similar example. We can conclude from example (25) that the null object refers to an inanimate entity, while the overt pronoun refers to a human. The impression from examples (23) and (24) above is confirmed: An animate object has to be marked by a morphological device, while an inanimate object does not need overt marking. However, the acceptability of null objects depends on more than just animacy. Examples (26a) and (26b) both feature a human direct object. In (26a) it is taken up by a pronoun. In (26b) there is a null object, despite the referent being human.

(26a) *O embaixador procura um diplomata*  
 DEF.M.SG ambassador look\_for.3SG INDEF.M.SG diplomat  
*que fala alemão, mas não encontra ele.*  
 REL speak.3SG German but NEG find.3SG 3SG.M  
 'The ambassador looks for a diplomat that speaks German, but he does not find him.'

(26b) *O embaixador procura um diplomata*  
 DEF.M.SG ambassador look\_for.3SG INDEF.M.SG diplomat  
*que fale alemão, mas não encontra Ø.*  
 REL speak.SBJ.3SG German but NEG find.3SG  
 'The ambassador looks for a diplomat that speaks German, but he does not find one.'

(Brazilian Portuguese)

When *um diplomata* 'a diplomat' is taken up by the pronoun *ele*, it is understood that it does not refer to any arbitrary diplomat, but to a specific one. The null object in (26b) triggers the interpretation that the ambassador needs a diplomat, but does not have one in mind yet, so he looks for any diplomat that is available, but finds none; it has an unspecific reading. The distinction in specificity is reflected by the mood of the relative clause: indicative in (26a) and subjunctive in (26b). I conclude that the acceptability of null objects in BP depends on animacy and specificity, the same dimensions that are relevant for *a*-marking in Spanish.

So far, all the examples featuring null objects had identical subjects in the sentence introducing the object and in the sentence with the null object. But a

null object is also possible in BP when a new subject is introduced in the same clause:

- (27) *João descascou a banana, mas Pedro não comeu.*  
 João peel.PST.3SG DEF.F.SG banana but Pedro NEG eat.PST.3SG  
 'João peeled the banana, but Pedro did not eat it.'

(Brazilian Portuguese; Cyrino 1994, 144)

With respect to this issue, there is an important difference between BP and EP: Example (27) is grammatical in BP, but ungrammatical in EP. In EP, where the 3<sup>rd</sup> person direct object clitics that have been lost in BP are still in frequent use, the acceptability of null objects is very restricted. Of all the examples of null objects above, only the first one in (23) is grammatical in EP, where two sentences share the same subject and object and only the verb changes. Also, it is important that the two sentences are coordinated; a null object in a subordinated sentence as in (28) is only possible in BP, but not in EP.

- (28) *Comprei o casaco depois que experimentei Ø.*  
 buy.PST.1SG DEF.M.SG jacket after that try\_on.PST.1SG  
 'I bought the jacket after trying it on.'

(Brazilian Portuguese; Cyrino 2016, 180)

An early study of null objects in EP is offered by Raposo (1986), who describes the phenomenon within the Government and Binding framework and analyzes it as a syntactic variable. While he mentions the acceptability of null objects in Chinese as occurring under almost the same circumstances as in EP, he assumes that within the Romance family, the null object is an exclusive feature of Portuguese which is thus not found in languages such as Spanish and French. Raposo distinguishes null objects from VP ellipsis, in that they occur, for instance, in answers to polar questions. Outlining the narrow limits of null objects in EP, he admits the acceptability of cases like (23), but also gives the following example:

- (29) *A Juana viu Ø na TV ontem.*  
 DEF.F.SG Juana see.PST.3SG in\_DEF.F.SG TV yesterday  
 'Juana saw it on TV yesterday.'

(European Portuguese; Raposo 1986, 373)

Unfortunately, Raposo does not provide the context for the sentence in (29), thus it is not clear what kind of antecedent the null object has, e.g. whether it is animate or specific. Raposo himself argues that the context is relevant for the

interpretation of null objects. Nevertheless, for many of his examples, the context remains unknown. His analysis is thus rather unsatisfying. Kato (2011) revisits Raposo's work and attempts to integrate null object clitics and null definite articles, which are homonymous in Portuguese, into a single category.

As mentioned above, Raposo (1986) assumed that the null object is an exclusive feature of Portuguese. Later research proved this hypothesis wrong. Portuguese is the Romance language in which null objects are most frequent, but in fact Spanish also displays them under certain conditions. Campos (1986) found a phenomenon resembling a null object in examples like (30) below and named it "indefinite object drop".

- (30) *Compraste café? – Sí, compré Ø.*  
 buy.PST.PF.2SG coffee yes buy.PST.PF.1SG  
 'Did you buy coffee? – Yes, I bought (some).'

(Spanish; Campos 1986, 354)

As was the case for the Portuguese null object, the second sentence, which in this case was the answer, does not show an overt direct object, but nevertheless is neither ungrammatical nor underspecified for its object; it is unambiguously understood that the object of 'buy' is 'coffee'. In an attempt similar to Raposo's analysis of Portuguese (1986), the same author classifies the dropped object as a variable, which may only substitute indefinite objects. In our terminology, all the examples of dropped direct objects are not only indefinite, but also unspecific, and additionally, all of them are inanimate. This reminds us, at first sight, of what we have seen in Portuguese. However, it is questionable whether the dropped objects as described by Campos are really equivalent to null objects. In all his examples, the subject and verb of the sentence containing the dropped object are identical to those of the sentence containing the antecedent, and the antecedent is always the direct object, too. The dropping of the indefinite object might thus be a simple case of VP ellipsis, common in all Romance languages.

While Campos (1986) looked at ES, recent work on null objects in Spanish concentrates on Latin American varieties. Schwenter (2006) and Cyrino (2016) both compare null objects in AS to those in BP. Regarding the findings of Campos, Cyrino claims that all varieties of AS allow for a sentence like (31).

- (31) *Quería comprar libros pero no encontraba.*  
 want.PST.1SG buy books but NEG find.PST.1SG  
 'I wanted to buy books but I couldn't find them.'

(American Spanish; Alamillo/Schwenter 2007 *apud* Cyrino 2016, 189)

Unlike (30), and like some of the examples of BP, (31) changes the verb from the first part of the sentence to the second. Thus I assume that it has, indeed, a null object. The object is inanimate and unspecific. The modified versions of this example below, which feature a specific (32a) and a definite (32b) object, are judged ungrammatical:

(32a) *Quería comprar unos libros pero no\* Ø encontraba.*  
 want.PST.1SG buy INDEF.M.PL books but NEG find.PST.1SG  
 ‘I wanted to buy some (specific) books but I couldn’t find them.’

(32b) *Quería comprar los libros pero no\* Ø encontraba.*  
 want.PST.1SG buy DEF.M.PL books but NEG find.PST.1SG  
 ‘I wanted to buy the books but I couldn’t find them.’

(American Spanish)

We see in these examples that null objects of AS are possible with unspecific antecedents only. This means that they are more restricted than in BP, but it suits our general assumptions about the relation between specificity and overt object marking.

Even within AS, there is considerable variation with respect to the acceptability of null objects. Paraguayan Spanish allows for definite inanimate antecedents to be taken up by null objects:

(33) *¿Dónde encontraste esa blusa? – Ø Compré*  
 where find.PST.2SG DEM.F.SG blouse – Ø buy.PST.1SG  
*en el mall.*  
 in DEF.M.SG mall  
 ‘Where did you find that blouse? – I bought it at the mall.’

(Paraguayan Spanish; Schwenter 2006, 30)

In allowing for examples like (33), Paraguayan Spanish is similar to BP. Just as BP, it has seen a loss of 3<sup>rd</sup> person direct object clitics (cf. Schwenter 2006, 30), which probably favours the extension of null objects to environments where they are not acceptable in other varieties of Spanish. The absence of direct object pronouns when the referent is inanimate has already been observed in Paraguayan Spanish by Palacios (1998). She reports on sporadic occurrences of null objects with animate antecedents (loc. cit., 435), but finds that there is a strong tendency for null objects to be used only for inanimate referents. While she attributes the origin of Paraguayan Spanish null objects to language contact with Guaraní, she also mentions that null objects with inanimate referents occur to a similar extent



in Spanish varieties which are in contact with Quechua, namely those spoken in Peru, Bolivia and Ecuador (loc. cit., 443ss.). I shall not expand on the details of these contact phenomena, which fall outside the scope of the present study. At this point it is sufficient for us to say that where there is a split in the use of null objects in a variety of Spanish, it follows the predictions of the marking scales introduced before. The differences between varieties of Spanish are no problem for my analysis as long as they do not contradict my basic assumptions about marking scales. The exact cut-off point for marking may change from variety to variety, even from speaker to speaker, as long as the order of the elements on the scale is not changed. So far, we have not seen any reason to assume that the scale itself is changed.

Apparently, null objects are possible in the Spanish language. However, our knowledge about the acceptability of null objects in ES is not yet sufficient. Schwenter and Cyrino only worked on American Spanish, and Campos' (1986) study does not provide enough robust evidence. New empirical research is thus needed on null objects in ES. This goes beyond the scope of this paper, but it is noted here as highly relevant. However, at the present state of research we can formulate the generalization that the frequency of null objects varies on two dimensions, namely from Spanish (less) to Portuguese (more) and from Europe (less) to America (more).

## 5 A common denominator

In the preceding sections, it has become clear that there is a notable common denominator in the object marking patterns described for Spanish and Portuguese. Prepositional DOM, clitic doubling, *leísmo* and null objects share some important properties that I will deal with in this Section. Obviously, all those phenomena essentially consist of heterogeneous marking of direct objects. An animacy scale is relevant for all of them, and for those that occur in a domain where both specific and unspecific objects are possible, a specificity distinction is also evident.

Some existing studies have already investigated similarities among *a*-marking, *leísmo* and null objects, mainly with the aim of subsuming either *leísmo* or null objects under the notion of DOM. Flores/Melis (2007) study *leísmo* from a DOM perspective. They describe the diachronic development of object marking in both the nominal and the pronominal domain, adopting the hypothesis that *leísmo* is a part of Spanish DOM. Fernández-Ordóñez (2012), in her work on the variation of pronominal paradigms in Spanish dialects, also considers *leísmo* to be a

DOM system. What is missing in these accounts is the comparative dimension. Fernández-Ordóñez at least includes the Basque language, but neither of these two studies offers a proper perspective on Portuguese. A comparative view on Spanish and Portuguese is, however, available in the works of Schwenter (2006; 2014, among others). Schwenter (2006) compares null objects in BP and AS to Spanish *a*-marking. He makes the important point that the null object, especially in the case of BP, is not an anomaly, but rather the standard unmarked case for anaphoric direct objects. Null objects, he argues, show several DOM properties, but are not ultimately classified as DOM. The step taken in Schwenter (2014) is that he names null objects and *a*-marking as “two kinds of differential object marking”. *Leísmo* is mentioned very briefly in this latter work, but not included properly in the classification of the object marking patterns. Clitic doubling is discussed as a DOM phenomenon in Melis (2018).

Thus far, existing studies only adequately included two of the four phenomena. What has been missing is the integration of all four in a single work. This has been done in the current paper. The most important question that has to be answered now is: Should we consider all those patterns of heterogeneous object marking instances of DOM, as has been done separately for *leísmo*, clitic doubling and null objects? If we want to stick to the narrow definition of Differential Object Marking as set out in Section 2, which essentially states that DOM consists of a contrast between direct objects with and without a morphological object marker, then we cannot subsume all phenomena described in this paper under the term of DOM. The only clear case is that of *a*-marking; there is no reason to doubt that the classification of Spanish and Portuguese prepositional object marking as DOM, which has been confirmed in many works, is correct. Clitic doubling, however, does not match the definition. There is a split between the presence and absence of a morphological marker, but all direct objects eligible for clitic doubling are already overtly marked by another means – either by *a* or by an object pronoun. *Leísmo* is not DOM in the sense of our definition, either. There is no alternation between overt object marking and no object marking, but rather between two different object markers: All pronominal direct objects are overtly marked, but some are marked like indirect objects. In the latter point there is a parallel to prepositional DOM: The *a*-marked direct objects are marked in the same way as indirect objects as well. In the sense of de Swart (2014), *leísmo* is a case of paradigmatic DOM. However, finding a functional explanation for the relevance of gender remains difficult. And finally, null objects are also not DOM in the sense of the classical definition: The essential opposition is one between marking and no marking, but not in the sense of presence or absence of an object marker. What is present or absent is not only the object marking, but the entire pronoun.

We see that clitic doubling, *leísmo*, null objects and *a*-marking differ in terms of important properties, and that they cannot be subsumed under the definition of DOM as given in Section 2. Yet there is still a common denominator here. Whenever the heterogeneous marking depends on the object itself, it can be predicted on the base of the same essential scale, represented in (34). The cut-off point for overt marking differs from variety to variety and from mechanism to mechanism, but the scale remains unchanged. The higher a direct object is ranked, the more atypical it is as an object, and the more likely it is to be overtly marked. Gender is only relevant with pronominal objects, (pro-)nominality is only relevant with animate objects, and animacy is only relevant with specific objects. The result is the following object marking scale for Ibero-Romance languages:

- (34) Ibero-Romance object marking scale:  
 specific, animate, pronominal, masculine  
 > specific, animate, pronominal, feminine  
 > specific, animate, nominal  
 > specific, inanimate  
 > unspecific

The remaining question is: Does the uniform analysis of Spanish and Portuguese object marking phenomena, as presented in this paper, allow us to subsume all of these under the same notion? Classifying all of them as DOM would require changes to the well-grounded and widely recognized definition of this term, which I regard as undesirable. Thus, we need a new term to generalize about *leísmo*, null objects, clitic doubling and prepositional DOM. In the previous paragraphs I spoke about patterns of “heterogeneous object marking”. This term, however, is not accurate enough, as the adjective “heterogeneous” does not imply that there is a predictable system behind the object marking mechanisms. The most central common denominator among the different phenomena, I argue, is the shared scale in (34). Thus, tentatively, I propose the term “scale-based object marking”.

It cannot be claimed, though, that the object marking scale presented above accounts for all direct object marking asymmetries in Spanish and Portuguese. Verbal factors, such as aspectual properties and lexical peculiarities, also influence object marking in the two languages. My proposal does not deny these effects, and neither does it exclude further linguistic mechanisms potentially impacting the presence or choice of an object marker, occasionally overriding the object marking scale. Yet for the cases occurring under the terms that were assumed as characteristic of DOM in the narrow sense in Section 2.1 – differential marking occurs within the same syntactic and semantic environment, that is, it is not caused only by word order or by lexical peculiarities of the predicate; and

it can be traced back to properties of the respective object referents – the scale makes valid predictions.

## 6 Conclusions

In the present paper I have presented a comparative analysis of prepositional DOM, clitic doubling, *leísmo* and null objects in Spanish and Portuguese. All four phenomena are found in both languages, although not necessarily in all varieties. Prepositional DOM is more frequent in Spanish than in Portuguese. *Leísmo* is common in European Spanish, found less frequently in American Spanish,<sup>11</sup> attested to a small extent in Brazilian Portuguese and not found in European Portuguese. Null objects are more frequent in Portuguese than in Spanish, and more frequent in America than in Europe. The exact marking patterns differ not only from language to language and from continent to continent; even local varieties within a country may show their own peculiarities in object marking. Yet all those patterns are based on the same scale presented in (34), in Section 5, which is decisive whenever the marking depends on the properties of the object itself.

As a generalization deriving from all those phenomena I have proposed the notion “scale-based object marking”. Whether the direct object of a transitive verb in the Ibero-Romance languages is marked more intensively than the default, be it by the presence of a morphological object marker, by choice of a different case form, or by overt realization of a pronoun, is determined by its position on a scale which ranks all potential objects with respect to the criteria of gender, (pro-)nominality, animacy and specificity. Scale-based object marking as described in this study is, then, a property of all Ibero-Romance languages.

As a next step in research on scale-based object marking in Spanish and Portuguese, broadly based empirical studies are needed to prove or disprove the generalization proposed in this paper. Corpus studies might be a useful start here, but will probably not be sufficient, since there is a lack of data on the spoken language for many varieties that would need to be included in any study. Fur-

---

<sup>11</sup> As an anonymous reviewer notes, the impression that *leísmo* is far less frequent in AS than in ES might be a result of data bias: In ES it is accepted within the rules of the language, which leads to high frequencies in written data as well as positive judgments by speakers. In AS, by contrast, it is not an accepted norm. Thus, it is hardly used in written texts, and speakers are less likely to judge it to be acceptable. An extensive empirical study would be necessary to avoid any such bias, but is not feasible within the scope of the present study.

thermore, acceptability judgments of speakers of different varieties seem to be an advisable means of verifying the relevance of the object marking scale.

## Bibliography

- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21 (2003), 435–483.
- Almeida, Gilce, *Quem te viu quem lhe vê. A expressão do objeto acusativo de referencia à segunda pessoa na fala de Salvador*, Salvador, PhD Thesis, Federal University of Bahia, 2009.
- Almeida, Gilce, *Prescrição gramatical e uso. O caso do pronome “lhe” no português brasileiro*, *Cadernos do Congresso Nacional de Linguística e Filologia* 15:5 (2011), 2398–2408.
- Bossong, Georg, *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen, Narr, 1985.
- Bossong, Georg, *Differential Object Marking in Romance and beyond*, in: Wanner, Dieter / Kibbee, Douglas A. (edd.), *New analyses in Romance linguistics. Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, April 7–9, 1988*, Amsterdam/Philadelphia, John Benjamins, 1991, 143–170.
- Bossong, Georg, *Die romanischen Sprachen. Eine vergleichende Einführung*, Hamburg, Helmut Busse, 2008.
- Campos, Héctor, *Indefinite object drop*, *Linguistic Inquiry* 17:2 (1986), 354–359.
- Comrie, Bernard, *Definite and animate direct objects. A natural class*, *Linguistica Silesiana* 3 (1979), 13–21.
- Comrie, Bernard, *Language universals and linguistic typology*, Oxford, Basil Blackwell, 1981.
- Cyrino, Sonia, *O objeto nulo no português do Brasil. Um estudo sintático-diacrônico*, Campinas, PhD Thesis, State University of Campinas, 1994.
- Cyrino, Sonia, *The null object in Romania Nova*, in: Kato, Mary/Ordóñez, Francisco (edd.), *The morphosyntax of Portuguese and Spanish in Latin America*, New York, Oxford University Press, 2016, 177–203.
- Cyrino, Sonia, *Reflexões sobre a marcação morfológica do objeto direto por “a” em português brasileiro*, *Estudos linguísticos e literários* 58 (2017), 83–103 (= 2017a).
- Cyrino, Sonia, *On animacy restrictions for the null object in Brazilian Portuguese*, in: Hellan, Lars/Malchukov, Andrej/Cennamo, Michela (edd.), *Contrastive studies in verbal valency*, Amsterdam, John Benjamins, 2017, 275–294 (= 2017b).
- Cyrino, Sonia/Lopes, Ruth, *Null objects are ellipsis in Brazilian Portuguese*, *The Linguistic Review* 33:4 (2016), 483–502.
- de Swart, Peter, *Prepositional inanimates in Dutch. A paradigmatic case of Differential Object Marking*, *Linguistics* 52:2 (2014), 445–468.
- Delille, Karl Heinz, *Die geschichtliche Entwicklung des präpositionalen Akkusativs im Portugiesischen*, Bonn, Romanisches Seminar der Universität Bonn, 1970.
- Fábregas, Antonio, *Differential Object Marking in Spanish. State of the art*, *Borealis* 2:2 (2013), 1–80.
- Fernández-Ordóñez, Inés, *Dialect areas and linguistic change. Pronominal paradigms in Ibero-Romance dialects from a cross-linguistic and social typology perspective*, in: De

- Vogelaar, Gunther/Seiler, Guido (edd.), *The dialect laboratory. Dialects as a testing ground for theories of language change*, Amsterdam, John Benjamins, 2012, 73–106.
- Flores, Marcela/Melis, Chantal, *El leísmo desde la perspectiva del “mercado diferencial del objeto”*, *Revista de Historia de la Lengua Española* 2 (2007), 83–107.
- García, Erica, *The role of theory in linguistic analysis: The Spanish pronoun system*, Amsterdam, North-Holland, 1975.
- García García, Marco, *Differentielle Objektmarkierung bei unbelebten Objekten im Spanischen*, Berlin/Boston, de Gruyter, 2014.
- García García, Marco, *Nominal and verbal parameters in the diachrony of Differential Object Marking in Spanish*, in: Seržant, Ilja/Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 209–242.
- Granda, Germán de, *Origen y formación del leísmo en el español del Paraguay. Ensayo de un método*, *Revista de Filología Española* 62:3 (1982), 259–283.
- Kabatek, Johannes, *Wohin strebt die differentielle Objektmarkierung im Spanischen?*, *Romanistisches Jahrbuch* 67:1 (2016), 211–239.
- Kato, Mary, *A natureza do objeto nulo e do nome nulo no Português Europeu e no Português Brasileiro*, *Revista Intercâmbio* 23 (2011), 13–27.
- López, Luis, *Indefinite objects. Scrambling, choice functions, and differential marking*, Cambridge, Mass., MIT Press, 2012.
- Marroquim, Mario, *A língua do nordeste (Alagoas e Pernambuco)*, São Paulo, Companhia Editora Nacional, 1934.
- Melis, Chantal, *Spanish indexing DOM, topicality, and the case hierarchy*, in: Seržant, Ilja/Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 97–128.
- Montaut, Annie, *The rise of Differential Object Marking in Hindi and related languages*, in: Seržant, Ilja/Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 281–313.
- Nascentes, Antenor, *O linguajar carioca*, Rio de Janeiro, Organização Simões, 1953.
- Nascentes, Antenor, *Lheísmo no português do Brasil*, *Revista Letras* 11 (1960), 108–113.
- Nascimento, Maria Edna Porangaba do, *A sintaxe do clítico “lhe” no português brasileiro*, Maceió, PhD Thesis, Federal University of Alagoas, 2010.
- Palacios Alcaine, Azucena, *Variación sintáctica en el sistema pronominal del español paraguayo. La elisión de pronombres objeto*, *Anuario de Lingüística Hispánica* 14 (1998), 431–454.
- Pereira de Sousa, Shirley/da Silva, Sunamita Sâmelia Simplício, *Alternância de acusativo e dativo no espanhol da Espanha e no português do Brasil. Uma análise comparativa entre o leísmo e o lheísmo*, *Revista do GELNE* 16:1 (2014), 63–78.
- Raposo, Eduardo, *On the null object in European Portuguese*, in: Jaeggli, Osvaldo/Silva-Corvalán, Carmen (edd.), *Studies in Romance linguistics*, Dordrecht, Foris, 1986, 373–390.
- Schäfer-Prieß, Barbara, *O acusativo preposicional na história da língua portuguesa*, in: Head, Brian, et al. (edd.), *História da língua e história da gramática. Actas do Encontro*, Braga, Universidade do Minho, 2003, 405–417.
- Schäfer-Prieß, Barbara/Schöntag, Roger, *Spanisch/Portugiesisch kontrastiv*, Berlin/Boston, de Gruyter, 2012.

- Schwenter, Scott, *Null objects across South America*, in: Face, Timothy/Klee, Carol, *Selected proceedings of the 8<sup>th</sup> Hispanic Linguistics Symposium*, Somerville, Mass., Cascadilla, 2006, 23–36.
- Schwenter, Scott, *Two kinds of Differential Object Marking in Portuguese and Spanish*, in: Amaral, Patrícia/Carvalho, Ana Maria (edd.), *Portuguese-Spanish interfaces. Diachrony, synchrony, and contact*, Amsterdam, John Benjamins, 2014, 237–260.
- Schwenter, Scott/Silva, Gláucia, *Overt vs. null direct objects in spoken Brazilian Portuguese. A semantic/pragmatic account*, *Hispania* 85:3 (2002), 577–586.
- Schwenter, Scott/Silva, Gláucia, *Anaphoric direct objects in spoken Brazilian Portuguese. Semantics and pragmatics*, *Revista Internacional de Lingüística Iberoamericana* 1:2 (2003), 99–123.
- Seržant, Ilja, *Weak universal forces. The discriminatory function of case in Differential Object Marking systems*, in: Schmidtke-Bode, Karsten, et al. (edd.), *Explanation in typology. Diachronic sources, functional motivations and the nature of evidence*, Berlin, Language Science Press, 2019, 149–178.
- Thun, Harald, *Personalpronomina für Sachen. Ein Beitrag zur romanischen Syntax und Textlinguistik*, Tübingen, Narr, 1986.
- von Heusinger, Klaus, *The diachronic development of Differential Object Marking in Spanish ditransitive constructions*, in: Seržant, Ilja/Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 315–344.





Anna Pineda

# The development of DOM in the diachrony of Catalan

(Dis)similarities with respect to Spanish

**Abstract:** The existence of Differential Object Marking (DOM) is well-established in a number of Romance languages and varieties, such as Spanish and Romanian, where its use extends to several types of direct objects. For other languages in the Romance family, like Catalan, DOM is often considered to be absent, except for personal pronouns and a few other cases – at least from the perspective of normative grammar. However, in most varieties of Catalan, DOM applies to human direct objects generally, including proper names, definites and some indefinites, and even occasionally extends to bare plurals or inanimates. Although an exhaustive dialectal survey on the exact prevalence of DOM has yet to be carried out, it is clear that it is widespread and features in many dialects. While one might initially assume that this is the result of the influence of Spanish, such instances (at least partially) might in fact have arisen from the internal evolution of Catalan. Crucially, instances of DOM were remarkably abundant in Old Catalan, although this has sometimes gone quite unnoticed. That is, instances of DOM with proper names and human NPs are found in earlier Catalan texts (in the 13<sup>th</sup> to 15<sup>th</sup> centuries), and increase quite significantly from the 16<sup>th</sup> century on, reaching very high percentages of occurrences in some texts. The aim of this paper is to offer an account of the emergence and development of DOM in Catalan over time, showing the commonalities with neighbouring Spanish, as well as the important differences that distinguish these two languages. This is a large corpus study, based on the *Corpus Informatizat del Català Antic*, and comprising the period from the first written texts to the 16<sup>th</sup> century, with some notes on the 17<sup>th</sup> century too.

**Keywords:** Differential Object Marking (DOM), Catalan, Spanish, language contact, diachronic linguistics, corpus study

---

**Acknowledgments:** This work has been supported by a Juan de la Cierva-*Incorporación* fellowship from Spain's Ministry of Economy and Competitiveness (IJCI-2016-30474) and research project funding from the same ministry (FFI2014-56968-C4-1-P), as well as by the Alexander von Humboldt Foundation (Humboldt Research Fellowship for experienced researchers). Many thanks also to the different anonymous reviewers as well as to the participants of the *Workshop on Differential Object Marking in Spanish – Diachronic change and synchronic variation*, held at the University of Zurich (Switzerland) on 4–5 June 2018, for their helpful comments. All errors and omissions are my own.

---

**Anna Pineda**, Sorbonne Université/University of Cologne, e-mail: pinedaicirera@gmail.com

# 1 Introduction

Differential Object Marking (DOM) is a well-established phenomenon in many languages across the world, such as Hebrew, Hindi and Turkish, and also in several Romance languages and varieties, such as Spanish, Sardinian and Romanian. It consists of the introduction of direct objects by using a specific, differential marker. This marker has often been identified with a preposition (hence the alternative name “prepositional accusative”), and in the case of Romance languages, among many others, the marking often coincides with that for indirect objects, that is to say, dative marking. DOM as a phenomenon continues to arouse interest in linguistics, both in more traditional descriptive accounts and in more theoretical ones (cf. e.g. Rohlfs 1971; 1973; Bossong 1991; 1998; Pensado 1995; Torrego 1998; Aissen 2003; Laca 2006; Leonetti 2008; Iemmolo 2010; López 2012).

The emergence of DOM has been attributed to many causes and functions. It is nevertheless interesting to highlight a specific function here; namely, its status as a mechanism that distinguishes highly particularized participants, when these are objects, which could otherwise be identified with the subject function. In other words, DOM marks objects that have properties typical of subjects. Thus, the appearance of the marking in a direct object is conditioned by the semantic properties of the affected element. In fact, animacy and definiteness of a direct object are properties which regulate the emergence and expansion of DOM (cf. Silverstein 1976; Dixon 1979). The two hierarchies proposed in the literature are set out below. Categories ranked higher are more likely to show DOM than those ranked lower.

(1) Animacy scale:

HUMAN > ANIMATE > INANIMATE

(2) Definiteness/Specificity scale:

PRONOUNS > PROPER NAMES > DEFINITE NPs > SPECIFIC INDEFINITE NPs >  
NON-SPECIFIC INDEFINITE NPs (> BARE NPs)

In the Catalan prescriptive tradition, the presence of DOM is restricted to the elements located at the top of the hierarchies in (1)–(2); these are mainly personal pronouns, plus a few other cases, which we will describe in detail in Section 2.

However, as already noted by Sancho Cremades (2002, 1737) concerning this aspect of the grammar, there is “a clear divergence between the spoken Catalan and the written Catalan which follows the normative grammar. In the language spoken throughout the linguistic domain there is a wide usage of the preposi-

tion *a* (and the forms *an*, *ana*, *amb*) in front of definite animate direct objects, as occurs in Spanish: *No he vist {a la Maria/als nois}* ‘I have not seen Maria/the boys’” [our translation]. In fact, DOM is very frequent today in most Catalan varieties (not only in colloquial registers), both in the spoken and written language – even in the press and institutional communications, this being visible in the digital era when many texts are often made public before a proof-reader has had the chance to revise them. Before exploring this further, let us see how the description and prescription of the Catalan language has evolved over the last century with respect to this aspect of the grammar.

The paper is structured as follows. In Section 2 we describe how prescriptive grammar has dealt with DOM, from Fabra (1918) until the appearance of the new Catalan normative grammar in 2016. Section 3 focusses on the controversial discussion regarding the nature of DOM, which may be seen either as the result of the influence of Spanish, or (at least in part) as the product of the internal evolution of Catalan. Evidence for the latter claim is provided in Section 4, where detailed diachronic data are presented, and a comparison with the evolution of DOM in Spanish is made. Section 5 briefly summarizes the main conclusions of the paper.

## 2 DOM in normative Catalan

In 2016 the Catalan Academy (Institut d’Estudis Catalans) published the new Catalan prescriptive grammar (GIEC 2016). Until then, Standard Catalan was regulated by Fabra’s (1918) *Gramàtica catalana* – Pompeu Fabra was the leading figure in the standardization of Catalan and the establishing of a prescriptive grammar. From Fabra’s work to the GIEC, the range of contexts where DOM is allowed has been progressively widened. In what follows, we provide a description of the relevant contexts of DOM usage in order to show how prescriptivism has evolved here.

### 2.1 Pronominal direct objects

In Standard Catalan, DOM is obligatory with strong personal pronouns, which must also be clitic doubled<sup>1</sup> (Fabra 1918, §112.I; GIEC 2016, §19.3.2.1, 19):

---

<sup>1</sup> In Catalan, clitic doubling of 1<sup>st</sup> and 2<sup>nd</sup> person pronouns is obligatory, whereas for 3<sup>rd</sup> person pronouns the situation is slightly different. In some dialects, such as Valencian, doubling of 3<sup>rd</sup>

- (3) a. *T' he vist \*(a) tu.*  
 CL.2SG.ACC have.1SG seen DOM you  
 'I have seen you.'
- b. *El Joan l' estima \*(a) ella.*  
 the Joan 3SG.F.ACC love.PRS.3SG DOM she  
 'Joan loves her.'

In addition, Fabra (1918, §112.III) also considered DOM to be allowed in front of certain “pronouns”; the examples he gave include the universal *tothom* ‘everybody’ (4) and [+human] *tots* ‘all’, as well as the relative [+human] *qual* (5):

- (4) *He salutad (a) tothom.*  
 have.1SG greeted DOM everybody  
 'I have greeted everybody.'
- (5) *La doctora Pons, (a) la qual tan admires,*  
 the doctor.F Pons, DOM the which so.much admire.PRS.2SG  
*és aquí.*  
 be.PRS.3SG here  
 'Doctor Pons, whom you admire so much, is here.'

Several other grammarians have enlarged the list of “pronouns” with which DOM was to be allowed in Standard Catalan, reaching a general consensus. In particular, the relative and interrogative *qui* ‘who’ (6) was added by Ruaix (1985, 174), *ningú* (7) was added by Wheeler et al. (1999, §14.1.1.1), *algú* was suggested by Solà (1994, 177) and *altri* was pointed out by Mestres et al (1995). Note that the cases of *qui* and *ningú* had already been observed by Fabra, not in his prescriptive grammar (1918), but in the one published in 1912 (Fabra 1912, §122), where he established that Catalan admitted the use of DOM for these cases.

- (6) a. *Trobaran (a) qui ho ha fet.*  
 find.FUT.3PL DOM who it have.3SG did  
 'They will find who did it.'

---

person pronouns is obligatory, whereas in other dialects it is not, under certain pragmatic conditions, as illustrated in the example below (Todolí 2002, §6.5.5.3a):

- (i) *També (l') avisaran a ell.*  
 also CL.ACC.M.3SG notify.FUT.3PL DOM he  
 'They will notify him too.'

b. (A) *qui has vist?*  
 DOM who have.2SG seen  
 ‘Who have you seen?’

(7) *No han detingut (a) ningú.*  
 No have.3PL arrested DOM nobody  
 ‘They have not arrested anybody.’

The range of pronominal direct objects allowing the use of DOM was enlarged by the new Catalan prescriptive grammar (GIEC 2016, §19.3.2.1, 19.3.2.2). For example, reference was made to direct objects corresponding to [+human] pronominal quantifiers *cada u* ‘every one’, *cadascú* ‘everyone’, *qualsevol* ‘any one’ (8), as well as the interrogatives *quin/-a/-s/-es* ‘which ones’ (9) and *quants* ‘how many’ (10):

(8) *Aquí premien (a) qualsevol.*  
 here reward.PRS.3PL DOM anyone  
 ‘Here they reward anyone.’

(9) (A) *quina dona has vist?*  
 DOM which.F woman have.2SG seen  
 ‘Which woman have you seen?’

(10) (A) *quants alumnes premiaran?*  
 DOM how.many.PL students reward.FUT.3PL  
 ‘How many students will they reward?’

Finally, note that when a pronominal direct object is coordinated with a non-pronominal one, as in (11), there is also the need for DOM (as a parallelism effect), as pointed out by Solà (1994, 166–167), this once more becoming the consensus view.

(11) *T’ he vist a tu i al metge.*  
 CL.ACC.2SG have.1SG seen DOM you and DOM.the doctor  
 ‘I’ve seen you and the doctor.’

## 2.2 Potentially ambiguous direct objects and other stylistically marked contexts

In structures where both the subject and the object appear in post-verbal position, DOM emerges as a useful mechanism to distinguish the two syntactic functions at play. This is the case for reciprocals (12) and comparative structures with the verb elided (13), two contexts for which Fabra (1918, §112.II) noted the need for DOM.

- (12) *S'abraçaven l' un \*(a) l' altre.*  
 REFL-hug.IPFV.3PL the one DOM the other  
 'They were hugging each other.'

- (13) *El perseguia com el gat \*(a) la rata.*  
 CL.ACC.3SG.M chase.IPFV.3SG like the cat DOM the rat  
 'He chased him like the cat chases the rat.'

This distinguishing function of DOM also becomes clear in several other contexts of potential ambiguity, which were already alluded to by Fabra (1912, §122), who claimed that the use of DOM seemed "perfectly admissible" in order to avoid any confusion. To exemplify the contexts of potential ambiguity, he in fact used a sentence with the interrogative *qui* 'who', which also happens to be a pronominal direct object (recall §2.1):

- (14) a. *Qui ha vist en Miquel?*  
 who has seen the Miquel?  
 Ambiguous: 'Who has Miquel seen?'/ 'Who has seen Miquel?'
- b. *A qui ha vist en Miquel?*  
 DOM who has seen the Miquel?  
 Non-ambiguous: 'Who has Miquel seen?'

In examples like these, DOM resolves the potential ambiguity between the object and subject interpretation.<sup>2</sup> Over the years some grammarians, especially Solà

---

<sup>2</sup> Fabra observed many other contexts of this type. Interestingly, as Xavier Rofes (p.c.) points out, in the earliest versions of his *Gramàtica catalana*, starting with the 1918 edition, Fabra devoted a whole page (which would disappear from the editions published in the 1930s) to offer a variety of resources, such as the use of the comma, changes of word order, and recourse the passive, to solve a number of contexts of potential ambiguity (*La fonètica precedeix la morfologia* 'Phonetics precedes morphology'/'Morphology precedes phonetics', *Qui ha vist en Joan?* 'Who has seen Joan?'/ 'Who has Joan seen?', *Demà tornarà l'estranger que visità en Joan* 'Tomorrow

(1994), have pointed out several other contexts where the absence of DOM would lead to a subject/object ambiguity, leading to the consensus that in such contexts DOM is recommended:

- (15) a. *El Barça serà campió si venç al*  
 the Barça be.FUT.3SG champion if beat.PRS.3SG DOM.the  
*Madrid.*  
 Madrid  
 ‘Barça will be the champion if they beat Madrid.’ (Without DOM, another interpretation would be possible: ‘Barça will be the champion if Madrid beats some other team.’)
- b. *Com penseu que afectarà el món d’ Internet*  
 how think.PRS.2PL that affect.FUT.3SG the world of Internet  
*a l’ economia mundial?*  
 DOM the economy world  
 ‘How do you think that the world of the Internet will affect the world economy?’ (Without DOM, another interpretation would be possible: ‘How do you think that the world economy will affect the world of the Internet?’)
- c. *L’ estimo com/més que a la meva mare.*  
 CL.ACC.3SG love.PRS.1SG like/more than DOM the my mother.  
 ‘I love him/her like/more than (I love) my mother.’  
 (Without DOM, another interpretation would be possible: ‘I love him/her like/more than my mother loves him/her.’)

The GIEC (2016, §19.3.2.4b, c) also offers a fine-grained description of the contexts of potential ambiguity where DOM should be used, with examples such as (16), and cases of reordering for stylistic reasons where, again, DOM is useful (17). The GIEC (2016, §19.3.2.4d) also establishes that the preposition is usually employed to mark the animate direct object that appears postponed to an argumental prepositional complement (especially if the latter has a complex structure), as in examples (18):

---

the foreigner whom Joan visited will be back’/‘Tomorrow the foreigner who visited Joan will be back’) with resources other than DOM. We can deduce from this that, if providing alternatives to DOM was seen as necessary, then DOM was indeed used by speakers.

- (16) a. *Diuen que rellevarà a l' alcalde una*  
 say.PRS.3PL that substitute.FUT.3SG DOM the mayor a  
*regidora.*  
 councillor  
 'They say that a councillor will substitute the mayor.' (Without DOM, another interpretation would be possible: 'They say that the mayor will substitute a councillor.')
- b. *És preocupant veure com ha enfonsat a l'*  
 is alarming see.INF how has sunk DOM the  
*acusat la teva declaració.*  
 accused your testimony  
 'It is alarming to see how your testimony has sunk the accused.'  
 (Without DOM, another interpretation would be possible: 'It is alarming to see how the accused has sunk your testimony.')
- c. *Aquest és l' objectiu que defineix a la*  
 this is the goal that defines DOM the  
*biolingüística.*  
 biolinguistics  
 'This is the goal that defines biolinguistics.' (Without DOM, another interpretation would be possible: 'This is the goal that biolinguistics defines.')
- (17) *A l' apatia habitual vencé l' esperança del triomf.*  
 DOM the apathy usual beat.PST.3SG the hope of.the victory  
 'The usual apathy was beaten by the hope of victory.'
- (18) a. *No va poder convèncer d' estudiar literatura*  
 no can.PST.3SG convince.INF of study.INF literature  
*medieval a les filles.*  
 medieval DOM the daughters  
 'He could not convince his daughters of studying medieval literature.'
- b. *Ha convidat a passar un cap de setmana*  
 has invited to spend.INF a end of week  
*en un hotel de muntanya a la teva*  
*germana.*  
 in a hotel of mountain DOM the your sister  
 'He has invited your sister to spend a weekend in a mountain hotel.'



## 2.3 Dislocated direct objects

One last context that was added by Solà (1994, 167) to the list requiring, or at least allowing, DOM was left-/right-dislocation (19) (cf. also Escandell Vidal 2007a; 2007b; 2009), again reaching consensus among grammarians (Sancho Cremades 2002, 1738, fn. 68).<sup>3</sup> The use of DOM in such contexts is also described in the GIEC (2016, §19.3.2.4a).

- (19) a. (A) *la meva mare, la visitaré demà.*  
 DOM the my mother CL.ACC.3SG.F visit.FUT.1SG tomorrow  
 ‘My mother, I will visit her tomorrow.’  
 b. *La visitaré demà, (a) la meva mare.*  
 CL.ACC.3SG.F visit.FUT.1SG tomorrow DOM the my mother  
 ‘My mother, I will visit her tomorrow.’

## 2.4 Summary

The Catalan prescriptive tradition, then, has gradually widened the range of contexts in which DOM is allowed or recommended. Table 1 summarizes the contexts for this:

However, as shown in Table 1, the normative grammar proscribes the use of DOM in most of its possible contexts, such as nominal phrases with human referents (*He vist a la teva cosina* ‘I have seen DOM your cousin’) or proper names (*He vist a la Maria* ‘I have seen DOM Maria’), which are not dislocated nor subject to any ambiguity between a subject and an object interpretation. In this regard,

---

<sup>3</sup> In Balearic Catalan, inanimate dislocated NPs can also show DOM:

- (i) a. *Colliu-les, a les peres, que ja són madures.*  
 pick.imp-CL.ACC.3PL.F DOM the pears, that already be.PRS.3PL ripe  
 ‘Pick the pears, as they are ripe.’ (Moll 1975, §239)  
 b. *Ja ho sé, [ən] això.*  
 already it know.prs.1sg dom this  
 ‘This I know.’ (Rosselló 2002, 1932, fn. 77)

These instances, which will not be addressed in this paper, are analyzed in detail by Escandell Vidal (2007a; 2007b; 2009), who offers an account where topicality is a key factor. Also, note that in example (ib) the preposition *a* appears in its phonetic transcription, reflecting one of the allomorphic variants that a shows across Catalan dialects, including also [ə]n[ə], [ə]m, [ə]m[ə], [ə]mb, [ə]mb[ə]. Cf. Albareda (2009) for a complete description of such variants.

**Table 1:** DOM in Standard Catalan (GIEC 2016).

	DOM in Standard Catalan
strong pronouns	compulsory
reciprocal structures	compulsory
comparative structures (elided verb)	compulsory
other pronouns	allowed
relatives, interrogatives	allowed
proper names	not allowed (except if dislocated or potential ambiguity)
human NPs	not allowed (except if dislocated or potential ambiguity)
animate NPs	not allowed (except if dislocated or potential ambiguity)
inanimate NPs	not allowed (except if potential ambiguity)

prescription thus departs from what occurs in formal and informal registers of spoken Catalan, and even to a large extent in written Catalan before a proof-reader has intervened. We will explore this in the next Section.

### 3 DOM in Catalan: evolution or interference? From Old Catalan to the 21<sup>st</sup> century

Despite the considerable gap that exists between prescriptivism and language use, it is useful to recall that Fabra was aware of the “real” scope of DOM in Catalan, as indirectly shown by some passages of his work.<sup>4</sup> Nowadays, like then,

<sup>4</sup> Evidence is found, for example, in the following passage extracted from his prescriptive grammar (Fabra 1918), where he claimed:

It is very useful to know how to distinguish an indirect object from a direct one, [...]. Thus, in *Hem escrit a la Maria* ‘We have written to Maria’, if we replace the complement *a la Maria* with a pronoun, this will be *li* ‘to her’ (*Li hem escrit* ‘We have written to her’); therefore the mentioned complement is an indirect object (and *Hem escrit a la Maria* ‘We have written to Maria’ is correct). On the other hand, in *Hem vist a la Maria* ‘We have seen DOM Maria’, if we replace the complement *a la Maria* ‘DOM Maria’ with a pronoun, this will be *l* ‘her’ (*L’hem vista* ‘We have seen her’); therefore the mentioned complement is a direct object (and it is thus appropriate to omit the preposition: *Hem vist la Maria*). (Fabra 1918, §113) [our translation]

the perception that any attentive speaker will have of the extent of DOM use is that the phenomenon keeps on spreading. We will now offer some examples of the phenomenon that we have collected recently. A clarification is in order: whereas there is an urgent need for a large-scale survey of the use of DOM across Catalan dialects and registers,<sup>5</sup> it is also true that any preliminary observation of

---

Fabra explained how to distinguish a direct object from an indirect one precisely because for many speakers both syntactic functions could formally coincide, that is to say, could be headed by the preposition *a*. In other words, many speakers felt – and still feel – that a sequence such as *Hem vist a la Maria* ‘We have seen DOM Maria’ is completely natural. This is why Fabra wanted to introduce the resource of pronominal substitution in order to detect the uses of the preposition which were not accepted in the prescriptive grammar. In other works, Fabra was very clear about the quite well-extended use of DOM that existed in Catalan. His grammar published in 1912 is a good example:

[...] in Catalan one says indistinctly *Hem vist la casa de ta germana* ‘We have seen your sister’s house’ and *Hem vist ta germana* ‘We have seen your sister’: the direct object is introduced in both cases without the preposition *a* [...] there has recently been a preference for the construction with *a*, up to the point that many writers consider the use of this preposition to be as well-extended as it is in Spanish. [...] thus, one often hears *Coneixes a la meva filla?* ‘Do you know DOM my daughter?’, but one also hears with the same frequency *Coneixes la meva filla?* ‘Do you know my daughter?’ (Fabra 1912, §122) [our translation]

From this passage, and in particular from expressions such as ‘one says indistinctly’ or ‘with the same frequency’, it is clear that the scope DOM had at that time was worthy of consideration.

In fact, Fabra was well aware that Catalan DOM was not to be seen as merely an influence from Spanish, but rather as a phenomenon that, although far more extended in Spanish, is present in very different (Romance) languages. While he did not express that view in his prescriptive grammar, as we saw in Section 2, he did make it clear at the First International Conference of the Catalan Language (Primer Congrés Internacional de la Llengua Catalana), in 1906, when he offered an amendment to the presentation given by another grammarian, Costa i Llobera, and claimed:

The preposition *a* has to be used in front of strong personal pronouns [...] and in some other cases. Its use in the accusative cannot be considered a ‘castilianism’, but rather Castilian [= Spanish] has seen a greater extension of its use than the other languages, since there are also uses of the *a* in Italian dialects such as Sicilian, Calabrese and Roman; the phenomenon is also found in the Engadine and in the Romanian language spoken in the mouth of the Danube (Fabra 2010, 877) [our translation]

5 For the particular case of Valencian Catalan, Sancho Cremades (1995, 199) claims that DOM in this variety shows up with the same pattern of distribution that it has in Spanish, and he gives the examples with proper names and definite NPs:

- (i) No veig            {a    Cast/als            xiquets}    des de    fa            sis anys.  
 not see.PRS.1SG {DOM    Cast/DOM-the    children}    from    ago            six years  
 ‘I haven’t seen Cast/the children for six years.’

the daily use of the Catalan language reveals that the scope of the phenomenon goes well beyond the prescriptivist norm, as shown in Table 2 below. Thus, in what follows, examples are provided from the media as well as other formal or institutional contexts, both spoken and written.

Our data show that across most varieties DOM is spreading (as a possibility) to human/animate direct objects, whether they are proper names (20) or definite NPs (21), a fact already pointed out by Sancho Cremades (2002, 1737), Aissen (2003, 451) and also Hualde (1992, 86–87; 237–241), who describes Catalan as a language with DOM extending to all human definites and claims that “[f]or most speakers, in spoken language all human direct objects are marked by the preposition *a*”. Interestingly, Naess (2004, 1188) takes Hualde’s data and classifies Catalan as a language with DOM depending on animacy/humanness.

- (20) *Hem vist a Puigdemont ferm i animat.*  
 have.1PL seen DOM Puigdemont firm and cheerful  
 ‘We have seen Puigdemont firm and cheerful.’  
 (*Diari de Tarragona*, 07/04/2018)

- (21) a. *Han detingut al president.*  
 have.3PL arrested DOM.the president  
 ‘They have arrested the president.’  
 (*El Nacional.cat*, 25/03/2018)

- b. *Estem parlant de protegir a les dones.*  
 be.PRS.1PL talking of protect.INF DOM the women  
 ‘We are talking about protecting women.’  
 (Central Catalan speaker, spoken in a formal context, Barcelona,  
 15/04/2019)

Importantly, it is also possible to find DOM with elements located lower down the animacy and definiteness scales provided in (1)–(2), such as with indefinites, not only specific ones (22), but also non-specific ones (23), and even with bare plurals (24), this contra Hualde (1992, 241), who claims that indefinites never take DOM.

- (22) a. *Després he vist a un grup d’estudiants que  
 then have.1SG seen DOM a group of students that  
 estaven cridant.*  
 be.IPFV.3PL yelling  
 ‘Then I’ve seen a group of students who were yelling.’  
 (*Diari de Girona* [Central Catalan], 24/03/2017)

- b. *Han trobat a una persona amb una ferida a la cara.*  
 have.3PL found DOM a person with a wound in the face  
 ‘They found someone with a wound in their face.’  
 (*Diari Més, Tarragona-Reus-Costa Daurada* [Central Catalan], 03/05/2018)
- (23) a. *Dies abans cinc homes s’ havien preparat per agredir a alguna dona en grup.*  
 days before five men REFL had.3PL prepared for attack.INF DOM some woman in group  
 ‘Some days before five men had prepared to attack some woman in a group.’  
 (*Elcritic.cat* [Central Catalan], 27/04/2018)
- b. *Tu mataries a un home per la vida d’ un altre?*  
 you kill.COND.2SG DOM a man for the life of an other  
 ‘Would you kill a man for another man’s life?’  
 (“El meu lament”, song by Ferran Palau [Central Catalan speaker], album *Santa ferida*, 2015)
- c. *Coneixeu a algun llicenciat en Filologia Catalana?*  
 know.PRS.2PL DOM some graduate in Filology Catalan?  
 ‘Do you know a graduate in Catalan philology?’  
 (Job offer via WhatsApp, within a group of Catalan philologists from different dialects)
- (24) a. *Els últims mesos hem vist que l’ Estat acusava a professors de delictes d’ odi.*  
 the last months have.1PL seen that the State accuse.IPFV.3SG DOM teachers of crimes of hate  
 ‘These past few months we have seen that the [Spanish] State accused teachers of hate crimes.’  
 (Central Catalan speaker, spoken in a formal context, Barcelona, 16/05/2018)

- b. *Vam iniciar uns cursos per formar a mestres en  
start.PST.1PL some courses to form DOM teachers in  
gramàtica i llengua de signes.  
grammar and language of signs*  
'We started some courses to train teachers in grammar and sign  
languages.'

(Central Catalan speaker, spoken in a formal context, Barcelona,  
7/11/2018)

Some speakers also find natural examples of DOM with direct objects whose referent corresponds to social institutions, such as political parties, or abstract names that can be interpreted as referring metonymically to humans (25), including names of places which probably allow for a collective reading (26). Moreover, one can even find examples with clearly inanimate direct objects, both definite and indefinite (and even bare) (27).

- (25) a. *Ciudadans serveix per parar a Podem.<sup>7</sup>  
Ciudadans serve.PRS.3SG to stop DOM Podem*  
'Ciudadans serve to stop Podem.'

(Central Catalan speaker, RAC1 radio station, spoken in a formal  
context, 18/05/2018)

- b. *Un partit que uneix a diferents sensibilitats.  
a party that unite.PRS.3SG DOM different sensitivities*  
'A party which unites different sensitivities.'

(Central Catalan speaker, RAC1 radio station, spoken in a formal  
context, 26/10/2018)

- (26) *Aquestes primàries, aquest gruix és el  
these primaries, this substantial.amount be.PRS.3SG the  
que pot fer a Barcelona gran.  
what can.3SG make DOM Barcelona great*  
'These primary elections, this substantial amount [of people voting] is what  
can make Barcelona great.'

(Central Catalan speaker, RAC1 radio station, spoken in a formal context,  
08/05/2018)

---

6 Ciudadans and Podem are political parties.

- (27) a. *Els determinants acompanyen als noms.*  
 the determiners accompany.PRS.3PL DOM.the nouns  
 ‘Determiners accompany nouns.’
- b. *El morfema flexiu de grau -íssim només és gramatical si flexiona a un adjectiu.*  
 the morpheme inflectional of degree -íssim only is grammatical if inflect.PRS.3PL DOM an adjective  
 ‘The inflectional degree morpheme -íssim is only grammatical if it inflects an adjective.’
- c. *Un morfema flexiu de grau pot modificar a un adjectiu, i l’ exclamatiu quin pot modificar a adjectius amb funció de nom.*  
 a morpheme inflectional of degree can.3SG modify.INF DOM an adjective, and the exclamative quin can modify.INF DOM adjectives with function of noun  
 ‘An inflectional degree morpheme can modify an adjective and the exclamative *quin* can modify adjectives which function as nouns.’  
 (Academic texts written by Central Catalan speakers who were fourth year Applied Languages students, Universitat Pompeu Fabra 2018)

In short, we observe that DOM in Catalan is spreading, at least for some speakers, from the elements located at the top of the animacy and definiteness hierarchies to the elements located lower down, and in some cases the phenomenon is even attested at the bottom of the hierarchies. Thus, even if there is still no exhaustive dialectal study of this particular syntactic phenomenon, one can observe, as Sancho Cremades (2002, 1737) pointed out, that there is a “a clear divergence” between what (a notable number of, if not most) speakers do and what the prescriptive grammar establishes. It is easy to attribute such a disparity to the view that DOM is the result of interference from Spanish, and therefore a phenomenon that one should avoid to ensure a more genuine Catalan, especially given the sociolinguistic context of Catalan being historically a minoritized language threatened by Spanish.

However, an accurate analysis of the diachrony of Catalan demonstrates that DOM, a grammatical device banned in most contexts by prescriptive grammars (recall Section 2), may in fact constitute a genuine feature of the language. Indeed, a notable use of DOM is attested in old times, before the sociolinguistic pressure from Spanish began in the 16<sup>th</sup> century, due to the dynastic union with Castile, the subsequent move of the court and the rise of Spanish political power. As Salvador/Pérez Saldanya (1993) stated, one should take into account that:

'If Catalan, before experiencing sociolinguistic pressure from Spanish, had evolved towards certain grammatical solutions which were born during the time of its configuration as a language, then considering the result [of DOM, our note] as a castilianism turns out to be problematic [...] some recent studies [...] suggest caution with regard to the position of normative grammar.' (Salvador/Pérez Saldanya 1993, 60) [our translation]

In sum, while Catalan normative grammar prescribes DOM to be relatively restricted in formal usages, in “colloquial” varieties of the language, DOM is used more widely. We use quotation marks for the term “colloquial” because this wider usage of DOM is not truly restricted to colloquial linguistic contexts (e.g. conversations with friends and family) but it is also very frequently found in all kinds of formal discourses (e.g. political, journalistic) or more broadly in the press (radio, TV, and newspapers, especially those published online, for which a proof-reader has normally not revised the texts). Thus, any observation of contemporary Catalan reveals that there is a much wider set of contexts where DOM is or can actually be used by most speakers, which clearly contrasts with what is prescribed for Standard Catalan. This is shown in Table 2.

**Table 2:** DOM in Standard Catalan (GIEC 2016) vs. “Colloquial” Catalan.

	DOM in Standard Catalan	DOM in “Colloquial” Catalan
STRONG PRONOUNS	compulsory	compulsory
RECIPROCAL STRUCTURES	compulsory	compulsory
COMPARATIVE STRUCTURES (elided verb)	compulsory	compulsory
OTHER PRONOUNS	allowed	attested
RELATIVES, INTERROGATIVES	allowed	attested
PROPER NAMES	not allowed (except if dislocated or potential ambiguity)	attested
HUMAN NPS	not allowed (except if dislocated or potential ambiguity)	attested
ANIMATE NPS	not allowed (except if dislocated or potential ambiguity)	attested
INANIMATE NPS	not allowed (except if potential ambiguity)	very infrequent, but possible



## 4 The diachrony of DOM in Catalan and Spanish: (dis)similarities

Bearing in mind the extension of DOM in the Romance area, it seems logical not to disregard the consideration that the current scope of this phenomenon in Catalan could (at least partially) result from the internal evolution of the language. Several studies, most of which focus on the 15<sup>th</sup> century, have shown that DOM was relatively abundant in Old Catalan, not only with personal pronouns (Meier 1947; 1948; Perera 1986; Adell 1994; Cabanes 1994; 1995). In the context of the project for a *Gramàtica del català antic*, Pineda (in press a; in press b) offers an exhaustive study of the emergence and spreading of DOM in the diachrony of Catalan, from the earliest texts (11<sup>th</sup>–13<sup>th</sup> centuries) to the end of the Old Catalan period, in the 16<sup>th</sup> century. The study is based on a corpus comprising 900,000 words, which is representative of diachronic, diatopic, diaphasic and diastratic variation of the language.<sup>7</sup> The conclusions of Pineda (in press a; in press b) indicate that DOM appeared with proper names (both person and deity names) and human definite NPs, in addition to personal pronouns. We summarize the results in Table 3, which also includes preliminary data from the 17<sup>th</sup> century, the initial period of Modern Catalan (cf. Pineda 2018; 2019):

**Table 3:** Overview of the evolution of DOM in Old Catalan (11<sup>th</sup>–16<sup>th</sup> centuries), and at the beginning of Modern Catalan (17<sup>th</sup> century).

Century	PERSONAL PRONOUNS	PERSONAL NAMES	DEITY NAMES	HUMAN DEFINITES	HUMAN INDEFINITES
11 <sup>th</sup>	0% (0/1)	(no occurrences)	(no occurrences)	0% (0/6)	(no occurrences)
13 <sup>th</sup>	17.9% (19/106)	16.2% (23/142)	0% (0/135)	2.4% (10/417)	1% (1/96)
14 <sup>th</sup>	34.5% (19/55)	1.9% (5/262)	0% (0/29)	2.8% (16/567)	2.7% (4/147)
15 <sup>th</sup>	61.1% (36/58)	12.8% (23/180)	21.8% (12/55)	9.7% (60/617)	7.3% (14/192)
16 <sup>th</sup>	97.3% (36/37)	64.4% (100/153)	92.6% (25/27)	26.8% (125/466)	11.8% (18/153)
17 <sup>th</sup>	100% (2/2)	90.6% (77/85)	93.5% (43/46)	69.9% (76/92)	22.4% (28/80)

<sup>7</sup> Data extracted from the *Corpus Informatitzat del Català Antic* (CICA), [www.cica.cat](http://www.cica.cat).

One can immediately see that DOM was already present in the language before interference from Spanish began in the early 16<sup>th</sup> century. From then on, the frequency of DOM with different types of direct objects increases rapidly, reaching very high percentages in the 17<sup>th</sup> century, when political and linguistic subordination was immense. Before drawing any conclusions from these observations, let us show in detail how DOM evolves within each class of direct objects, and see the similarities and differences that Catalan presents with respect to Spanish. We ought to say that, due to space restrictions, the data are presented here in a streamlined way, but those readers seeking for a much more exhaustive and detailed account are referred to Pineda (in press a; in press b).

In order to organize the presentation of the data in this Section, we will combine the two scales in (1)–(2). Starting with personal pronouns, located at the higher positions of the animacy and definiteness scales, our data show that DOM is not systematic until the 16<sup>th</sup> century. Thus, examples with and without DOM, as in (29) and (28) respectively, even within the same text, coexist for a long time. On the other hand, as shown in Table 3, DOM becomes absolutely systematic in the 16<sup>th</sup> century. At that time, there is only one instance without the marking attested in our corpus (30):

- (28) a. *pare, no vules mi pregar d' aquesta cosa*  
 father no want.IMP.2SG me pray.INF of this thing  
 'Father, don't you want to pray me to do that.' (13<sup>th</sup> c., *Vides*, 37)
- b. *que tu ams tu matex*  
 that you love.SBJV.PRS.2SG you self  
 'That you love yourself.' (13<sup>th</sup> c., Llull, *Doctrina* I, 95)
- c. *féu tornar mi*  
 do.PST.3SG return me  
 'He made me return.' (14<sup>th</sup> c., *Clams* II, 221)
- (29) a. *lo fil de na Godoi qui venc contra ela e a la porta sua ab espasa treita e*  
*requerí a ela*  
 requested DOM her  
 'Godoi's son, with his sword out, came to her, at her door, and requested her.'  
 (13<sup>th</sup> c., *Clams* I, 72)

- b. *aquei dit Matheu a él pagar no*  
 that mentioned Matheu DOM him pay.inf no  
*volia ni pagat no l' avia*  
 want.IPFV.3SG nor paid no CL.ACC.M.3SG had  
 'That guy, Matheu, didn't want to pay him and didn't pay him.'  
 (13<sup>th</sup> c., *Clams* I, 83)
- c. *enassí con tu ab dejunis as trebalat a tu,*  
 just like you with fastings has worked DOM you  
*axí és mester qu' él se trebal*  
 in.this.way is necessary that he REFL work.SBJV.PRS.3SG  
 'Just like you have worked yourself with fastings, he needs to work himself.'  
 (13<sup>th</sup> c., *Vides*, 29)

- (30) *ý encara pensava no trobaria nosaltres*  
 and still think.IPFV.3SG no find.COND.3SG us  
 'And she even thought that she would not find us.'  
 (16<sup>th</sup> c., Liori i Requesens, *Epistolaris*, letter 13, written in  
 Molins de Rei, Catalonia)

This situation clearly differs from Spanish where, according to several studies, personal pronouns take DOM systematically, without exceptions, from the earliest texts in the 12<sup>th</sup> century (Pensado 1995, 19; Company 2002, 207–208; von Heusinger/Kaiser 2005, 35–36; 41; Laca 2006, 426; 469).

Regarding personal pronouns, it is worth clarifying that reflexive pronouns (*si*, *si mateix* 'oneself') behave differently with respect to the spread of DOM, disallowing DOM until the 14<sup>th</sup> century (31). As Table 4 shows, they only started to be marked in the 15<sup>th</sup> century (32), and have subsequently continued as such:

**Table 4:** Evolution of DOM with reflexives in Old Catalan.

Century	REFLEXIVE PERSONAL PRONOUNS
11 <sup>th</sup>	0% (0/1)
13 <sup>th</sup>	0% (0/18)
14 <sup>th</sup>	0% (0/35)
15 <sup>th</sup>	54.5% (6/11)
16 <sup>th</sup>	(no occurrences)

- (31) a. *que ·l pagès, meynspresan sí, amàs més*  
 that the peasant underestimating himself love.SBJV.PST more  
 ‘That the peasant, underestimating himself, would love more.’  
 (14<sup>th</sup> c., *Diàlegs*, f. 8r)
- b. *volén més si ociure que venir viu en*  
 wanting more himself kill.INF than come.INF alive in  
*mans de sos enemichs*  
 hands of his enemies  
 ‘Preferring to kill himself than come alive in his enemies’ hands.’  
 (14<sup>th</sup> c., *Agramont, Regiment*, 63a)
- (32) *aquests hòmens de leys fan richs a*  
 these men of laws make.PRS.3PL rich.PL DOM  
*sí mateyx e destroexen tota Anglaterra*  
 themselves and destroy.PRS.3PL all.F.SG England  
 ‘These men of the law are enriching themselves and destroying all  
 England.’  
 (15<sup>th</sup> c., *Martorell, Tirant*, 197)
- Turning to human proper names, i.e. personal names, Table 3 shows that these could take DOM as early as in the 13<sup>th</sup> century, as shown in examples (33):
- (33) a. *Maria Domingo près a na Joanna, [...] ab*  
 Maria Domingo take.PST.3SG DOM the Joanna with  
*Abdues les mans als cabels*  
 both the hands in.the hair  
*Lit.* ‘Maria Domingo<sub>i</sub> grasped Joanna<sub>i</sub> by her hair<sub>i</sub> [...] with her own hands<sub>i</sub>.’  
 ‘Maria Domingo<sub>i</sub> grasped Joanna’s [...] hair with her own hands<sub>i</sub>.’  
 (13<sup>th</sup> c., *Clams I*, 46)
- b. *Aquel encalsava ab lo coltel treit al dit Apariçi*  
 that chase.IPFV.3SG with the knife out DOM.the mentioned Apariçi  
 ‘That one was chasing the mentioned Apariçi while holding a knife.’  
 (13<sup>th</sup> c., *Clams I*, 50)
- c. *[ells] tenien en terra a Michel Mercer*  
 they have.IPFV.3PL in ground DOM Michel Mercer  
 ‘They kept/had Michel Mercer on the ground.’  
 (13<sup>th</sup> c., *Clams I*, 76)

Interestingly, geolectal differences can be observed at this time, with Valencian texts offering the highest number of occurrences of DOM with proper names, as seen in Table 5.

**Table 5:** Evolution of DOM with human proper names in Old Catalan.

Century	PERSONAL NAMES WITH DOM (ALL DIALECTS)	PERSONAL NAMES WITH DOM (VALENCIAN CATALAN)
11 <sup>th</sup>	(no occurrences)	(no occurrences)
13 <sup>th</sup>	16.2% (23/142)	19.3% (17/88)
14 <sup>th</sup>	1.9% (5/262)	4.1% (4/98)
15 <sup>th</sup>	12.8% (23/180)	9.5% (2/21)
16 <sup>th</sup>	65.4% (100/153)	89.2% (58/65)

In the 13<sup>th</sup> century, 17 of the 23 total occurrences of DOM with human proper names correspond to Valencian texts. At the same time, these 17 occurrences represent a slightly higher percentage of DOM within Valencian texts (19.3%, vs. 16.3% for all Catalan). Similarly, in the 14<sup>th</sup> century, 5 out of 6 occurrences of DOM with human proper names are found in the Valencian area, where the percentage of DOM is more than twice than in Catalan as a whole (5.1% vs. 2.3%).<sup>8</sup> This dialectal pattern seems not to hold in the 15<sup>th</sup> century, as Valencian texts seem to show less DOM than the Catalan mean. However, we argue that this is due to the particular passage of the Valencian texts under consideration: for example, only two occurrences of human proper names in direct object function were found in *Tirant*, one of the two Valencian texts from the 15<sup>th</sup> century in our corpus. This extremely low number of human proper names in direct object function certainly affected the

<sup>8</sup> On the other hand, if one compares the numbers of the 13<sup>th</sup> and the 14<sup>th</sup> centuries, it one may find surprising the relatively elevated proportion of DOM in the 13<sup>th</sup> c. with respect to the 14<sup>th</sup> c. In this regard, in addition to the possible influence of the textual typologies of the works of one century and the other, we must highlight two aspects: first, 5 of the occurrences of the 13<sup>th</sup> c. belong to the *Llibre dels Fets* by Jaume I, a work which has been counted belonging to the 13<sup>th</sup> c. for its date of composition, but which has been analyzed using a copy of the 14<sup>th</sup> century; and second, 16 of the examples of the 13<sup>th</sup> century correspond to a work (*Clams I*) which, in addition to its Valencian adscription, contains a major proportion of proper names in DO function, with a total of 77.

overall figures for that century. This is corroborated by Perera's (1986) observation on *Tirant*:

'When the direct object is a personal proper name, in *Tirant* there is an absolute predominance of the construction with preposition [= DOM]; cases where this type of direct object does not bear preposition are very scarce.' (Perera 1986, 65) [our translation]

Moreover, 9 of the 23 occurrences of DOM in the 15<sup>th</sup> century are from *Curial* (34), a work whose dialectal adscription has been subject to a long-lasting debate, and in our study, following the CICA's criterion, it has been considered as belonging to general Catalan (i.e., as not belonging to any dialect in particular). However, its Valencian adscription is the one that seemed to be supported by more evidence, as recently argued by Soler (2017). Finding a meaningful number of DOM in the text thus supports Soler's claim.

- (34) a. *Mas l' emperador tenia prop sí a Curial*  
 but the emperor have.IPFV.3SG near himself DOM Curial  
 'And the emperor had Curial nearby.'  
 (15<sup>th</sup> c., *Curial*, 62)
- b. *e víu que volia ociure a Othó*  
 and see.PST.3SG that want.IPFV.3SG kill.INF DOM Othó  
 'And he saw that he wanted to kill Othó.'  
 (15<sup>th</sup> c., *Curial*, 67)
- c. *lo rey de Navarra, loctinent general del senyor*  
 the king of Navarre, lieutenant general of.the lord  
*rey féu penjar a n Guillem Arús, pagès*  
 King make.PST.3SG hang DOM the Guillem Arús, peasant  
*del terme de Terraça*  
 of.the area of Terrassa  
 'The king of Navarre, general lieutenant of the lord king, had Guillem Arús hung, a peasant of the area of Terrassa.'  
 (15<sup>th</sup> c., Safont, *Dietari*, 95)
- d. *condescendria en favorir al dit*  
 condescend.COND.3SG in favor.INF DOM.the mentioned  
*micer Pere Ram*  
 mister Pere Ram  
 'He would condescend to favour the above-mentioned gentleman Pere Ram.'  
 (15<sup>th</sup> c., *Epistolari* IIa, letter 25, written in València)

- e. *per què no m' has dat a Rachel?*  
 why not CL.DAT.1SG have.PRS.2SG given DOM Rachel  
 'And why haven't you given Rachel to me?'

(15<sup>th</sup> c., Sant Vicent, *Sermons* IV, 12)

To explain the substantially different patterns attested in the Valencian texts, several factors need to be considered. First, this was a late area within the Catalan linguistic domain: the language was taken there once the land was conquered in 1229–1245. In addition, a great portion of repopulators came from Aragon, and therefore the potential influence of Aragonese could have played a role – although at this point we cannot provide any further details on DOM in that variety. In addition, being a lateral area within the Catalan linguistic domain, the Valencian territories were probably more subject to the influence of Spanish, whose penetration into the Catalan-speaking area would soon become more pronounced. Recall from Table 3 that from the 16<sup>th</sup> century onwards, once the influence of Spanish reaches high levels, DOM becomes much more frequent in general (35), and the dialectal constraints (Valencian vs. other varieties) seem to be blurred:

- (35) a. *Are deslīga[n] a Susanna y líga[n] los vells*  
 now untie.PRS.3PL DOM Susanna and tie.PRS.3PL the old.PL  
 'Now they are untying Susanna and tying up the old fellows.'  
 (16<sup>th</sup> c., *Consueta de Susanna*, 431)
- b. *fonch int[errogat] si coneix a Joan*  
 be.PST.3SG asked whether know.PRS.3SG DOM Joan  
*Palou, fill de Ant[oni]*  
 Palou son of Antoni  
 'He was asked whether he knew Joan Palou, Antoni's son.'  
 (16<sup>th</sup> c., *Illes XVI–28*, 458)
- c. *No fou poca consolació de trobar a Lloÿset,*  
 not be.PST.3SG little consolation of find.INF DOM Lloÿset  
*com aribàrem, tan bonico y sanet*  
 as arrive.PST.1PL so pretty and healthy  
 'It was not of little consolation to find Lloÿset so pretty and healthy,  
 when we arrived.'  
 (16<sup>th</sup> c., Liori i Requesens, *Epistolaris*, letter 13, written in Molins de  
 Rei, Catalonia)

- d. *Déu ha derrocat y destruït a Job*  
 God have.PRS.3SG demolished and destroyed DOM Job  
 ‘God has demolished and destroyed Job.’  
 (16<sup>th</sup> c., Conques, *Job*, 81)
- e. *esquarteraren a Johan Martí, de Campanar*  
 quarter.PST.3PL DOM Johan Martí from Campanar  
 ‘They quartered Johan Martí, from Campanar.’  
 (16<sup>th</sup> c., *Antiquitats I*, 82)

If we compare the evolution of DOM with personal names in Catalan with the evolution in Spanish, substantial differences emerge, since in Spanish DOM with human proper names was nearly systematic from the very beginning (Laca 2006, 443, among others).

Let us focus now on deity names (*Déu* ‘God’, *Satanàs* ‘Satan’, *Sant Joan* ‘Saint John’, etc.). Table 3 shows that, as with personal proper names, DOM was usual in Old Catalan, particularly from the 15<sup>th</sup> century onwards, especially among Valencian authors (36), and was subsequently generalized to all dialects (37).<sup>9</sup>

- (36) a. *tots veurem clarament a Déu*  
 all.PL see.FUT.1PL clearly DOM God  
 ‘We will all see God clearly.’  
 (15<sup>th</sup> c., Sant Vicent, *Sermons IV*, 9)
- b. *nostre Senyor Déus ha mès capità a sent*  
 Our Lord God have.PRS.3SG put captain DOM saint  
*Miquel sobre tots los àngels.*  
 Miquel above all.PL the angels  
 ‘Our Lord God has put Saint Miquel as captain over all the angels.’  
 (15<sup>th</sup> c., Sant Vicent, *Sermons IV*, 71)
- c. *él amà a Déu sobiranament*  
 he love.PST.3SG DOM God very much  
 ‘He loved God very much.’  
 (15<sup>th</sup> c., Sant Vicent, *Sermons IV*, 82)

<sup>9</sup> In turn, deity nouns (those containing determiners or possessives, such as *Nostre Senyor* ‘Our Lord’, *el Senyor* ‘the Lord’, *la Verge* ‘the Virgin’, etc.) pattern with human definite NPs. For the distinction between deity nouns and deity names, and the validity of this in different languages, cf. Caro Reina (2020) and references therein.



- (37) a. *Deffensen a Déu provant que, com a just, no*  
 defend.PRS.3PL DOM God proving that, as fair, no  
*castiga a nengú sens culpa*  
 punish.PRS.3SG DOM nobody without fault  
 ‘They defend God, proving that since he is fair, he doesn’t punish anybody who is not guilty.’

(16<sup>th</sup> c., Conques, *Job*, 37)

- b. *offenien greument a Déu*  
 offend.IPFV.3PL gravely DOM God  
 ‘They used to offend God gravely.’

(16<sup>th</sup> c., *Antiquitats I*, 57)

Turning now to human definite NPs, Table 3 shows that, although the absence of marking is general in the earlier period of Old Catalan, DOM has a modest but not insignificant presence, especially from the 15<sup>th</sup> century onwards and mainly in texts from the Valencian area (39).<sup>10</sup> In the 14<sup>th</sup> century, only very scarce occurrences of DOM with human definite NPs are attested (38).

- (38) a. *Volem anar a la vila e estaylarem al*  
 want.PRS.1PL go.INF to the town and isolate.FUT.1PL DOM.the  
*rey, que no y pusca entrar ost.*  
 king that not there can.SBJV.PRS.3SG enter.INF army  
 ‘We want to go to the town and isolate the king so that no army can enter there.’

(14<sup>th</sup> c., Jaume I, *Fets*, f. 39v)

- b. *e regonech al dit mercader*  
 and recognize.PST.3SG DOM.the mentioned trader  
 ‘And he recognized the mentioned trader.’

(14<sup>th</sup> c., *Epistolari Ic*, letter 90, written in València)

<sup>10</sup> As an anonymous reviewer notes, throughout the 15<sup>th</sup> century the Valencia area was more influential than Catalonia in economic and political terms, and gave rise to many of the greater literary achievements (e.g. the chivalric novel *Tirant lo Blanc*) and authors (e.g. the poets Ausiàs March and Jordi de Sant Jordi) in the period. This might well have influenced literary production in the other areas of the Catalan-speaking countries, although it is difficult to make a similar claim when it comes to personal letters or other text types. While we agree that an analysis of the textual transmission at that time would be very interesting, it is outside the scope of the current paper.

- (39) a. *lo rey près per la mà a l' ermità*  
 the king take.PST.3SG by his hand DOM the hermit  
 'The king took the hermit by his hand.'  
 (15<sup>th</sup> c., Martorell, *Tirant*, 102)
- b. *yo penjaré als jurats*  
 I hang.FUT.1SG DOM.the jurors  
 'I will hang the jurors.'  
 (15<sup>th</sup> c., *Epistolari* IIa, letter 10, written in València)

Finally, as observed with the previous categories, there is a sharp change in the 16<sup>th</sup> century, when the presence of DOM with human definite NPs increases significantly (40), still with greater frequency in the Valencian texts (41):

- (40) a. *lo prevera interrogarà primerament a l' home*  
 the priest interrogate.FUT.3SG firstly DOM the man  
 'The priest will first interrogate the man.'  
 (16<sup>th</sup> c., *Illes XVI–23*, 444)
- b. *Per què donchs sosté als affligits?*  
 why then sustain.PRS.3SG DOM.the heartbroken.PL  
 'Why then does he support the heartbroken?'  
 (16<sup>th</sup> c., Conques, *Job*, 42)
- c. *animant als cavallés romans*  
 cheer.up.GER DOM.the knights Roman  
 'Cheering up the Roman knights.'  
 (16<sup>th</sup> c., *Grandeses*, 151)
- (41) a. *Jo crech que a[u]rà vists a mos*  
 I think.PRS.1SG that have.FUT.3SG<sup>11</sup> seen.M.PL DOM my  
*fills, puy sòn estats en Barcelona*  
 children since be.3PL been.M.PL in Barcelona  
 'I think that you will have seen my children, since they have been to Barcelona.'  
 (16<sup>th</sup> c., Liori i Requesens, *Epistolaris*, letter 12, written in València)
- b. *Ý per ospedar a estos senyós*  
 and for host.INF DOM these men  
 'And because I was hosting these men.'  
 (16<sup>th</sup> c., Liori i Requesens, *Epistolaris*, letter 11, written in Molins de Rei, Catalonia)

The overall evolution of DOM with human definite NPs in the diachrony of Catalan is substantially different from what has been observed for Spanish, where DOM with this type of object has had a significant presence from the earliest texts and exceeds 50% of cases as early as the 14<sup>th</sup> century (Laca 2006, 443).

Lower in the hierarchies that predict the emergence and extension of DOM cross-linguistically (1)–(2), we find human indefinites. In Old Catalan, these types of objects typically appear without DOM, as shown in Table 3. In those few cases where the marking is present (42), the object is usually specific, as predicted in the hierarchies. Example (42c) is particularly illustrative here, in that it shows the contrast in the use of DOM for the first indefinite, the referent of which is extensively specified by a series of appositions, and the second one (‘another Moor captain’), which is left unspecified and therefore bears no DOM:

- (42) a. *eyl ssí víu a ·I· ssarayn apallat Haxec*  
 he EMPHASIS see.PST.3SG DOM a Saracen called Haxec  
 ‘He did see a Saracen called Haxec.’  
 (14<sup>th</sup> c., *Cort*, 203)
- b. *fonch interrogat si coneix ell [...] a*  
 be.PST.3SG asked whether know.PRS.3SG he DOM  
*una dona inculpada de bruxa*  
 a woman accused of witch  
 ‘He was asked whether he knows a woman who had been accused of  
 being a witch.’  
 (15<sup>th</sup> c., *Lleida* IId, f. 4r)

---

11 This verb, inflected in the 3<sup>rd</sup> person singular, refers to a 2<sup>nd</sup> person (polite form *vostè*).

- c. *lo duch de Segorb sentencià en Segorb*  
 the duke of Segorb sentence.PST.3SG in Segorb  
*a hun moro, que era alamí de Algar,*  
 DOM a Moor who be.IPFV.3SG officer of Algar  
*que-s deya Caravan, qui era lo principal*  
 who-REFL say.IPFV.3SG Caravan who be.IPFV.3SG the principal  
*de la muntanya de Spatan ab dos fills seus;*  
 of the mountain of Spatan with two sons his  
*y hun alter moro capità.*  
 and a other Moor captain

‘The duke of Segorb sentenced in Segorb a Moor, who was officer of Algar, who was named Caravan, who was the principal of the mountain of Spatan, with two of his sons; and another Moor captain.’

(16<sup>th</sup> c., *Antiquitats I*, 11)

In this respect, then, Old Catalan behaves like Old Spanish, where indefinites used to lack DOM (Laca 2006, 443, 458–460).

As for bare (singular and plural) NPs, at the bottom of the definiteness/specificity scale (2), DOM is not an option in Old Catalan. There are only three examples in our corpus, from the 16<sup>th</sup> century and from Valencian texts. In all three cases, other factors (such as word order alterations) co-occur and may explain the use of DOM. For example, in (43) we see the presence of a complement breaking the adjacency between the verb and the direct object. On the other hand, in Spanish, DOM with this type of objects appears “sporadically since the 15<sup>th</sup> century” [our translation] (Laca 2006, 445).

- (43) *en Barcelona no admètan per la vida a home*  
 in Barcelona not admit.PRS.3PL for the life DOM man  
*perquè sia molt rich si altres calitats convenients ad aquell estament li falten*  
 ‘In Barcelona they don’t admit at all man only because he is very rich, if he lacks any other of the required qualities.’

(16<sup>th</sup> c., *Despuig, Col·loquis*, 125)

If we now focus on the animacy scale (1), we conclude that, with the exception of objects with human referents, Old Catalan disallows DOM. Accordingly, whereas in the diachrony of Spanish DOM with animate direct objects was possible from the 14<sup>th</sup> century (Laca 2006, 444–445), this is not the case in Catalan. Only very few examples are found, all of which are from the 16<sup>th</sup> century when, as previously mentioned, there was a significant Spanish influence on Catalan. Additionally, note that in example (44), other factors, including the absence of a verb and

the parallelism with an obligatorily marked personal pronoun, are also relevant factors:

- (44) *ý à· ns dotat a nosaltres de més capacitat*  
 and has CL.ACC.1PL endowed DOM us of more capacity  
*ý saber que a les bèsties del camp*  
 and wisdom than DOM the beasts of.the countryside  
 ‘[God] has endowed us with more capacity and wisdom than he has  
 endowed the beasts of the countryside.’  
 (16<sup>th</sup> c., Conques, *Job*, 86)

Another difference between DOM in Catalan and Spanish relates to inanimate proper names, that is, place names. In Spanish ‘from the earliest texts DOM appears optionally with proper names referring to inanimates, in particular place names’ (Laca 2006, 450) [our translation], reaching 38% of occurrences in the *Cid* and 100% (from a small total amount) in the *Quijote* and *Lucanor* (cf. also Caro Reina 2020, 245–247). In Old Catalan, however, DOM with place names is not possible, barring a few examples, all of which come from the 16<sup>th</sup> century, in the Valencian area:

- (45) a. *conquistà Almeria, a Tortosa ý la*  
 conquer.PST.3SG Almeria, DOM Tortosa and the  
*antiga ý populosa Lleyda*  
 ancient and populated Lleida  
 ‘He conquered Almeria, Tortosa and the ancient and populated Lleida.’  
 (16<sup>th</sup> c., Despuig, *Col·loquis*, 96)
- b. *lo ardit que tingueren los grechs per a*  
 the trick that have.PST.3PL the Greek.PL to  
*pèndrer a Troya*  
 take.INF DOM Troy  
 ‘The trick that the Greek had to take Troy.’  
 (16<sup>th</sup> c., Despuig, *Col·loquis*, 119)
- c. *E après de haver saquejat a Gandia*  
 and after of have.INF sacked DOM Gandia  
*ý Oliva, los de Oriola tornaren-se-n ab molta roba que havien pres.*  
 and Oliva  
 ‘And after having sacked Gandia and Oliva, the people from Oriola went  
 back with a lot of clothes they had taken.’  
 (16<sup>th</sup> c., *Antiquitats I*, 69)

With other types of inanimates, DOM in Old Catalan is very marginal (46) and is limited to contexts in which its absence would result in ambiguity between the subject and the object interpretation. The marginality of DOM with inanimates (not including place names) is thus similar in Old Catalan and Old Spanish, where such examples were also very rare (Laca 2006, 450).

- (46) a. *lo voler amant, ahirant al membrar*  
 the will love.GER, irritate.GER DOM.the memory  
*e al entendre.*  
 and DOM.the understanding  
 ‘Loving the will, irritating the memory and the understanding.’  
 (13<sup>th</sup> c., Llull, *Doctrina* I, 151)
- b. *La sapiència y felicitat [...] vens y*  
 the wisdom and happiness beat.PRS.3SG and  
*avantaja també a les perles y pedres precioses*  
 overtake.PRS.3SG also DOM the pearls and gems precious  
 ‘The wisdom and happiness beats and overtakes also the pearls and  
 precious gems.’  
 (16<sup>th</sup> c., Conques, *Job*, 74)

In sum, our corpus study shows that the emergence and development of DOM in the diachrony of Catalan follows the predictions in terms of prominence based on the person/animacy and definiteness/specificity hierarchies, going from personal pronouns to categories lower down the hierarchies in (1)–(2).<sup>12</sup> This evolution thus resembles in general terms the evolution of DOM in Spanish. However, the paths of extension of DOM in both languages diverge in many respects too, such as chronology and extension to a wider or narrower range of categories. Broadly speaking, DOM always consolidates and extends to different categories earlier in Old Spanish than in Old Catalan.<sup>13</sup> At the same time, it is remarkable – as an anonymous reviewer suggests – that not even very low categories, such as

**12** In particular, Pineda (in press a) concludes that following combined scale of definiteness and animacy is the one regulating the emergence and expansion of DOM in the diachrony of Catalan:

(i) PERSONAL PRONOUNS > HUMAN PROPER NAMES (personal names) > HUMAN DEFINITE NPs > HUMAN INDEFINITE NPs > DEITY NAMES > RELATIVES AND INTERROGATIVES [excluded from this paper] > INANIMATE PROPER NAMES (place names) > ANIMATE NPs > INANIMATE NPs > HUMAN BARE NPs

**13** Differences are also found in terms of the role and importance of the different semantic and contextual triggering factors (e.g. lexical type of verb, dislocation, etc.), which are analyzed in (Pineda in press a; b).

indefinites and inanimates, are entirely free of DOM. Thus, although DOM is far less extended than in Spanish in quantitative terms, it is not the case to the same extent in qualitative terms.

This study sheds light on the question of whether DOM in Catalan has undergone a process of narrowing – as argued by Dalrymple/Nikolaeva (2011, 212–213), based on the allegedly reduced use of DOM according to prescriptive grammar – or, alternatively, if the contemporary non-prescriptive uses in Section 3 (examples 20–27) represent the continuation of medieval patterns. We have shown that DOM was a syntactic resource present in the language in the 13<sup>th</sup>–15<sup>th</sup> centuries, long before the influence of Spanish was relevant. It is therefore difficult to attribute the spread of this phenomenon exclusively to an external influence – recall Salvador’s and Pérez Saldanya’s words at the end of Section 3. From the late 15<sup>th</sup> to the early 16<sup>th</sup> century, the sociolinguistic pressure from Spanish unquestionably played a significant role. It is during that time that the percentage of DOM increases exponentially across all types of objects, both with personal pronouns and for proper names and human NPs. We conclude that DOM is (at least partially) a phenomenon native to Catalan, and that the influence of Spanish, from the late 15<sup>th</sup> and the early 16<sup>th</sup> century onwards, was probably more quantitative than qualitative. It is impossible to determine how DOM would have evolved in Catalan in the absence of any influence from Spanish, but in any case one should take into account that: (i) DOM is a phenomenon that exists in many very different languages of the world, including the Romance family; (ii) DOM already existed in Catalan before the Spanish influence began; and (iii) in the languages of the world the cases of retraction of a syntactic phenomenon like this one are particularly scarce, whereas the general pattern is the progressive enlargement of the set of contexts with DOM following the hierarchies in (1)–(2) (cf. Bossong 1991, 152–153; Aissen 2003, 472, fn. 33 and also Dalrympe/Nikolaeva 2011, 211–215).

## 5 Concluding remarks

We have provided a comprehensive picture of Differential Object Marking (DOM) in Catalan, analyzing its diachronic evolution as well as its scope in present-day language, and we have been able to draw significant conclusions from our discussion.

Following the criteria established by Thomason (2001, 93–94) to determine whether a linguistic change can be considered the result of exogenous factors or of endogenous development, the presence of DOM in Old Catalan texts which precede the impact of Spanish leads us to the conclusion that DOM is (at least partially) a

phenomenon native to Catalan (the increase of DOM with different objects in the 16<sup>th</sup> and the 17<sup>th</sup> century owing to the influence of Spanish notwithstanding).

At the same time, however, and in accordance with the fact that the phenomenon was already present in the language before the period of external influence of Spanish, what occurred in Catalan could be a case of what Heine/Kuteva (2003, 539; 2010, 86) term “replica grammaticalization” – we thank Itxaso Rodríguez-Ordóñez (p.c.) for having suggested this possibility to us. This type of grammaticalization is triggered by language contact, where a language adopts not a category or a grammatical concept that it did not have until then, but a grammaticalization process existing in another language: in the case under study, Catalan (where DOM already existed but with many constraints) would have adopted the grammaticalization process of DOM existing in Spanish and would thus have made of DOM a grammatical resource with a much wider scope. This means that because of the contact with Spanish, Catalan could have given to DOM a new grammatical sense, as a marker for a wider range of DOs.

Although there are no studies of DOM in Catalan after the 17<sup>th</sup> century, and a systematic Catalan dialect survey remains to be carried out, our observations on spontaneous speech and language usage in the media over the past few years indicate that Catalan DOM has continued to spread towards less-prototypical categories: current uses of DOM with indefinites (both specific and non-specific), bare plurals or even some inanimate NPs can be found, as exemplified in Section 3.<sup>14</sup> Thus, we argue that the diachronic and contemporary data provided in this study call into question Dalrymple/Nikolaeva’s (2011, 212–213) suggestion that the evolution of DOM in Catalan is an instance of the “narrowing” of DOM.<sup>15</sup> We hope that this paper has shed some light on the nature of DOM in Catalan, one of the most controversial – and interesting – phenomena in the (descriptive and prescriptive) grammar of this language.

---

**14** A similar evolution is attested in Neapolitan: in present-day use, and from the 19th century, DOM extends to animate and individuated direct objects, whereas in Old Neapolitan the phenomenon was far more restricted, subject to several lexical, syntactic, semantic and pragmatic factors (Ledgeway 2009, 831; cf. also Sornicola 1997).

**15** When reviewing the existing literature on DOM in Old Catalan, also Escandell Vidal (2009, 842), citing Bossong (1983/1984, 11), suggested: “The diachronic data suggest that DOM is not necessarily a transfer from Spanish and that the normative ban against it has given rise to a system in which ‘the formerly far more widespread use of the preposition a with the direct object has been reduced to the absolute minimum (the free pronoun)’ [...]. If this view is correct, what the standardization did was to promote an artificial regression of DOM”.



## Bibliography

- Adell, Marc-Vicent, *Problemàtica de l'acusatiu preposicional en català. Notes al voltant de Roís de Corella*, in: Massot i Muntaner, Josep (ed.), *Miscel·lània Germà Colón*, vol. 1, Barcelona, Publicacions de l'Abadia de Montserrat, 1994, 93–109.
- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21 (2003), 435–483.
- Albareda i Valls, Cristina, *Les variants al·lomòrfiques de la preposició “a” en algunes varietats del català. Descripció i anàlisi*, MA Thesis, Universitat Autònoma de Barcelona, 2009.
- Bosson, Georg, *Differential Object Marking in Romance and beyond*, in: Wanner, Dieter/ Kibbee, Douglas A. (edd.), *New analyses in Romance linguistics. Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, April 7–9, 1988*, Amsterdam/Philadelphia, John Benjamins, 1991, 143–170.
- Bosson, Georg, *Le marquage différentiel de l'objet dans les langues d'Europe*, in: Feuillet, Jack (ed.), *Actance et valence dans les langues de l'Europe*, vol. 20, Berlin/New York, de Gruyter, 1998, 193–258.
- Cabanes Fitor, Vicent, *Anàlisi de la construcció d'objecte directe de persona en català (segles XIII–XVI). Estudi del sintagma “article + substantiu”*, in: Todolí i Cervera, Júlia, et al. (edd.), *Estudis sobre morfologia valenciana*, Alcoi, Marfil, València, Generalitat Valenciana, 1994, 77–136.
- Cabanes Fitor, Vicent, *Anàlisi de la construcció d'objecte directe de persona en català (segles XIII–XVI). Estudi del nom propi*, in: Martines Peres, Vicent/Martines Peres, Josep/Ponsoda Sanmartín, Joan Josep (edd.), *A sol post. Estudis de llengua i literatura*, Alcoi, Marfil, 1995, 47–90.
- Caro Reina, Javier, *Differential Object Marking with proper names in Romance languages*, in: Kempf, Luise/Nübling, Damaris/Schmuck, Mirjam (edd.), *Linguistik der Eigennamen*, Berlin/Boston, de Gruyter, 2020, 225–258.
- Corpus Informatitzat del Català Antic (CICA)*, Joan Torruella (dir.), with Manel Pérez Saldanya and Josep Martines. [www.cica.cat](http://www.cica.cat) [last access: 06.05.2021].
- Dalrymple, Mary/Nikolaeva, Irina, *Objects and information structure*, Cambridge, Cambridge University Press, 2011.
- Dixon, Robert M. W., *Ergativity*, *Language* 55:1 (1979), 59–138.
- Escandell Vidal, Victoria, *Topics from Ibiza. Differential Object Marking and clitic dislocation*, in: Kaiser, Georg/Leonetti, Manuel (edd.), *Proceedings of the workshop “Definiteness, Specificity and Animacy in Ibero-Romance Languages”*, Konstanz, Universität Konstanz, 2007a, 23–43.
- Escandell Vidal, Victoria, *Acusatiu preposicional i dislocació amb clític*, *Caplletra* 42 (2007b), 185–218.
- Escandell Vidal, Victoria, *Differential Object Marking and topicality. The case of Balearic Catalan*, *Studies in Language* 33:4 (2009), 832–885.
- Fabra, Pompeu, *Gramàtica catalana*, Barcelona, Institut d'Estudis Catalans, 1918/1933.
- Fabra, Pompeu, *Gramàtica de la llengua catalana*, in: Mir, Jordi/Solà, Joan (edd.), *Pompeu Fabra, Obres completes*, vol. 1, Barcelona, Institut d'Estudis Catalans, 1912/2005.
- Fabra, Pompeu, *Converses filològiques*, in: Mir, Jordi/Solà, Joan (edd.), *Pompeu Fabra, Obres completes*, vol. 7, Barcelona, Institut d'Estudis Catalans, 2010.

- GIEC = Institut d'Estudis Catalans, *Gramàtica de la llengua catalana*, Barcelona, Institut d'Estudis Catalans, 2016.
- Heine, Bernd/Kuteva, Tania, *On contact-induced grammaticalization*, *Studies in Language* 27:3 (2003), 529–572.
- Heine, Bernd/Kuteva, Tania, *Contact and grammaticalization*, in: Hickey, Raymond (ed.), *The handbook of language contact*, Oxford, Wiley-Blackwell, 2010, 86–105.
- Hualde, José Ignacio, *Catalan*, London, Routledge, 2012.
- Iemmolo, Giorgio, *Topicality and Differential Object Marking. Evidence from Romance and beyond*, *Studies in Language* 34:2 (2010), 239–272.
- Laca, Brenda, *El objeto directo. La marcación preposicional*, in: Company Company, Concepción (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, vol. 1, México DF, UNAM/Fondo de Cultura Económica, 2006, 423–475.
- Ledgeway, Adam, *Grammatica diacronica del napoletano*, Tübingen, Niemeyer, 2009.
- Leonetti, Manuel, *Specificity in clitic doubling and Differential Object Marking*, *Probus* 20:1 (2008), 33–66.
- López, Luis, *Indefinite objects. Scrambling, choice functions, and differential marking*, Cambridge, Mass., MIT Press, 2012.
- Meier, Harri, *O problema do acusativo preposicional no catalão*, *Boletim de Filologia* 8 (1947), 237–260.
- Meier, Harri, *Sobre as origens do acusativo preposicional nas línguas românicas*, in: Meier, Harri (ed.), *Ensaio de filologia românica*, Lisboa, Edição da “Revista de Portugal”, 1948, 115–164.
- Mestres, Josep M., et al. (edd.), *Manual d'estil. La redacció i l'edició de textos*, Vic, Eumo, 1995.
- Næss, Åshild, *What markedness marks. The markedness problem with direct objects*, *Lingua* 114:9 (2004), 1186–1212.
- Pensado, Carmen, *El complemento directo preposicional. Estado de la cuestión y bibliografía comentada*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 11–59.
- Perera i Parramon, Joan, *Contribució a l'estudi de les preposicions en el “Tirant lo Blanch” (primera part)*, *Llengua & Literatura* 1 (1986), 51–109.
- Pineda, Anna, *Acostament al marcatge diferencial d'objecte als inicis del català modern*, *eHumanista/IVITRA* 14 (2018), 570–596.
- Pineda, Anna, *Aspectes de la transitivitat en els inicis del català modern*, *Caplletra* 66 (2019), 207–236.
- Pineda, Anna, *El complement directe (I). El marcatge diferencial d'objecte*, in: Martines, Josep/Pérez, Saldanya/Manuel Rigau, Gemma (edd.), *Gramàtica del català antic*, Amsterdam/Philadelphia, John Benjamins, in press a.
- Pineda, Anna, *El complement directe (II). Aspectes de la transitivitat*, in: Martines, Josep/Pérez, Saldanya/Manuel Rigau, Gemma (edd.), *Gramàtica del català antic*, Barcelona, John Benjamins, in press b.
- Rohlf, Gerhard, *Autour de l'accusatif prépositionnel dans les langues romanes*, *Revue de Linguistique Romane* 35 (1971), 312–334.
- Rohlf, Gerhard, *Panorama de l'accusatif prépositionnel en Italie*, *Studii și Cercetări Lingvistice* 24 (1973), 617–621.
- Ruaix i Vinyet, Josep, *El català/2. Morfologia i sintaxi*, Moia, Ed. Ruaix, 1985.

- Salvador, Vicent/Pérez Saldanya, Manuel, *Transitivité et interférence linguistique. La construction A + complément d'objet direct en espagnol et en catalan*, *Contrastes. Revue de linguistique contrastive* (1993), 39–67.
- Sancho Cremades, Pelegrí, *La categoria preposicional*, Valencia, Universitat de València, 1995.
- Sancho Cremades, Pelegrí, *La preposició i el sintagma preposicional*, in: Solà, Joan, et al. (edd.), *Gramàtica del català contemporani*, vol. 2, Barcelona, Empúries, 2002, 1749–1768.
- Silverstein, Michael, *Hierarchy of features and ergativity*, in: Dixon, Robert M.W. (ed.), *Grammatical categories in Australian languages*, Canberra, Australian Institute of Aboriginal Studies, 1976, 112–171.
- Solà, Joan, *Sintaxi normativa. Estat de la qüestió*, Barcelona, Empúries, 1994.
- Soler, Abel, *Enyego d'Àvalos, autor de Curial e Güelfa*, *Estudis Romànics* 39 (2017), 137–165.
- Sornicola, Rosanna, *L'oggetto preposizionale in siciliano antico e in napoletano antico. Considerazioni su un problema di tipologia diacronica*, *Italienische Studien* 18 (1997), 66–80.
- Thomason, Sarah G., *Language contact. An introduction*, Washington, Georgetown University Press, 2001.
- Todolí, Júlia, *Els pronoms*, in: Solà, Joan, et al. (edd.), *Gramàtica del català contemporani*, vol. 2, Barcelona, Empúries, 2002, 1337–1433.
- Torrego, Esther, *The dependencies of objects*, Cambridge, Mass., MIT Press, 1998.
- von Heusinger, Klaus/Kaiser, Georg A., *Affectedness and Differential Object Marking in Spanish*, *Morphology* 21:3 (2011), 593–617.
- Wheeler, Max W./Yates, Alan/Dols, Nicolau, *Catalan. A comprehensive grammar*, London/New York, Routledge, 1999.



Senta Zeugin

# DOM in Modern Catalan varieties

An empirical study based on acceptability judgment tasks

**Abstract:** This paper examines the current situation of Differential Object Marking (DOM) in the major Catalan varieties, focusing on the animacy of the direct object. In the past, Catalan standard grammar has advocated for very conservative rules for applying DOM, and thus a substantial divergence between spoken language and standard grammar was created. Nowadays, standard Catalan seems to accept DOM more readily, but still falls short of the wide range of uses found in Catalan dialects. Hence, this paper relies on experimental methods (i.e. an acceptability judgment task with a 2x3 within-subjects design) to analyze the situation found in the major Catalan dialects, i.e. Central Catalan, North-Western Catalan, Valencian and Majorcan. The experiment focuses on direct objects in definite noun phrases and the test sentences were manipulated for presence/absence of DOM and three degrees of animacy (human, animal, inanimate). Results suggest that for human direct objects in noun phrases the use of DOM is as well accepted as the unmarked version, which contradicts what is claimed by Catalan standard grammars. However, the unmarked version is slightly preferable in the case of animals and clearly preferable for inanimate objects.

**Keywords:** Differential Object Marking (DOM), Catalan, experimental linguistics, acceptability judgment tasks, diatopic variation

## 1 Introduction

For some languages, research into DOM is already quite advanced, whereas for others, much ground remains to be covered. One such language is Catalan. Although an increasing number of studies on DOM have been published in recent years, there is still more to discover about the phenomenon. This is one of the

---

**Acknowledgements:** This article is based largely on the author's MA-Thesis (Zeugin 2017). I'd like to thank two anonymous reviewers as well as Albert Wall, Johannes Kabatek and John Barlow for their comments and suggestions. I am also very grateful to all the native speakers who reviewed the material for the experiment as well as everyone who participated in the experiment itself.

---

Senta Zeugin, University of Zurich, [senta.zeugin@uzh.ch](mailto:senta.zeugin@uzh.ch)

main reasons why the experiment described in this article takes Catalan and its main varieties as its focus.

Historically, the view of Catalan authors on DOM, which is often seen as a prototypical feature of neighbouring Spanish, has been rather biased and less than favourable. For the most part it was regarded as a mere alien structure, existing only because of language contact, and thus, further studies were not deemed necessary.<sup>1</sup> However, this view has not been shared by everyone, and there have been a few early defenders of DOM as an inherent Catalan feature. Meier (1945) shows that already in the earlier stages of Catalan, DOM had spread to varying degrees throughout its territories, citing examples from authors such as Bernat Metge and Sant Vicent Ferrer (both 14<sup>th</sup>/15<sup>th</sup> century) and Jacint Verdaguer (19<sup>th</sup> century). Geographical differences can be explained, as Müller (1971) notes, as a result of the gradual loss of the Latin case system (favouring the accusative as the universal case), which for Catalan seemed to start in the south of the territory and spread slowly towards the north. Thus, according to Müller (1971), it should come as no surprise that a text from Valencia (Sant Vincent Ferrer) contains more occurrences of DOM than a contemporary work from Barcelona (Metge). The author goes on to argue that the late starting point for the loss of the Latin case system seems to lie at the root of the much narrower expansion of DOM usage in Catalan compared to Spanish, where this loss began about a century earlier. Further confirmation of the autonomous origins of Catalan DOM can be found in more recent studies, for example by Irimia/Pineda (2019). The authors underline fundamental differences between Spanish and Catalan DOM and cite examples from texts as early as the 14<sup>th</sup> century.

As a result of this change of perception, the study of Catalan DOM has become more detailed and objective over the years. For example, Rohlfs (1971) and later Pensado (1995), Wheeler/Yates/Dohls (1999) and Sancho Cremades (2008) present detailed characterizations, adding numerous examples of DOM in contexts that were not described in earlier works. Thus, they demonstrate, contrary to what until then had often been claimed by Catalan linguists and to what Catalan normative works suggested, that DOM was actually being used much

---

**1** One of the defenders of this viewpoint was Par (1923, 153), who stated that: “En català modern jamay lo règim directe porta la preposició ‘a’, quan aquell es un nom. Mes ab los pronoms personals tònic s’hi es introduhida, sens dubte per influencia castellana; barbarisme qui es acceptat, àdhuch per qualques gramàtics” (‘In Modern Catalan, the direct object in the form of a noun is never marked by the preposition *a*. However, it is used with strong personal pronouns, without a doubt due to Castilian influence; a barbarism that is, nonetheless, accepted by some grammarians’ [translation SZ]). Badia Margarit’s (1962) opinion is not quite so extreme and, similarly to Pompeu Fabra (1933), he describes various permissible contexts for DOM. However, both authors do still urge to avoid it and to use any other means whenever possible.

more widely. The view that DOM in Catalan is nothing more than a contamination from Spanish has been questioned for some time and is considered outdated. In separating the two phenomena and recognizing the historic and synchronic individuality of Catalan DOM, a need for a detailed description of its evolution and present occurrence arises. And it is in this latter area of investigation that the current paper seeks to contribute to the bigger picture of DOM in Catalan in general and its major dialects in particular.

For this purpose, an experiment was conducted, consisting of an acceptability judgment task in which participants were asked to evaluate 60 sentences, presented to them in a random order. Before describing the experiment, some background information on the situation of DOM in the major Catalan dialects will be provided. Therefore, Section 2 contains an overview of the status of DOM in standard Catalan as well as a brief introduction to the diatopic study of DOM in four major Catalan dialects. The remaining Sections will focus on the experiment itself, starting with its method and general setup in Section 3, and the design of the test sentences in Section 4. Subsequently, Section 5 covers the procedure of the experiment as well as relevant information about the participants themselves. Lastly, in Section 6, the results from the different parts of the experiment will be discussed, in each case by first focusing on the regions individually and then looking at all the dialects together. Section 7 then presents conclusions, including some observations on possible future research.

## 2 DOM: standard Catalan versus Catalan dialects

When studying the behaviour of dialects, what is predominantly of interest is the spoken language, considering that this is where diatopic variation usually manifests itself. In order to establish such variations, the dialects themselves can be compared. An additional source of insight is the contrast of each dialect with the established standard version of the respective language. In most languages, if not all, there is a certain divergence between standard and everyday language as a natural result of their individual purposes. Additionally, in the case of Catalan, the standard language has served as a means of distinguishing Catalan itself more clearly from neighbouring Spanish, as can be seen, for example, in the works of Fabra (1933) and Badia Margarit (1962). Thus, differences between the standard and the spoken language might be more pronounced in Catalan<sup>2</sup> than

---

<sup>2</sup> Sancho Cremades (2008) confirms the existence of a considerable gap between standard and spoken language in his chapter in the *Gramàtica del català contemporani*.

in other languages, and this suggests that it might be useful to take a brief look at the process of linguistic normalization and its important role in the creation of standard Catalan.<sup>3</sup>

During a long time the normative version of Catalan, i.e. standard Catalan, was defined in three works by Pompeu Fabra from the beginning of the 20<sup>th</sup> century: the *Normes ortogràfiques*, published in 1913, the *Gramàtica catalana* from 1918 and the *Diccionari general de la llengua catalana* from 1931 (cf. Costa Carreras 2001). Especially the grammar from 1918, and later on its updated version from 1933, was regarded as normative up until very recently, as Costa Carreras (2001) notes.

Fabra deemed only a very limited number of contexts acceptable territory for DOM, and suggested to avoid it whenever possible. Among the very limited permitted uses in Fabra (1933) is the marking of strong personal pronouns, elements like *tots*, *tothom* and *el qual*, as well as expressions of reciprocity (*l'un a l'altre*) plus cases of possible ambiguity. A rather problematic aspect of Fabra's standard Catalan was the fact that he based it predominantly on the Central Catalan dialect and thus did not pay heed to the wide spectrum of Catalan dialects.<sup>4</sup> Furthermore, standard Catalan, having its origins in about 1918, did not receive a

---

**3** Linguistic normalization (*normalització lingüística*) is a phrase coined by Catalan sociolinguists, referring to the joint processes of establishing a linguistic standard, or norm, for a certain language (*normativització*), and the extension of this language to all registers and contexts (*extensió social*) (cf. Kremnitz 1979; Vallverdú 1979; Lagarde 2009). In the case of Catalan, the process of normalization was not without its difficulties. At the beginning of the 18<sup>th</sup> century after the War of the Spanish Succession, Catalan was substituted by Spanish as the official language of Catalonia. During the following two centuries Catalan first experienced a significant decline in usage, and later on near the end of the 19<sup>th</sup> century, led by a cultural renaissance, a surge in popularity (cf. Vallverdú 1979). Fueled by constant political tension, the Catalan-speaking, antimonarchist bourgeoisie saw the potential of conserving and reanimating their mother tongue, in order to gain the support of the working class population. With this historical background, the linguistic normalization of Catalan started at the beginning of the 20<sup>th</sup> century, particularly with the works of Pompeu Fabra. Shortly after, during the dictatorship of Primo de Rivera (1923–1930) and from 1939 on under Franco, any languages other than Castilian Spanish were prohibited (cf. Kremnitz 1979). In spite of all the setbacks, the Catalan normalization process never came to a complete stop. Still, it wasn't until the 1970s that it was able to run much more smoothly. Furthermore, this historical background explains some of the political tension that still surrounds Catalan today and the insistence of some of the older grammars on distancing Catalan as far as possible from Spanish.

**4** Fabra's initial idea was to create a much more inclusive standard, striving to use not only the Central dialect, but to let it be influenced and moulded by the varieties of the other regions, all the while getting rid of any vestiges of Spanish (cf. Fabra 1908). Colón Domènech (2009, 23) indicates that, while Fabra's intentions might have been good initially, in the end he developed a standard based only on one dialect: "Malgrat una pretesa voluntat de fer un estàndard acumulatiu, l'estàndard en realitat és restrictiu; és a dir, 'particular'" ('Despite the intention of creating a cumulative standard, the standard is actually restrictive, or rather, 'particular' [translation SZ]).



comprehensive update for almost a century until the publication in 2016 of the *Gramàtica de la llengua catalana*, the new normative grammar of the *Institut d'Estudis Catalans* (IEC).

As far as DOM is concerned, this new normative work generally seems to follow in the footsteps of Fabra, stating that, as a rule, the direct object does not require a preposition. Also, in most of the permitted contexts for DOM, it appears to concur with Fabra, while providing much more detailed explanations and pointing out several exceptions. The authors of this recent grammar include a number of new permitted contexts for DOM, for example, when the order of a sentence has been changed by dislocations. It appears that with this new grammar, the normative authorities (nowadays the IEC) have slowly come around to recognizing the existence and validity of DOM as a Catalan phenomenon. The authors even go so far as to refer to the presence of diatopic variation, albeit without providing further details. Nevertheless, a look at the regional grammars and studies on the subject suggests that the updated standard language is still somewhat removed from the real situation of the spoken language throughout the territory.



**Illustration 1:** The Catalan dialects<sup>5</sup>.

<sup>5</sup> Illustration 1: Wikimedia CC BY-SA 2.5; annotations of dialect names [SZ]. Names of the dialects are consistent with those used in this article. In the case of Majorcan it should be noted that this

In the remainder of this Section, a brief overview of the situation of DOM in the four major Catalan dialects will be presented. As already mentioned, the Central Catalan dialect was the general basis for the development of the Catalan standard in the early 20<sup>th</sup> century. An analysis of spoken Central Catalan can be found in Escandell Vidal (2009). Her study illustrates that even the most closely related dialect diverges to a certain degree from the standard language, in the case of DOM most often concerning its use with proper names, human definite noun phrases and occasional non-human animate ones.<sup>6</sup> Still, the behaviour of Central Catalan is somewhat similar to standard Catalan, whereas especially in Majorcan and Valencian the extension of DOM is much greater than what is considered permissible in the general Catalan standard and also in the respective regional standards.

For the Balearic dialect, Moll (1979) proposes similar guidelines to those of Fabra's grammar, while at the same time introducing one of the earliest recognized examples of DOM with inanimate objects,<sup>7</sup> a use that certainly goes beyond any normative prescriptions. In more recent studies, such as Wheeler/Yates/Dols (1999) or Escandell Vidal (2009), the Majorcan situation is addressed in even greater detail. The latter author, echoing a previous study by Rohlf's (1971), emphasizes that DOM in Majorcan is found primarily in dislocated positions. According to Escandell Vidal (2009), it seems that not only intrinsic factors of the direct object, such as animacy and definiteness, have an impact on the use of DOM in Majorcan, but also others like topicality and information structure.

In contrast to Majorcan, to which many studies concentrating on DOM are dedicated, for North-Western Catalan there is much less information available. One of the few studies, by Boladeras Taché (2011), is very inconclusive concerning the contexts where DOM can be found. Nonetheless, the author does allude to one notable aspect, which itself probably hinders a more precise examination: the fact that in North-Western Catalan marked and unmarked objects, in combination with the definite article, can sound the same.

On Valencian, on the other hand, much more information can be found, and it tends to suggest a range of contexts for DOM, which goes beyond the uses permitted in the new normative grammar by the IEC. There are examples such as *Hem vist com perseguia un policia a un lladre* (AVL 2006, 303) that show strong

---

doesn't include the dialects spoken on Ibiza or Menorca. Considering that all participants in the experiment in question came from the island of Majorca, the more generic term Balearic Catalan seemed inappropriate in this instance.

<sup>6</sup> Escandell Vidal (2009, 840) cites examples like *Les monges no estimen a/ana les nenes* and *Veuré a la Maria*.

<sup>7</sup> The example from Moll (1979, 203) is: *Colliules, a les peres que ja són madures*.

parallels to the evolution found in Spanish DOM. Sancho Cremades (1995) confirms the similarity of Valencian to Spanish DOM, and adds that it is usually associated with direct objects denoting a person, the capacity of an object to feel the action described by the verb, or topicality. The only restriction noted by Sancho Cremades (1995, 199) is the need for the object to be clearly determined (*no veig als xiquets* [. . .] vs. *he trobat uns xiquets* [. . .]). However, considering the example cited above, even this restriction doesn't always seem to hold. A further interesting aspect is the fact that as early as the 15<sup>th</sup> century, there were occurrences of DOM in Valencian, as several examples found in the Sermons of Sant Vicent Ferrer demonstrate (*Senyor, yo bé am mon pare, mas més am a vós* cf. Meier 1945, 239).

### 3 The experiment: main aspects and methodology

As noted in the Introduction, the experiment discussed in this paper consisted of an acceptability judgment task that examined the present use of DOM in Catalan. Thus, the judgments of native Catalan speakers concerning the presence or absence of DOM in given sentences were studied. One of the important aspects of a more detailed description of DOM usage nowadays, and therefore one of the three main aspects investigated in this experiment, is diatopic variety. The geographic distribution of Catalan-speaking regions is considerable. The largest territory can be found in Spain, but there are also areas in France (Roussillon) and Italy (Alghero, Sardinia), resulting in contact situations with at least four different romance languages. Combined with the relative isolation of the Balearic Islands, this leads to a variety of different geographical situations as well as different language contact situations. The possible regional differences are confirmed by some of the previously cited literature, e.g. regional grammars, focusing on the major dialects of Catalan.

Consequently, 'regional variation' as a first major aspect was examined by distributing the experiment across the regions of the six major Catalan dialects. In addition to the regional varieties, the experiment focused on two further aspects: 'animacy' and, to a lesser degree, 'position'. Including both of these factors in the same test sentence was not feasible, because it would have rendered it impossible to determine which one was more influential. Therefore, the experiment was divided into two parts, each centring around one main aspect and thus avoided any conflict between them. Furthermore, since the sentences for both parts differed in structure, it was expected that participants would not easily deduce what linguistic aspect the experiment was focusing on. Ideally, participants evaluated the test sentences spontaneously and without thinking about them extensively.

As mentioned, a total of 60 randomly allocated sentences were given to each participant, divided into two groups of 30, that is, for the two parts of the experiment. Both parts were designed using the Latin Squares method, with six individual conditions distributed across six lists and five repetitions, leading to 30 test sentences per experiment part.

The first part was created around ‘animacy’, one of the main parameters generally used to determine different DOM-configurations.<sup>8</sup> The focus lay on the animacy of the direct object in noun phrases. In accordance with the basic animacy scale shown below (cf., for example, von Heusinger/Kaiser 2005, 37), the degree of animacy of the direct object was manipulated to see whether this had an effect on the acceptability of the test sentences.

Animacy scale: HUMAN > ANIMATE > INANIMATE

By combining this scale with the presence or absence of the DOM marker *a*, six categories (conditions) were created, each of which was represented by the same number of test sentences. The distribution of the conditions is shown in Table 1 below.

**Table 1:** Conditions for ‘animacy’.

	<b>+HUMAN</b>	<b>–HUMAN, +ANIMATE</b>	<b>–ANIMATE</b>
<b>DOM</b>	Condition 1	Condition 2	Condition 3
<b>∅</b>	Condition 4	Condition 5	Condition 6

Whereas the main focus of the whole experiment was on this first part about the animacy of the direct object in combination with DOM, the sentences from the second part were mostly used as distractors. With this goal in mind, the chosen sentence structure clearly differed from the one in the first part, as will be seen later on. Nevertheless, in an effort to glean further information on the behaviour of the phenomenon in Catalan, DOM was also included. Thus, the third main aspect to be considered was ‘position’, or, more specifically, the influence of a change in the position of the pronominal direct object and its resulting combination with clitic doubling and DOM on the acceptability of the test sentence. As presented in Table 2, the first positional category was topicalization, bearing in mind that the literature suggests it to be highly susceptible to co-occurrence

<sup>8</sup> Cf., for example, Bossong (1991), von Heusinger/Kaiser (2005) or Aissen (2003).

with DOM.<sup>9</sup> Secondly, dislocation to the right was chosen, which seems to be a frequent partner with DOM, particularly in Majorcan (as indicated by Escandell Vidal 2009). Unfortunately, due to the design of the experiment and the need for a rigorous uniformity of the structure of the test sentences, it was not possible to use a total dislocation to the right, as would have been the ideal case. This has to be taken into account when analyzing the results of the second part of the experiment. In Catalan, both types of dislocation go hand in hand with the introduction of a strong personal pronoun, which is where DOM comes into play. As a third category for the factor ‘position’, the most basic sentence type was used, with neither dislocation nor clitic doubling, and thus also without DOM. In combining these three positional categories with a change between 2<sup>nd</sup> and 3<sup>rd</sup> person singular pronouns, six conditions were again created, as can be seen in Table 2.

**Table 2:** Conditions for ‘position’.

	<b>Topicalization with clitic doubling</b>	<b>Dislocation to the right with clitic doubling</b>	<b>Neither dislocation nor clitic doubling</b>
<b>2<sup>nd</sup> SG</b>	Condition 11	Condition 12	Condition 13
<b>3<sup>rd</sup> SG</b>	Condition 14	Condition 15	Condition 16

Following the characteristics that they represent, each condition shown in Tables 1 and 2 corresponds to a sentence, or rather, a variation of a test sentence. Consequently, a test sentence always had six variations. Naturally, a participant was not asked to evaluate all of these six variations but just one of them, thus avoiding interference caused by lexical repetition and cross-sentence judgments. As can be seen in Table 3, for the Latin Squares method with six conditions per test sentence, six different lists have to be created, each containing only one of the conditions per test sentence. Furthermore, every condition needs to appear the same number of times in each list, which guarantees the most equal distribution. Consequently, after six sentences, all of them with six variations, every list includes each condition exactly once.

In order to achieve reliable results, it is necessary to gather as much data as possible, thus making it tempting to present participants with a huge number of test sentences. Doing so would be risky and most likely result in the loss of many participants due to boredom. Therefore, a compromise between maximal data gain and minimal loss of participants had to be reached. Accordingly, a total of

<sup>9</sup> Cf., for example, Rohlfes (1971) and Pensado (1995).

**Table 3:** Latin Squares.

	List 1	List 2	List 3	List 4	List 5	List 6
<b>Sentence 1</b>	c1	c2	c3	c4	c5	c6
<b>Sentence 2</b>	c6	c1	c2	c3	c4	c5
<b>Sentence 3</b>	c5	c6	c1	c2	c3	c4
<b>Sentence 4</b>	c4	c5	c6	c1	c2	c3
<b>Sentence 5</b>	c3	c4	c5	c6	c1	c2
<b>Sentence 6</b>	c2	c3	c4	c5	c6	c1

60 sentences (each with its six respective variations) were created, 30 per experiment part. In other words, for each of the experiment parts Table 3 was repeated five times, leaving every participant to evaluate each condition five times. Once created, the sentence variations were randomly distributed among the six lists, following the pattern seen in Table 3, and each participant was again randomly presented with only one of these lists, containing 60 sentences in total. Consequently, for each variation of a test sentence to be evaluated at least once (360 overall, 180 per experiment part), six different participants were needed.

## 4 The experiment: sentence construction

In accordance with the two major aspects under investigation, ‘animacy’ and ‘position’, two sentence groups were designed with the aim of acquiring as much information about DOM in the Catalan varieties as possible. At the same time, the intention was to control for as high a number of other presumably influential factors as possible. In the following, both sentence groups will be discussed in detail. An important aspect here is the fact that the language used for all test sentences is a more or less standard version of Catalan. Taking into account every regional variation of the major Catalan dialects when constructing the sentence catalogue would not have been practical. It would have meant either designing six individual catalogues, sacrificing the possibility of comparing the results on a statistical level, or using an amalgamation of diatopical traits that could only have led to confused participants. A second aspect of a more general nature concerns word choices. Throughout both experiment parts, no verb was used more than once and none of the subjects or objects of the sentences belonging to the ‘animacy’ part were repeated at any point. Furthermore, only transitive verbs were used.

Concerning the first experiment part, as pointed out, the aspect under scrutiny was the animacy of the direct object in noun phrases and its influence on the perceived acceptability of a given sentence with or without DOM. Following the basic animacy scale presented in Section 3, the direct objects chosen denote humans, animals and inanimate objects. All the sentences of this group exhibit exactly the same basic structure, as can be seen in Table 4 below, consisting of an animated subject, a transitive verb in *perfet perifràstic*, a direct object in the form of a noun phrase and a circumstantial complement. Thus, the only manipulated part of the sentence was the direct object itself in terms of its degree of animacy and presence/absence of the preposition *a*, which marks DOM.

**Table 4:** Examples of test sentences for experiment part ‘animacy’.

	<b>+HUMAN</b>	<b>–HUMAN, +ANIMATE</b>	<b>–ANIMATE</b>
<b>DOM</b>	El director va filmar <u>a la meva nora</u> per al documental sobre la vila.	El director va filmar <u>a la meva mula</u> per al documental sobre la vila.	El director va filmar <u>a la meva granja</u> per al documental sobre la vila.
∅	El director va filmar <u>la meva nora</u> per al documental sobre la vila.	El director va filmar <u>la meva mula</u> per al documental sobre la vila.	El director va filmar <u>la meva granja</u> per al documental sobre la vila.
<b>DOM</b>	El guàrdia va transportar <u>al meu nét</u> fins a l'estació de ferrocarrils.	El guàrdia va transportar <u>al meu gat</u> fins a l'estació de ferrocarrils.	El guàrdia va transportar <u>al meu sac</u> fins a l'estació de ferrocarrils.
∅	El guàrdia va transportar <u>el meu nét</u> fins a l'estació de ferrocarrils.	El guàrdia va transportar <u>el meu gat</u> fins a l'estació de ferrocarrils.	El guàrdia va transportar <u>el meu sac</u> fins a l'estació de ferrocarrils.

As far as the different categories of direct objects are concerned, it should be highlighted that for the human objects only terms denoting family members were used. On the one hand, this helped to keep the sentence structure as balanced as possible. On the other hand, it has been shown that, traditionally, they are more often found with DOM than other nouns. The choices for the animals of the second category of direct objects were made with the intention of avoiding the more unusual ones. Hence, the selection was limited to animals that are often found as (more or less) common pets or on a farm. Finally, for the category of inanimate direct objects, the nouns selected all designate material objects, thus steering clear of abstract concepts that can sometimes be attributed a certain amount of animacy due to metonymy.

As can be seen in Table 4, each sentence contains the possessive as part of the direct object in order to avoid possible influences of other languages such as, for example, Sardinian where the mere presence of the definite article can block the occurrence of DOM.<sup>10</sup> The co-occurrence of possessive and definite article as seen in these sentences is an inherent characteristic of Catalan grammar.

Generally speaking, it should be emphasized that for each of these blocks of six variations of a test sentence, the direct objects were kept as similar in form as possible. Hence, the direct objects used within each of the blocks of six share the same gender and number of syllables. Each object, as well as each subject, was used only once throughout the experiment and the complements were not repeated. Furthermore, in an effort to keep possible interferences as small as possible, the gender combinations of subject and object were distributed equally.<sup>11</sup>

As was the case for the test sentences of the ‘animacy’ part, those for the part about the position of the direct object were also designed with care. Although this experiment part was not as central to the study as the first one, the sentences were still designed with an eye to maintaining the possibility of attaining some information about DOM in Catalan. Consequently, the sentences were all constructed with the following basic structure: omitted subject, direct object in pronominal form (depending on the condition with or without a type of dislocation and DOM), transitive verb in *perfet perifràstic*, and circumstantial complement.

**Table 5:** Examples of test sentences for experiment part ‘position’.

	<b>Topicalization with clitic doubling</b>	<b>Dislocation to the right with clitic doubling</b>	<b>Neither dislocation nor clitic doubling</b>
<b>2. SG</b>	<u>A tu et</u> van saludar des del cafè del jardí botànic.	<u>Et</u> van saludar <u>a tu</u> des del cafè del jardí botànic.	Et van saludar des del cafè del jardí botànic.
<b>3. SG</b>	<u>A ell e </u> van saludar des del cafè del jardí botànic.	<u>E </u> van saludar <u>a ell</u> des del cafè del jardí botànic.	El van saludar des del cafè del jardí botànic.
<b>2. SG</b>	<u>A tu et</u> van vèncer en la final del torneig d’escacs.	<u>Et</u> van vèncer <u>a tu</u> en la final del torneig d’escacs.	Et van vèncer en la final del torneig d’escacs.
<b>3. SG</b>	<u>A ella la</u> van vèncer en la final del torneig d’escacs.	<u>La</u> van vèncer <u>a ella</u> en la final del torneig d’escacs.	La van vèncer en la final del torneig d’escacs.

<sup>10</sup> Blasco Ferrer (1986) and Bossong (1982) both highlight this link between presence of the definite article and absence of DOM in Sardinian.

<sup>11</sup> From a total of 30 sentences, 7 have both feminine subject and object, 7 both masculine subject and object, and 8 feminine subject and masculine object, and vice versa.



Compared to those of the first part, these sentences feature a different basic sentence structure, but still allow for the possibility of observing the behaviour of DOM with personal pronouns. Hence, they were regarded as adequate considering the aims of the experiment. As a second manipulation, a change between 2<sup>nd</sup> and 3<sup>rd</sup> person singular pronouns was introduced, in order to see whether there were any observable differences in acceptability. In addition, the 3<sup>rd</sup> person singular forms alternated between feminine and masculine pronouns. Similarly to the sentences of the ‘animacy’ part, each verb occurs only once throughout the whole experiment and the complements vary as well. Finally, it should be noted that, although the second category shows a clear movement to the right of the direct object, it does not constitute a full movement to the right-most position in the sentence, as would be the prototypical case. Examples of this more extreme movement have been found in several Catalan dialects (cf. Rohlfs 1971; Escandell Vidal 2009). Unfortunately, including this specific type of sentence was not possible due to the design of the experiment and the need to keep the overall uniformity of the test sentences intact.

One further detail remains to be mentioned here. All test sentences were designed by the article’s author, who happens not to be a native speaker of Catalan. Nonetheless, prior to conducting the experiment, several native speakers from various regions checked them for their grammatical correctness.

## 5 The experiment: proceedings and participants

The experiment described in this article was designed for online participation and was set up using OnExp, a free program from the University of Göttingen. Participants began the experiment with a short introduction, in which they were familiarized with the general aims of the survey (without mentioning explicitly the exact topic under investigation). In a second step, they were introduced to the three stages of the experiment itself. To start with, participants were asked to provide some personal information, including age, origin, and speaking habits; this information was given anonymously. They were then shown the evaluation system used for the test sentences and were given the opportunity to become accustomed to it by practicing on eight exercise sentences designed to exemplify the different acceptability levels. The evaluation system consisted of an acceptability scale ranging from 1 to 7, with 1 being equal to ‘not at all natural’, and 7 ‘totally natural’.



**Illustration 2:** Acceptability Scale.

After these introductory sections, participants continued with the third phase: the experiment itself. They were shown each test sentence, one after another, in a random order, and were asked to give their opinion on its acceptability, using the scale shown above.

Before turning to the results of the experiment, some details about the participants should be discussed briefly. While some participants were obtained through personal contacts, the majority was reached via social media. The link leading to the survey was published on Twitter and in numerous groups and sites on Facebook. In addition, multiple cultural associations and platforms in each of the regions were contacted, as well as universities and language schools. The only prerequisite was that a participant should be a native speaker of one of the major dialects of Catalan.

From the six major regions a total of 305 participants completed the survey. Out of these, unfortunately, some had to be eliminated for various reasons. Some had more than one major dialect as a first language, making it impossible to attribute their responses clearly to just one region. Others skipped several test sentences, leading to them being rejected for the overall results. Some evaluated every sentence with the same extremely high or low value (always giving a 7, for example), leading to a distortion of the final overall scores.<sup>12</sup> Having eliminated these cases, it was also necessary to level the number of participants per list within every region, in order to avoid giving certain sentences more or less weight in the final regional analysis. Thus, a total of 260 participants was finally achieved. In the following Table the distribution of participants among regions and according to the more important demographic factors is displayed.

As can be seen, for the four regions of the bigger dialects, a sufficiently high number of participants were found to allow for a reliable analysis of their responses. Sadly, this was not the case for Alguerese and Rossellonese, where only a very small number of participants were registered, thus allowing for no more than the reflection of a few trends and some mostly speculative observations on their use of DOM. For the purpose of this paper, then, these have been excluded from further study, and will not be discussed in the following Section.

On the whole, it can be observed that the distribution of female and male participants is rather balanced (with the exception of Valencian, where approximately three quarters were male, and Majorcan, where two thirds were female). A significant demographic aspect was the question of whether the participants

---

<sup>12</sup> Due to their particular configuration of conditions, some of the test sentences were certain to receive very low or very high evaluations. Therefore, if a participant attributed the exact same value to every sentence, the possibility that he didn't take the experiment seriously was very high and thus he was excluded.

**Table 6:** Demographic distribution of participants.

	Total	Gender		Philoling <sup>13</sup>		Age			Education		Mother tongue	
		F	M	No	Yes	≤25	26–50	51≤	University	other	Cat.	Dial. + dialect
<b>Total</b>	260	125	135	213	46	44	162	52	207	53	148	112
<b>Alguerese</b>	11	5	6	10	1	0	2	9	9	2	3	8
<b>Central Cat.</b>	90	39	51	73	17	14	57	19	74	16	52	38
<b>Majorcan</b>	72	48	24	67	4	17	49	5	51	21	45	27
<b>N-W Catalan</b>	36	18	18	22	14	6	20	9	33	3	28	8
<b>Rossellonese</b>	9	3	6	8	1	0	5	4	7	2	2	7
<b>Valencian</b>	42	12	30	33	9	7	29	6	33	9	18	24

were students of linguistics or philology in general, which could have led them to deduce the nature of the experiment more easily. As is shown in columns 4 and 5, the great majority of participants, overall and also within each region, had not attended courses pertaining to these fields of study. While most participants were between 26 and 50 years old, it is an interesting fact that for Alguerese 9 out of 11 participants were older, some considerably. Again, the number of participants here is too small to draw any definitive conclusions.

Turning to the level of education of the participants, it should be noted that, both overall and within each region, there seems to be a very similar distribution of those having completed, or being in the process of completing, university studies. Overall 79.6% of participants attended university while 20.4% did not. This general distribution is reflected to a greater or lesser degree in each region (Majorcan exhibits a slightly lower percentage of university studies, and North-Western Catalan a slightly higher one).

The final category reflected in Table 6 concerns the native language of the participants. They were able to select from the following options: the six major Catalan dialects, plus Spanish, Italian, French and Sardinian. As noted above, when participants picked multiple Catalan dialects as their native language they had to be excluded from the final analysis. As for the total numbers, 56.9% (148 out of 260) only chose a Catalan dialect as their native language, whereas 43.1% (112 out of 260) indicated one or more of the other romance languages as well. Considering the individual regions, an interesting finding emerged: when participants from Central Catalan, Majorcan and North-Western Catalan were asked

<sup>13</sup> Philoling: Abbreviation [SZ] indicating if participants were students of linguistics and/or philology in general.

for their mother tongue, a majority chose only their dialect. Valencian speakers, however, tended to choose not only their variety, but one of the other Romance languages as well (mainly Spanish).

## 6 The experiment: results

As noted above, each of the two main parts of the experiment should be considered separately, and for this reason the results here will be examined in different sections. The main focus will be on the first part of the experiment, containing the test sentences designed for ‘animacy’, but there will also be a brief description of the results for ‘position’. Due to the reasons described earlier, only the results from the four most important dialects (Central Catalan, North-Western Catalan, Majorcan and Valencian) will be shown, as well as the overall results and comparison of these four regions.

Before discussing the results themselves, a few details about the methods used in the analysis of the data should be noted. The main part of the analysis was carried out using the programs Excel and R (R Core Team 2019), because of the possibilities they offer for the calculation of mean values for each of the conditions and for the comparison of these based on different parameters. The results presented in the following are, as a rule, based on all the data of one region and the perceived effects therein. Furthermore, in addition to the descriptive analysis, a further analytical method was used, this as a first approach to a statistical analysis of the collected data. The model chosen was a repeated measure ANOVA (Analysis of Variance),<sup>14</sup> which made it possible to ascertain whether or not the possible effects detected by the descriptive approach could be confirmed as statistically significant. When such effects could be confirmed to be statistically significant, then we can say that there was a very high probability that they had an impact on the acceptability of the different types of sentences tested in the experiment. Taking into account two of the main sources for possible bias, two sets of ANOVAs were performed for every region – one focusing on the participants (Subject-Analysis) and one for the test sentences (Item-Analysis). For a first statistical analysis of the collected data, ANOVA was deemed an appropriate choice, allowing for a general impression of the data’s statistical value. In future work, alternative models could be consulted as a means of achieving an even bigger picture.

---

<sup>14</sup> The ANOVAs were conducted in R, using the *ez* package (Lawrence 2016).

A final aspect to note is that for the overall results, when combining data from the four main regions, each mean value was calculated based on the 1,200 judgments that the respective condition had received (as a result of the total of 240 participants, each evaluating every condition five times).

## 6.1 Animacy

In this Section, the results of the four regions will be introduced separately in order to better illustrate the possible effects of the manipulated factors on the perceived acceptability of the test sentences. The factor of regional variance will then be discussed by presenting the overall results concerning the first main aspect, ‘animacy’, including a comparison of the four major Catalan dialects. Before starting with the dialect with the highest participation (Central Catalan), a reminder of the focus of the first experiment seems useful. Its main manipulated aspect was the degree of animacy of the direct object in combination with the presence or absence of DOM. The sentences presented to the participants in a completely random order were of the following type: *El guàrdia va transportar al/ el meu nét/gat/sac fins a l'estació de ferrocarrils.*

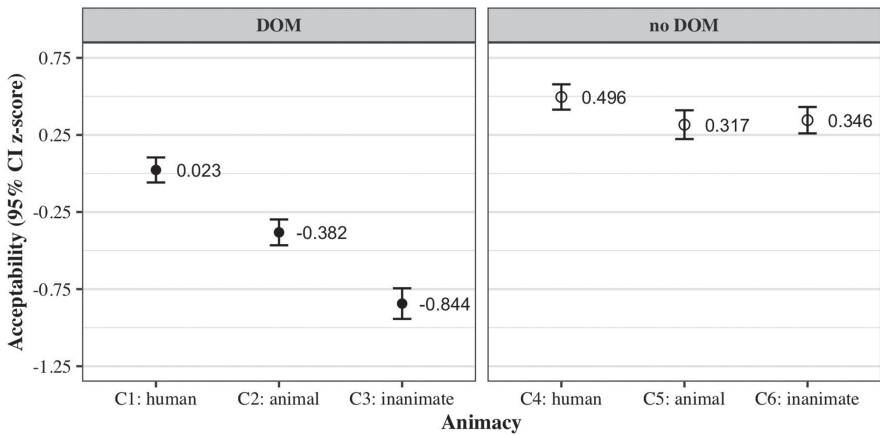
### 6.1.1 Regional results for ‘animacy’

In the following Graph, the results obtained from the participation of 90 Central Catalan speakers are illustrated, equalling 450 separate evaluations of each condition. All the values depicted in this Graph, and in those that follow, are means of the aggregated z-scores calculated from the acceptability judgments that each condition had received.<sup>15</sup> They are shown with a Confidence Interval of 95%, represented by the error bars of each mean value. On the horizontal axis the six conditions are represented, showing the three degrees of animacy (human, animal, inanimate) and split into two blocks according to the presence or absence of the second factor, DOM.

Focusing on the presence/absence of DOM, overall, Central Catalan speakers clearly prefer the sentences without the preposition *a*, since C4–C6 show higher values than C1–C3. Nevertheless, the three less-favoured conditions still seem

---

<sup>15</sup> The decision to use mean z-scores instead of normal mean judgments is based on the fact that they allow for a fairly accurate comparison of the participants, taking into account the possible individual interpretations of the acceptability scale.



**Graph 1:** Central Catalan: results for ‘animacy’.

to be at the least to some extent acceptable, bearing in mind that especially C1 (humans), but also C2 (animals), reach relatively high mean z-scores. Considering the fact that all three pairs of objects (C1 and C4, C2 and C5, C3 and C6) show mean values whose error bars more and more clearly do not overlap, there is a high probability that the presence/absence of DOM has an effect on the acceptability of a given sentence. Although Central Catalan speakers tend towards the versions without DOM, Graph 1 indicates that mainly for human objects (C1 vs. C4) there is a relatively high acceptance of the version with DOM.

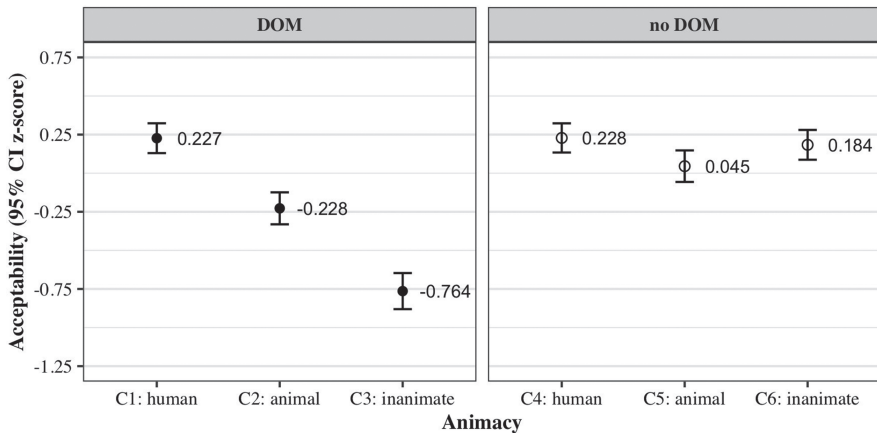
Furthermore, it appears that the participants regard the animacy of the direct object as an influential factor for the acceptability of DOM. Looking at the conditions containing DOM (C1–C3), the human object is distinctly favoured over animals, which in turn are preferred over inanimate marked objects. Since once again neither mean values nor error bars overlap, it can be assumed that the direct object’s animacy plays an important part when deciding about the acceptability of a sentence.

The two ANOVAs confirmed both of the effects described for Central Catalan. Furthermore, not only are the effects of DOM and ‘animacy’ manipulation separately statistically significant, but their interaction is as well, implying that they interactively influence the acceptability of any given sentence.<sup>16</sup>

#### 16 ANOVA results for Central Catalan:

1. Subject-Analysis for the effects of DOM:  $F(1, 89) = 179.483, p < .001$  and ‘animacy’:  $F(2, 178) = 57.317, p < .001$  as well as their interaction:  $F(2, 178) = 36.327, p < .001$  on the acceptability of the test sentences.

For the analysis of the Majorcan data, the evaluations of 72 native speakers were taken into account, resulting in a total of 360 judgments per condition. An important detail here is the fact that one participant pointed out that he would have given higher votes if the sentences had included the proper definite articles of his dialect (*es, sa*). Earlier, an explanation was given as to why it was not feasible to incorporate such regional variations. Even so, it is worth noting that Majorcan speakers' acceptance of sentences containing DOM could be even higher than it already is, considering the following Graph.



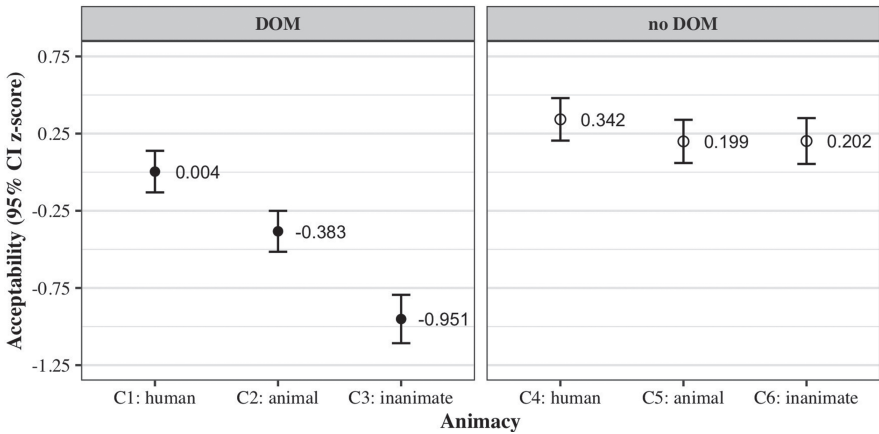
**Graph 2:** Majorcan: results for ‘animacy’.

Again, the factor ‘animacy’ appears to have a certain amount of influence over the perceived acceptability of the sentences including DOM, since the mean values of C1–C3 clearly differ from each other and their error bars do not overlap at all. Unsurprisingly, the human object was preferred over the animals and inanimate objects. Remarkably, the z-scores for human objects with DOM (C1) and for the same type of object without DOM (C4) are nearly the same (with a difference of 0.001). This might suggest that Majorcan speakers do not have a preference for one or the other, and thus do not care whether or not a human object is marked. They seem to find both versions equally acceptable. Even for the objects representing animals (C2 and C5) the gap between the two values is quite small. Only when looking at the inanimate objects, a strong partiality for the version without DOM (C3 vs. C6) can be seen.

2. Item-Analysis for the effects of DOM:  $F(1, 29) = 267.394, p < .001$  and ‘animacy’:  $F(2, 58) = 29.354, p < .001$  as well as their interaction:  $F(2, 58) = 26.432, p < .001$  on the acceptability of the test sentences.

ANOVAs were used to determine whether these possible effects found in the descriptive approach were of any statistical value. The calculations confirmed, both for ‘animacy’ and for DOM, the very high probability of an effect on the acceptability of a given sentence, showing these, as well as their interaction, to be statistically highly significant.<sup>17</sup>

Continuing with the data from North-Western Catalan, it should be noted that this was the region with the smallest number of participants (apart from Rosellonese and Alguerese), amounting only to 36, which still renders 180 evaluations per condition. Looking at the mean values depicted in Graph 3, speakers of North-Western Catalan seem to accept DOM with objects denoting humans far more often than when they refer to animals.



**Graph 3:** North-Western Catalan: results for ‘animacy’.

The lowest mean value is attributed to the inanimate objects marked by *a*. In addition, it is highly likely that ‘animacy’ has an effect on the acceptability of DOM for this region because the three values (C1–C3) show plainly visible differences and their error bars do not overlap at all. The three conditions representing absence of DOM (C4–C6) are at a relatively high (and similar) level, indicating that ‘animacy’

**17 ANOVA results for Majorcan:**

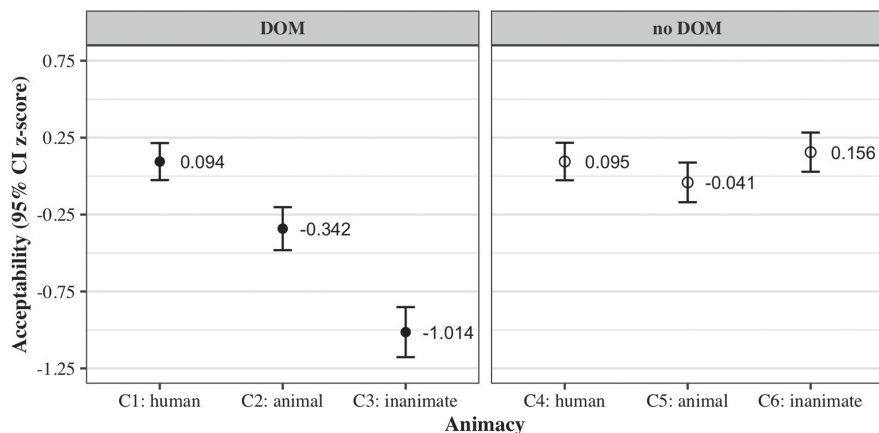
1. Subject-Analysis for the effects of DOM:  $F(1, 71) = 52.534, p < .001$  and ‘animacy’:  $F(2, 142) = 61.369, p < .001$  as well as their interaction:  $F(2, 142) = 52.984, p < .001$  on the acceptability of the test sentences.
2. Item-Analysis for the effects of DOM:  $F(1, 29) = 55.894, p < .001$  and ‘animacy’:  $F(2, 58) = 28.429, p < .001$  as well as their interaction:  $F(2, 58) = 26.39, p < .001$  on the acceptability of the test sentences.



does not come into play regarding the acceptance of sentences without DOM. The non-marked object is clearly favoured overall by North-Western Catalan speakers, with the human objects being the only conditions where the values are somewhat close to each other (C1 vs. C4). Furthermore, the error bars of each pair do not overlap, indicating that presence/absence of DOM very probably affects the acceptability of a sentence.

These tendencies are confirmed by both ANOVAs, producing a statistically very significant result for DOM and for ‘animacy’, as well as their combination.<sup>18</sup> Thus, it seems highly probable that both separately and interactively they have an impact on the acceptance of the sentences.

Lastly, for Valencian a total of 42 participants answered the questionnaire, yielding 210 evaluations per condition overall and allowing for a fairly reliable analysis of the data.



**Graph 4:** Valencian: results for ‘animacy’.

Valencians seem to distinguish clearly between the individual degrees of animacy when the object is combined with DOM (C1–C3). Considering their z-scores, we note that the human object is deemed quite acceptable, followed

<sup>18</sup> ANOVA results for North-Western Catalan:

1. Subject-Analysis for the effects of DOM:  $F(1, 35) = 41.411, p < .001$  and ‘animacy’:  $F(2, 70) = 25.891, p < .001$  as well as their interaction:  $F(2, 70) = 15.365, p < .001$  on the acceptability of the test sentences.

2. Item-Analysis for the effects of DOM:  $F(1, 29) = 106.245, p < .001$  and ‘animacy’:  $F(2, 58) = 26.484, p < .001$  as well as their interaction:  $F(2, 58) = 9.229, p < .001$  on the acceptability of the test sentences.

by animals, which are still seen as somewhat acceptable. Then there is a considerable gap before the marked inanimate objects, these being regarded as decidedly less normal. Given that the error bars do not overlap, there appears to be an effect of animacy manipulation on the acceptability of the sentences. As was the case in the other regions, the conditions without DOM (C4–C6) are judged to be quite equally acceptable by Valencian speakers overall.

Similarly to Majorcan, Valencians do not exhibit a clear preference when the direct object involves humans and presence/absence of DOM (C1 vs. C4). The respective z-scores differ only by the smallest of margins (0.001), indicating the participants' indifference. For the other two categories, the values show that Valencians lean towards the versions without DOM (although in case of animals only slightly). For C2 and C5 (animals) as well as C3 and C6 (inanimate things) the error bars do not overlap, indicating a possible effect of DOM manipulation on the acceptance of a given sentence.

For Valencian, the ANOVA calculations gave statistically significant results for both the effect of 'animacy' and DOM, as well as their interaction, thus confirming the high probability of their influence on the acceptability of the test sentences.<sup>19</sup>

### 6.1.2 Overall results for 'animacy' and a comparison of the four regions

Having discussed the results for each region separately, including the possible and probable effects of 'animacy' and DOM on the acceptability of sentences, we turn now to the overall results and a comparison of the four regions.

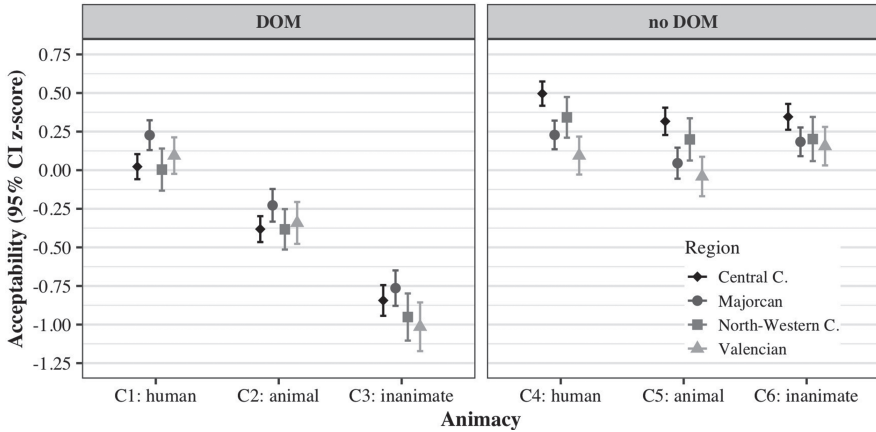
The comparison of the results of the four regions shows that all of them follow a similar general pattern of acceptance, suggesting no large regional effect on the acceptability of the sentences. Nonetheless, some variation can be found. The participants evidently favour human marked objects (C1) over animals (C2) and over inanimate objects (C3), whereas the conditions without DOM (C4–C6) all appear to have quite similar levels of acceptance. Interestingly, Majorcan and Valencian are the dialects that more freely accept DOM not only with humans but

---

<sup>19</sup> ANOVA results for Valencian:

1. Subject-Analysis for the effects of DOM:  $F(1, 41) = 40.534, p < .001$  and 'animacy':  $F(2, 82) = 27.768, p < .001$  as well as their interaction:  $F(2, 82) = 34.642, p < .001$  on the acceptability of the test sentences.

2. Item-Analysis for the effects of DOM:  $F(1, 29) = 91.891, p < .001$  and 'animacy':  $F(2, 58) = 18.248, p < .001$  as well as their interaction:  $F(2, 58) = 24.895, p < .001$  on the acceptability of the test sentences.



**Graph 5:** Overall comparison of the results for ‘animacy’.

also with animals. Overall, Majorcan shows the highest mean values, even for the inanimate objects (C3). This confirms to some extent what can be found in the literature on the subject. According to several authors, in Majorcan even inanimate direct objects with Differential Object Marking can be found.<sup>20</sup>

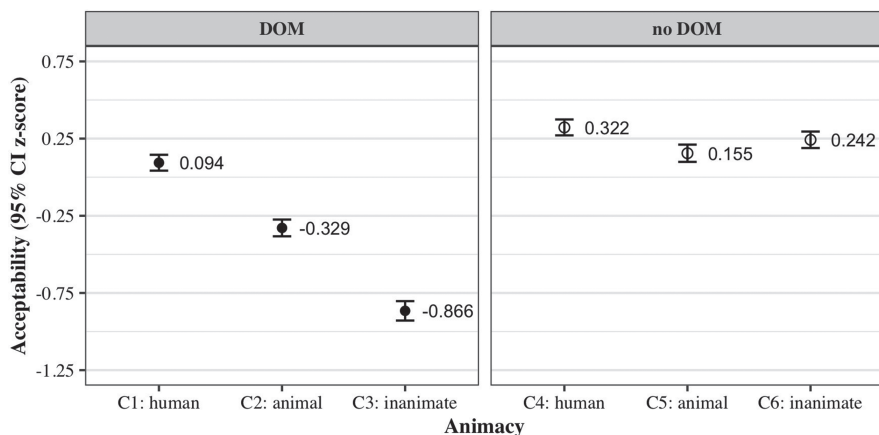
Central and North-Western Catalan behave very similarly to one another, not only in their lower acceptance of sentences with DOM, but also in their higher acceptance of those without (C4–C6). Furthermore, both dialects clearly favour the non-marked objects in every category, whereas Majorcan and Valencian display a different behaviour. As noted earlier, and illustrated in Graph 5, these latter two dialects do not seem to mind whether human direct objects are marked or not – they accept both versions equally (C1 and C4 are virtually at the same level).

A curious detail is the fact that all four dialects seem to consider non-marked animals (C5) slightly less acceptable than their human and inanimate counterparts (C4 and C6). This might be due to the animals chosen for the test sentences, although when designing the experiment, the intention was to avoid less common animals. This aspect could be explored further by dividing the animal category into subcategories and thus observing whether there was any correlation between animal type and acceptability level.

<sup>20</sup> Examples of these objects with DOM are found, among others, in Rohlfs (1971, 323): *a ses patates, else pelarem/else pelarem, a ses patates* and Wheeler/Yates/Dols (1999, 462): *A ses tovalloles, posa les dins es calaix*.

Generally speaking, and as it pertains to the acceptability of DOM, the four Catalan varieties appear to behave more or less as might be expected<sup>21</sup> – favouring the most animated objects. Nonetheless, it is worth noting that whereas the literature on Catalan DOM usually emphasizes the importance of ambiguity as a deciding factor in the use of DOM, these results suggest that it actually might not be so important, at least for the human object. None of the sentences tested during the experiment show any ambiguity whatsoever as to which elements hold the function of subject and direct object. Yet participants still judged the sentences with DOM to be highly acceptable – sometimes even as acceptable as the non-marked ones, as in the case of Majorcan and Valencian.

Before moving on to the second part of the experiment, a few details about the overall results for ‘animacy’ ought to be noted. In Graph 6 the mean z-scores represent the combination of the evaluations from all four regions, which resulted in a total of 240 participants, equalling 1,200 judgments per condition.



**Graph 6:** Overall results for ‘animacy’.

Unsurprisingly, the general shape and form of the overall distribution of mean values is similar to the regional ones. The overall results present the same scalar distribution for the marked objects (C1–C3), with human objects being seen as more acceptable than animals and inanimate ones. The error bars do not overlap, which is a strong indication of the possible effect of ‘animacy’ on perceived acceptability.

<sup>21</sup> From a general theoretical viewpoint, it seems that the higher up on the animacy-scale, the more frequently a direct object is marked (cf. Bossong 1991; Aissen 2003).

Regarding the three pairs (C1 and C4, C2 and C5, C3 and C6) globally, there is also a more or less clear preference for the unmarked object, though for the human object (C1 vs. C4) the predilection is not as pronounced as for the others. Looking at the respective error bars for all pairings, the fact that they do not overlap suggests an effect of DOM on the acceptability of a given sentence. An interesting finding in terms of this global distribution is the fact that C1 and C4 have very similar levels of acceptance, as indicated earlier. This implies that in nominal clauses with human direct objects, DOM is rather well accepted among Catalan speakers on the whole. In fact, they accept it nearly as much as the version that is prescribed by the normative grammar – the one without DOM. Thus, it can be concluded that in this type of sentence, participants did not mind whether or not the object was marked; they seemed to find both options normal.

For the overall analysis, the ANOVAs yielded statistically significant values, and thus confirmed the high probability of an effect of DOM and of ‘animacy’, as well as their interaction.<sup>22</sup> In this global study of the four regions, the ANOVAs were used to test for another effect: regional variance. Although, as seen in Graph 5 under the descriptive approach, some slight regional differences can be perceived, the statistical analysis does not show a significant value for this effect in both Subject- and Item-Analysis.<sup>23</sup> This implies, with very high probability, that the factor ‘region’ did not have a huge impact on the evaluations and hence that the acceptability of DOM in this type of sentence is not subject to many diatopic factors.

As discussed earlier, the figures shown here only include the results for the four most prevalent Catalan dialects and thus exclude Rossellonese and Alguerese. Considering their geographic situation, it is possible that with further data from these two regions there could be a higher diatopic impact on the acceptability of DOM. But at the moment this remains no more than speculation, and would have to be investigated with the help of more participants.

---

**22** ANOVA results for overall analysis:

1. Subject-Analysis for the effects of DOM:  $F(1, 236) = 253.784, p < .001$  and ‘animacy’:  $F(2, 472) = 149.916, p < .001$  as well as their interaction:  $F(2, 472) = 127.685, p < .001$  on the acceptability of the test sentences.

2. Item-Analysis for the effects of DOM:  $F(1, 29) = 246.266, p < .001$  and ‘animacy’:  $F(2, 58) = 49.692, p < .001$  as well as their interaction:  $F(2, 58) = 47.062, p < .001$  on the acceptability of the test sentences.

**23** ANOVA results for the overall analysis:

1. Subject-Analysis for the effect of ‘region’:  $F(3, 236) = 6.025, p < .001$  and its interaction with DOM and ‘animacy’:  $F(6, 472) = 1.693, p = 0.121$  on the acceptability of the test sentences.

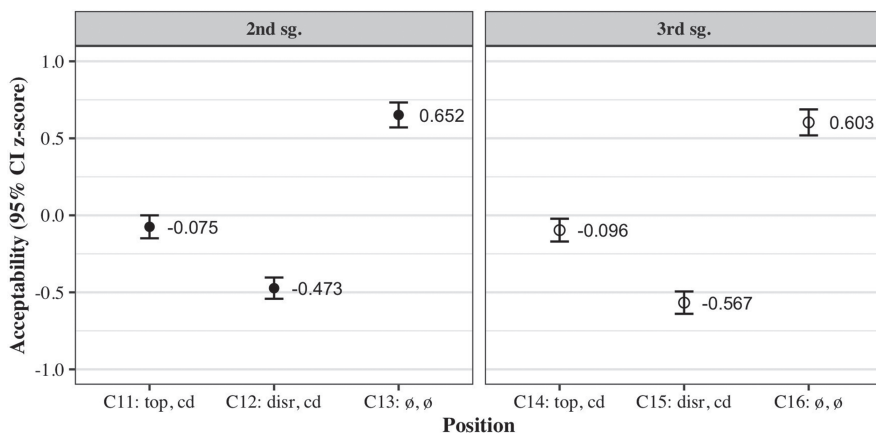
2. Item-Analysis for the effect of ‘region’:  $F(3, 87) = 1.816, p = 0.150$  and its interaction with DOM and ‘animacy’:  $F(6, 174) = 1.125, p = 0.349$  on the acceptability of the test sentences.

## 6.2 Position

As noted above, the second experiment part focused on ‘position’, and participants were presented with sentences of the following structure: *A tu et van vèncer/Et van vèncer a tu/Et van vèncer en la final del torneig d’escacs*. Here, the main manipulated factors were the position of the direct object given as a personal pronoun, and its corresponding clitic doubling combined with a change between 2<sup>nd</sup> and 3<sup>rd</sup> person singular pronouns.

### 6.2.1 Regional results for ‘position’

The graphs for these results are very similar in layout to those in the previous Section; the only difference lies in the horizontal axis, on which the corresponding conditions (C11–C16) were plotted with their respective combination of the manipulated characteristics.<sup>24</sup> First, in Graph 7 the results for Central Catalan are presented, where participants did not seem to show any concern for the change in personal pronouns. When comparing the respective pairs of conditions (C11 vs. C14, C12 vs. C15 and C13 vs. C16) very little difference can be seen in the levels of their mean z-scores, implying that the change between 2<sup>nd</sup> and 3<sup>rd</sup> person singular does not have any effect on the acceptability of a sentence.



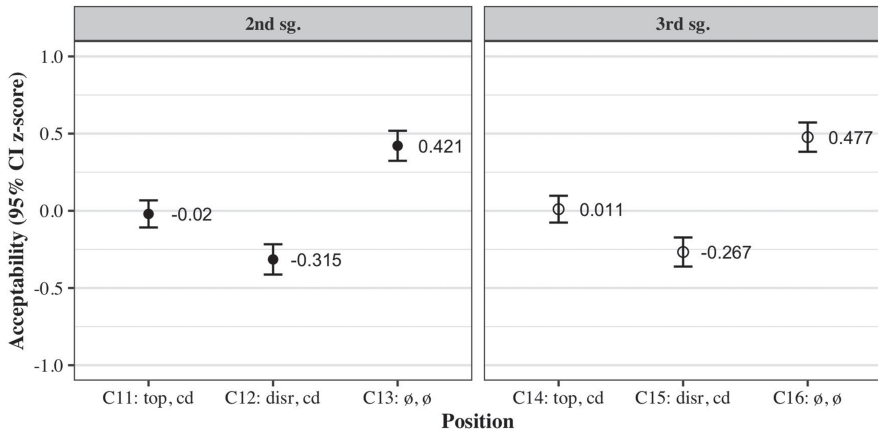
Graph 7: Central Catalan: Results for ‘position’.

<sup>24</sup> Abbreviations on the horizontal axis: topicalization (top), dislocation to the right (disr), clitic doubling (cd), absence of any type of dislocation and thus also clitic doubling (ø). The boxes show the split between 2<sup>nd</sup> and 3<sup>rd</sup> person singular pronouns.

On the other hand, when looking at the position of the direct object and the corresponding clitic doubling, there appears to be a certain effect on acceptability, since the values are clearly separated and their error bars do not overlap (comparing C11, C12, C13 and C14, C15, C16). Central Catalan speakers overwhelmingly prefer the sentences without any type of dislocation (C13 and C16). This in itself is not very surprising, considering that C13 and C16 represent the most baseline kind of sentence tested in this part of the experiment, consisting of short sentences without any type of modification. Of the two dislocations, Central Catalan speakers lean towards the sentences with topicalization, although they apparently accept both versions containing DOM to a somewhat limited degree.

The conducted ANOVAs indicate, as expected, that the change in the personal pronoun did not have an effect on the acceptability of the test sentences (as reflected by its non-significant result). However, the statistically significant results confirm the probability of an effect of the factor ‘position’.<sup>25</sup>

Continuing with the Majorcan data, the following Graph again plainly highlights that overall participants favoured the most basic type of test sentence



**Graph 8:** Majorcan: Results for ‘position’.

**25** ANOVA results for Central Catalan:

1. Subject-Analysis for the effects of ‘position’:  $F(2, 178) = 191.386, p < .001$  and ‘ppron’:  $F(1, 89) = 2.483, p = .119$  as well as their interaction:  $F(2, 178) = 0.487, p = .615$  on the acceptability of the test sentences.

2. Item-Analysis for the effects of ‘position’:  $F(2, 58) = 201.39, p < .001$  and ‘ppron’:  $F(1, 29) = 0.474, p = .497$  as well as their interaction:  $F(2, 58) = 0.316, p = .73$  on the acceptability of the test sentences.

without any manipulation. Nevertheless, the other two possible sentence structures show relatively high z-scores as well.

The change in personal pronoun does not seem to affect acceptability, given that the respective pairs of mean values are on very similar levels and their error bars overlap. Regarding the different types of position of the direct object, it is possible that there is a certain effect. Comparing only the two dislocations (both with DOM), Majorcan speakers show a slight preference for topicalization (C11 and C14) while they do not reject C12 and C15 as much as the Central Catalan speakers. This would confirm, at least partly, what was found in Escandell Vidal (2009) about DOM in Majorcan, although she points out that in this dialect dislocation to the right is even more accepted than it appears to be the case for the current study. This is very likely due to the fact that in the experiment presented here, the sentences contained only partial movements to the right and not prototypical dislocations to the extreme right of the sentence, due to the need for strict uniformity of sentence structures.<sup>26</sup>

By conducting ANOVAs, the possible effect of ‘position’ can be confirmed, given that a statistically significant result was found. Meanwhile, the change in personal pronoun and the interaction of the two factors did not seem to have an effect on the acceptability of the test sentences.<sup>27</sup>

Similarly, to the Majorcan and Central Catalan data, the speakers of North-Western Catalan prefer C13 and C16 without any manipulations, but the difference to the other two sentence variations is not as great as it was for the Central region. The average values, however, are sufficiently separate to assume the existence of an effect of the direct object’s position on the acceptability of the sentences (comparing C11, C12, C13 and C14, C15, C16). Both types of dislocation (with clitic doubling and DOM) seem to be deemed quite acceptable by the North-Western Catalan speakers, with a slight preference towards topicalization. As was the case for the first two dialects, the changed personal pronouns evidently appear to have no effect on the judgments of the participants.

---

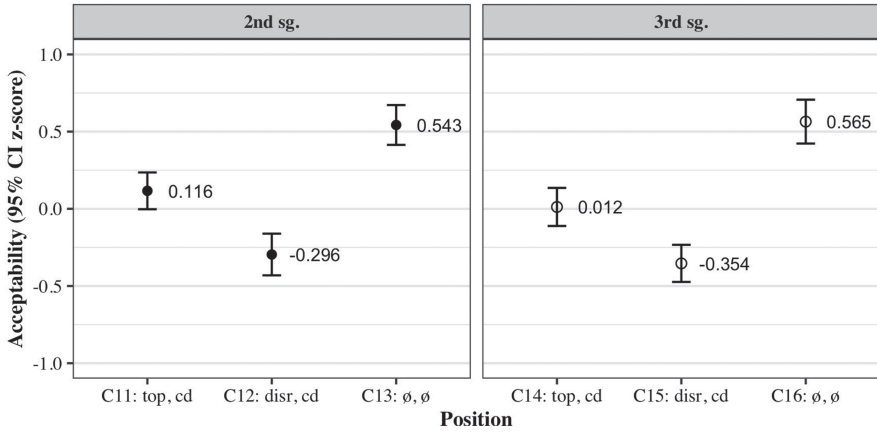
**26** A further experiment, including sentences with dislocations to the extreme right, would show more clearly to what degree Majorcans accept this type of dislocation.

**27** ANOVA results for Majorcan:

1. Subject-Analysis for the effects of ‘position’:  $F(2, 142) = 68.357, p < .001$  and ‘ppron’:  $F(1, 71) = 0.986, p = .324$  as well as their interaction:  $F(2, 142) = 0.029, p = .971$  on the acceptability of the test sentences.

2. Item-Analysis for the effects of ‘position’:  $F(2, 58) = 117.522, p < .001$  and ‘ppron’:  $F(1, 29) = 0.766, p = .389$  as well as their interaction:  $F(2, 58) = 0.034, p = .966$  on the acceptability of the test sentences.





**Graph 9:** North-Western Catalan: Results for ‘position’.

This is confirmed by the non-significant result that the ANOVAs yield for this factor. For the aspect ‘position’, then, the model confirms the very high probability of an effect on the acceptability of the sentences.<sup>28</sup>

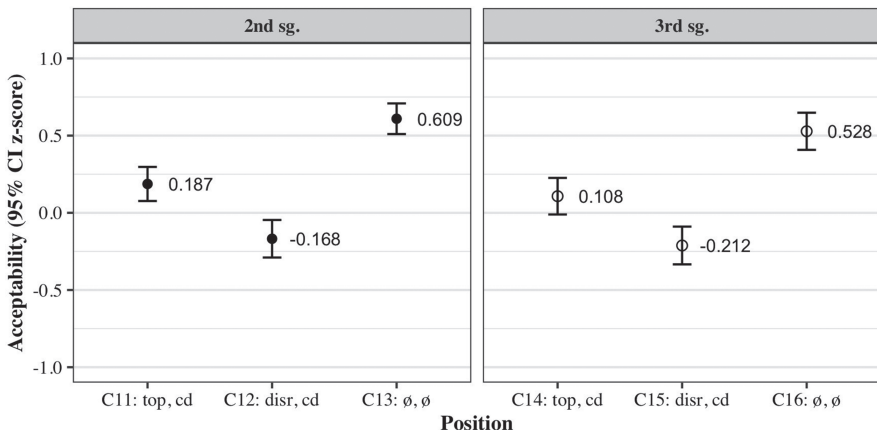
Valencian speakers, apparently, judged all conditions to be more or less acceptable, as can be seen in the following Graph. Again, the participants did not seem to pay attention to the change in personal pronoun, since the three respective pairs of mean values are at very similar levels.

On the other hand, ‘position’ did appear to have a certain influence on the acceptability of the sentences, given that in neither group the values, or their error bars, overlap. Finally, it seems that Valencian speakers accepted both types of dislocation (with clitic doubling and DOM) quite readily, although with a tendency towards topicalization. As was the case with the other regions, participants favoured C13 and C16, which is unsurprising bearing in mind that they represent short sentences without any manipulation.

**28** ANOVA results for North-Western Catalan:

1. Subject-Analysis for the effects of ‘position’:  $F(2, 70) = 40.835, p < .001$  and ‘ppron’:  $F(1, 35) = 0.63, p = .433$  as well as their interaction:  $F(2, 70) = 0.513, p = .601$  on the acceptability of the test sentences.

2. Item-Analysis for the effects of ‘position’:  $F(2, 58) = 69.143, p < .001$  and ‘ppron’:  $F(1, 29) = 0.544, p = .467$  as well as their interaction:  $F(2, 58) = 0.662, p = .52$  on the acceptability of the test sentences.



**Graph 10:** Valencian: Results for ‘position’.

Furthermore, both ANOVAs once again only confirmed the effect of the change in position on the acceptability of the test sentences.<sup>29</sup>

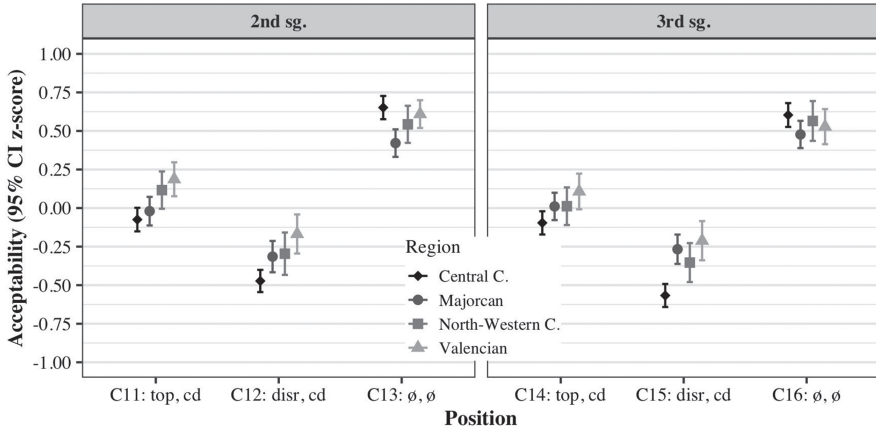
### 6.2.2 Overall results for ‘position’ and a comparison of the four regions

Having discussed the regional results concerning the factor ‘position’ separately, we will now take a look at the four dialects together. One of the most obviously notable things here is the fact that these distributions are not quite as uniform and continuous as those in the comparison graph for the ‘animacy’ results (Section 6.1.2, Graph 5), which suggests a possible regional effect.

As we have already seen in the regional analysis, the change in personal pronoun (between 2<sup>nd</sup> and 3<sup>rd</sup> person singular) did not seem to have any effect on the acceptability of the sentences, considering that the respective z-scores are all at very similar levels (this is even better illustrated in Graph 12, containing the global results).

<sup>29</sup> ANOVA results for Valencian:

1. Subject-Analysis for the effects of ‘position’:  $F(2, 82) = 44.876, p < .001$  and ‘ppron’:  $F(1, 41) = 1.67, p = .203$  as well as their interaction:  $F(2, 82) = 0.076, p = .926$  on the acceptability of the test sentences.
2. Item-Analysis for the effects of ‘position’:  $F(2, 58) = 77.204, p < .001$  and ‘ppron’:  $F(1, 29) = 0.473, p = .497$  as well as their interaction:  $F(2, 58) = 0.125, p = .883$  on the acceptability of the test sentences.



**Graph 11:** Overall comparison of the results for ‘position’.

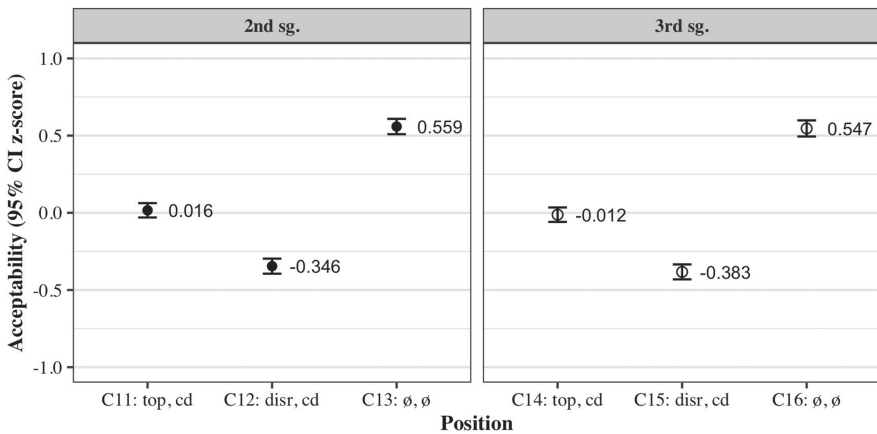
On the other hand, an interesting finding here is that all four major Catalan dialects concur as to which sentences they prefer the most: C13 and C16, both without either form of dislocation and both only differing in the form of the personal pronoun. The mean values for C16 are almost identical, whereas for C13 they are also very close. This predilection for the most basic sentence is, of course, not unexpected, although it still seems quite intriguing that the results of all four regions are so similar.

Concerning the other sentence variations (all conditions with DOM), there is slightly more regional variation. Graph 11 illustrates that Valencians, compared to the other speakers, tended to give the test sentences higher evaluations, whereas the Central Catalan participants, overall, rated them the lowest. However, the exceptions here are C13 and C16. Interestingly, for these two conditions the averages of Central Catalan surpass those of the other regions, whereas for every other condition they are clearly lower than the rest. In other words, the difference between the two conditions with low average values (C11 and C12) and C13 is far greater for Central Catalan than for the other regions. This is also true for the comparison of C16 to C14 and C15. Of all the regions, participants from Central Catalan appear to be the firmest in their preference for the sentences that are free of any type of dislocation of the direct object.

As for the conditions with topicalization (C11 and C14) and dislocation to the right (C12 and C15), Graph 11 illustrates that the distribution of Valencian and Central Catalan is more or less parallel, while Majorcan and North-Western Catalan present more variation. As can be seen, for C11 (topicalization) the North-Western z-scores are somewhat higher than those for Majorcan, whereas Majorcan presents slightly higher values for C15 (dislocation to the right). For C12

and C14, both dialects have very similar mean scores. A final point to raise here is the fact that of the two possibilities with dislocations (and DOM), all the regions clearly lean towards the sentences including topicalization. Again, there is a high probability that this is due in part to the fact that the dislocations to the right were not total, but only partial ones. Interestingly, these partial dislocations still attain a certain degree of acceptance, bearing in mind that they are not the prototypical dislocations to the right, for which there are numerous examples, as indicated in Rohlf (1971) and Escandell Vidal (2009), among others.

Following the comparison of the four regions, Graph 12 below depicts the overall mean values for the combination of the total of 240 participants throughout the four regions (equalling 1,200 judgments per condition). The Graph illustrates perfectly that the change in personal pronoun did not appear to have any effect on the acceptability of the individual conditions, given that the corresponding averages are virtually the same (C11 vs. C14, C12 vs. C15, C13 vs. C16). This, not surprisingly, is confirmed by the ANOVAs, which did not show a statistically significant value for the change in personal pronoun.<sup>30</sup>



Graph 12: Overall results for 'position'.

### 30 ANOVA results for overall analysis:

1. Subject-Analysis for the effects of 'position':  $F(2, 472) = 268.8, p < .001$ , 'ppron':  $F(1, 236) = 1.633, p = .203$  and 'region':  $F(3, 236) = 6.025, p < .001$  as well as the interaction between 'region' and 'position':  $F(6, 472) = 4.718, p < .001$  on the acceptability of the test sentences.
2. Item-Analysis for the effects of 'position':  $F(2, 58) = 239.729, p < .001$  and 'ppron':  $F(1, 29) = 0.367, p = .55$  and 'region':  $F(3, 87) = 20.872, p < .001$  as well as the interaction between 'region' and 'position':  $F(6, 174) = 7.343, p < .001$  on the acceptability of the test sentences.

On the other hand, looking at the position of the direct object, there does seem to be an effect. Within both groups (C11, C12, C13 and C14, C15, C16) the averages differ clearly and their error bars do not overlap. Hence, it is possible that the position of the direct object may have a certain influence on the acceptability of a given sentence. The sentences, it appears, were more acceptable when they contained a topicalization of the direct object than when there was a dislocation to the right. However, as can be seen in Graph 12, in overall terms, participants preferred the most basic version without dislocation.

These findings derived from the descriptive approach are confirmed by the statistical analysis by means of the two ANOVAs. The resulting values for the factor ‘position’ are statistically highly significant, suggesting that there is indeed such an effect on the evaluation of the sentences. The final aspect explored through the ANOVAs was the possible effect of the factor ‘region’ on the acceptability of the sentences. As implied in Graph 11 by the visible differences in the four distributions of the mean values, the statistically significant results of both ANOVAs confirm that the factor ‘region’ is highly likely to have an impact on the acceptability of a given sentence. Furthermore, the combination of ‘region’ and ‘position’ is confirmed to have had some influence on the perceptions of the various conditions presented to the participants.

## 7 Conclusions

The aim of this article has been to illustrate that Catalan in general, and its dialects in particular, are an independent part of the puzzle constituted by the individual types of DOM found in languages across the world. After a long period of being dismissed as a mere Spanish imposition, DOM is now beginning to be seen as an inherent part of the Catalan language itself. Nonetheless, there still are certain discrepancies between the uses permitted by Catalan normative grammars and the current situation found in Catalan dialects, as the experiment described in this article was able to highlight. Based on the results discussed in the previous Sections, it can be said that all Catalan dialects, to a greater or lesser extent, accept DOM in more contexts than the standard language commonly permits.

Furthermore, the factor ‘animacy’ appears to have quite a substantial impact on the acceptability of DOM with noun phrases, both for Catalan in general as well as for each of the dialects. In each case there was a pronounced scalar distinction in the perceived acceptability of the direct objects’ three animacy levels, with the human object reaching a considerably high level throughout the regions. In addition, in this type of sentence, and contrary to what standard Catalan dictates, it

appears that the total lack of ambiguity does not impede speakers' acceptance of DOM in the least. The only major regional variation in this part of the experiment was found for the human objects. Both Valencian and Majorcan seem to perceive the use of DOM with human direct objects in noun phrases as facultative, accepting its presence and absence at equally high levels. Central and North-Western Catalan, by contrast, generally prefer the non-marked version.

Concerning the second factor, 'position', on the whole the position of the direct object had an effect on the acceptability of the sentences, with all the dialects tending towards the most basic sentences as the most acceptable ones and topicalization as the more acceptable version containing DOM. Interestingly, in this part of the experiment, the regional differences were more pronounced than in the first part.

In conclusion, both main factors analyzed in this experiment, 'animacy' and 'position', appear to have an effect on the acceptability of DOM in the Catalan varieties, both on a general as well as on a regional level. Moreover, from a statistical point of view, the ANOVAs confirm the existence of both effects, as well as an interaction effect between DOM and 'animacy'. The third main factor, regional variance, was only found to be of statistical significance for 'position', thus indicating that in some contexts diatopic factors have an impact on the acceptability of DOM in Catalan, whereas in others they do not.

As mentioned in the opening sections of this paper, some of the literature suggests that Catalan and Spanish DOM share several characteristics. The results of the present experiment seem to indicate that this is true to a certain degree, especially for the three marked objects in the analysis of 'animacy', which show a similar distribution to Spanish.<sup>31</sup> However, the results also reaffirm the inherently independent evolution of Catalan DOM, considering that the three unmarked objects in the 'animacy' section behave in a wholly dissimilar way to Modern Spanish DOM (cf. Wall 2015). In this study the results suggested that in Modern Spanish, unmarked human objects in particular are considered not at all acceptable; and this, in turn, underlines the individuality of the Modern Catalan DOM configuration.

There are several ways in which the current research might be extended and complemented. Based on the existing data, the possible influence of some of the demographic aspects described could be explored, although for statistically rel-

---

**31** That is not to say that the Catalan configuration is solely due to Spanish influence, although geographic proximity suggests at least some interaction between the two languages (as can be expected in any situation of adjacent languages). Rather, it seems that the evolution of Catalan, for this type of marked objects, is slowly going in the same direction as that of Spanish, reflecting a perfectly normal evolution of a DOM system.

evant conclusions a larger dataset from each region would be required. Indeed, in terms of collecting more data, further material on Rossellonese and Algerese would help to complete the whole picture of DOM in the Catalan dialects. This would be especially interesting, taking into account that in these regions it might be possible to observe influences from other Romance languages which do not show DOM, or which do so only on a more limited scale. As noted earlier on, further subcategories of the animacy of direct objects, such as multiple animal categories or further inanimate categories, could also be investigated, as well as other categories for the second part of the experiment, including total dislocations to the right. In conclusion, it is safe to say that there is still far more to explore towards gaining a complete picture of DOM in Catalan, both in general and in all its different dialects.

## Bibliography

- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21 (2003), 435–483.
- AVL = Acadèmia Valenciana de la Llengua, *Gramàtica normativa valenciana*, València, Publicacions de l'Acadèmia Valenciana de la Llengua, 2006.
- Badia Margarit, Antoni M., *Gramàtica catalana*, Madrid, Gredos, 1962.
- Blasco Ferrer, Eduardo, *La lingua sarda contemporanea. Grammatica del logudorese e del campidanese. Norma e varietà dell'uso. Sintesi storica*, Cagliari, Della torre, 1986.
- Boladeras Taché, Josep M., *Les preposicions en català nord-occidental*, Lleida, Pagès, 2011.
- Bosson, Georg, *Der präpositionale Akkusativ im Sardischen*, in: Winkelmann, Otto/Braisch, Maria (edd.), *Festschrift für Johannes Hubschmid zum 65. Geburtstag*, Bern/München, Francke, 1982, 579–599.
- Bosson, Georg, *Differential Object Marking in Romance and beyond*, in: Wanner, Dieter/Kibbee, Douglas A. (edd.), *New analyses in Romance linguistics. Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, April 7–9, 1988*, Amsterdam/Philadelphia, John Benjamins, 1991, 143–170.
- Colón Domènech, Germà, *Català. Del general al particular*, in: Kabatek, Johannes/Pusch, Claus D. (edd.), *Variació, poliglòssia i estàndard. Processos de convergència i divergència lingüístiques en català, occità i basc*, Aachen, Shaker, 2009, 21–32.
- Costa Carreras, Joan, *Norma i procés d'estandardització*, in: Pradilla, Miquel Àngel (ed.), *Societat, llengua i norma. A l'entorn de la normativització de la llengua catalana*, Benicarló, Alambor, 2001, 41–55.
- Escandell Vidal, Victoria, *Differential Object Marking and topicality. The case of Balearic Catalan*, *Studies in Language* 33:4 (2009), 832–885.
- Fabra, Pompeu, *Sobre diferents problemes pendents en l'actual català literari*, *Anuari de l'Institut d'Estudis Catalans MCMVII* (1908), 352–369.
- Fabra, Pompeu, *Gramàtica catalana*, Barcelona, Institut d'Estudis Catalans, <sup>7</sup>1933.

- Institut d'Estudis Catalans, *Gramàtica de la llengua catalana*, Barcelona, Institut d'Estudis Catalans, 2016.
- Irimia, Monica A./Pineda, Anna, *Differential Object Marking and scales. Insights from Romance diachrony*, Proceedings of the Linguistic Society of America 4:57 (2019), 1–15, doi. org/10.3765/plsa.v4i1.4561 [last access: 29.04.2019].
- Kremnitz, Georg (ed.), *Sprachen im Konflikt. Theorie und Praxis der katalanischen Soziolinguisten*, Tübingen, Gunter Narr, 1979.
- Lagarde, Christian, *De la norma a la queta. L'evolució de la normativització a Catalunya*, in: Kabatek, Johannes/Pusch, Claus D. (edd.), *Variació, poliglòssia i estàndard. Processos de convergència i divergència lingüístiques en català, occità i basc*, Aachen, Shaker, 2009, 313–326.
- Meier, Harri, *O problema do acusativo preposicional no catalão*, Boletim de Filologia 8 (1945), 237–260.
- Moll, Francesc de B., *Gramàtica catalana. Referida especialment a les Illes Balears*, Palma de Mallorca, Moll, 1979.
- Müller, Bodo, *Das morphemmarkierte Satzobjekt der romanischen Sprachen (Der sogenannte präpositionale Akkusativ)*, Zeitschrift für romanische Philologie 87:5 (1971), 477–519.
- Par, Anfós, *Sintaxi catalana. Segons los escrits en prosa de Bernat Metge (1398)*, Halle, Niemeyer, 1923.
- Pensado, Carmen, *La creación del complemento directo preposicional y la flexión de los pronombres personales en las lenguas románicas*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 179–233.
- Rohlf's, Gerhard, *Autour de l'accusatif prépositionnel dans les langues romanes*, Revue de Linguistique Romane 35 (1971), 312–334.
- Sancho Cremades, Pelegrí, *La categoria preposicional*, Valencia, Universitat de València, 1995.
- Sancho Cremades, Pelegrí, *La preposició i el sintagma preposicional*, in: Solà, Joan, et al. (edd.), *Gramàtica del català contemporani*, vol. 2, Barcelona, Empúries, 2008, 1689–1796.
- Vallverdú, Francesc, *La normalització lingüística a Catalunya*, Barcelona, Laia, 1979.
- von Heusinger, Klaus/Kaiser, Georg A., *The evolution of Differential Object Marking in Spanish*, in: von Heusinger, Klaus/Kaiser, Georg A./Stark, Elisabeth (edd.), *Proceedings of the workshop "Specificity and the Evolution/Emergence of Nominal Determination Systems in Romance"*, Konstanz, Fachbereich Sprachwissenschaft der Universität Konstanz, 2005, 33–69.
- Wall, Albert, *The role of variation in the processing of Differential Object Marking (DOM) in Spanish*, Poster at the 48<sup>th</sup> annual meeting of the SLE, Leiden, 2015.
- Wheeler, Max W./Yates, Alan/Dols, Nicolau, *Catalan. A comprehensive grammar*, London/New York, Routledge, 1999.
- Zeugin, Senta, *La Marca Diferencial del Objeto en las variedades catalanas. Animación y posición del objeto directo como factores de influencia en la aceptabilidad de la MDO*, MA Thesis, University of Zurich, 2017.



Diego Romero Heredero

# Telicity and Differential Object Marking in the history of Spanish

**Abstract:** This paper explores the influence of telicity on Differential Object Marking (DOM) from a diachronic perspective. Previous research has frequently pointed to telicity as one of the verbal factors that trigger DOM in Spanish. However, this assumption has not been tested empirically. The present study provides a corpus analysis examining the question of whether the assumed impact of telicity on DOM in Modern Spanish, which has been observed in synchronic studies, also holds for earlier stages of Spanish. The corpus-based study covers the 14<sup>th</sup>, 16<sup>th</sup> and 20<sup>th</sup> centuries, and is the first empirical study that addresses the relationship between telicity and DOM with human direct objects. The results challenge previous analyses, suggesting that telicity has no significant impact on DOM in the periods surveyed. These findings lead us to conclude that telicity alone does not trigger DOM, which raises the question of whether it is possible that the effect of telicity observed in previous studies could be the consequence of the interaction of telicity with other verbal factors such as affectedness or agentivity.

**Keywords:** aspectual class, corpus analysis, definiteness, Differential Object Marking (DOM), language change, lexical aspect, telicity, verbal factors

---

**Acknowledgements:** This paper was presented at the Workshop “Differential Object Marking in Spanish – diachronic change and synchronic variation” (University of Zurich, Zurich, June 4<sup>th</sup>–5<sup>th</sup>, 2018). I would like to thank the organizers of the workshop, Larissa Binder, Johannes Kabatek, Philipp Obrist and Albert Wall, for the great space for discussion they gave me, and the audience of the workshop for their comments. I would also like to thank Javier Caro Reina, Marco García García, Klaus von Heusinger and Manuel Leonetti for their comments and invaluable support, without which this article would certainly not have been possible. The research for this paper has been funded by the Deutsche Forschungsgemeinschaft (DFG – German Research Foundation) SFB 1252 (Project-ID 281511265) *Prominence in language* within the project B04, *Interaction of nominal and verbal features for Differential Object Marking*, at the University of Cologne.

---

**Diego Romero Heredero** University of Cologne, e-mail: d.romeroheredero@uni-koeln.de

# 1 Introduction

Differential Object Marking (DOM) is a grammatical phenomenon conditioned by the interaction of various factors (cf. García García 2018 for an overview). These factors are not only related to the noun phrase (NP) having the function of direct object, but also to other aspects of the context in which it appears. In the early 1950s, Fernández Ramírez had already pointed out with regard to Spanish that “the nature of the verb and the nature of the noun or pronoun that functions as a direct object influence this phenomenon” (1951, 151). In this regard, Laca (2006) distinguishes between local and global factors. Local factors are those concerned with the properties of the NP that occupies the object position, while global factors are those concerned with the context in which it appears (Laca 2006, 430–431). More precisely, the animacy of the direct object, its degree of referentiality and the specificity of the referent are considered local factors (Company Company 2002; Leonetti 2004; Laca 2006). On the other hand, global factors are those related to the semantics of the verb, such as telicity, affectedness and agentivity (Torrego 1998; 1999; von Heusinger 2008; 2011; García García 2014; Romero Heredero *in press*); the presence of secondary predication referring to the object (Laca 2006); the topicality of the object and clitic doubling (Melis 1995; Torrego 1998; Laca 2006; Ormazábal/Romero 2013).

The influence of local factors has been widely documented in Spanish, not only from a synchronic point of view, but also from a diachronic perspective (Company Company 2002; Laca 2006; von Heusinger 2008). In contrast, global factors, and more specifically those related to the lexical properties of the verb, have not received as much attention. Nevertheless, while some studies deal with affectedness and agentivity, establishing the extent to which these factors influence DOM (Pottier 1968; von Heusinger 2008; 2011; García García 2018), telicity continues unattended, with the exception of the work of Torrego (1998; 1999), which will be presented in Section 2.

The aim of this paper is to shed light on the factor of telicity drawing on diachronic data. It will be discussed whether telicity has been a trigger for DOM in earlier stages of Spanish, as assumed for Modern Spanish, and whether the diachronic evolution of DOM supports the claims about the effect of telicity at present. Thus, the impact of telicity on DOM with human direct objects is addressed for the first time from a diachronic perspective and with an empirical methodology.<sup>1</sup>

The paper is structured as follows: Section 2 deals with the relationship between lexical aspect and case marking, which has been studied in many lan-

---

<sup>1</sup> For a more detailed version of the analysis that also includes affectedness and the interaction between telicity and affectedness, cf. Romero Heredero (*in press*).

guages and, more specifically, addresses the effect of telicity on DOM which has been suggested for Spanish by previous studies; Section 3 presents the diachronic corpus study conducted for the 14<sup>th</sup>, 16<sup>th</sup> and 20<sup>th</sup> centuries; finally, Section 4 summarizes all the findings derived from the corpus study and the most important ideas raised in the discussion about the influence of telicity on DOM.

## 2 Telicity, case-marking and DOM

Telicity is a property of the verbal phrase (VP) implying the existence of an end point, after which the event designated by the predicate terminates or does not continue to take place (Verkuyl 1972; Dowty 1979; Tenny 1994; Krifka 1998). Hopper and Thomson (1980, 252) point to telicity as one of the factors linked to a high degree of transitivity in the clause, arguing that an action described by a telic VP is more effectively transferred to the object than one lacking an endpoint.

The relationship between case-marking and telicity or, more broadly, between case-marking and aspect, is a phenomenon that is well documented in many languages (Richardson 2012, 965). This relationship has been suggested for Slavic languages (Russian, Borer 2005, among others; Belarusian, Ukrainian, Czech, Slovak, Polish, Bosnian/Croatian/Serbian, Richardson 2007), Germanic languages (German, Leiss 2000, among others; Icelandic, Svenonius 2002), and even more prominently, for Finno-Ugric languages (Finnish, Kiparsky 1998, among others; Estonian, Tamm 2007, among others; Hungarian, Csirmaz 2006). Although, to a lesser extent, the case-aspect relationship has also been defended for other languages such as Scottish Gaelic (Ramchand 1997), Hindi (Mohanani 1984) or Spanish (Torrego 1998; 1999), among others.

One of the most frequently cited languages that exhibits this case-aspect relationship is Finnish, which presents an accusative-versus-partitive opposition for the internal argument which seems to be linked to an aspectual contrast (Richardson 2012, 965). The nature of that opposition has been described in terms of boundedness, telicity, and sometimes both. Kiparsky's examples in (1) exemplify this relationship. In Finnish the accusative has been characterized in terms of resultativity, as illustrated in (1) by the verb *to shoot*. This verb can assign different cases, giving rise to different aspectual interpretations. While in (1a) the verb has a partitive object and denotes an activity ('to shoot at'), in (1b) the verb is followed by an accusative object and describes an accomplishment ('to shoot dead') (Kiparsky 1998, 266–267). This example shows that accusative case in Finnish is associated with a telic character of the event, while partitive case is related to an atelic reading.

- (1) Case-marking in Finnish
- a. *Ammu-i-n karhu-a*  
shoot-PST-1SG bear-PART  
'I shot at the/a bear.'
  - b. *Ammu-i-n karhu-n*  
shoot-PST-1SG bear-ACC  
'I shot the/a bear.'

(Kiparsky 1998, 266–267)

With regard to Spanish, Torrego (1998; 1999) are the only studies devoted to the interaction between DOM and lexical aspect. Torrego (1999) describes the relationship between DOM and telicity<sup>2</sup> according to the assumptions in (2).

- (2) Interaction between telicity and DOM in Spanish
- a. Telic verbs impose DOM on their direct object.
  - b. DOM triggers a telic interpretation on atelic verbs.

(Torrego 1999, 1787–1790)

The first assumption (2a) refers to the requirement of DOM by telic verbs. Following Torrego, telic verbs, which comprise achievements (e.g. *encontrar* 'to find') and accomplishments (e.g. *construir* 'to build'), describe delimited situations, and this delimitation confers them a specific character. The direct object of telic verbs moves to a certain syntactic position which implies the interpretation of the NP as the boundary of the event and, therefore, as specific. In that position the object receives a certain type of case that implies *a*-marking in Spanish. Therefore, telic verbs require DOM due to the specific character of the situations they describe and, for this reason, intrinsically telic verbs, such as *insultar* 'to insult', obligatorily require their direct object to be marked with DOM, as can be observed in (3), where the sentence without *a*-marking would be ungrammatical.

- (3) DOM with telic verbs  
*Marta insultó \*(a) un compañero.*  
'Marta insulted a colleague.'

(Torrego 1999, 1787)

The second assumption (2b) is aimed at the interpretation of atelic verbs. In contrast to telic verbs, these verbs do not require DOM. However, their interpreta-

---

<sup>2</sup> Torrego refers to the inherent telicity of certain aspectual classes of verbs, not to telicity as a property of the VP.

tion changes if the direct object is differentially *a*-marked. Following Torrego, *a*-marking forces a reading of the object as specific and this implies a telic reading of the event. The example in (4a) shows that activity verbs, such as *besar* ‘to kiss’, do not require DOM. Nevertheless, the reading is telic when the direct object is *a*-marked, as can be observed in (4b), where the verb admits a temporal prepositional phrase (PP) introduced by the preposition *en* ‘in’ (telicity test). This reading is not obtained in (4c), where the temporal PP is rejected.

- (4) DOM with atelic verbs
- a. *Besaron (a) un niño.*
  - b. *Besaron a un niño en un segundo.*
  - c. \**Besaron un niño en un segundo.*
- ‘They kissed a boy (in a second).’

(Torrego 1999, 1788–1789)

The two assumptions presented above represent the only systematic attempt so far to describe the impact of telicity on DOM in Spanish. However, they are based on a rather small number of examples and subsequent studies have suggested that speakers’ intuitions do not always support them (Delbecque 2002, 90; Fábregas 2013, 25). Furthermore, it should be noted that these assumptions have not been empirically tested. Moreover, they only address the issue from a synchronic perspective.

The following section provides a diachronic corpus analysis investigating the impact of telicity on DOM from 14<sup>th</sup> to 20<sup>th</sup>-century European Spanish.

### 3 Diachronic corpus analysis

In order to study the effect of telicity on DOM in the 14<sup>th</sup>, 16<sup>th</sup> and 20<sup>th</sup> centuries, I conducted a corpus analysis based on the *Corpus del Diccionario histórico de la lengua española* (CDH). Section 3.1 presents the hypotheses according to the assumptions of the previous studies described in Section 2. Section 3.2 is dedicated to the design of the study. Section 3.3 contains the results according to definiteness, telicity and aspectual class of the verbs. Finally, Section 3.4 discusses the results.

### 3.1 Hypothesis

As mentioned in Section 1, some diachronic studies on factors such as referentiality or affectedness have shown that their impact on DOM was already present in previous stages of Spanish. However, telicity has been studied only from a synchronic perspective so far. This leads to the question of whether this factor also had any impact on previous stages of Spanish. To address this question, I have conducted a corpus study encompassing the 14<sup>th</sup>, 16<sup>th</sup> and 20<sup>th</sup> centuries with the hypothesis reported in (5).

(5) Hypothesis for DOM and telicity

The occurrence of DOM is more frequent with human direct objects of telic predicates than with human direct objects of atelic predicates.

This hypothesis is based on the notion of telicity as a property of the VP, as introduced in Section 2. This remark is relevant since previous studies that have addressed the relationship between telicity and DOM have generally based their hypotheses on a more restricted concept of telicity that refers only to the semantics of the verb (inherent telicity). The problem with this approach arises when categorizing certain verbs according to their inherent telicity in terms of Vendler (1957) (cf. Marín 1999 for discussion). In (6) the verb *to run* can be considered either atelic (activity verb) or telic (accomplishment verb) depending on the argument structure it exhibits.

(6) *To run* with different argument structure

- a. *Mary runs.* (one argument; activity)
- b. *Mary runs a mile.* (two arguments; accomplishment)

Considering this problem, it seems safer to take the definition of telicity as a property of the VP, and to classify all the VPs according to their behaviour in the test that has traditionally been used to distinguish telic from atelic predicates. As can be seen in (7), those VPs that admit a PP headed by the preposition *in* with a delimiting function will be considered telic, while those that do not admit it will be classified as atelic.

(7) Telicity test

- a. \**Mary runs in an hour.*
- b. *Mary runs a mile in an hour.*

### 3.2 Study design

The analysis is based on the *Corpus del Diccionario histórico de la lengua española* (CDH). It comprises the entries corresponding to the *CDH nuclear* as well as a selection of texts from both the *Corpus diacrónico del español* (CORDE) and the *Corpus de referencia del español actual* (CREA). The CDH allows for the search of single words (or a combination of words) according to word class, author, work, time span, text type, country and text subject. Unfortunately, this corpus is only partially annotated and, although it allows searching by word type, it is not possible to search for specific syntactic patterns such as DOM.

The search is restricted to European Spanish of the 14<sup>th</sup>, 16<sup>th</sup> and 20<sup>th</sup> centuries. The sources for European Spanish are extensive enough for a diachronic study since they contain a total of 7,745,250 words for the 14<sup>th</sup> century, 49,797,748 words for the 16<sup>th</sup> century and 113,509,174 words for the 20<sup>th</sup> century. The time span selection is motivated by Octavio de Toledo's (2016, 61–64) observation that the 18<sup>th</sup> and 19<sup>th</sup> centuries are less represented than the rest of the periods in the CORDE and, consequently, in the CDH. Moreover, the same author argues that these centuries constitute a period of great linguistic variation contrary to what had been previously defended (cf. Kabatek 2016 for discussion). With this in mind, it has been decided to avoid them and to select three periods whose representation in the corpus is wider and which present more internal stability.

The study is based on 3,200 instances following a 2x2x3 factorial design that takes into account telicity, definiteness and century as independent variables, and the presence/absence of DOM as dependent variable. The distribution of instances according to the independent variables has been structured as follows (cf. Table 1): as far as telicity is concerned, half of the instances contain a telic VP and the other half show an atelic VP. With regard to definiteness, 75% of the cases have a definite direct object NP and the remaining 25% contain an indefinite direct object NP. Finally, regarding the temporal distribution, the 14<sup>th</sup> century represents 20% of instances; the 16<sup>th</sup> century 30%; and the 20<sup>th</sup> century 50%. This configuration allows an easy comparison between the results obtained for telic and atelic VPs, being this the main goal of the study, since both categories are equally represented. As regards the distribution of cases according to definiteness and century, the configuration reflects the availability of instances found in the corpus for the different categories.

All instances extracted from the corpus to complete the distribution described above contain a full definite or indefinite human direct object NP in postverbal position (SVO). Thus, collective nouns, animate non-human NPs, inanimate NPs, proper names, bare nouns and left dislocations are excluded.

**Table 1:** Distribution of instances extracted from the CDH according to century, telicity and definiteness.

		14 <sup>th</sup> century	16 <sup>th</sup> century	20 <sup>th</sup> century	TOTAL
[+ telic]	definite NP	240	360	600	1,200
	indefinite NP	80	120	200	400
[- telic]	definite NP	240	360	600	1,200
	indefinite NP	80	120	200	400
TOTAL		640	960	1,600	3,200

The collection of the 3,200 examples is based on the selection of ten inherently telic verbs and ten atelic verbs, following Vendler's classification. This selection has not implied the assignment of the instances of each verb to a certain category, but it has been just the starting point, since the telicity of each case has been evaluated individually using the test described in Section 3.1. Then, the search for instances has been carried out, starting from those verbs, until achieving the pre-established quantity of instances for each category.<sup>3</sup> When the first ten verbs have not been enough to reach this goal, other verbs that meet the necessary conditions have been added (cf. Appendix for a complete list of the verbs used for the analysis). Possible changes in meaning or argument structure of the verbs over time have been taken into account using the *Diccionario del castellano del siglo XV en la Corona de Aragón* (DiCCA-XV). Further, it has been checked that the contribution of examples of each verb amounts to a maximum of 10%, thus ensuring that each category is comprised of at least 10 verbs. This measure avoids a situation where the result corresponding to each category reflects the behaviour of single verbs that have higher representation in the corpus.

Having obtained all the examples, they have been described individually according to the presence or absence of DOM (dependent variable) and the aspectual class of the verb (inherent telicity). With regard to the latter, it has been annotated according to the most common argument structure of each verb, taking into account the potential problems this may entail (cf. Section 3.1). However, this information may be interesting to check the assumptions of previous studies.

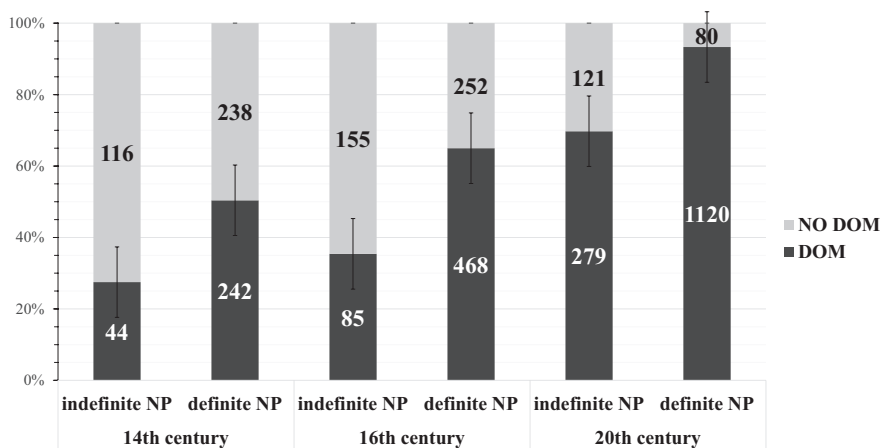
<sup>3</sup> The concept of "category" in this context refers to the different sets of examples that are generated by the factorial design of the search, e.g. definite direct objects in telic VPs from the 16<sup>th</sup> century.



### 3.3 Results

This section describes the results of the corpus-based study according to definiteness, telicity and aspectual class. The results are presented in five different figures which allow us to observe the evolution of the impact on DOM of the different factors.

The first part of this section is devoted to definiteness. Although the hypothesis of this work does not directly refer to this factor, definiteness is essential in order to understand the evolution of DOM in Spanish. In addition, another reason for including the results of definiteness here is to provide an overview of the evolution of DOM which allows us to check whether this data points in the same direction as the results obtained in previous studies. Figure 1 illustrates the occurrence of DOM according to century (14<sup>th</sup>/16<sup>th</sup>/20<sup>th</sup>) and definiteness (definite NP/indefinite NP). With regard to indefinite direct object NPs, the presence of DOM shows a relative frequency of 28% (44/160) in the 14<sup>th</sup> century, 35% (85/240) in the 16<sup>th</sup> century and 70% (279/400) in the 20<sup>th</sup> century. As far as definite direct object NPs are concerned, the occurrence of DOM has a relative frequency of 50% (242/480) in the 14<sup>th</sup> century, 65% (468/720) in the 16<sup>th</sup> century and 93% (1120/1200) in the 20<sup>th</sup> century. Although there has been a general increase in the presence of DOM over time, this data shows that definite human direct object NPs have always favoured the presence of DOM over indefinite human direct object NPs.



**Figure 1:** DOM according to century and definiteness.

Note also that these numbers largely confirm the results obtained in previous diachronic corpus-based studies (Company Company 2002; Laca 2006; von Heusinger/Kaiser 2011).

Figures 2 and 3 present the results for telicity according to definite and indefinite NPs, respectively. In the case of definite NPs, the relative frequency of DOM with regard to telicity is distributed as follows: the occurrence of DOM shows a relative frequency of 54% (129/240) with atelic predicates and 47% (113/240) with telic predicates in the 14<sup>th</sup> century. In the 16<sup>th</sup> century, the use of DOM registers a relative frequency of 66% (237/360) for atelic predicates and 64% (231/360) for telic predicates. Finally, the presence of DOM has a relative frequency of 95% (568/600) for atelic predicates and 92% (552/600) for telic predicates in the 20<sup>th</sup> century. The occurrence of DOM increases over time as shown in the results according to definiteness. However, Figure 2 does not show any significant effect of telicity on DOM with definite direct objects in any of the three periods.

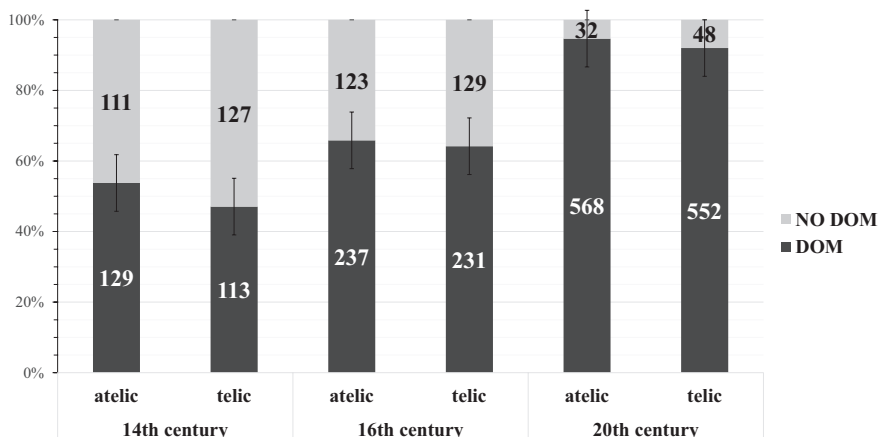


Figure 2: DOM with definite NPs according to century and telicity.

Examples of the lack of DOM from 20<sup>th</sup>-century Spanish with definite human direct objects are given in (8), where both a direct object of an atelic VP, such as *recordar la hortelana* ‘to remember the gardener’, and a direct object of a telic VP, such as *abatir sus caciques* ‘to overthrow their chiefs’ are not differentially marked.

(8) Definite NPs without DOM in 20<sup>th</sup>-century Spanish

- a. *¿Recuerda usted la pobre hortelana enferma que vimos en la ermita aquella tarde?* (atelic predicate)

‘Do you remember the poor sick gardener we saw at the hermitage that afternoon?’

(20<sup>th</sup> century, Blasco Ibáñez, *Entre naranjos*)

- b. *La civilización que [...] abatió sus caciques.* (telic predicate)

‘The civilization that [...] overthrew their chiefs.’

(20<sup>th</sup> century, Lopetegui/Zubillaga, *Historia de la Iglesia en la América española desde el descubrimiento hasta el siglo XIX*)

Turning to indefinite NPs, i.e. to Figure 3, the relative frequency of DOM is distributed as follows: the occurrence of DOM has a relative frequency of 35% (28/80) with atelic predicates and 20% (16/80) with telic predicates in the 14<sup>th</sup> century. In the 16<sup>th</sup> century, the use of DOM registers a relative frequency of 41% (49/120) for atelic predicates and 30% (36/120) for telic predicates. Finally, the presence of DOM shows a relative frequency of 69% (137/200) for atelic predicates and 71% (142/200) for telic predicates in the 20<sup>th</sup> century. As was observed in the case of definite direct objects, we can see the increase in the relative frequency of DOM with indefinite direct objects over time. Furthermore, in the case of indefinite NPs, the 14<sup>th</sup> and 16<sup>th</sup> centuries show a slight tendency to favour the occurrence of DOM in atelic VPs, however the effect is not significant. Consequently, the assumed hypothesis of the influence of telicity on DOM could not be confirmed.

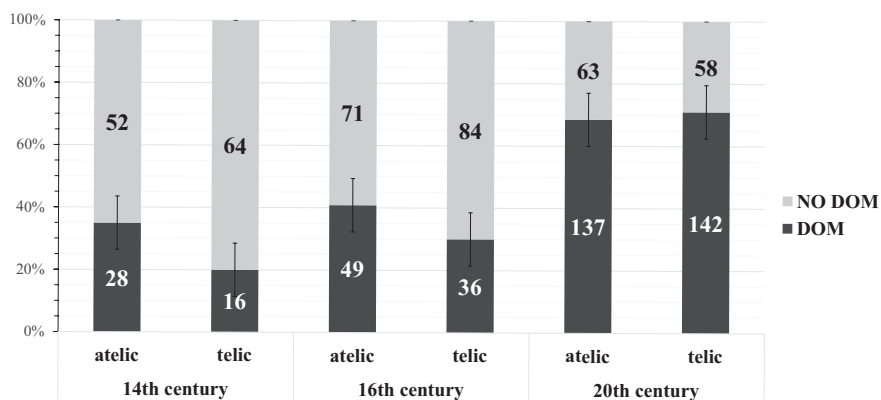


Figure 3: DOM with indefinite NPs according to century and telicity.

Examples of the presence of DOM from 14<sup>th</sup>-century Spanish with indefinite human direct objects are given in (9), where all examples present *a*-marking in their direct object regardless of whether the predicate is telic or atelic.

(9) Indefinite NPs with DOM in 14<sup>th</sup>-century Spanish

- a. *Juan Núñez amava a un cavallero que desían Gonçalo Gómez.* (atelic predicate)  
 ‘Juan Núñez loved a knight who was called Gonçalo Gómez.’  
 (14<sup>th</sup> century, Anonymous, *Crónica del muy valeroso rey don Fernando el cuarto*)
- b. *En la manyana ella encontro a vn ciudadano antigo.* (telic predicate)  
 ‘In the morning she found an old citizen.’  
 (14<sup>th</sup> century, Fernández de Heredia, *Traducción de Vidas paralelas de Plutarco, III*)

Figures 4 and 5 summarize the results for aspectual class according to definite and indefinite NPs, respectively. It is important to remember that the number of instances contained in each category is not balanced in this case since aspectual class is only addressed indirectly in this study. With regard to definite NPs, the relative frequency of DOM is distributed as follows: the occurrence of DOM presents in the 14<sup>th</sup> century a relative frequency of 61% (43/70) with states, 49% (86/176) with activities, 44% (28/64) with accomplishments and 50% (85/170) with achievements. In the 16<sup>th</sup> century, the use of DOM exhibits a relative frequency of 64% (63/98) for states, 65% (179/277) for activities, 60% (61/102) for accomplishments and 68% (165/243) for achievements. Lastly, the presence of DOM registers a relative frequency of 90% (150/166) with states, 96% (452/472) with activities, 93% (125/134) with accomplishments and 92% (393/428) with achievements in the 20<sup>th</sup> century. Although the increase in DOM over time is again visible in Figure 4, the aspectual class shows no significant effect on the occurrence of *a*-marking with definite objects.

Examples of the lack of DOM from 20<sup>th</sup>-century Spanish with definite human direct objects are given in (10), where both the direct object of an activity verb, such as *guiar* ‘to guide’, and the direct object of an achievement verb, such as *hallar* ‘to find’, are not differentially marked.

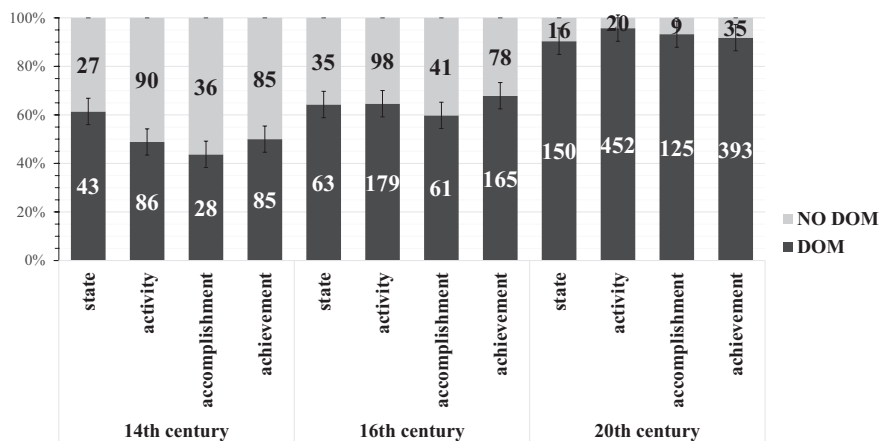


Figure 4: DOM with definite NPs according to century and aspectual class.

(10) Definite NPs without DOM in 20<sup>th</sup>-century Spanish

- a. *Y guiaré los ciegos por camino que no sabían.* (activity)  
 ‘And I will lead the blind by a way which they did not know.’  
 (20<sup>th</sup> century, Anonymous, *Biblia Reina-Valera*)
- b. *entre pescadores halló Cristo sus primeros secuaces.* (achievement)  
 ‘Among fishermen, Christ found his first followers.’  
 (20<sup>th</sup> century, Pardo Bazán, *San Francisco de Asís. Siglo XIII*)

Turning to indefinite NPs, Figure 5 shows the following distribution: the presence of DOM in the 14<sup>th</sup> century has a relative frequency of 23% (5/22) for states, 40% (23/58) for activities, 9% (1/11) for accomplishments and 22% (15/69) for of achievements. In the 16<sup>th</sup> century, DOM registers a relative frequency of 20% (7/35) for states, 50% (43/86) for activities, 33% (12/36) for accomplishments and 28% (23/83) for achievements. Finally, DOM shows a relative frequency of 40% (23/58) for states, 82% (122/149) for activities, 86% (31/36) for accomplishments and 66% (103/157) for achievements in the 20<sup>th</sup> century. Like in the rest of the figures, an increase of DOM can be observed in relation to the different periods. Moreover, while the variation between aspectual classes is almost imperceptible in Figure 4, in the case of indefinite NPs a higher variation can be noticed. However, the differences between aspectual classes have turned out not to be significant. Thus, the data presented in this paper cannot confirm the relationship between inherently telic verbs and DOM proposed in previous studies.

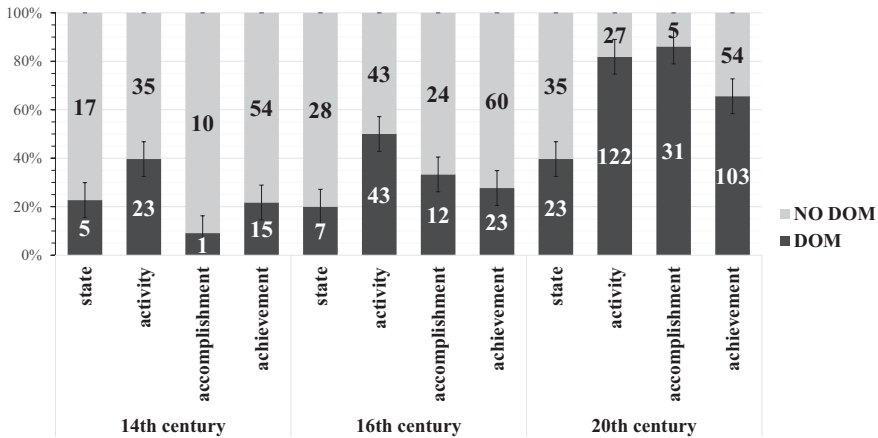


Figure 5: DOM with indefinite NPs according to century and aspectual class.

Examples of the presence of DOM from 14<sup>th</sup>-century Spanish with indefinite human direct objects are given in (11), where both present *a*-marking in their direct object regardless of aspectual class.

- (11) Indefinite NPs with DOM in 14<sup>th</sup>-century Spanish
- a. *non guardauan a otras mugeres sinon a sus madres.* (activity)  
‘They did not respect other women apart from their mothers.’  
(14<sup>th</sup> century, Don Juan Manuel, *Libro de los estados*)
  - b. *et clamo a vn cauallero que era aguazil suyo.* (achievement)  
‘And he called a knight who was his sheriff.’  
(14<sup>th</sup> century, Fernández de Heredia, *Gran crónica de España III*)

### 3.4 Discussion

The results of the corpus-based study presented in the previous section do not confirm the hypothesis proposed in Section 3.1 repeated herein (12).

- (12) Hypothesis for DOM and telicity
- The presence of DOM is more frequent with human direct objects of telic predicates than with human direct objects of atelic predicates.

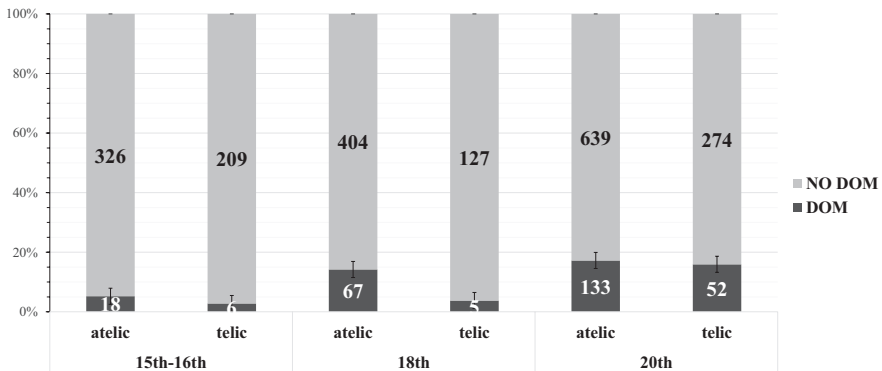
The corpus analysis has revealed that telicity had no impact on the presence of DOM in previous centuries. Similarly, aspectual class does not show any clear relationship with the use of *a*-marking. These results question Torrego’s (1999,

1787) first assumption (cf. 2a), since the inherent telicity of accomplishments and achievements neither implies DOM in the 20<sup>th</sup> century, nor triggers its occurrence in previous stages of Spanish.

However, it is interesting to note that with indefinite direct objects a slight variation can be observed which does not exist for definite direct objects. This can be seen in the results concerning both telicity and aspectual class. Although these are not significant differences, it means that telicity or aspectual class might have a slight effect on DOM, perhaps as a consequence of its interaction with other factors, but only on indefinite direct objects. Therefore, if such a slight variation can be explained by the interaction of telicity with some other factor, the effect of telicity is subordinated to definiteness and does not act directly on direct human objects, contrary to what previous studies defended.

Although this study is the only diachronic empirical study that addresses the impact of telicity on DOM for human direct objects, indirect evidence supporting the results presented above can be found in Barraza (2008), which deals with the relationship between DOM and telicity for inanimate direct objects. It is important to note that her concept of telicity refers to the aspectual class of verbs and not to the whole VP. Thus, she groups under the category of “telic” those instances which contain accomplishment and achievement verbs, and brings together under the label of “atelic” those whose verb is a state or an activity. Her study is based on a corpus comprised of 2,260 instances of inanimate direct objects which fall into three different periods (15<sup>th</sup>–16<sup>th</sup> centuries, 18<sup>th</sup> century and 20<sup>th</sup> century). The results obtained from her corpus analysis are shown in Figure 6. Interestingly, they point in the same direction as the results of the study I have carried out. Following Barraza, telicity does not have a relevant impact on DOM with inanimate direct objects. However, a minimal effect can be observed that favours the occurrence of DOM with inherently atelic verbs in previous stages of Spanish. This pattern matches the slight tendency that has been described above for indefinite direct objects in this study.

It is also interesting to comment on some aspects related to Torrego’s second assumption (2b), which holds that DOM triggers a telic interpretation in atelic verbs (1999, 1788–1790). I should start by pointing out the difficulty of addressing this issue with a corpus-based study, since it is necessary to have adverbials that explicitly reflect the aspectual reading of the verb. However, the example in (13a) allows us to address this issue, as it contains an activity verb, *maltratar* ‘to abuse’, with an adverbial that reinforces the atelic reading. Even so, following Torrego, it could be argued that, since the object is marked, the reading of the predicate is telic and that the PP headed by *durante* ‘for’ leads to an iterative reading of the telic event. However, in (13b) we observe that if the PP indicating duration is replaced by another PP that delimits the event, the result is ungrammatical. This indicates



**Figure 6:** DOM with inanimate NPs according to century and telicity (adapted from Barraza 2008).

that the predicate *maltrató a su hijo* behaves like an atelic predicate, regardless of DOM, since it does not admit adverbials delimiting the event. Also, note that the example in (13a) would be ungrammatical without DOM. Hence, DOM cannot trigger a change in the reading of the verb if its occurrence is mandatory.

(13) Effect of DOM with an atelic verb on the aspect of the VP

- a. *Un escritor huido de la URSS maltrató a su hijo durante el último año de convivencia.*

‘A writer who fled the USSR abused his son during the last year of living together.’

(20<sup>th</sup> century, Manuel Vázquez Montalbán, *La soledad del manager*)

- b. \**Un escritor huido de la URSS maltrató a su hijo en un mes/año.*

‘A writer who fled the USSR abused his son in one month/year.’

In addition, it is possible to find another argument against Torrego’s second assumption in the example presented in (14). In this case, an activity verb such as *arrastrar* ‘to drag’, which is intrinsically atelic, should get a telic reading when its direct object has DOM. At first sight, this seems to be exactly the case, because no problem arises when the telicity test is applied, as shown in (14a). However, (14b) shows that telicity is not derived from DOM but from the PP indicating the goal of the event, i.e. *hasta las gradas de la catedral* ‘to the cathedral steps’. If this goal PP disappears, as in (14b), the predicate stops behaving like a telic one and therefore does not admit a PP headed by *en* ‘in’.



- (14) Effect of DOM with an atelic verb on the aspect of the VP
- a. *Arnal arrastró a su amigo hasta las gradas de la catedral (en dos minutos/\*durante dos minutos).*  
‘Arnal dragged his friend to the cathedral steps (in two minutes/\*for two minutes).’  
(20<sup>th</sup> century, Wenceslao Fernández Flórez, *Fantasmas*)
  - b. *Arnal arrastró a su amigo (\*en dos minutos/durante dos minutos).*  
‘Arnal dragged his friend (\*in two minutes/for two minutes).’

With regard to the effect of DOM on atelic verbs, it might also be interesting to comment on the patterns of the verb *rodear*. As shown in (15), this verb has two readings in Spanish, one dynamic (‘circle’/‘go around’) and one stative (‘surround’), i.e. telic and atelic respectively.

- (15) Aspectual readings of *rodear* (‘go around’/‘surround’)
- a. *Los que del castillo baxaron rodearon al emperador por todas partes.*  
(dynamic and telic)  
‘Those from the castle circled the emperor everywhere.’  
(16<sup>th</sup> century, Sierra, *Espejo de príncipes y caballeros*, 2<sup>a</sup> parte)
  - b. *En la sala, el grupo familiar rodeaba al patrón.* (stative and atelic)  
‘In the living room, the family group was surrounding the boss.’  
(20<sup>th</sup> century, Valle-Inclán, *Tirano Banderas*)

Since *rodear* ‘go around/surround’ inherently offers two aspectual readings, the aspectual value of the VP will be given by factors different from the verbal head. This situation makes it easier to check not only which factors trigger a change in the lexical aspect, but also whether the occurrence of DOM itself triggers an aspectual change, as proposed in Torrego’s second assumption (1999, 1788–1790). In this study, all instances of direct objects of this verb are *a*-marked and most of them are definite objects; therefore, without being able to disentangle definiteness from the rest of factors, it is difficult to determine the effect of secondary factors that affect the use of DOM. However, the example presented in (16) includes an indefinite object. Following Torrego, when the verb exhibits its telic meaning, the object has to be *a*-marked. This does not seem to be the case due to the sentence context (*estaban sentadas diez esclavas blancas* ‘ten white slaves sat’), which indicates that it is a static situation. On the other hand, if the verb takes its atelic meaning, such as having an object with DOM, it should obtain a telic reading. However, this does not happen either, since the sentence context only allows for the stative reading, as we have just mentioned.

- (16) Effect of DOM on the aspect of the VP with *rodear* ('go around'/'surround')  
*Una pradera en la que estaban sentadas diez esclavas blancas, rodeando a una joven.*  
 'A meadow where ten white slaves sat, surrounding a young woman.'  
 (20<sup>th</sup> century, Blasco Ibáñez, *Traducción de las mil y una noches*)

The reported instances question whether DOM functions as a trigger for a telic interpretation on atelic verbs. As demonstrated, atelic verbs keep their natural reading in spite of *a*-marking and, in case of presenting a telic interpretation, this does not seem to arise from DOM.

## 4 Conclusions

Drawing on diachronic data from a corpus-based study, this paper has contributed to the understanding of the factors that trigger Differential Object Marking (DOM) in Spanish. The aim of this study was to investigate the impact of telicity on DOM from a diachronic perspective and the results have revealed that, in contrast to other factors, such as affectedness (von Heusinger/Kaiser 2011; Romero Heredero in press) and agentivity (García García 2018), whose influence has been proven from a diachronic point of view, telicity has not produced any clear effect on the evolution of DOM.

As stated above, this work is the first empirical study that addresses the impact of telicity on DOM with human direct objects and, in addition, the first to address this verbal factor from a diachronic perspective.

The contribution of this study points to the conclusion that telicity can no longer be considered an independent factor that has a significant influence on the occurrence of DOM in Spanish, as it has been proposed by previous studies. Nevertheless, it remains open to the possibility that the slight variation found in the case of indefinite direct object NPs can be explained by the interaction of telicity with other factors such as affectedness and agentivity.

In recent years, it has been shown that the occurrence of DOM correlates not only to the highest degree of transitivity, but also to the lowest (cf. García García 2018 and Fábregas 2013 for discussion). Therefore, it has been suggested that the different factors configuring transitivity favour the use of *a*-marking, either by enhancing the transitivity of the clause, e.g. human affected direct objects, or by reducing it to the minimum, e.g. inanimate agentive direct objects. Following this approach, it has been assumed that telicity, like affectedness, increases the degree of transitivity and thus favours the use of DOM. However, according to the

results obtained in this study, the idea arises that not all factors linked to transitivity somehow influence *a*-marking, and this is the case for telicity.

The analysis of the verbal factors involved in transitivity continues to be a promising field of study that needs further research to resolve the complex situation that the most recent studies reveal.

## Corpus

CDH = Real Academia Española, *Corpus del Diccionario histórico de la lengua española*, 2013 [online], <https://apps.rae.es/CNDHE> [last access: 19.07.2021].

## Bibliography

- Barraza Carbajal, Georgina, *Marcación preposicional de objeto directo inanimado*, in: Company Company, Concepción/Moreno de Alba, José G. (edd.), *Actas del VII Congreso Internacional de Historia de la Lengua Española. Mérida (Yucatán), 4–8 septiembre de 2006*, Madrid, Arco, 2008, 341–352.
- Company Company, Concepción, *El avance diacrónico de la marcación prepositiva en objetos directos inanimados*, in: Bernabé Pajares, Alberto, et al. (edd.), *Presente y futuro de la lingüística en España. La Sociedad Lingüística, 30 años después. Actas del II Congreso de la Sociedad Española de Lingüística*, vol. 2, Madrid, Consejo Superior de Investigaciones Científicas, 2002, 146–154.
- Delbecque, Nicole, *A construction grammar approach to transitivity in Spanish*, in: Davidse, Kristin/Lamiroy, Béatrice (edd.), *The nominative & accusative and their counterparts*, Amsterdam/Philadelphia, John Benjamins, 2002, 81–130.
- Dowty, David, *Word meaning and Montague grammar. The semantics of verbs and times in generative semantics and in Montague's PTQ*, Dordrecht, Reidel, 1979.
- Fábregas, Antonio, *Differential Object Marking in Spanish. State of the art*, *Borealis* 2:2 (2013), 1–80.
- Fernández Ramírez, Salvador, *Gramática española*, vol. 4, Madrid, Arco, 1986 [1951].
- García García, Marco, *Differentielle Objektmarkierung bei unbelehten Objekten im Spanischen*, Berlin/Boston, de Gruyter, 2014.
- García García, Marco, *Nominal and verbal parameters in the diachrony of Differential Object Marking in Spanish*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 209–242.
- Hopper, Paul J./Thompson, Sandra A., *Transitivity in grammar and discourse*, *Language* 56:2 (1980), 251–299.
- Kabatek, Johannes, *Wohin strebt die differentielle Objektmarkierung im Spanischen?*, *Romanistisches Jahrbuch* 67:1 (2016), 211–239.

- Kiparsky, Paul, *Partitive case and aspect*, in: Butt, Miriam/Geuder, Wilhelm (edd.), *The projection of arguments. Lexical and compositional factors*, Stanford, CSLI, 1998, 265–308.
- Krifka, Manfred, *The origins of telicity*, in: Rothstein, Susan (ed.), *Events and grammar*, Dordrecht, Springer, 1998, 197–235.
- Laca, Brenda, *El objeto directo. La marcación preposicional*, in: Company Company, Concepción (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, vol. 1, México DF, UNAM/Fondo de Cultura Económica, 2006, 423–475.
- Leonetti, Manuel, *Specificity and Differential Object Marking in Spanish*, *Catalan Journal of Linguistics* 3:1 (2004), 75–114.
- Lleal, Coloma (dir.), *DiCCA XV – Diccionario del castellano del siglo XV en la Corona de Aragón*, <http://ghcl.uib.edu/diccxv/> [last access: 12.02.2020].
- Marín, Rafael, *Una propuesta para el tratamiento de la información aspectual en HPSG*, *Procesamiento del lenguaje natural* 24 (1999), 70–82.
- Melis, Chantal, *El objeto directo personal en “El Cantar de Mio Cid”. Estudio sintáctico-pragmático*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 133–163.
- Octavio de Toledo y Huerta, Álvaro S., *Aprovechamiento del CORDE para el estudio sintáctico del primer español moderno (ca. 1675–1825)*, in: Kabatek, Johannes (ed.), *Lingüística de corpus y lingüística histórica iberorrománica*, Berlin/Boston, de Gruyter, 2016, 57–89.
- Ormazabal, Javier/Romero, Juan, *Object clitics, agreement and dialectal variation*, *Probus* 25:2 (2013), 301–344.
- Real Academia Española, *Diccionario de la lengua española*, 23 ed. [online version 23.4] <https://dle.rae.es> [last access: 19.07.2021]
- Richardson, Kylie, *Case*, in: Binnick, Robert I. (ed.), *The Oxford handbook of tense and aspect*, Oxford, Oxford University Press, 2012, 960–985.
- Romero Heredero, Diego, *Marcado diferencial de objeto y semántica verbal en español*, Berlin/Boston, de Gruyter, in press.
- Tenny, Carol, *Aspectual roles and the syntax-semantics interface*, Dordrecht, Kluwer Academic, 1994.
- Torrego, Esther, *The dependencies of objects*, Cambridge, Mass., MIT Press, 1998.
- Torrego, Esther, *El complemento directo preposicional*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española. Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*, vol. 2, Madrid, Espasa Calpe, 1999, 1779–1805.
- Vendler, Zeno, *Verbs and times*, *Philosophical Review* 66:2 (1956), 146–160.
- Verkuyl, Henk J., *On the compositional nature of the aspects*, Dordrecht, Reidel, 1972.
- von Heusinger, Klaus, *Verbal semantics and the diachronic development of DOM in Spanish*, *Probus* 20:1 (2008), 1–31.
- von Heusinger, Klaus/Kaiser, Georg A., *Affectedness and Differential Object Marking in Spanish*, *Morphology* 21:3 (2011), 593–617.
- von Heusinger, Klaus, *The diachronic development of Differential Object Marking in Spanish ditransitive constructions*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 315–344.
- von Heusinger, Klaus/García García, Marco/Montrul, Silvina, *Microvariation of Differential Object Marking in Latin American Spanish on the basis of judgement tests*, forthcoming.

## Appendix: Verbs included in the study

The following table contains all the verbs used for the analysis carried out in this work according to aspectual class. The subscript that follows some verbs distinguish their different meanings. These meanings have been checked using the *DLE* and the *DiCCA XV*.

<b>STATE</b>		<b>449</b>
abatir <sub>2</sub>	to depress	1
admirar <sub>1</sub>	to admire	36
amar <sub>1</sub>	to love	119
conocer <sub>1</sub>	to know	107
identificar <sub>2</sub>	to compare	12
preciar	to appreciate	5
recordar <sub>2</sub>	to bear in mind	23
recordar <sub>3</sub>	to remind	14
rodear <sub>1</sub>	to surround	38
tener <sub>1</sub>	to have	94
<b>ACTIVITY</b>		<b>1,218</b>
acorrer	to help	4
acribillar	to riddle	10
admirar <sub>2</sub>	to amaze	11
admirar <sub>3</sub>	to contemplate	9
agitar	to shake	6
agraviar	to offend	4
alimentar	to feed	2
amar <sub>2</sub>	to make love	3
amparar	to protect	1
apalear	to beat	43
apedrear	to stone	29
apretar	to press	38
arrastrar	to drag	61
asediar	to besiege	18
atender	to serve	13
atormentar	to torment	6
ayudar	to help	5
azotar	to flog	14
capitanear	to captain	1

(continued)

<b>STATE</b>		<b>449</b>
castigar	to punish	23
catar <sub>1</sub>	to look	4
cohechar	to force	2
combatir	to combat	48
consolar	to console	6
contemplar	to contemplate	11
contristar	to sadden	2
criar	to raise	79
empujar	to push	47
enseñar	to educate	5
evocar	to remember	17
forzar	to force	3
guardar	to protect	2
guiar	to guide	105
honrar	to honour	15
hostigar	to harass	1
humillar	to humiliate	3
incitar	to encourage	1
investigar	to investigate	4
llevar	to take [sb]	2
maltratar	to abuse	122
martirizar	to torment	4
menospreciar	to despise	10
oír	to hear	119
oprimir	to oppress	30
pendrar	to hold	3
percibir	to perceive	2
proveer	to provide	1
punir	to punish	14
remolcar	to tow	1
sacudir	to shake	12
señorear	to domineer	1
servir	to serve	2
sitiar	to besiege	4
transportar <sub>1</sub>	to transport	37
ver	to see	120

(continued)

<b>STATE</b>		<b>449</b>
vigilar	to guard	52
zarandear	to shake	26
<b>ACCOMPLISHMENT</b>		<b>383</b>
conocer <sub>2</sub>	to make love	6
conocer <sub>4</sub>	to meet	42
despedazar	to tear apart	36
destrózar	to destroy	30
identificar <sub>3</sub>	to ask for the documentation	5
identificar <sub>4</sub>	to prove the identity	5
neutralizar	to neutralise	12
olvidar	to forget	78
quemar	to burn	79
recordar <sub>1</sub>	to recall	15
recordar <sub>4</sub>	to remind	1
rodear <sub>2</sub>	to circle	15
traer	to bring	60
transportar <sub>2</sub>	to transport	1
<b>ACHIEVEMENT</b>		<b>1,150</b>
abatir <sub>1</sub>	to take down	46
arrollar	to mow down	44
catar <sub>2</sub>	to find	1
clamar	to call	16
conocer <sub>3</sub>	to recognise	40
derribar	to knock down	89
descartar	to reject	8
descubrir <sub>1</sub>	to discover	38
descubrir <sub>2</sub>	to uncover	9
detectar	to detect	9
eliminar	to eliminate	31
encerrar	to lock [sb] up	1
encontrar <sub>1</sub>	to find	145
encontrar <sub>2</sub>	to hit	30
enviar	to send	17
hallar	to find	147

(continued)

<b>STATE</b>		<b>449</b>
herir	to wound	2
identificar <sub>1</sub>	to recognise	54
llamar	to call	32
localizar	to locate	62
matar	to kill	111
mencionar	to mention	36
nombrar	to name	52
reconocer	to recognise	1
tener <sub>2</sub>	to give birth	5
vencer	to defeat	122
<b>TOTAL</b>		<b>3,200</b>



# Javier Caro Reina, Marco García García and Klaus von Heusinger

## Differential Object Marking in Cuban Spanish

**Abstract:** Recent research has shown that Differential Object Marking (DOM) is less frequent in some varieties of Latin American Spanish than in European Spanish. This is the case in Caribbean Spanish, which includes Cuban, Dominican, and Puerto Rican Spanish. We will investigate whether these varieties have preserved an older language stage or, rather, whether this is a more recent development resulting from DOM retraction. In this paper, we will focus on DOM in Cuban Spanish. Following on from Alfaraz (2011), who studied DOM on the basis of sociolinguistic interviews, we will examine DOM both from a diachronic and from a synchronic perspective. The diachronic approach is based on a corpus analysis encompassing the 19<sup>th</sup> and 20<sup>th</sup> centuries while the synchronic approach is based on grammaticality judgment tasks. The corpus analysis points to a slight retraction which evolved with indefinite human nouns. The results of the grammaticality judgment tasks reveal that Cuban Spanish speakers accept the absence of DOM with definite human nouns, which is unacceptable in European Spanish. They also rate the absence of DOM with indefinite human nouns as highly acceptable, as opposed to their European counterparts. We compare the findings provided from the corpus analysis and the judgment task by discussing the importance of considering both production and acceptability data. Thus, this paper makes an important empirical and theoretical contribution to the patterns of DOM in Caribbean Spanish.

---

**Acknowledgements:** This paper was presented at the Workshop “Diachrony of Differential Object Marking” (Institut National des Langues et Civilisations Orientales, Paris, November 16–17, 2017). We would like to thank the audience of the workshop for their comments. We would also like to thank Johannes Hofmann for the fieldwork in Havana, Julio Manero González for his help with the corpus analysis, and Diego Romero Heredero for his help with the design and analysis of the judgment task questionnaires. Last but not least, we would like to express our special gratitude to two anonymous reviewers for very useful comments and to the editors, Johannes Kabatek, Philipp Obrist, and Albert Wall, for their patience and great editorial work. The research for this paper has been funded by the German Research Foundation (DFG) as part of the SFB 1252 *Prominence in Language* (Project-ID 281511265) in the project B04 *Interaction of nominal and verbal features for Differential Object Marking* at the University of Cologne.

---

**Javier Caro Reina**, University of Cologne, e-mail: javier.caroreina@uni-koeln.de

**Marco García García**, University of Cologne, e-mail: marco.garcia@uni-koeln.de

**Klaus von Heusinger**, University of Cologne, e-mail: klaus.vonheusinger@uni-koeln.de

**Keywords:** animacy, Caribbean Spanish, corpus analysis, Cuban Spanish, definiteness, Differential Object Marking, European Spanish, judgment tasks, spoken data

## 1 Introduction

The term Differential Object Marking (DOM) is used to describe the phenomenon by which case marking of the direct object depends on certain semantic-pragmatic conditions such as animacy, referentiality, and topicality, as well as agentivity, affectedness, and telicity (Bossong 1985; Aissen 2003; García García/Primus/Himmelmann 2018; Witzlack-Makarevich/Seržant 2018, among others). DOM in Spanish is a very well attested and widely studied phenomenon. Traditionally, research has concentrated on European Spanish, where DOM has experienced a considerable expansion from Old to Modern European Spanish (Laca 2006; von Heusinger/Kaiser 2011; García García 2018), as will be shown in Section 2. However, there is a growing body of research that has examined the patterns of DOM in varieties of Spanish spoken in Argentina (Barrenechea/Orecchia 1977; Dumitrescu 1997; Tippets 2010; 2011; Montrul 2013; Hoff/Díaz-Campos 2015; Hoff 2018), Cuba (Alfaraz 2011), Mexico (Dulme 1986; Company Company 2002a; Lizárraga Navarro/Mora-Bustos 2010; Tippets 2010; 2011; Ordóñez/Treviño 2016), Peru (Mayer/Delicado Cantero 2015), Venezuela (Domínguez et al. 1998; Balasch 2011), Uruguay (Barrios 1981), and USA (Montrul 2014). In these studies, DOM has been approached from various perspectives, such as language attitudes (Hoff/Díaz-Campos 2015), heritage languages (Montrul 2014), language change (Company Company 2002a; 2002b), and language variation. With regard to language variation, most studies are embedded in the *Habla Culta* project (see Lope Blanch 1986). This is the case with Alfaraz (2011) for Cuban Spanish, Domínguez et al. (1998) and Balasch (2011) for Venezuelan Spanish, and Tippets (2010; 2011) for Buenos Aires, Madrid, and Mexico City Spanish.

Interestingly, the occurrence of DOM has been found to differ in varieties of European and Latin American Spanish, suggesting two opposed tendencies: DOM expansion and DOM retraction. In this respect, Caribbean Spanish has been reported to exhibit a lower frequency of DOM than other varieties of Spanish (Jiménez Sabater 1975, 169–170; Álvarez Nazario 1992, 237; López-Morales 1992, 141; Lunn 2002; Alba 2004, 140–141; Bullock/Toribio 2009; Alfaraz 2011, among others).<sup>1</sup>

---

<sup>1</sup> Additionally, similar patterns have also been observed in Bolivia (Mendoza Quiroga 1992, 459), La Palma (Régulo Pérez 1970, 82–83), and Venezuela (Domínguez et al. 1998).

An example from Dominican Spanish is given in (1), where the definite human direct objects *esa hija* ‘that daughter’ and *las personas de Francia* ‘the people from France’ are not differentially marked. Note that in European Spanish the absence of DOM would result in an ungrammatical utterance.

(1) Lack of DOM in Dominican Spanish

a. *Luba quería mucho ∅ esa hija.*

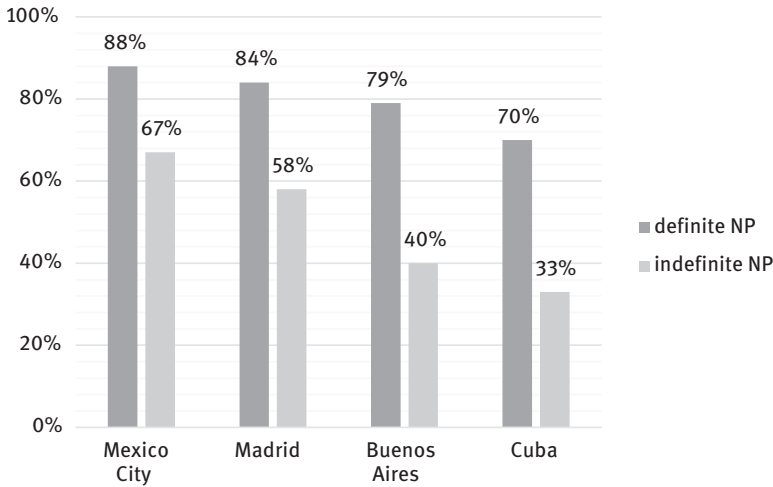
‘Luba loved that daughter very much.’

b. *Para entender ∅ las personas de Francia.*

‘In order to understand the people from France.’

(Bullock/Toribio 2009, 59)

Figure 1 illustrates the occurrence of DOM with definite and indefinite human direct objects in the varieties of Spanish spoken in Mexico City, Madrid, Buenos Aires, and Cuba.<sup>2</sup> With regard to definite human direct objects, DOM has a relative frequency of 88% (153/174) in Mexico City Spanish, 84% (87/104) in Madrid Spanish, 79% (162/205) in Buenos Aires Spanish, and 70% (168/240) in Cuban Spanish. With regard to indefinite human direct objects, DOM has a relative fre-



**Figure 1:** DOM with definite and indefinite human direct objects in selected varieties of Spanish (Tippets 2010, 134, 147, 156; Alfaraz 2011, 224).

<sup>2</sup> The data from Mexico City, Madrid, and Buenos Aires are taken from Tippets (2010), who used the corpora of the *Habla culta de la Ciudad de México* (1971), *Habla culta de Madrid* (1981), and *Habla culta de Buenos Aires* (1987), respectively. Similarly, the data from Cuba is partly based on *Habla culta de Miami* (1968–1969) (cf. Section 3 for details).

quency of 67% (32/48), 58% (21/36), 40% (21/53), and 33% (10/30), respectively. We can observe that DOM occurs more frequently in Mexican Spanish than in the other varieties of Spanish. More importantly, DOM is less frequent in Cuban Spanish than in the other varieties of Spanish.

While there have been a series of sociolinguistic studies on phonological, morphosyntactic, and lexical variation in Cuban Spanish (cf. Cuza 2017 for a comprehensive overview), DOM has not received much attention, with the exception of the work of Alfaraz (2011), which will be presented in Section 3. The aim of the present study is to give a detailed account of DOM in Cuban Spanish drawing on both diachronic and synchronic data. We will address the question of whether the lower incidence of DOM in Cuban Spanish constitutes a remnant from older stages of Spanish or, rather, whether it constitutes a recent development pointing to a retraction of DOM. This is followed by a methodological discussion of the adequate means for investigating variation in DOM, including sociolinguistic interviews, corpus analyses, and grammaticality judgments.

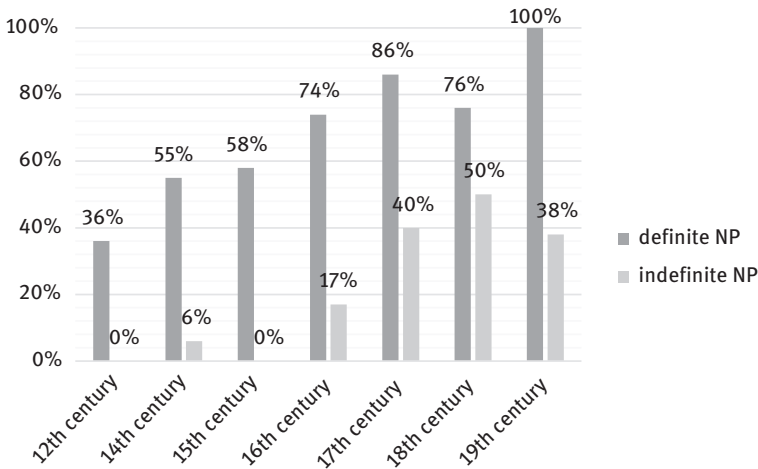
The paper is structured as follows: Section 2 deals with the diachronic development of DOM in European Spanish. Section 3 presents the study conducted by Alfaraz (2011) on the basis of spontaneous speech. Section 4 contains the diachronic corpus analysis conducted for 19<sup>th</sup>- and 20<sup>th</sup>-century Cuban Spanish. Section 5 is dedicated to the grammaticality judgment tasks for Cuban and European Spanish. Section 6 summarizes the main findings of our study and concludes by discussing DOM expansion and retraction.

## 2 Diachronic development of DOM in European Spanish

The diachronic development of DOM has been examined in a number of empirical studies which have assessed the relevance of animacy and definiteness (Company Company 2002a; Laca 2006), topicality (Melis 1995; Pensado 1995), affectedness (von Heusinger 2008; von Heusinger/Kaiser 2011), and telicity (Romero Heredero, this volume). More recently, the development of DOM has been considered with regard to both monotransitive and ditransitive constructions (Ortiz Ciscomani 2005; 2011; von Heusinger 2018). While most studies have concentrated on human and animate direct objects, some have been devoted to inanimate direct objects (Company Company 2002b; Barraza Carbajal 2008; García García 2014; 2018).

Laca (2006) provides the most fine-grained analysis. For this reason, we will refer to her findings in the ensuing sections. Before summarizing her results, we will address two critical empirical issues. First, Laca's (2006) corpus covers a large period of time reaching from the 12<sup>th</sup> to the 19<sup>th</sup> century. However, her data base is rather small, being generally confined to just one or two text samples per century. This is all the more problematic when we try to account for differences between European and Latin American Spanish since the data for each linguistic area is only based on one text per century (cf. Kabatek 2016, 216, 230 for further critical aspects). Second, Laca's (2006) findings cannot be directly compared to other diachronic studies such as Company Company (2002a) and von Heusinger/Kaiser (2011). The comparison with Company Company (2002a) is difficult because the author does not distinguish between definite and indefinite human direct objects. Similarly, the diachronic study of von Heusinger/Kaiser (2011) is less fine-grained since it only contains data from the 15<sup>th</sup>, 17<sup>th</sup> and 19<sup>th</sup> centuries.

In the absence of a more extensive and comparable corpus study, we take Laca's (2006) findings as a point of departure for our diachronic investigation (cf. Section 4). Where possible, her data will be complemented and compared with the findings from other empirical studies. On the basis of Laca's (2006) corpus study, Figure 2 shows the diachronic development of DOM with human direct objects according to definiteness and century.



**Figure 2:** DOM with human direct objects according to definiteness and century in European Spanish (adapted from Laca 2006, 442–443).

Note that in contrast to the more fine-grained distinctions put forward by Laca (2006), Figure 2 only includes the results regarding full lexical NPs (e.g. *una/la mujer* ‘a/the woman’). That is, it excludes pronouns, proper names, NPs without lexical heads (e.g. *los más conocidos* ‘the best known’), definite-like NPs with universal quantifiers (e.g. *cada persona* ‘each person’), indefinite-like NPs with existential (or weak) quantifiers (e.g. *algunas personas* ‘some people’), and bare nouns (e.g. *personas* ‘people’). In addition, it only contains the occurrences of DOM in European Spanish. Laca’s (2006) findings can be summarized as follows: First, there is a clear rise of DOM with both definite and indefinite NPs, whereby the percentages of DOM are clearly and constantly higher with definite NPs than with indefinite NPs. Second, with regard to definite NPs, DOM increases greatly and ends up becoming categorical. More specifically, we observe 36% (13/36) of DOM in the 12<sup>th</sup> century, 55% (36/66) in the 14<sup>th</sup> century, 58% (38/65) in the 15<sup>th</sup> century, 74% (26/35) in the 16<sup>th</sup> century, 86% (117/136) in the 17<sup>th</sup> century, 76% (22/29) in the 18<sup>th</sup> century, and 100% (28/28) in the 19<sup>th</sup> century. Third, with regard to indefinite NPs, DOM also increases considerably, though it never becomes categorical. The development seems to begin between the 15<sup>th</sup> and 16<sup>th</sup> centuries (cf. also von Heusinger/Kaiser 2011, 611). We observe a sharp rise of DOM, reaching 17% (1/6) in the 16<sup>th</sup> century, 40% (21/53) in the 17<sup>th</sup> century, and 50% (8/16) in the 18<sup>th</sup> century. Later, there is a slight decrease to 38% (3/8) in the 19<sup>th</sup> century.<sup>3</sup>

In summary, Laca’s (2006) results show that in Old Spanish, DOM is optional with definite human direct objects, but absent from indefinite human direct objects. In Modern European Spanish, by contrast, DOM is obligatory with definite human direct objects, but optional with indefinite human direct objects. As we will see in Section 3, spoken Cuban Spanish resembles 16<sup>th</sup>-century European Spanish. This raises the question of whether Modern Cuban Spanish has retained this prior language stage. This issue will be discussed in more detail in Section 4.

### 3 Alfaraz’s (2011) spoken data

In this section, we will report on the study carried out by Alfaraz (2011), who examined DOM in Cuban Spanish on the basis of recordings made in Miami in

---

<sup>3</sup> In this respect, Laca (2006, 460) argues that the relatively high percentage of DOM in the 18<sup>th</sup> century is due to the disproportionately high number of causative constructions compared to previous centuries. These constructions have been shown to have a positive influence on the frequency of DOM (cf. García García 2018, 235–336).

the 1960s and 1990s. These two sets of data allow for a real time study. The first corpus consists of a subset of the recording collected in 1968–1969 for the *Habla Culta* project. It is based on 24 speakers aged between 30 and 50. The second corpus was collected by Alfaraç in 1996–98. It comprises 26 speakers that were classified into two age groups (30–43 and 62–77), which allowed for an apparent time study. Importantly, both corpora are comprised of (semi-)directed interviews that were conducted with monolingual speakers of Spanish upon their arrival in the USA. Therefore, we can exclude a contact-induced change resulting from contact with English (cf. Carter/Lynch 2015 for Miami Cuban Spanish).

Alfaraç (2011) found in her two data sets a total of 502 human direct objects, of which 368 (73%) contained DOM. She further analyzed the instances of DOM according to linguistic and social factors. The linguistic factors include referentiality (pronoun, proper name, definite NP, indefinite specific NP, and non-specific NP) and word order (postverbal, preverbal). The social factors include time period (1968–1969, 1996–1998) and age group (30–43 and 62–77 years for the 1990s period).

With regard to referentiality, pronouns and proper names are always differentially marked, both of which have a relative frequency of 98% (94/96 and 88/90, respectively). In contrast, DOM gradually decreases with definite NPs, indefinite specific NPs, and non-specific NPs, which have a relative frequency of 70% (168/240), 33% (10/30), and 17% (8/46), respectively. As for indefinite NPs, the author distinguishes between specific and non-specific NPs. The latter category is not homogeneous since it is comprised of non-specific indefinite NPs and bare nouns. For this reason, we will exclusively refer to specific indefinite NPs when talking about indefinite NPs. In summary, DOM was found to occur more frequently with definite rather than indefinite NPs (70% vs. 33%), as shown in Figure 3.

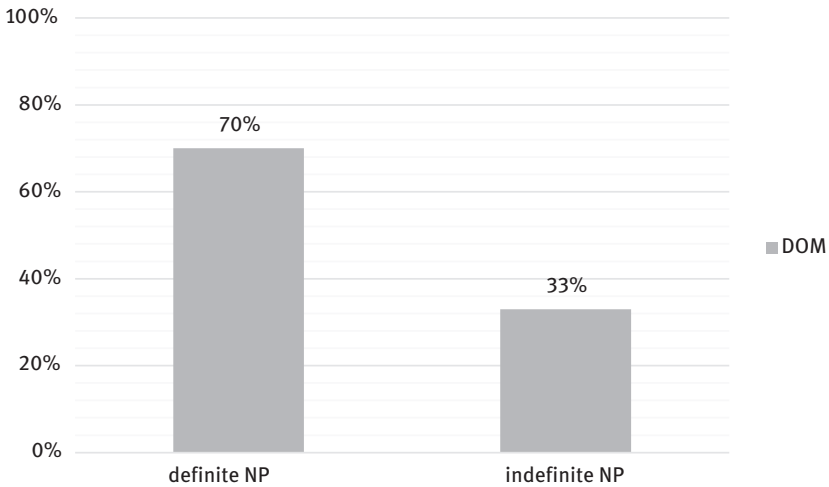
Examples from the sample are shown in (2), where the definite direct objects *la vieja aquella* ‘that old woman’ and *la abuela de Tetico* ‘Tetico’s grandmother’ are not differentially marked. The same applies even for the left-dislocated direct object *esa gente* ‘those people’. In this respect, Cuban Spanish differs from European Spanish, where *a*-marking is required in all of these cases.

(2) Lack of DOM with definite human direct objects in Cuban Spanish

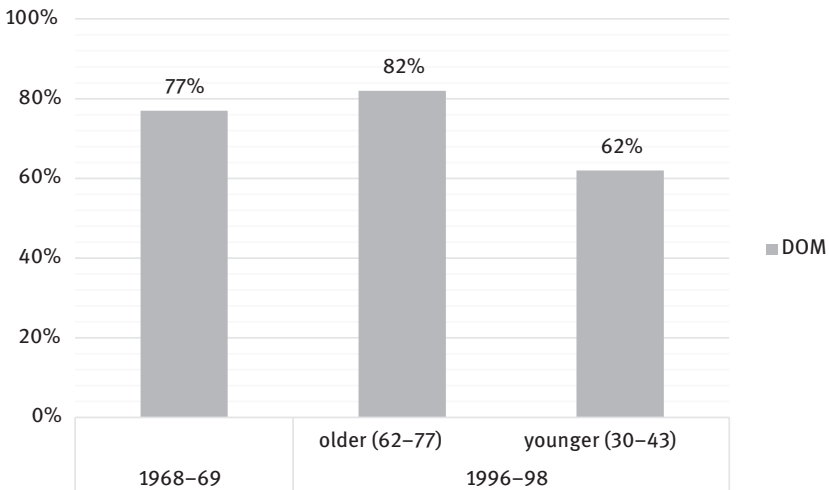
- a. *Tú no viste Ø la vieja aquella fajándose con el viejo aquel.*  
‘You didn’t see that old woman fighting with that old man.’
- b. *Ella no conoció Ø la abuela de Tetico.*  
‘She didn’t meet Tetico’s grandmother.’
- c. *Ø Esa gente tú la manipulas.*  
‘You’re manipulating those people.’

(Alfaraç 2011, 228–229)

Interestingly, the occurrence of DOM varies according to time period, as depicted in Figure 4. In the first period (1968–1969), DOM was found to occur in 77% of cases involving definite and indefinite NPs. In the second period (1996–1998), however, the author observed a decrease in the occurrence of DOM. More specifically, the



**Figure 3:** DOM with human direct objects according to definiteness in Cuban Spanish (adapted from Alfaraz 2011, 224).



**Figure 4:** DOM in Cuban Spanish with all human direct objects according to age group (adapted from Alfaraz 2011, 224).



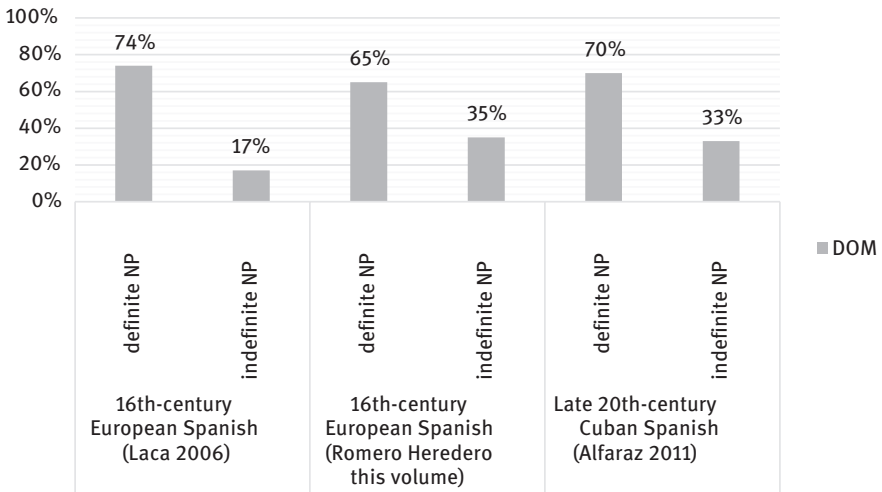
younger generation employed DOM less frequently than the older one (62% vs. 82%). In other words, the results gleaned from the real time and apparent time studies show DOM retraction. Note that the results provided by Alfaraz (2011, 224) do not allow for the combination of age group with definiteness, which would have resulted in a more fine-grained picture of the development of DOM across generations according to definite and indefinite NPs.

## 4 Diachronic corpus analysis

In order to study the patterns of DOM in Cuban Spanish in the 19<sup>th</sup> and 20<sup>th</sup> centuries, we conducted a corpus analysis based on the *Corpus diacrónico del español* (CORDE). Section 4.1 presents the hypotheses according to the patterns of DOM described in Section 2 and Section 3. Section 4.2 is dedicated to the study design. Section 4.3 contains the results according to definiteness and animacy in SVO sentences. Section 4.4 discusses the results.

### 4.1 Hypotheses

The patterns of DOM laid out in Section 2 and Section 3 enable us to detect similarities between 16<sup>th</sup>-century European Spanish and Modern Cuban Spanish. More specifically, Modern Cuban Spanish resembles 16<sup>th</sup>-century European Spanish with respect to definiteness, as illustrated in Figure 5. The values for 16<sup>th</sup>-century European Spanish are based on Laca (2006, 442) and Romero Heredero (this volume) (cf. also Keniston 1937, 10–11, 14 for DOM with definite and indefinite human direct objects in 16<sup>th</sup>-century Castilian prose). With regard to definite human direct objects, DOM has a relative frequency of 74% (26/35) and 65% (468/720) in 16<sup>th</sup>-century European Spanish and 70% (168/240) in Modern Cuban Spanish. With regard to indefinite human direct objects, DOM has a relative frequency of 17% (1/6) and 35% (85/240) in 16<sup>th</sup>-century European Spanish and 33% (10/30) in Modern Cuban Spanish. The different percentages found with indefinite human direct objects in 16<sup>th</sup>-century European Spanish (17% vs. 35%) result from the number of tokens examined by Laca and Romero Heredero (1/6 vs. 85/240) (cf. Section 2 for a critical discussion).



**Figure 5:** DOM in 16<sup>th</sup>-century European Spanish and 20<sup>th</sup>-century Cuban Spanish with definite and indefinite human direct objects (Laca 2006, 442; Romero Heredero, this volume; Alfaraz 2011, 224).

The similarities between 16<sup>th</sup>-century European Spanish and Modern Cuban Spanish (especially with definite human direct objects) raises the question of whether Cuban Spanish is undergoing a process of retraction or whether it has just retained a prior language stage. In order to answer this question, we carried out a diachronic corpus-based study that involves an analysis of DOM in the 19<sup>th</sup> and 20<sup>th</sup> centuries. Our hypotheses are summarized in (3). Note that scholars such as Pérez Guerra (1992, 489) have explained the lower occurrence of DOM in Dominican Spanish in terms of a remnant feature. However, this assumption has not been empirically tested yet.

(3) Hypotheses for 19<sup>th</sup>- and 20<sup>th</sup>-century Cuban Spanish

H1: Cuban Spanish is undergoing DOM retraction.

H2: Cuban Spanish has retained a prior language stage of DOM.

## 4.2 Study design

The analysis is based on the *Corpus diacrónico del español* (CORDE). The corpus allows searching for single words (or combination of words) according to author, work, time span, text type (book, journal, etc.), country, and genre (cf. Octavio de

Toledo y Huerta 2006 for a critical discussion). In this respect, the corpus differs from others such as the *Corpus del español*, which does not allow for a diachronic search according to country. The sources for Cuban Spanish are well suited for a diachronic study since they contain a total of 883,618 words for the 19<sup>th</sup> century and 1,499,345 words for the 20<sup>th</sup> century.<sup>4</sup> Unfortunately, the CORDE is not annotated. Therefore, it is not possible to search for specific syntactic patterns such as DOM. Since verbal factors such as affectedness have proved to have an impact on the occurrence of DOM (cf. von Heusinger/Kaiser 2011; Romero Heredero in press), we manually searched for transitive verbs with high and low affectedness, i.e. predicates with a higher and a lower preference for DOM, in order to obtain a balanced data set with respect to this verbal factor. Note, however, that affectedness will not be treated in this study.

The verbs with high affectedness are the following (the number of tokens are given in brackets): *cuidar* ‘to take care of’ (6), *ejecutar* ‘to execute’ (2), *golpear* ‘to hit’ (7), *herir* ‘to hurt’ (3), *humillar* ‘to humiliate’ (7), *matar* ‘to kill’ (38), *violar* ‘to rape’ (1), and *violentar* ‘to force’ (1). The verbs with low (or no) affectedness are *buscar* ‘to look for’ (38 tokens), *conocer* ‘to know’ (10), *contemplar* ‘to contemplate’ (14), *entender* ‘to understand’ (4), *escuchar* ‘to listen to’ (7), *esperar* ‘to wait for’ (10), *mirar* ‘to look at’ (87), *observar* ‘to observe’ (12), *oír* ‘to hear’ (14), and *ver* ‘to see’ (35).<sup>5</sup> The number of tokens involving verbs with high and low affectedness amounts to 65 and 231, respectively. The search was conducted by means of regular expressions.<sup>6</sup>

Altogether, we found 296 instances of full definite and indefinite human direct object NPs. For comparison, Alfaraz (2011) has 270 tokens (excluding non-specific NPs) while von Heusinger/Kaiser (2011) have 423 tokens. By contrast, Laca (2006)

---

4 In contrast to other Caribbean Spanish varieties, Cuban Spanish is well represented in the corpus. For comparison, the corpus contains a total of 393,119 and 134,287 words for Puerto Rico and the Dominican Republic respectively for the 19<sup>th</sup> and 20<sup>th</sup> centuries.

5 With regard to the perception verbs *escuchar* ‘to listen’, *mirar* ‘to look at’, *oír* ‘to hear’, and *ver* ‘to see’, we also looked at ACI constructions, which had a frequency of 1, 3, 5, and 6 tokens respectively. These cases were always differentially marked (cf. García García 2018, 235–237 for DOM with ACI constructions).

6 Regular expressions allow for the substitution of one or more characters. For example, the string *mat\** results in 1,483 tokens from 72 different documents in the time span between 1800 and 1975. The tokens include inflected forms such *mataba* ‘s/he was killing’, *mató* ‘s/he killed’, *matando* ‘killing’, etc. However, they also contain other forms such as *matanza* ‘carnage’, *materia* ‘matter’, *matrimonio* ‘marriage’, etc. which had to be excluded manually. Of the 1,483 tokens, only 38 involved instances of *matar* ‘to kill’ with definite and indefinite human NPs as their direct objects.

has a total of 775 tokens, which are distributed along the centuries as follows: 42 (12<sup>th</sup> c.), 97 (14<sup>th</sup> c.), 76 (15<sup>th</sup> c.), 181 (16<sup>th</sup> c.), 189 (17<sup>th</sup> c.), 85 (18<sup>th</sup> c.), and 105 (19<sup>th</sup> c.). The instances of human direct objects found in the corpus were subsequently classified according to verb, affectedness (high vs. low), century (19<sup>th</sup> vs. 20<sup>th</sup> century), year, DOM (presence vs. absence), definiteness (definite vs. indefinite NP), author, and work. Appendix 1 gives an overview of the sources that contained instances of DOM with human direct objects in combination with the verbs selected. The table in Appendix 1 is arranged according to century, author, and record. The examples taken from the corpus are cited according to year, author, and a shortened name of the record (e.g. 1966/Lezama/Paradiso). We did not distinguish specific from non-specific indefinite direct objects. Our search was restricted to direct object NPs with human referents in SVO sentences. That is, we excluded collective nouns (*gente* ‘people’, *multitud* ‘crowd’, etc.),<sup>7</sup> animate non-human NPs,<sup>8</sup> inanimate NPs, proper names, bare nouns, left dislocations, and impersonal constructions.

### 4.3 Results

The results of the corpus analysis of Cuban Spanish are arranged in Figure 6 according to definiteness (definite vs. indefinite NP) and century (19<sup>th</sup> vs. 20<sup>th</sup> century). With regard to definite human direct objects, DOM has a relative frequency of 95%, both in the 19<sup>th</sup> (120/126) and 20<sup>th</sup> (118/124) centuries. With regard to indefinite human direct objects, DOM has a relative frequency of 56% (9/16) and 43% (13/30) in the 19<sup>th</sup> and 20<sup>th</sup> centuries, respectively. That is, the occurrence of DOM has remained stable with definite human direct objects (95%) while it has experienced a slight decrease with indefinite human direct objects (56% > 43%). Examples of lack of DOM with definite human direct objects are given in (4), where the NPs *los doce prisioneros* ‘the twelve prisoners’ and *el padrino de la boda* ‘the best man’ are not differentially marked.

<sup>7</sup> Compare *aguantar a la gente* ‘to put up with people’ (1938/Serpa/Contrabando) to *esperar la gente de mister Bourton* ‘to wait for Mr Bourton’s people’ (1938/Serpa/Contrabando).

<sup>8</sup> For example, *matar un gallo blanco* ‘to kill a white cock’ (1906/Ortiz/Brujos) vs. *matar a un cerdo* ‘to kill a pig’ (1906/Ortiz/Brujos).

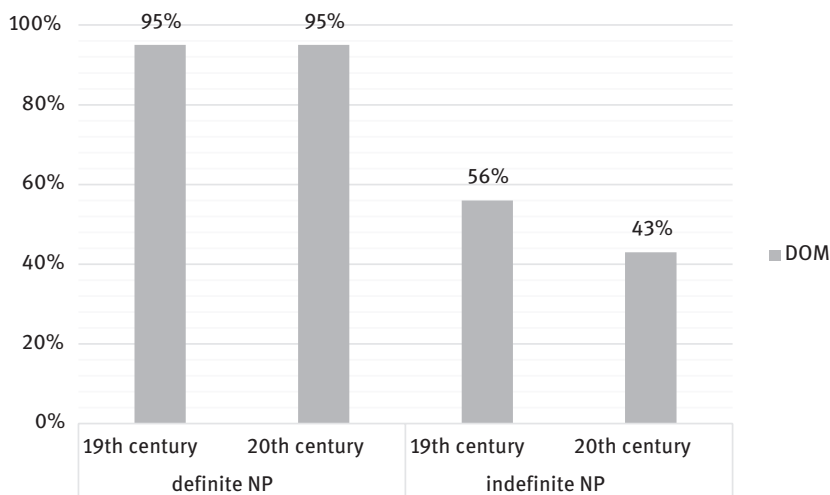


Figure 6: DOM in Cuban Spanish according to definiteness and century (CORDE).

(4) Lack of DOM with definite human direct objects

- a. *Aquiles mató con su mano Ø los doce prisioneros* (1889/Martí/Edad).  
'Achilles killed the twelve prisoners with his own hands.'
- b. *Sonriéndose marchó hacia la sala para buscar Ø el padrino de la boda* (1966/Lezema/Paradiso).  
'Smiling he went to the hall to look for the best man.'

Examples of lack of DOM with indefinite human direct objects are shown in (5), where the NPs *otra madre* 'another mother' and *un hombre* 'a man' are not differentially marked.

(5) Lack of DOM with indefinite human direct objects

- a. *Conque ya sabes... a buscar Ø otra madre* (1884/Ortega/Cleopatra).  
'You already know... Go and look for another mother.'
- b. *Yo maté Ø un hombre* (1938/Serpa/Contrabando).  
'I killed a man.'

In order to examine inter-speaker variation, we further looked at the instances of DOM with human direct objects according to author (cf. Table 1). For example, Gómez de Avellaneda consistently uses DOM with both definite and indefinite human direct objects. However, in Insúa, Lezama Lima, and Serpa, the presence of DOM is more frequent with definite human direct objects while the absence of DOM

is more frequent with indefinite human direct objects. Notably, none of the authors generally avoided DOM with both definite and indefinite human direct objects.

**Table 1:** DOM with human direct objects according to author and definiteness.

	Definite NP		Indefinite NP	
	DOM	no DOM	DOM	no DOM
Anónimo	5	0	0	0
Auber Noya	15	1	0	1
Augier	4	0	0	0
Bobadilla	9	0	0	0
Cabrera Infante	7	2	1	2
Carlos Felipe	14	0	2	1
Carpentier	9	0	2	1
El Cucalambé	6	0	0	1
Fernández Retamar	0	0	1	0
Gómez de Avellaneda	56	0	3	0
Guillén	1	0	0	0
Heredia	5	1	0	1
Hostos	6	2	2	1
Insúa	11	0	1	3
Lachatañaré	4	1	1	1
Lezama Lima	39	1	2	6
Loynaz	3	0	1	0
Martí	22	1	3	1
Ortega Munilla	13	1	1	2
Ortiz	3	2	1	0
Serpa	6	1	1	3

The results gained from Table 1 point to the existence of variation within some authors. Examples of variation involving definite and indefinite human direct objects are given in (6) and (7), respectively.

- (6) Inter-speaker variation with definite human direct objects (taken from 1938/Serpa/Contrabando)
- a. *na más que mató a su mujer.*  
'Just for killing his wife.'

- b. *¿busca Ø los guardacostas?*  
 ‘Is he looking for the coastguards?’
- (7) Inter-speaker variation with indefinite human direct objects (taken from 1966/Lezama Lima/Paradiso)
- a. *se sentó en un café para esperar a un amigo, que le soportaba sus crisis*  
 ‘He sat down in a café to wait for a friend, who tolerated his crisis.’
- b. *conozco Ø un profesor de estética que nos visitó hace pocos meses.*  
 ‘I know an aesthetics teacher who visited us a couple of months ago.’

Let us turn to the hypotheses postulated in the previous section. The corpus analysis provides evidence that in 19<sup>th</sup>-century Cuban Spanish, DOM is much more frequent than in 16<sup>th</sup>-century European Spanish. Recall from Section 2 that in 16<sup>th</sup>-century European Spanish, DOM occurs with definite and indefinite NPs with a relative frequency of 74%/65% and 17%/35%, respectively (cf. Figure 5). By contrast, in 19<sup>th</sup>-century Cuban Spanish, DOM occurs with definite and indefinite NPs with a relative frequency of 95% and 56%, respectively. Thus, H1 is not borne out for 19<sup>th</sup>-century Cuban Spanish. In other words, 19<sup>th</sup>-century Cuban Spanish has not retained a prior language stage. This result therefore challenges the assumption held by scholars such as Pérez Guerra (1992, 489) that Caribbean Spanish preserves the patterns of DOM as found in prior stages of European Spanish. This issue will be discussed in more detail in the ensuing section.

#### 4.4 Discussion

In this section, we will first address DOM retraction and then critically discuss the implications derived from the type of language data (written vs. spoken). The corpus search in texts written by Cuban authors in the 19<sup>th</sup> and 20<sup>th</sup> centuries has revealed a slight decrease of DOM with indefinite human direct objects. Table 2 summarizes the occurrence of DOM with definite and indefinite human direct objects in 16<sup>th</sup>-century European Spanish (Laca 2006; Romero Heredero, this volume), 19<sup>th</sup>- and 20<sup>th</sup>-century written Cuban Spanish (CORDE), and spoken Modern Cuban Spanish (Alfaraz 2011). Assuming that the patterns of DOM in 16<sup>th</sup>-century European Spanish also applied for 16<sup>th</sup>-century Cuban Spanish, it follows that there is a rise of DOM between the 16<sup>th</sup> and 20<sup>th</sup> centuries (e.g. 74%/65% > 95% in the case of definite NPs). In this respect, Cuban Spanish resembles European Spanish, which has also experienced DOM expansion (cf. Figure 2). However, we could detect a slight decrease of DOM with indefinite human direct objects from the 19<sup>th</sup> to the 20<sup>th</sup> century (56% > 43%). Considering Modern

Cuban Spanish on the basis of Alfaraz's spoken data, this tendency continues to develop with indefinite human direct objects (43% > 33%) while DOM also begins to decrease with definite human direct objects (95% > 70%), which suggests an ongoing change involving DOM retraction. In this respect, Cuban Spanish differs from Modern European Spanish. The diachronic picture that emerges from Table 2 further suggests that in Cuban Spanish, DOM retraction began with indefinite human direct objects and subsequently expanded to definite human direct objects. These findings support evidence for the patterns of DOM retraction, which affects less prominent categories (indefinite NPs) prior to more prominent ones (definite NPs). In this sense, Cuban Spanish seems to be another instance of DOM retraction within the Romance language group, as has been reported for Portuguese (Delille 1970) and Catalan (Dalrymple/Nikolaeva 2011, 212).

**Table 2:** Diachronic overview of DOM with definite and indefinite human direct objects.

	Definite NP	Indefinite NP
16 <sup>th</sup> -century European Spanish (Laca 2006)	74% (26/35)	17% (1/6)
16 <sup>th</sup> -century European Spanish (Romero Heredero, this volume)	65% (468/720)	35% (85/240)
19 <sup>th</sup> -century Cuban Spanish (CORDE)	95% (120/126)	56% (9/16)
20 <sup>th</sup> -century Cuban Spanish (CORDE)	95% (118/124)	43% (13/30)
Modern Cuban Spanish (Alfaraz 2011)	70% (168/240)	33% (10/30)

A word of caution, however, should be that we are comparing two different types of sources. On the one hand, we have written corpora, which are associated with a formal (or standard) style. On the other, we have spontaneous speech, which is associated with a casual (or informal) style. As a consequence, written language can be assumed to be more averse to language innovations such as DOM retraction whereas spoken language is probably more progressive in this respect. This might be the reason for the higher frequency of DOM found in our corpus search based on written language (CORDE) compared to the lower frequency of DOM attested in the spoken data used by Alfaraz (2011).

In order to deepen our understanding of this kind of variation, we complemented the production studies (corpus, interviews) with a grammaticality judgment task. While production data only provides information about the more acceptable form in a given context, acceptability judgments also offer insights into less commonly used forms in that context. We think that such information will help us to observe the diachronic development of this variation in much clearer light.



## 5 Judgment tasks

We designed a questionnaire in order to assess the acceptability grade regarding the presence and absence of DOM with human definite and indefinite direct objects both in Cuban and European Spanish. Section 5.1 formulates the hypotheses according to the patterns of DOM in Modern Cuban and Modern European Spanish previously described in Section 2 and Section 3, respectively. Section 5.2 gives a detailed account of the study design. Section 5.3 presents the results of the acceptability judgments of DOM and lack of DOM according to animacy and definiteness. Section 5.4 discusses the results.

### 5.1 Hypotheses

In Section 4, we studied the use of DOM in Spanish records written by Cuban authors in the 19<sup>th</sup> and 20<sup>th</sup> centuries. We then compared the results of our corpus analysis to the findings of Alfaraz (2011), Laca (2006), and Romero Heredero (this volume). We found that the Cuban corpus of written language behaved very much like the European corpus of written language, but quite differently from the Cuban corpus of spoken language of Alfaraz (2011). This provided, on the one hand, evidence for H1 that in Cuban Spanish, DOM underwent retraction and, on the other hand, evidence against H2 that Cuban Spanish has retained a prior language stage. In this section, we provide additional empirical evidence for H1, namely data from a grammaticality judgment task. In this task, participants had to decide how acceptable they found a human direct object with or without DOM. Grammaticality judgment tests offer a different empirical perspective than corpus analyses. As for specific grammatical constructions such as DOM, in a corpus we usually find only one form in a given sentence, let's say the more acceptable form, which in the case of constructions with human definite direct objects will be those with DOM. By contrast, in grammatical judgment tasks we obtain a graded evaluation between competing forms, i.e. between standard and less standard forms, which in the case of sentences with definite human direct objects correspond to those with and without DOM, respectively.

According to our hypothesis H1, we predict the following results from the grammaticality judgments. If there is DOM retraction, we expect Cuban Spanish speakers to rate the absence of DOM (noDOM) with human definite and indefinite direct objects better, i.e. with higher acceptability values, than their European counterparts (P1a). In addition, we also predict that Cuban Spanish speakers will rate the presence of DOM worse, i.e. with lower acceptability values, than their European counterparts (P1b). Since we quantify over judgments, we thus expect

a higher average acceptability value for cases of noDOM and a lower average acceptability value for DOM instances from Cuban Spanish speakers in comparison with their European Spanish counterparts.

(8) Predictions for the grammaticality judgments tasks based on H1

H1: In Cuban Spanish, DOM underwent retraction.

P1a: Speakers of Cuban Spanish will show higher acceptability values for noDOM than speakers of European Spanish.

P1b: Speakers of Cuban Spanish will show lower acceptability values for DOM than speakers of European Spanish.

## 5.2 Study design

The questionnaire employed for Cuban and European Spanish was comprised of general information on the sociolinguistic background of the participants (age, gender, education level, first and second language), instructions with four examples illustrating how to fulfil the judgment task, the judgment task itself, and final comments. The judgment tasks had an approximate duration of 15 minutes. The questionnaires for European and Cuban Spanish differed slightly from each other since the vocabulary had to be adapted to lexical variation (e.g. *celular* and *móvil* for ‘mobile phone’ in Cuban and European Spanish, respectively).

The grammaticality judgment task consisted of a Likert scale ranging from 1 (unacceptable) to 6 (totally acceptable). The questionnaire included 16 test items, 8 with definite and 8 with indefinite human direct objects in SVO sentences.<sup>9</sup> In addition, we provided 16 fillers which served as control items. The test items displayed a direct object, which was employed once with and once without DOM. The two different versions appeared in different item lists (questionnaire A and questionnaire B) such that the participants could only see a single version of the same direct object. The two experimental item lists were pseudo-randomized in different orders for questionnaires A and B before being distributed to the participants. Examples of items with and without DOM both with definite and indefinite human direct objects are shown in (9) and (10), respectively (cf. Appendix 2 for the complete list of test items employed for definite and indefinite human direct objects). The verbs were selected according to affectedness. More specifically, the verbs with high affectedness are *cuidar* ‘to take care of’, *golpear* ‘to hit’,

---

<sup>9</sup> This questionnaire further contained 32 items which tested inanimate direct objects in SVO sentences as well as animate and inanimate direct objects in clitic-doubling constructions. For the present study, these conditions have not been considered.

*matar* ‘to kill’, and *lesionar* ‘to injure’. The verbs with low affectedness are *acusar* ‘to accuse’, *denunciar* ‘to report’, *oír* ‘to hear’, and *ver* ‘to see’. Finally, the filler sentences were comprised of 8 grammatical (e.g. *Francisco renunció al puesto de trabajo* ‘Francisco refused the job’) and 8 ungrammatical control sentences (e.g. \**José le llevó Juan al libro* ‘José brought Juan to the book’).

- (9) DOM condition with definite and indefinite human direct objects
- a. *Patricio lesionó al portero en la discoteca.*  
‘Patricio injured the doorman at the nightclub.’
  - b. *Alberto lesionó a un defensa en el partido de la semana pasada.*  
‘Alberto injured a defender during last week’s match.’
- (10) noDOM condition with definite and indefinite human direct objects
- a. *Patricio lesionó Ø el portero en la discoteca.*  
‘Patricio injured the doorman at the nightclub.’
  - b. *Alberto lesionó Ø un defensa en el partido de la semana pasada.*  
‘Alberto injured a defender during last week’s match.’

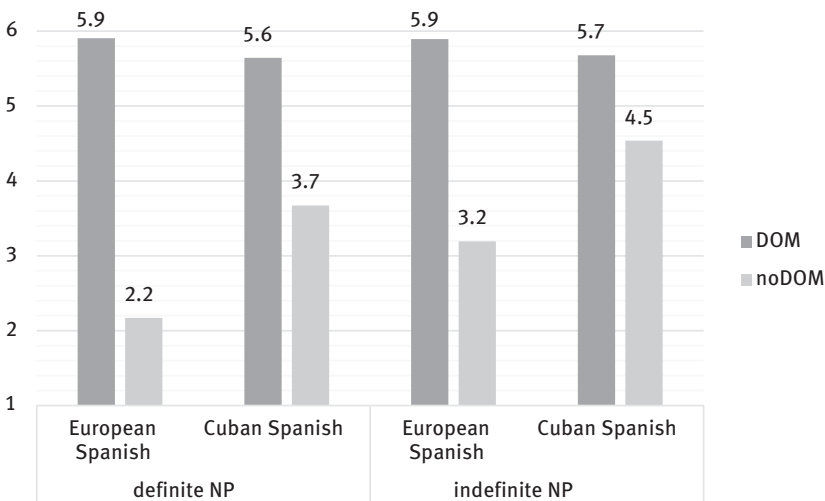
With regard to the distribution of the questionnaire, we employed two different methods depending on the country in question. For Cuban Spanish, the questionnaires were handed out to the participants by a student from the University of Cologne in a university classroom of Havana. For European Spanish, the questionnaires were distributed electronically by means of the platform Google Forms. The access link was made available on the websites of universities and social networks. We obtained a total of 214 filled-out questionnaires, of which 75 were from Cuba (38 for questionnaire A and 37 for questionnaire B) and 139 from Spain (82 for questionnaire A and 57 for questionnaire B). After revising the filled-out questionnaires, we had to remove 16 participants since the answers to the control fillers (both grammatical and ungrammatical) deviated considerably from the expected values in more than 20% of the answers. Thus, the number of valid questionnaires amounted to 62 for Cuba (33 for questionnaire A and 29 for questionnaire B) and 136 for Spain (79 for questionnaire A and 57 for questionnaire B).

### 5.3 Results

Figure 7 summarizes the results of the questionnaire for European and Cuban Spanish. The DOM condition with definite human direct objects (cf. ex. 9a above) is highly acceptable in both varieties (European Spanish: 5.9 vs. Cuban Spanish 5.6). Interestingly, the DOM condition for indefinite human direct objects (cf. ex.

9b) is also very acceptable in both varieties (5.9 vs. 5.7). Thus, prediction P1b, according to which speakers of Cuban Spanish should show lower acceptability values for DOM than speakers of European Spanish, is not supported by the questionnaire study.

As for the noDOM condition, we observe more variation. Definite human direct objects without DOM (cf. ex. 10a) are not acceptable in European Spanish (2.2) while they are much more acceptable in Cuban Spanish (3.7). We observe a very similar pattern for the lack of DOM with human indefinite direct objects (cf. ex. 10b): In European Spanish it is not as acceptable as in Cuban Spanish (3.2 vs. 4.5). Hence, prediction P1a, according to which speakers of Cuban Spanish should exhibit higher acceptability values for noDOM than speakers of European Spanish, is clearly confirmed by our questionnaire study.



**Figure 7:** Acceptability values for DOM and noDOM with definite and indefinite human direct objects in European and Cuban Spanish (1 = unacceptable, 6 = totally acceptable).

## 5.4 Discussion

As for the DOM condition, the results of the judgment task experiment confirm the assumptions in the literature and the observations from the corpus studies in Section 4. Definite human direct objects with DOM are always rated as perfect forms, both in European and Cuban Spanish. This supports the assumption that DOM is obligatory with definite human direct objects. Contrary to pre-

diction P1b, however, we found no difference between European and Cuban Spanish, even though Alfaraz (2011) mentioned some examples with definite direct objects without DOM (cf. ex. 2 above). The very high acceptability of DOM with definite direct objects in Cuban Spanish suggests that if there is retraction, it is optional since the forms with DOM are fully acceptable. We observe a very similar pattern for DOM with indefinite direct objects: They are rated as totally acceptable both in European and in Cuban Spanish. At first glance, this is surprising since corpus studies point to a clear difference between the distribution of DOM with definite and indefinite direct objects, the latter showing a much lower frequency of DOM (cf. Figures 1 and 2 as well as Table 2 in Sections 1, 2 and 4.4, respectively). This contrast is not reflected in the acceptability study.

We can account for these different results by assuming that in corpora we usually find the more acceptable form for a certain context whereas the data from the questionnaire study shows whether or not a form is acceptable. Since DOM is optional with indefinites, or more specifically, obligatory with human specific indefinites, but optional with human non-specific indefinites, participants always rated DOM with human indefinite direct objects with very high acceptability values. Thus, the high acceptability of DOM with human indefinite direct objects does not contradict the results from the corpus analysis. Note that, as for definite direct objects, we do not find a difference between European and Cuban Spanish with respect to the acceptability of DOM with indefinite direct objects. This suggests that, if there is retraction in Cuban Spanish, it is also optional for indefinite direct objects.

While the acceptability of DOM is always very high in both varieties, the acceptability of noDOM cases differs strongly and therefore allows for interesting observations. The lack of DOM (noDOM condition) with definite human direct objects is ungrammatical in European Spanish and should therefore be rated very low, which is actually the case (2.2). Speakers of Cuban Spanish, however, rate this construction much higher (3.7). This is consistent with the observation in Alfaraz (2011) that noDOM is found much more often in Cuban Spanish than in European Spanish (cf. Table 2). The relatively high acceptability of noDOM in Cuban Spanish can be viewed as indirect evidence of DOM retraction in this variety.

The same differences between Cuban and European Spanish concerning the acceptability of noDOM with definite direct objects are also attested for noDOM with indefinite direct objects. In European Spanish, indefinite direct objects without DOM receive medium-range grammaticality scores (3.2) while in Cuban Spanish they are rated as quite acceptable (4.5). It is surprising that European speakers of Spanish rate this construction as only halfway acceptable. If DOM

is optional with indefinite direct objects, we would expect the lack of DOM to be much more acceptable, as is the case with speakers of Cuban Spanish. We think, however, that our examples of direct objects in simple transparent sentences clearly provide instances of specific indefinites. We further assume with the literature that direct objects without DOM cannot receive a specific interpretation (Leonetti 2004, 98–99).<sup>10</sup> Under this assumption, the rather low rating seems to reflect the mismatch between a specific interpretation of the direct object and its realization with a form that is restricted to non-specific meanings, at least in European Spanish. For Cuban Spanish, the quite acceptable ratings for noDOM with indefinites suggest that the mentioned requirement of DOM with specific indefinite direct objects does not hold. Be this as it may, we see a clear difference between European and Cuban Spanish. As has been shown, the absence of DOM is much more acceptable in Cuban than in European Spanish. Again, this points to a higher flexibility of DOM in Cuban Spanish and to a first step towards retraction.

In summary, the acceptability study confirms prediction P1a that Cuban Spanish speakers show higher acceptability values for noDOM cases than European Spanish speakers, but not prediction P1b that Cuban Spanish speakers exhibit lower acceptability values for DOM cases than European speakers. We still take this as support for our hypothesis H1 that, in Cuban Spanish, DOM underwent retraction. We would also like to assert that the different empirical methods complement each other. The spoken and written data presented in Sections 3 and 4 show that the distribution of DOM clearly differs between Cuban and European Spanish. This contrast is not mirrored in the questionnaire study for the acceptability of DOM, but rather for the acceptability of noDOM. From a more general point of view, we would like to stress that corpus data and judgment data are different methods which may unveil underlying contrasts. Still, combining both may provide a broader empirical coverage: They involve different types of data (production vs. acceptability) and thus provide different types of evidence that might or might not point into the same direction, as is the case with DOM retraction in Cuban Spanish.

---

**10** As an anonymous reviewer correctly points out, the question of whether an indefinite direct object is interpreted as specific or non-specific can be disambiguated by context. For example, this can be achieved by adding to the test items a further sentence containing a modal operator indicating the epistemic (non-)specificity of the indefinite direct object in question, such as 'I know X/I do not know X'. For our next experiments, we will introduce this modification in order to control for (epistemic) specificity.

## 6 Conclusions and discussion

Drawing on data from spontaneous speech, a diachronic corpus analysis, and grammaticality judgment tasks, this paper has given a synchronic and diachronic account of Differential Object Marking (DOM) in Cuban Spanish. The data from spontaneous speech reveals that DOM is less frequent in Cuban Spanish than in Buenos Aires, Madrid, and Mexico City Spanish. The lower frequency of DOM as compared to other varieties of Spanish raises the question of whether Cuban Spanish has retained a prior language stage or, rather, underwent DOM retraction. In this respect, Caribbean Spanish has previously been assumed to preserve a prior language stage of European Spanish (Pérez Guerra 1992, 489). More specifically, spoken Modern Cuban Spanish resembles written 16<sup>th</sup>-century European Spanish, especially with respect to DOM with definite human direct objects. Our diachronic corpus analysis has conclusively shown that DOM in written Modern Cuban Spanish is not a remnant of a prior language stage since it experienced a clear expansion between the 16<sup>th</sup> and 19<sup>th</sup> centuries (cf. Section 4.4 for details).

The data from spontaneous speech from the 20<sup>th</sup> century (Alfaraz 2011) together with the written corpus data from the 19<sup>th</sup> and 20<sup>th</sup> centuries (CORDE) favour the hypothesis that in Cuban Spanish, DOM underwent retraction. However, DOM retraction seems to be a rather recent phenomenon which began to develop in the 20<sup>th</sup> century. Moreover, it is much more evident in Alfaraz's (2011) spoken data from spontaneous speech than in our written corpus data from CORDE, where we could only detect a slight decrease of DOM with indefinite direct objects (from 56% to 43%), but not with definite direct objects. These differences might be due to the fact that spontaneous speech represents an informal style whereas written texts reflect a rather formal, more conservative language use (cf. Kock/De Mello 1997 for discussion on the *Habla Culta*). Indirect evidence that DOM retraction is a recent development in Cuban Spanish comes from other Caribbean varieties such as Puerto Rican Spanish. For example, López-Morales (1992, 141) indicates that DOM is less frequently found among young speakers.

While spontaneous speech and written language constitute instances of language production, grammaticality judgment tasks allow us to carefully evaluate the rate of acceptability with the presence and absence of DOM. In addition, they are diagnostic tools for detecting language change. In this respect, the judgment tasks conducted for Cuban and European Spanish have provided further evidence for the hypothesis of DOM retraction in Cuban Spanish. Interestingly, we could not find any difference between European and Cuban Spanish for the acceptability of sentences with DOM since both speakers of European and Cuban Spanish rated the test items with DOM as totally acceptable, both with human definite

and with human indefinite direct objects. However, we could observe clear differences with respect to the acceptability of sentences lacking DOM. Speakers of Cuban Spanish rated the absence of DOM with human definite and indefinite direct objects as highly acceptable, as opposed to their European counterparts. Similar observations have been made by Vaquero (1978), who conducted acceptability judgment tasks with university students of Puerto Rico, showing that the absence of DOM received an extremely positive evaluation, at least with human indefinite direct objects.

DOM variation, including both expansion and retraction, remains a promising research field, especially if we extend the empirical focus to more varieties within and beyond the Caribbean area, and if we attempt to model the factors conditioning variation. In order to take on this challenge, we can conclude from the present study that it is crucial to consider both production and acceptability data, and to analyze not only the conditions for the presence of DOM, but also those for the absence of DOM.

## Corpus

CORDE = Real Academia Española, *Banco de datos (CORDE) [en línea]. Corpus diacrónico del español*, <http://www.rae.es>, [last access: 01.02.2019].

## Bibliography

- Aissen, Judith, *Differential Object Marking. Iconicity vs. economy*, *Natural Language and Linguistic Theory* 21:3 (2003), 435–483.
- Alba, Orlando, *Cómo hablamos los dominicanos. Un enfoque sociolingüístico*, Santo Domingo, Grupo León Jimenes, 2004.
- Alfaraz, Gabriela G., *Accusative object marking. A change in progress in Cuban Spanish?*, *Spanish in Context* 8:2 (2011), 213–234.
- Álvarez Nazario, Manuel, *El habla campesina del país. Orígenes y desarrollo del español en Puerto Rico*, Río Piedras, Universidad de Puerto Rico, 1992.
- Balasz, Sonia, *Factors determining Spanish Differential Object Marking within its domain of variation*, in: Michnowicz, Jim/Dodsworth, Robin (edd.), *Selected proceedings of the 5<sup>th</sup> workshop on Spanish sociolinguistics*, Somerville, Mass., Cascadilla, 2011, 113–124.
- Barraza Carbajal, Georgina, *Maricación preposicional de objeto directo inanimado*, in: Company Company, Concepción/Moreno de Alba, José G. (edd.), *Actas del VII Congreso Internacional de Historia de la Lengua Española. Mérida (Yucatán), 4–8 septiembre de 2006*, Madrid, Arco, 2008, 341–352.



- Barrenechea, Ana María/Orecchia, Teresa, *La duplicación de objetos directos e indirectos en el español hablado en Buenos Aires*, in: Lope Blanch, Juan Manuel (ed.), *Estudios sobre el español hablado en las principales ciudades de América*, Ciudad de México, UNAM, 1977, 351–381.
- Barrios, Graciela, *Comportamiento de la preposición “a” con complementos directos en el español de Montevideo*, in: Elizaincín, Adolfo (ed.), *Estudios sobre el español del Uruguay*, Montevideo, Universidad de la República, 1981, 13–25.
- Bosson, Georg, *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen, Narr, 1985.
- Bullock, Barbara E./Toribio, Almeida Jaqueline, *Reconsidering Dominican Spanish. Data from the rural Cibao*, *Revista Internacional de Lingüística Iberoamericana* 7:2 (2009), 49–73.
- Carter, Phillip M./Lynch, Andrew, *Multilingual Miami. Current trends in sociolinguistic research*, *Language and Linguistics Compass* 9:9 (2015), 369–385.
- Company Company, Concepción, *El avance diacrónico de la marcación prepositiva en objetos directos inanimados*, in: Bernabé Pajares, Alberto, et al. (edd.), *Presente y futuro de la lingüística en España. La Sociedad Lingüística, 30 años después. Actas del II Congreso de la Sociedad Lingüística*, vol. 2, Madrid, CSIC, 2002, 146–154 (= 2002a).
- Company Company, Concepción, *Grammaticalization and category weakness*, in: Wischer, Ilse/Diewald, Gabriele (edd.), *New reflections on grammaticalization*, Amsterdam/Philadelphia, John Benjamins, 2002, 201–215 (= 2002b).
- Cuza, Alejandro, *Introduction. New facts in Cuban Spanish research*, in: Cuza, Alejandro (ed.), *Cuban Spanish dialectology. Variation, contact, and change*, Washington DC, Georgetown University Press, 2017, xix–xxvii.
- Dalrymple, Mary/Nikolaeva, Irina, *Objects and information structure*, Cambridge, Cambridge University Press, 2011.
- Delille, Karl Heinz, *Die geschichtliche Entwicklung des präpositionalen Akkusativs im Portugiesischen*, Bonn, Romanisches Seminar der Universität Bonn, 1970.
- Domínguez, Carmen Luisa, et al., *Observaciones sobre el uso de la preposición “a” en el objeto directo. Un estudio sobre el español de Mérida*, *Letras* 59 (1998), 89–120.
- Dulme, Erika Ehnis, *El uso de la preposición “a” ante objeto directo en el habla popular de la ciudad de México*, in: Moreno de Alba, José G. (ed.), *Actas del II Congreso Internacional sobre el Español de América, Ciudad de México, 27–31 de enero de 1986*, México, UNAM, 1986, 404–407.
- Dumitrescu, Domnita, *El parámetro discursivo en la expresión del objeto directo lexical. Español madrileño vs. español porteño*, *Signo y Seña* 7 (1997), 305–354.
- García García, Marco, *Differentielle Objektmarkierung bei unbelehten Objekten im Spanischen*, Berlin/Boston, de Gruyter, 2014.
- García García, Marco, *Nominal and verbal parameters in the diachrony of Differential Object Marking in Spanish*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 209–242.
- García García, Marco/Primus, Beatrice/Himmelmann, Nikolaus P., *Shifting from animacy to agentivity*, *Theoretical Linguistics* 44:1–2 (2018), 25–39.
- von Heusinger, Klaus, *Verbal semantics and the diachronic development of DOM in Spanish*, *Probus* 20 (2008), 1–31.
- von Heusinger, Klaus, *The diachronic development of Differential Object Marking in Spanish ditransitive constructions*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 315–344.

- von Heusinger, Klaus/Kaiser, Georg A., *Affectedness and Differential Object Marking in Spanish*, *Morphology* 21:3–4 (2011), 593–617.
- Hoff, Mark R., *Are Argentines “a”-blind? Acceptability of “a”-marked inanimate direct objects*, in: MacDonald, Jonathan E. (ed.), *Contemporary trends in Hispanic and Lusophone linguistics. Selected papers from the Hispanic Linguistic Symposium 2015*, Amsterdam/Philadelphia, John Benjamins, 2018, 121–142.
- Hoff, Mark R./Díaz-Campos, Manuel, *Conciencia y actitudes. El caso de la “a” acusativa en el español argentino*, *Signo y Seña* 28 (2015), 89–110.
- Jiménez Sabater, Max, *Más datos sobre el español de la República Dominicana*, Santo Domingo, INTEC, 1975.
- Kabatek, Johannes, *Wohin strebt die differentielle Objektmarkierung im Spanischen?*, *Romanistisches Jahrbuch* 67:1 (2016), 211–239.
- Keniston, Hayward, *The syntax of Castilian prose. The sixteenth century*, Chicago, University of Chicago Press, 1937.
- Kock, Josse de/De Mello, George (edd.), *Lengua escrita y habla culta en América y España. Diez casos*, Salamanca, Universidad de Salamanca, 1997.
- Laca, Brenda, *El objeto directo. La marcación preposicional*, in: Company Company, Concepción (ed.), *Sintaxis histórica del español. Primera parte: La frase verbal*, vol. 1, México, UNAM/Fondo de Cultura Económica, 2006, 423–475.
- Leonetti, Manuel, *Specificity and Differential Object Marking in Spanish*, *Catalan Journal of Linguistics* 3 (2004), 75–114.
- Lizárraga Navarro, Glenda Zoé/Mora-Bustos, Armando, *Variación en la marcación diferenciada de objeto en español*, *Forma y Función* 23:1 (2010), 9–38.
- Lope Blanch, Juan Manuel, *El estudio del español hablado culto. Historia de un proyecto*, México, UNAM, 1986.
- López-Morales, Humberto, *El español del Caribe*, Madrid, MAPFRE, 1992.
- Lunn, Patricia, *Tout se tient in Dominican Spanish*, in: Lee, James F./Geeslin, Kimberly L./Clements, J. Clancy (edd.), *Structure, meaning, and acquisition in Spanish. Papers from the 4<sup>th</sup> Hispanic Linguistics Symposium*, Somerville, Mass., Cascadilla, 2002, 65–72.
- Mayer, Elisabeth/Delicado Cantero, Manuel, *Continuity and innovation in Peruvian Spanish. Pragmatics and contact in (Differential) Object Marking*, in: Sessarego, Sandro/González-Rivera, Melvin (edd.), *New perspectives on Hispanic contact linguistics in the Americas*, Madrid, Vervuert, 2015, 99–120.
- Melis, Chantal, *El objeto directo personal en “El Cantar de Mio Cid”. Estudio sintáctico-pragmático*, in: Pensado, Carmen (ed.), *El complemento directo preposicional*, Madrid, Visor, 1995, 133–163.
- Mendoza Quiroga, José G., *Aspectos del castellano hablado en Bolivia*, in: Hernández Alonso, César (ed.), *Historia y presente del español de América*, Ponce, Junta de Castilla y León, 1992, 437–500.
- Montrul, Silvina, *La marcación diferencial del objeto directo en el español de Argentina. Un estudio experimental*, in: Colantoni, Laura/Rodríguez Louro, Celeste (edd.), *Perspectivas teóricas y experimentales sobre el español de la Argentina*, Madrid, Vervuert, 2013, 207–228.
- Montrul, Silvina, *Structural changes in Spanish in the United States. Differential Object Marking in Spanish heritage speakers across generations*, *Lingua* 151, Part B (2014), 177–196.

- Octavio de Toledo y Huerta, Álvaro S., *Aprovechamiento del CORDE para el estudio sintáctico del primer español moderno (ca. 1675–1825)*, in: Kabatek, Johannes (ed.), *Lingüística de corpus y lingüística histórica iberorrománica*, Berlin/Boston, de Gruyter, 2016, 57–89.
- Ordóñez, Francisco/Treviño, Esthela, *Agreement and DOM with impersonal “se”*. A comparative study of Mexican and Peninsular Spanish, in: Kato, Mary/Ordóñez, Francisco (edd.), *The morphosyntax of Portuguese and Spanish in Latin America*, Oxford, Oxford University Press, 2016, 236–258.
- Ortiz Ciscomani, Rosa María, *Los objetos concurrentes y la bitransitividad en el español en perspectiva diacrónica*, in: Eddington, David (ed.), *Selected proceedings of the 7<sup>th</sup> Hispanic Linguistics Symposium*, Somerville, Mass., Cascadilla Press, 2005, 192–202.
- Ortiz Ciscomani, Rosa María, *Construcciones bitransitivas en la historia*, México, UNAM, 2011.
- Pérez Guerra, Irene, *Aportación a un tema en debate en el Caribe hispánico. El arcaísmo del español dominicano*, in: Ariza Viguera, Manuel (ed.), *Actas del II Congreso Internacional de Historia de la Lengua Española*, Madrid, Pabellón de España, 1992, 483–490.
- Régulo Pérez, Juan, *Notas acerca del habla de la isla de la Palma*, *Revista de Historia Canaria* 32 (1970), 12–174.
- Romero Heredero, Diego, *Marcado diferencial de objeto y semántica verbal en español*, Berlin, de Gruyter, in press.
- Tippets, Ian, *Differential Object Marking in Spanish. A quantitative variationist study*, Ohio, Ohio State University PhD thesis, 2010.
- Tippets, Ian, *Differential Object Marking. Quantitative evidence for underlying hierarchical constraints across Spanish dialects*, in: Ortiz-López, Luis A. (ed.), *Selected proceedings of the 13<sup>th</sup> Hispanic Linguistics Symposium*, Somerville, Mass., Cascadilla, 2011, 107–117.
- Vaquero, María, *Enseñar español, pero ¿qué español?*, *Boletín de la Academia Puertorriqueña de la Lengua Española* 6:1 (1978), 127–146.
- Witzlack-Makarevich, Alena/Seržant, Ilja A., *Differential Argument Marking. Patterns of variation*, in: Seržant, Ilja A./Witzlack-Makarevich, Alena (edd.), *Diachrony of Differential Argument Marking*, Berlin, Language Science Press, 2018, 1–40.

## Appendix 1: CORDE sources

This appendix contains the CORDE sources employed for the diachronic corpus analysis (cf. Section 4). The sources are arranged according to century, author, and record. The number of words is given in brackets.

### Sources

Century	Author	Record
19 <sup>th</sup> century	Auber Noya	<i>Ambarina</i> (97,206)
	El Cucalambé	<i>Consecuencias de una falta</i> (9,744); <i>Poesías completas</i> (60,283)
	Gómez de Avellaneda	<i>Dolores</i> (23,973); <i>Dos mujeres</i> (83,824); <i>El artista barquero o los cuatro 5 de junio</i> (68,640); <i>El cacique de Turmequé</i> (17,617); <i>Espatolino</i> (49,301); <i>La baronesa de Joux</i> (12,553); <i>La dama de Amboto</i> (2,262); <i>La flor del ángel</i> (6,935); <i>La montaña maldita</i> (4,165); <i>La ondina del lago azul</i> (10,590); <i>La velada del helecho o el donativo del diablo</i> (19,160); <i>Poesías</i> (43,221); <i>Una anécdota de la vida de Cortés</i> (4,802)
	Heredia	<i>Carta de Boston</i> (621); <i>Poesías</i> (36,951); <i>Teatro 3. Una ópera de Rossini</i> (506)
	Hostos	<i>La peregrinación de Bayoán</i> (85,248)
	Martí	<i>La Edad de Oro</i> (66,806); <i>Lucía Jerez</i> (36,099)
	Ortega Munilla	<i>Cleopatra Pérez</i> (47,098)
20 <sup>th</sup> century	Anónimo	<i>¡Oh, mío Yemayá!</i> (27,527)
	Augier	<i>Prosa varía</i> (126,005)
	Bobadilla	<i>A fuego lento</i> (64,073)
	Cabrera Infante	<i>Tres tristes tigres</i> (137,073)
	Carlos Felipe	<i>De película</i> (8,574); <i>El Chino</i> (18,061); <i>El travieso Jimmy</i> (25,885); <i>Réquiem por Yarini</i> (22,989)
	Carpentier	<i>El siglo de las luces</i> (117,871)
	Fernández Retamar	<i>Fervor de la Argentina</i> (42,768)
	Guillén	<i>El son entero</i> (5,144)
	Insúa	<i>El negro que tenía el alma blanca</i> (70,841)
	Lachatañaré	<i>El sistema religioso de los lucumí y otras influencias africanas en Cuba</i> (75,304); <i>Manual de santería</i> (16,181)
	Lezama Lima	<i>Paradiso</i> (220,057)

(continued)

Loynaz	<i>Jardín</i> (104,393)
Ortega Munilla	<i>Los tres sorianitos</i> (22,669)
Ortiz	<i>Contrapunteo cubano del tabaco y el azúcar</i> (208,650); <i>Los negros brujos</i> (80,325)
Serpa	<i>Contrabando</i> (69,565)

## Appendix 2: Test items of the questionnaire

This appendix lists the test items of the European Spanish version of the questionnaire employed for the judgment tasks (cf. Section 5). The items are arranged according to definite and indefinite human direct objects. Note that they exhibit DOM although the presence and absence of DOM was altered and pseudo-randomized in the questionnaires.

Test items with definite human direct objects:

1. *Gonzalo mató al mafioso en el parque.*  
'Gonzalo killed the mobster in the park.'
2. *Patricio lesionó al portero en la discoteca.*  
'Patricio injured the doorman at the nightclub.'
3. *Carolina golpeó al futbolista en el vestuario.*  
'Carolina hit the soccer player in the locker room.'
4. *Elena cuidó al herido durante dos semanas.*  
'Elena took care of the injured person for two weeks.'
5. *Irene acusó al secretario general en la asamblea.*  
'Irene accused the general secretary at the assembly.'
6. *Cecilia denunció al ladrón en la comisaría de policía.*  
'Cecilia reported the thief at the police station.'
7. *Josefina vio al cura en el parque.*  
'Josefina saw the priest in the park.'
8. *Julio oyó al basurero desde la cocina.*  
'Julio heard the garbage collector from the kitchen.'

Test items with indefinite human direct objects:

9. *Pablo mató a un rehén durante el secuestro.*  
'Pablo killed a hostage during the kidnapping.'

10. *Alberto lesionó a un defensa en el partido de la semana pasada.*  
'Alberto injured a defender during last week's match.'
11. *Andrea golpeó a un fotógrafo a la salida del restaurante.*  
'Andrea hit a photographer at the restaurant's exit.'
12. *Javier cuidó a un enfermo con gran profesionalidad.*  
'Javier took care of an injured person very professionally.'
13. *Laura acusó a un trabajador en la reunión.*  
'Laura accused a worker at the meeting.'
14. *Carlos denunció a un camarero tras una disputa.*  
'Carlos reported a waiter after an argument.'
15. *Ana vio a un político en la tienda del barrio.*  
'Ana saw a politician at the neighbourhood's shop.'
16. *Esther oyó a un bebé en el apartamento de al lado.*  
'Esther heard a baby in the apartment next door.'

# Index

- acceptability 10–11, 49, 65, 69, 71, 72, 77, 80, 81, 86, 87, 88, 90–93, 95–97, 141, 146, 148, 192, 213, 218–219, 229, 231, 232, 234, 235, 238, 253, 279–281, 285, 286, 289, 291, 292, 294–308, 310–312, 339, 354–356, 358–362
- Acceptability Judgment Task 66, 67, 88, 90, 93, 95, 96, 194, 196, 204, 215, 234, 238, 282, 296, 297, 299–302, 306, 307, 311, 312, 339, 354–362
- accusative 3, 5, 10, 21, 26, 30, 31, 32, 34, 44–46, 104, 107–109, 111–114, 116, 117, 131, 132, 142–144, 152, 154, 155, 163, 168, 182, 210, 214, 220, 222–229, 244, 253, 280, 317
- double accusative 143, 144, 152, 154, 155, 163
- affectedness 4, 57, 58, 74, 75, 79, 80, 84, 104, 109, 142, 143, 152, 153, 182, 194, 197, 200, 299, 315, 316, 320, 332, 333, 340, 342, 349, 350, 354, 356, 357
- agentivity 4, 58, 59, 65, 69, 72–74, 76, 78–80, 83, 85, 87, 88, 97, 98, 111, 143, 146, 154, 156, 315, 316, 332, 340
- agent–patient–asymmetry 76, 82, 84, 85, 90, 95
- ambiguity 41, 46, 59, 248, 249, 251, 272, 282, 302, 312
- ambiguous 80, 142, 184, 248
- ambitransitive 112
- analogy 45, 51, 229
- animacy 4, 5, 7, 21, 23, 24, 27, 30, 31, 34, 39–41, 46, 47, 50, 52, 53, 56, 58, 59, 66, 72, 73, 76, 79, 83, 87, 90, 97, 98, 103, 104, 106–109, 113–118, 120, 121, 124–127, 131–135, 141, 143, 144, 147, 152–156, 160, 161–163, 168, 213, 216–220, 222–228, 230–232, 234, 235, 237, 238, 244, 245, 249, 252, 254, 257, 258, 260, 270, 272, 274, 279, 284–286, 288–291, 294–303, 308, 311–313, 316, 321, 340, 342, 347, 350, 355, 356
- animal 27, 32, 40, 120, 141, 155, 218, 279, 289, 295–298, 300–302, 313
- argument 74, 76, 80, 82, 83, 98, 137, 155, 320, 322
- argument realization 73
- argument selection 74
- argument marking 103, 108, 110–112, 114, 117, 132
- argument structure 104, 118, 143, 155, 166, 320, 322
- argumenthood 173, 174, 180, 181, 196, 204, 208, 209
- aspect 78, 237, 316–319, 322, 323, 326–332, 335
- atelic 317–322, 324–326, 328–332
- attitudes 119
- bare nominals 120, 174, 175, 181–183, 192–196, 203–206, 208, 209, 243, 244, 254, 256, 270, 272, 274, 321, 344, 345, 350
- bound variable see *variable*
- boundedness 317
- case marking 6, 29, 41–44, 56–58, 74, 83, 103–105, 107, 110–114, 116, 117, 119, 174, 183, 200, 208, 210, 214–216, 222, 224, 238, 280, 316–318, 340
- causation 74, 75, 79, 84, 145, 156
- causative construction 66, 68–71, 77–80, 84, 85, 145, 156, 182, 192, 194, 196, 344
- change of state predicate 74, 79, 84
- clitic 25, 43, 71, 106, 108, 109, 114, 115, 117, 120, 217, 220, 225, 228–230, 232–234
- clitic doubling 4, 5, 58, 71, 103, 104, 106–109, 114–118, 119–121, 131–134, 147, 183, 184, 202, 220–223, 235–238, 245, 286, 287, 290, 304–307, 316, 356
- clitic–left–dislocation 70, 71, 81, 86, 89, 104, 115, 116, 120, 121, 123–126, 129
- clitic–right–dislocation 104, 115, 116, 120, 121, 123, 124, 126, 127, 129, 130
- constraints on DOM 59, 69, 71, 72, 77, 85, 86, 98, 109

- construction 4, 7, 11, 26, 28, 34, 42, 45, 47,  
 49, 54, 66, 68, 69–71, 77, 78, 96, 108,  
 111, 143, 163, 166, 169, 182, 194, 342,  
 344, 349, 350, 355  
 coordination 157, 223, 232, 247  
 corpus 4, 9, 10, 11, 65, 68, 69, 71, 87, 89,  
 140, 143, 156, 144, 238, 243, 259, 260,  
 264, 270, 272, 315, 317, 319–324, 328,  
 329, 332, 339, 341, 342, 343, 345,  
 347–350, 353–355, 358–361  
 dative 32, 34, 41, 43–45, 83, 86, 104, 114,  
 202, 214, 217, 220, 225–229, 244  
 definiteness 4, 5, 21, 23, 24, 26, 29–31, 34,  
 39–43, 46, 47, 50–53, 55, 56, 58, 59,  
 66, 67, 69, 72, 73, 85, 88–90, 92–94,  
 96, 103, 106–109, 113, 114, 116–118,  
 120–124, 130–135, 141, 147, 155, 163,  
 167, 168, 180, 206–208, 216–220, 233,  
 234, 243–245, 253, 254, 256, 257, 259,  
 260, 266–270, 272, 279, 284, 290, 297,  
 321–327, 329, 331, 339, 341–359, 361,  
 362, 367  
 deity 259, 266, 267, 271, 272  
 diachrony 3, 7, 9, 10, 26, 40–42, 52, 55, 56,  
 58, 65, 115, 140, 214, 217, 224, 229, 235,  
 243, 245, 257, 259, 269, 270, 272, 274,  
 316, 317, 319–321, 324, 329, 332, 339,  
 342, 343, 347–349, 354, 361  
 dialect 8, 22, 32, 44, 88, 104, 105, 134, 228,  
 235, 243, 245, 246, 251, 253, 255, 257,  
 263–266, 279, 281–285, 288, 291–295,  
 297, 300, 301, 303, 306, 308–313  
 diatopic 7, 8, 227, 259, 281, 283, 285, 288,  
 303, 312  
 differentiality 22, 23, 30, 31, 35  
 differential goal marking 6, 57  
 differential subject marking 5  
 disambiguation 41, 46, 47, 76, 184, 223,  
 224, 360,  
 dislocation 44, 47, 70, 81, 96, 103, 107–109,  
 114, 117–120, 126, 131, 132, 273, 283,  
 284, 287, 305–307  
 – left dislocation 53, 54, 67, 69, 71, 72, 85,  
 86, 88, 104, 112, 113, 116, 140, 251, 321,  
 345, 350  
 – right dislocation 69, 85, 86, 87, 88,  
 104, 113, 116, 290, 291, 304, 306,  
 309–311, 313  
 ditransitive 144, 155, 163, 342  
 elicitation 10, 118, 119, 141, 146, 150, 152,  
 157–159, 163–167, 169  
 event 57, 58, 73–76, 79, 80–83, 142, 180,  
 191, 317–319, 330  
 experiencer 65, 75, 78, 95, 97  
 experiment 4, 8, 10, 11, 96, 97, 141,  
 146–148, 150, 152, 153, 157, 159,  
 160, 165, 166, 170, 279–281, 284–295,  
 301, 302, 304–306, 311–313, 356,  
 358, 360  
 flagging 4, 6, 8  
 frequency 7, 8, 10, 11, 27, 30, 52, 53, 56, 76,  
 88, 103, 105, 109, 124, 128, 130–135,  
 141, 144, 146, 149, 214, 217, 224, 230,  
 232, 233, 235, 238, 245, 253, 258, 260,  
 265, 268, 287, 302, 320, 323–328,  
 339, 340, 341, 342, 344, 345, 347, 349,  
 350–353, 354, 359, 361  
 genericity 51, 69, 85, 86, 88–94, 96,  
 97, 100  
 global factors of DOM 7, 47, 98, 167, 316  
 goal 330, 331  
 grammaticalization 4–6, 9, 39, 40, 42,  
 46–49, 50, 51, 53–56, 59, 65, 72, 86, 69,  
 98, 105, 168, 274  
 heritage language 9, 106, 140, 148–150, 159,  
 168, 340  
 hierarchy (see also *scale*) 41, 51, 53, 59, 108,  
 126, 132, 147, 216  
 human 32, 39–43, 49, 51–59, 67, 69, 72,  
 79, 83, 108, 109, 114, 120, 121, 126, 127,  
 141, 156, 161, 162, 216, 218, 220, 221,  
 223–227, 229–231, 243, 244, 246, 247,  
 251, 252, 254, 256, 258, 259, 262–264,  
 266–270, 272, 273, 279, 284, 286, 289,  
 295–298, 299–303, 311, 312, 315, 316,  
 320, 321, 323, 324, 326, 328, 329, 332,  
 339, 341–359, 361, 362, 367



- inanimate 7, 11, 24, 26, 31, 39, 49, 53, 56,  
 58, 72–74, 76, 79, 83, 90, 108, 109,  
 115, 120, 124–126, 130, 133, 140–145,  
 147, 152–155, 160–163, 167–170, 200,  
 216–218, 220, 221, 223, 225, 227, 230,  
 231, 233, 234, 237, 243, 244, 251, 252,  
 256, 258, 271–274, 279, 284, 286, 289,  
 295–298, 300–302, 313, 321, 329, 330,  
 332, 342, 350, 356
- indefiniteness 24, 26, 29–31, 40, 50–52, 55,  
 69, 85, 87, 88, 91, 92, 96, 107, 115, 120,  
 122, 123, 133, 140, 144, 145, 147, 152, 155,  
 156, 163, 166, 167, 168, 174–179, 181–190,  
 192–197, 190, 198, 203, 204, 206, 208,  
 209, 216, 218, 219, 233, 243, 244, 254,  
 256, 259, 269, 270, 272–274, 321–329,  
 331, 332, 339, 341, 343–362, 359, 360, 367
- indexing 4, 57, 104, 222
- individuation 42, 50, 51, 53, 55, 59, 72, 88
- information structure 7, 39, 46, 48, 49, 59,  
 71, 88, 96, 104, 106–109, 114–117, 126,  
 129, 130, 133, 158, 281, 284, 286, 288,  
 290, 291, 322, 354, 356
- inherence 23, 28, 30, 31, 35, 98, 106, 142,  
 230, 318, 320, 322, 327, 329, 331
- interaction 65, 68, 72, 77, 78, 81–85,  
 88–96, 99
- language change 9, 26, 27, 49, 51, 65, 268,  
 274, 345, 354, 361
- language contact 3, 8, 9, 26, 27, 29, 30,  
 104–107, 109, 116, 117, 130–132, 134,  
 135, 140, 141, 146–150, 158, 165, 166,  
 169, 227, 234, 235, 274, 280, 285, 345
- left dislocation *see dislocation*
- leísmo 6, 58, 147, 150, 213–15, 225–229,  
 235–238
- local factors of DOM 7, 167, 316
- marking *see argument marking, case marking,*  
*differential goal marking, differential*  
*subject marking, topic-marking*
- monotransitive 342
- morphemmarkiertes Satzobjekt 22
- Nominal DOM 25, 30, 44, 49, 220, 222, 223, 235
- non-agentive 78, 79
- non-agreeing 109
- non-ambiguous 248
- non-prototypical 5, 6, 65, 71, 76, 87, 97
- non-specific 50, 51, 55, 85, 96, 115, 144, 152,  
 161, 167, 174, 176, 177, 181, 183–186,  
 188, 195, 211, 216, 219, 244, 254, 274,  
 345, 349, 350, 359, 360
- normative 225, 227, 229, 243–245, 251, 258,  
 275, 280, 282–284, 303, 311
- null object 6, 110, 213–215, 229–241
- object-experiencer 65, 66, 68
- object-experiencer-psych-verbs 66, 68, 69,  
 70–72, 77–81, 84, 85, 88–98, 92, 95,
- onomasiological 4
- overt pronoun 6, 110, 215, 231, 238, 241
- partitive 25, 176, 317
- patient 57, 58, 65, 69, 73, 74, 87, 95, 97, 104,  
 107, 108, 111, 115, 118, 120, 127–135,  
 142, 143
- periphery *see sentence periphery* 53, 66, 67,  
 69, 77, 85, 87, 88, 97, 100, 218
- place name 271, 272
- pragmatics 41, 45, 49, 51, 54, 59, 70, 72, 82,  
 117, 246, 274, 340
- preposition 3–5, 10, 21, 22, 25, 28, 32, 34,  
 41–44, 53, 55, 90, 117, 131, 166, 183,  
 213–216, 217, 222, 224, 225, 244,  
 235–238, 244, 249, 251–254, 264, 280,  
 283, 289, 295, 319, 320
- prominence 42, 48, 51, 54, 55, 59, 65, 72,  
 114, 272
- pronoun 23, 32, 39–42, 44–53, 54, 57,  
 59, 63, 67, 69, 77, 81, 85–90, 92–94,  
 96, 111, 114, 115, 117, 147, 168, 204,  
 214–217, 220–231, 235–238, 243–246,  
 248, 252, 253, 258–261, 271–273, 280,  
 282, 286, 287, 290, 291, 304–310, 316,  
 344, 345
- proper name 23, 32, 39, 42, 49–54, 59, 85,  
 87–89, 92, 93, 96, 115, 147, 168, 204,  
 216, 221, 243, 244, 251, 253, 254, 258,  
 259, 262–264, 266, 271–273, 284, 297,  
 321, 344, 345, 350
- proto-agent 65, 73–76, 79, 80, 82–85, 90,  
 95, 97

- proto-patient 73–75, 79, 80, 82–85, 88, 90, 95, 98  
 prototypical 10, 26, 47, 58, 69, 73, 76, 80, 84, 87, 90, 95, 97, 167  
 quantifier 173, 175, 178, 197, 199, 202, 206, 208, 247, 344  
  
 reference 23, 29, 30, 34, 35, 39, 41, 42, 47–49, 51, 54, 59, 65, 72, 85–88, 90, 92, 93, 96, 104, 115, 116, 141, 150, 154, 176, 207, 209–211, 219, 220, 223, 231, 298, 316, 320, 340, 345  
 referent 7, 39, 40, 47, 48, 50, 51, 53–56, 58, 75, 91, 152, 153, 167, 176, 184, 216, 220, 221, 224, 226, 227, 229, 231, 234, 238, 251, 256, 269, 270, 316, 350  
 reflexive 203, 261  
 register 70, 71, 80  
 resultativity 317  
 retraction of DOM 273, 339, 340, 342, 347, 348, 353–356, 359–362  
 reversible predicate 75, 141, 143, 153, 154, 160, 162, 163, 165–170  
 reversible-converse predicate 75  
 reversible-symmetrical predicate 75  
 roles see *semantic roles*  
  
 scale 216, 224, 225, 235, 237, 238, 302, 311  
 – acceptability scale 90, 291, 292, 295  
 – affectedness scale 142, 143  
 – animacy scale 40, 47, 120, 114, 131–134, 147, 168, 216, 220, 244, 254, 260, 272, 286, 289  
 – definiteness scale 47, 50, 56, 58, 114, 131–134, 147, 168, 216, 218, 220, 244, 254, 260, 270, 272  
 – Likert scale 146, 356  
 – referentiality scale 85, 87, 88, 92, 93, 96  
 – specificity scale 244, 270  
 – topic acceptability scale 49  
 scope 173–176, 178, 179, 181, 183, 184, 188–191, 195–198, 200–203, 205, 206, 208, 210, 211  
 semantic roles 46, 57, 65, 68, 71–73, 76–78, 83, 84, 86, 97, 104, 107, 108, 116, 118, 120, 127–129, 131–133, 135, 143  
 – (see also *agentivity, agent-patient-asymmetry, causation, experiencer, goal, object-experiencer, object-experiencer-psych-verbs, patient, proto-agent, proto-patient, theme*)  
 semasiological 4  
 sentence periphery 53, 66, 67, 69, 77, 85, 87, 88, 97, 100, 218  
 singularity 42, 52, 53, 55, 59  
 sociolect 10  
 sociolinguistics 10, 70, 81, 257, 258, 273, 282, 339, 342, 356  
 specificity 26, 47, 50, 55, 75, 85, 91, 96, 103, 106, 115, 116, 142, 144, 145, 150, 152, 153, 160, 161, 166, 167, 173–179, 181–188, 195–197, 198, 200–203, 205, 206, 208, 210, 211, 213, 216–222, 230–232, 234, 235, 237, 238, 244, 254, 269, 270, 272, 274, 316, 318, 319, 345, 350, 359, 360  
 spoken language 65, 85, 88, 97, 141, 142, 146, 215, 227, 228, 230, 238, 244, 245, 252, 254–256, 279, 281, 283, 284, 341, 344, 353–355, 360, 361  
 standard language 66, 71, 72, 76, 86, 149, 227, 245, 246, 252, 258, 279, 281–284, 288, 311, 354, 355  
 standardization 245, 275  
 stereotype 8  
  
 telicity 4, 104, 315–333, 340, 342  
 thematic distinctness 73, 74, 76, 81, 84, 85  
 thematic role see *semantic role*  
 theme 74, 84, 104, 107, 108, 111, 116, 118, 120, 127–133, 133  
 third wave 3, 4, 7, 10  
 topicality 41–49, 53, 54, 56, 59, 72, 86, 88, 96, 98, 103, 104, 111, 113, 115, 184, 251, 284, 285, 316, 340, 342  
 topicalization 26, 39, 42–49, 53–55, 59, 86, 88, 116, 286, 287, 290, 304–307, 309–312  
 topic-marker 42, 59, 71  
 topic-worthiness 41, 109, 222  
 transitivity 34, 41, 43, 56–58, 66, 73, 74, 76–78, 80, 81, 84, 88–95, 97, 99, 107,

- 108, 110–112, 120, 132, 134, 141, 142,  
145, 152, 159, 163, 165–169, 214, 216,  
222, 229, 238, 288–290, 317, 332,  
333, 349
- typology 3–7, 9, 21, 22, 28, 42, 85, 103, 105,  
106, 109, 110, 117, 130, 134, 135
- variable 199, 202, 232, 233, 321, 322  
– bound variable 51, 197, 198
- variance 303, 312
- variation 4, 7, 8, 26, 52, 58, 59, 103, 104, 106,  
109, 115, 116, 130, 133, 134, 139, 141, 146,  
147, 160–166, 168–170, 201, 220, 227,  
234, 235, 243, 259, 279, 281, 283, 285,  
287, 288, 300, 309, 312, 321, 327, 329,  
332, 342, 351–354, 356, 358, 362
- written language 70, 88, 119, 142, 238,  
243–245, 252, 254, 354, 355, 360, 361

