Differential Object Marking and the Lexical Semantics of Verbs in Spanish

Klaus von Heusinger & Georg A. Kaiser

http://www.itg.uni-stuttgart.de/vonHeusinger/
georg.kaiser<at>uni-konstanz.de

1. Introduction*

In Spanish, the direct object can be accompanied by the marker a, which is homophone to the preposition a ‘to’; hence, such direct objects are often called ‘prepositional accusative’, which is an type of Differential Object Marking (DOM). DOM is widespread among the languages of the world. Bossong (1985) mentions that at least 300 known languages exhibit DOM in one way or other. DOM or a-marking of the direct object in Spanish is a well studied subject (e.g. Brugè & Brugger 1996, Torrego 1999, Leonetti 2004, Pensado 1995 with a annotated bibliography), but there is no overall account of the different parameters that determine DOM, nor is there a theory that explains the interaction of the parameters. Two main approaches to DOM in general are currently under discussion: The Ambiguity Thesis and the Transitivity Thesis. The Ambiguity Thesis (Comrie 1975, Moravcsik 1978, Croft 1988, Bossong 1985, Aissen 2003) assumes that languages that do not distinguish subject and direct object tend to develop extra markers to indicate direct objects if they are too similar to typical subjects. These approaches focus on the properties of the direct object such as animacy, definiteness, specificity and topicality. The Transitivity Thesis (Hopper & Thompson 1980, Naess 2004), on the other side, assumes that a direct object is marked if it is a “good” argument in a transitive sentence. These approaches focus on semantic properties of the verb such as telicity, aspectuality and thematic information of the argument role of the direct object such as volitionality and agency. De Hoop & Narasimhan (2005) modify the Transitivity-Thesis and use the concept of “Strength of an Argument”. According to them, DOM-languages mark strong arguments in direct object positions. We will develop this thesis further and account for the strength by analyzing the interaction of the properties of the direct object with the lexical semantics of the verb. Most synchronic research on DOM or a-marking in Spanish focuses on the properties of the direct object, while the verbal semantics has been less well investigated. Moreover, diachronic studies focus exclusively on these properties and do not investigate the verb class. In our study we investigate the role of the verb class for the diachronical development of a-marking in Spanish. Our findings clearly show the importance of the verbal semantics and they also motivate the variations that can be found in DOM marking, both in synchronic and diachronic data.

DOM in Spanish, as in other languages, can only be explained by the interaction of several parameters. We assume the following three families of parameters for DOM: (i) the properties of the direct object, (ii) its competition with other arguments in the sentence (mainly the subject), and (iii) the lexical semantics of the verb. The properties of the argument include animacy, definiteness, specificity and topicality. Even though these categories originate on

* We would like to thank the audience at the International Workshop “Definiteness, Specificity and Animacy in Ibero-Romance Languages (Nereus III)” in Alcalá de Henares (Madrid) in October, 7th, 2006, and at the International Workshop “New Perspectives on Morphological Change” in Berlin in October, 28th, 2006 for constructive comments, and the organizers Manuel Leonetti and Elisabeth Stark for organizing these two inspiring events. We are also indebted to Victoria Escandell-Vidal for her valuable comments on a previous version of this paper. Special thanks go to Bruce Mayo for his detailed and helpful comments on our English. A substantial part of the corpus searches were undertaken by our student assistants Heidrun Bohnet, Annika Deichsel, Katharina Ruchti and Helga Szilagyi. This research was supported by the German Science Foundation (SFB 732 “Incremental Specification in Context”, for the first author, and in the SFB 471 “Variation and Change in the Lexicon”, for the second author).
different grammatical levels, they converge to an overall category “referential status”, which is often described in terms of degrees of “individualization”. The competition with the subject triggers DOM: if the direct object is too similar to a typical subject (animate, definite, topical), a DOM language tends to mark the object in order to disambiguate between the two arguments. At the same time the lexical semantics of the verb determine the role of the argument in the described event. Sentences with verbs that prefer animate direct objects are more likely to mark animate objects low on the Referentiality Scale than sentences with verbs that prototypically take inanimate objects. We show that the subtle interaction between the verbal semantics, on the one hand, and the referential properties of the noun, on the other, can explain the variation in DOM that we find. In particular, we try to account for diachronic variation in Spanish. Diachronic studies show that DOM in Spanish spread from personal pronouns and proper names to definite and finally indefinite noun phrases (all human/animate). Our data suggests that besides this general picture, the lexical semantics of the verb is a driving force in the diachronic evolution of DOM. In a detailed analysis we compare a particular piece of literature (selected chapters of the Bible) from different periods (14th, 16th and 20th century) and regions (including an American bible translation of the 20th century). Since we are dealing with the same text, this detailed comparison allows us to control the contextual settings. It also shows that the properties of the direct object do not suffice to explain the variation and the diachronic development of DOM. We therefore extend our analysis to different verb classes and extend our corpus search to the whole Bible, and in a another step we use an even broader text corpus (the electronically available Corpus del Español from the 12th to the 19th century). The findings confirm our original hypothesis that the verb class is a main parameter for DOM in Spanish.

2. Parameters of Differential Object Marking in Spanish

As noted above, differential Object Marking (DOM) in Spanish is expressed by the marker a, which is homophone to the preposition a ‘to’ and the dative marker a.¹ DOM in Spanish is determined by three families of parameters: (i) referential properties, (ii) competition between the arguments in a sentence, and (iii) transitivity properties of the verb, i.e. the lexical semantics of the verb. The following subsections are devoted to these parameters.

2.1 Nominal properties

Some categories of a noun are animacy, referentiality (definiteness and specificity), and topicality.² These properties derive from different types of information: Animacy is a lexical (or conceptual) property, specificity is a referential property, definiteness a discourse pragmatic one, and topicality a property of information structure. Still, all these properties interact and yield a more general concept of “referential status”, which corresponds to the often mentioned category “individuation”. Each particular parameter can be expressed by a scale of two or more values. A language cuts across the scale at one particular point – the language specific-transition point.

Silverstein (1976) has discussed the role of animacy for case-marking, see also Comrie (1975). We assume that animacy is a lexicalized conceptual category, i.e. speakers categorize

¹ For want of space we cannot discuss the role of clitic doubling for DOM, but see Suñer (1988), Brugé & Brugger (1996) and Parodi (1998) for clitic doubling and von Heusinger & Kaiser (2003) for the relation between a-marking and clitic doubling as expressing DOM. Leonetti (2004:100) states that the conditions licensing clitic doubling are a subset of the conditions that license a-marking. He refers to Bleam (1999:199), who points out that “the semantic properties which give rise to clitic doubling form a subset of the semantic properties which give rise to the prepositional accusative [...]”. We expect that our findings for a-marking are also relevant for clitic doubling, but we leave this for further research (see Leonetti (this volume)).

² There are additional referential properties such as number, collectivity, concreteness etc. that influence the “individualization” of an argument.
objects they speak about according to different values of animacy. The Animacy Scale (1) distinguishes three values. Depending on the language, we find a transition point between human and animate or between animate and inanimate:

(1) Animacy Scale:
     human > animate > inanimate
     + human | animate | inanimate
     - human

Spanish seems to take the first option and distinguishes between +human and –human. While the main information for animacy is encoded in the lexicon, additional contextual effects may shift the animacy marking to the higher or lower pole of the scale (see Weissenrieder 1990):

(2) (a) Vi *(a) la / una mujer.
       saw-1SG to the / a woman
(b) Vi *(a) la / una mesa.
       saw-1SG to the / a table
       ‘I saw the / a table.’

The main parameter, however, for the individualization of a noun is referentiality as expressed in the Referentiality Scale3 which combines definiteness and referentiality. This scale marks personal pronouns (Pro) most strongly, followed by proper names, definite noun phrases, specific indefinite noun phrases, and nonspecific indefinite noun phrases, with non-argumental nouns at the lowest end. The version presented in (3) combines different types of such Scales (see Comrie 1975, Bossong 1985, Croft 1988, Aissen 2003 and others). In particular, we have added the non-argumental noun slot, which is crucial for the description of DOM in Spanish.4

(3) Referentiality Scale (extended version of Aissen 2003:437: “Definiteness Scale”):
     personal pronoun > proper noun > definite NP > indefinite specific NP > indefinite non-specific NP > non-argumental

<table>
<thead>
<tr>
<th>Pro &gt;</th>
<th>PN &gt;</th>
<th>Def &gt;</th>
<th>Spec &gt;</th>
<th>–Spec &gt;</th>
<th>–Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;e,t&gt;-type</td>
</tr>
<tr>
<td>argument-status (e-type)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The scale can be cut in two at different transition positions. For contemporary European Spanish we assume that the cut is below non-specific indefinites (–Spec) and above non-argumentals (–Arg). In other words, DOM in Spanish (for animate direct objects) indicates that the noun is an argument rather than predicate that might be incorporated. (Leonetti 2004). Bleam (1999) formulates this distinction in terms of the type of noun: argument type e, vs. non-argumental type of <e,t>. The definite noun phrase in (4a) and the indefinite (specific) noun phrase in (4b) must be marked by a. Even the non-specific indefinite noun phrase in (4c) may optionally marked with a. The non-specificity is triggered by the subjunctive sepa in the

---

3 Contrary to our use of “Definiteness Scale” in earlier papers and the use of Aissen (2003), we prefer here to name the scale under discussion “Referentiality Scale” following Croft (2003:130). The advantage of this terminology is that it allows for a description that is independent from the otherwise well known and well grammaticalized definiteness. The terminology also allows us to include non-argumental direct objects, whether we call them incorporated or not.

4 We have suppressed many other categories such as possessives (depending on the language type: above definites), universals (pattern with definites), partitives (between definites and specific indefinites), different kinds of specific indefinites (see Hapxelmach 1997) etc. This linear scale does not provide a slot for non-referential definites (see section 4.3 for an example of this kind).
relative clause. Even the indefinite (non-specific) pronoun *alguien* in its possible non-specific reading takes *a*. Only the predicate <e,t>-type meaning of (4e) does not allow *a*:

(4) (a) Vi *<a> la mujer.*
    saw-1SG to the woman
    ‘I saw the woman.’
(b) Vi *<a> una mujer.*
    saw-1SG to a woman
    ‘I saw a woman.’
(c) Necesitan *<a> un ayudante que sepa inglés.*
    need-3PL to a assistant that speak-SUBJ.3SG English
    ‘They need an assistant who knows English’
(d) Está buscando *a alguien.*
    is looking to someone
    ‘(S)he is looking for someone’
(e) El dentista necesita *<a> un ayudante.*
    the dentist needs DOM a assistant
    ‘The dentist needs an assistant’

Topicality is a notion of information structure and therefore somewhat vague and difficult to test in corpora. Here we adopt the notion of “aboutness-topic”, i.e. the topic of a sentence is that piece of information that the sentence is about. Topics can be syntactically or intonational marked; the latter is hard to detect in corpora. The former would be a good test. We assume that a left moved direct object is topical, but we cannot tell whether a direct object close to the verb is topical or not.

(5) Topicality
    
    | topical | – topical |
    |---------|---------|

We assume that there are only topical and non-topical direct objects. The indefinite direct object right of the verb may optionally take *a*, while the left-moved one in (6b) must take it (cf. Leonetti 2004:86).

(6) (a) Ya *conocía <a> muchos estudiantes.*
    already knew-1SG to many students
    ‘I already knew many students’
(b) *(A) muchos estudiantes, ya los conocía.*
    (to) many students, already them knew-1SG
    ‘Many students I already knew’

We can summarize the conditions for DOM in Spanish in (7): Spanish has to mark the direct object if it is human and an e-type argument. In all other cases the marker must not be used.

(7) DOM in Modern Spanish: Simplified description:

<table>
<thead>
<tr>
<th>Standard Spanish</th>
<th>+ Arg</th>
<th>– Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td>+human</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>-human</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
2.2 Disambiguation and Competition

One approach to DOM in Spanish is to assume that a-marking disambiguates between subject and object. If the object is too similar to the subject, a-marking is necessary to indicate objecthood. It is interesting that a-marking is obligatory (or at least optional) even in those cases where we have two inanimate arguments (Weissenrieder 1991:146) (see also Torrego Salcedo 1999, García-García (this volume)).

(8) (a) ...que los gerundios modifican al sujeto.
that the gerunds modify to-the subject
‘...that the gerunds modify the subject’
(b) En esta receta, la leche puede sustituir al huevo.
In this recipe the milk can replace to-the egg
‘In this recipe egg can replace the milk.’

There is also a second aspect of competition and disambiguation: There are other constituents that are marked with a and therefore compete with the direct object. The indirect object in Spanish is obligatorily marked by a. As a consequence, if only one constituent is marked, it must be the indirect object (Torrego 1999:1784):

(10) (a) Perseguía al guardia el ladrón.
pursued to-the policeman the thief
‘The thief pursued the policeman’
(b) Perseguía el guardia al ladrón.
pursued the policeman to-the thief
‘The policeman pursued the thief’

In ditransitive sentences and in sentences with nouns marked with the preposition a, DOM-marking of the direct object can create ambiguity of the grammatical functions, as in (11a). In order to avoid this ambiguity, it often occurs that direct objects – but never indirect or prepositional objects – appear unmarked in these sentences, as in (11b), even though they would be otherwise marked. This avoidance of a is only a “stylistic rule” and is, according to Real Academia Española (1973:374f), mostly applied in cases when both objects are full nouns:

(11) (a) Ha sido forzoso dejar al conde en rehenes al enemigo.
has been compelling leave to-the count as hostage to-the enemy
(b) Ha sido forzoso dejar el conde en rehenes al enemigo.
has been compelling leave the count as hostage to-the enemy
‘It has been compelling to leave the count as hostage to the enemy’

Disambiguation and competition are important factors for a-marking in Spanish, in particular for non-core cases, as in (8). However, they do not furnish an overall account or parameter able to explain fully the distribution and variation of a-marking.

2.3 Transitivity and the lexical semantics of verbs

DOM-marking in Spanish depends not only on the referential properties of the direct object and its competition with the subject or other arguments in the sentence, but also on the lexical properties of the verb. This has been noted in descriptive grams of Spanish (Bello 1847:567-570, Fernández Ramírez 1951:151-190 and others). Particular approaches to describe DOM in terms of verb classes have been undertaken by Bolinger (1953), Fish
(1967), Pottier (1968), Leonetti (2004) and others (see also an overview in Pensado 1995 and Torrego 1999). These observations fit nicely into the more general approach of Transitivity of Hopper & Thompson (1980). They argue that the categories given below are ordered in a particular way: Languages prefer to mark categories with high transitivity values morphologically, rather than the lower values. Spanish, for example, marks the direct object with a if it is high in individuation (see point 10 in (12)). At the same time it also marks a telic event by this means (see example (13)), or affectedness (see examples (17)-(18) below). Hopper & Thompson (1980) account for the particular alignment of values shown in (12) by assuming that all high transitive values contribute to the discourse salience of the event described by the verb and its arguments.

(12) Parameters of Transitivity (Hopper & Thompson 1980)

<table>
<thead>
<tr>
<th></th>
<th>High transitivity</th>
<th>Low transitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Participants</td>
<td>Two participants or more (A and O)</td>
<td>one participant</td>
</tr>
<tr>
<td>2. Kinesis</td>
<td>Action</td>
<td>Nonaction</td>
</tr>
<tr>
<td>3. Aspect</td>
<td>Telic</td>
<td>Atelic</td>
</tr>
<tr>
<td>4. Punctuality</td>
<td>Punctual</td>
<td>Nonpunctual</td>
</tr>
<tr>
<td>5. Volitionality</td>
<td>Volitional</td>
<td>Nonvolitional</td>
</tr>
<tr>
<td>6. Affirmation</td>
<td>Affirmative</td>
<td>Negative</td>
</tr>
<tr>
<td>7. Mode</td>
<td>Realis</td>
<td>Irrealis</td>
</tr>
<tr>
<td>8. Agency</td>
<td>A high in potency</td>
<td>A low in potency</td>
</tr>
<tr>
<td>9. Affectedness of O</td>
<td>O totally affected</td>
<td>O not affected</td>
</tr>
<tr>
<td>10. Individuation of O</td>
<td>O highly individuated</td>
<td>O nonindividuated</td>
</tr>
</tbody>
</table>

For Spanish, different aspects of the lexical semantics of verbs have been described with respect to a-marking, including telicity, volitionality, type of causation, mode, agency and affectedness. For lack of space we can only refer to some of these aspects below.

According to Torrego (1999:1786f), telicity functions as a strong parameter for DOM in Spanish, for a-marked direct objects are obligatory with telic verbs, such as insultar (‘insult’):

(13) Marta insultó *(a) un compañero.
      Marta insulted-3SG to a colleague
      ‘Marta insulted a colleague’

DOM-marking can express subtle differences in the semantics of events. Torrego (1999:1788) notes the difference between (14a) and (14b): In (14a) the object is not identified nor identifiable, while in (14b) it is known (at least for the speaker) and more individualized. Moreover, in (14b) the subject is more strongly involved in the event, and the object constitutes an independent entity. This is also shown in (15) where the predicative llorando ‘crying’ is a secondary predicate. In (15a) it can only be predicated of the subject, while in (15b) it can be applied to the subject or the (highly individualized) object (Torrego 1999:1789).

(14) (a) Besaron un niño.
      kissed-3PL a child
(b) Besaron a un niño.
      kissed-3PL to a child
      ‘They kissed a child’
(15) (a) Besaron un niño llorando.  
   kiss-3PL a child crying
   ‘They kiss a child while they were crying’
   (secondary predication: subject)
(b) Besaron a un niño llorando.  
   kiss-3PL to a child crying
   ‘They kiss a child while they were crying’ OR: ‘They kiss a crying child’

According to Torrego (1999:1786) the verb matar ‘to kill’ strongly prefers a with direct objects, but does not require it. Thus the presence or absence of the marker a with animate objects expresses a very subtle difference in the kind of event: (16a) with the marked object is understood as expressing direct causation, while in (16b) without the marker expresses indirect causation.5

(16) (a) Han matado a un buscador de oro.  
   have-3PL killed to a searcher of gold
   ‘They have killed a gold searcher’
   (direct causation)
(b) Han matado un buscador de oro.  
   have-3PL killed a searcher of gold
   ‘They have killed a gold searcher’
   (indirect causation)

Torrego (1999:1791) discusses affectedness of the object by the event expressed in the verb as a very strong a-trigger in Spanish, as illustrated in the next examples with eventive or stative verbs. The affectedness can relate to physical circumstances, as in (17), or to psychological ones, as in (18). The (b)-examples without the marker are ungrammatical.

(17) (a) Golpearon a un extranjero.  
   beat-3PL to a stranger
   (b) *Golpearon un extranjero.  
   beat-3PL a stranger
   ‘They have beaten a stranger’

(18) (a) Odia a un vecino.  
   hate-3SG to a neighbor
   (b) *Odia un vecino.  
   hate-3SG a neighbor
   ‘(S)he hates a neighbor’

Contemporary Spanish has lexicalized this contrast: a whole class of verbs obligatorily take a with animate objects, such as saludar (‘greet’), odiar (‘hate’), insultar (‘insult’), castigar (‘punish’), sobornar (‘bribe’) or atacar (‘attack’) with animate objects, while other verbs like encontrar (‘find’), buscar (‘look for’), esconder (‘hide’) or ver (‘see’) allow for both options (with human/animate objects). According to Leonetti (2004:84) the marker a is fully lexicalized with verbs of the first class and does not express any further referential property. Nevertheless, the marker still indicates some kind of referential status, as Leonetti (2004:99)

5 Note that Torregro’s judgements in (14)-(15) and in (16) are not shared by all Spanish speakers. Victoria Escandell-Vidal (p.c.) points out that the difference in (16) becomes more acceptable with another object:
(i) (a) Has matado a un campeón.  
   has-2SG killed DOM a champion
   ‘you have killed a champion’
(i) (a) Has matado un campeón.  
   has-2SG killed a champion
   ‘you have killed the carrier of a champion’
notes with respect to bare nouns in object position. The verbs of the class that have a lexicalized marker (cf. (19a)) cannot take bare nouns, as in (20a). On the other side, the class of verbs that exhibit a certain optionality of the marker, as in (19b), also allow for bare nouns, as in (20b):

(19) (a) \{odiarse/ admirarse/ despreciarse/ amar/ aborrecer/ soportar\} *(a) una persona  
     ‘to hate/ admire/ despise/ love/ detest/ put up with… a person’  
     (b) \{llevar/ curar/ contratar/ describir/ encontrar/ ver\} *(a) una persona  
     ‘to take/ cure/ hire/ describe/ find/ see a person’

(20) (a) \{*odiarse/ *admirarse/ *despreciarse/ *amar/ *aborrecer/ *soportar\} personas  
     ‘to hate/ admire/ despise/ love/ detest/ put up with… persons’  
     (b) \{llevar/ curar/ contratar/ describir/ encontrar/ ver\} personas  
     ‘to take/ cure/ hire/ describe/ find/ see a person’

Affectedness itself seems to be a complex category that consists of subproperties such as the animacy of the object, the agency of the subject, the involvement of the object, and the aspect or akctionsart of the verb. We cannot speculate about the contribution of each of the mentioned factors or their interaction. We only state that we can distinguish verb classes according to affectedness and arrange them on a scale. To our knowledge, Pottier (1968:87) was the first to propose such a Scale of Affectedness, with different verb classes ranked according to the degree of the affectedness of their direct object.

(21) Scale of Affectedness (Pottier 1968:87)

\[
\begin{array}{cccc}
\text{matar} & \text{ver} & \text{considerar} & \text{tener} \\
‘kill’ & ‘see’ & ‘consider’ & ‘have’
\end{array}
\]

Affectedness is an intuitively valid category, but it is very difficult to give it a clear definition and apply it to various verbs. An additional complication is that other factors can interact with it. Therefore, we simplify the Scale of Affectedness and assume that the particular ranking is triggered by the specification of the verb to its object with respect to animacy. The verb matar ‘to kill’ has a strong tendency to take animate objects and is high in affectedness, while ver ‘see’ has no restriction with respect to animacy. Considerar ‘consider’ prefers an inanimate object and tener ‘have’ is an existential verb (see Bolinger 1953, Brugè & Brugger 1996:38, fn. 40 for the definition of “existential verbs”). It is also important to note that not the actual animacy is relevant, but the expected one.

(22) Scale of Affectedness and expected animacy of the object

\[
\begin{array}{cccc}
[+ \text{animate}] & [+/- \text{animate}] & [+/- \text{animate}] & [\text{neutral}] \\
\text{matar} & \text{ver} & \text{considerar} & \text{tener} \\
‘kill’ & ‘see’ & ‘consider’ & ‘have’
\end{array}
\]

This classification is quite coarse and many more subclasses are expected. The verbal class that strongly affects its objects (see (19a)) would be allocated above matar. In section 4.1 this classification will modified and then be the base for our corpus search.
3. Diachronic development of DOM

3.1 Diachronic development along the Referentiality Scale

Like Modern Standard Spanish, Old Spanish exhibits DOM. However, as shown in several diachronic studies (Melis 1995, Laca 2002, 2006), the use of DOM in Old Spanish is less frequent and determined by slightly different conditions than in Modern Spanish. The main results of these studies, summerized in von Heusinger & Kaiser (2005), are repeated here briefly and illustrated with some examples from the Cantar de mio Cid (following Melis 1995 and Laca 2006).

(Strong) object personal pronouns carry obligatory DOM in Old Spanish, as in (23). Human/animate proper names acting as direct object nouns are obligatory marked by DOM in Old Spanish, as in (24). Human/animate definite direct object NPs are not obligatorily marked by DOM in Old Spanish, as in (25). Human/animate non-definite direct object NPs are never marked by DOM in Old Spanish, as in (26) (cf. Laca 2006:444):

(23) e ssi fuèredes vençidos, non rebteses a nós \(\text{(Cid, 3566)}\)

and if would-2PL defeated not blame-IMP.2PL to us
‘but if you are defeated you are not to blame us’

(24) Matastes a Bucar & arrancamos el canpo \(\text{(Cid, 2458)}\)

killed-2SG to Bucar and rupture-1PL the field
‘you killed Bucar and and we have won the battle’

(25) (a) Reçiba a mis yernos commo elle pudier mejor \(\text{(Cid, 2637)}\)

receive-IMP.2SG to my sons-in-law as he could-3.SG better
‘Let him give to my sons-in-law the finest possible welcome’

(b) Ca yo case sus fijas con yfantes de Carrion \(\text{(Cid, 2956)}\)

for I married.1SG. his daughters with Infantes of Carrion
‘for I married his daughters to the Infantes of Carrion’

(26) Tanto traen las grandes ganancias, muchos gañados de ovejas e de vacas \(\text{(Cid, 480-481)}\)

very brought.3PL the big wealths many herds of sheep and of cows
‘They brought such great wealth, many herds of sheep and cows’

Comparing these facts in Old Spanish to the situation in Modern Spanish, we can state that there is a crucial difference in the marking of definite object NPs and specific indefinite NPs, both animate and human ones. According to Laca’s (2006) research, 36 percent of all animate definite object NPs are marked with DOM. In Modern Spanish, as already shown, these objects always appear with a. This difference is illustrated in (27) and (28), where the original version of El Cantar de Mio Cid is contrasted to a translation in Modern Spanish (cf. Laca 2006:455, Melis 1995:143):

(27) Old Spanish:

(a) En braços teneides mis fijas tan blancas commo el sol. \(\text{(Cid, 2333)}\)

in arms have-2.PL my daughters as white as the sun
‘In your arms you hold my daughters, as white as the sun’

(b) Escarniremos las fijas del Campeador. \(\text{(Cid, 2551)}\)

will-humiliate-1PL the daughters of the Battler
‘We shall humiliate the Battler’s daughters’
(28) Modern Spanish:
(a) tenéis a mis hijas, tan blancas como el sol, en vuestros brazos
have.2.PL to my daughters as white as the sun in your arms
(Cantar de mio Cid. Translation A. Reyes. Madrid: Espasa Calpe 1976)
‘In your arms you hold my daughters, as white as the sun’  (Cid, 2333)
(b) y podremos escarnecer a las hijas del Campeador. (Cid, 2551)
and will-can1.PL humiliare to the daughters of the Battler
(Cantar de mio Cid. Translation A. Reyes. Madrid: Espasa Calpe 1976)
‘We shall humiliate the Battler's daughters’

We can state that DOM has expanded considerably towards the right of the Referentiality Scale (while the Animacy Scale has not been affected from this extension, so far): the development goes from obligatory DOM for pronouns and proper nouns and optional DOM for definite nouns in Old Spanish to obligatory DOM for specific NPs and optional DOM for non-specific indefinite NPs in Modern Spanish. This holds for all animate direct objects (Laca 2006). Table (29) gives an informal representation of this development:


<table>
<thead>
<tr>
<th>+ animate</th>
<th>personal pronoun &gt;</th>
<th>proper noun &gt;</th>
<th>&gt; definite NP</th>
<th>&gt; indefinite spec. NP</th>
<th>&gt; indefinite non spec. NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Spanish (Cid)</td>
<td>+</td>
<td>+</td>
<td>± (36%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Evolution</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Modern (Standard) Spanish</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>±</td>
</tr>
</tbody>
</table>

3.2 Triggering conditions for the emergence and the development of DOM

Given this variation between Old and Modern Spanish with respect to the use of DOM, it is natural to ask which factors determine this variation. Melis (1995) and Laca (2006) point out that one of the most relevant factors for the use a in these cases are structures with syntactic topicalisation. In his study on El Cantar de Mio Cid, Melis (1995:134) observes that direct object NPs occurring in canonical word order, i.e. in postverbal position, are in general not employed with DOM, while preposed direct object NPs are. This observation is confirmed by Laca’s study. On the one hand, she observes that in the part of El Cantar de Mio Cid which she investigated, 80 percent of all animate definite object NPs used without a appear in the canonical postverbal position (see the examples in (27)). On the other hand, Laka (2006:455) notes that 73 percent of the definite object NPs used with a are either preposed, doubled by a co-referent clitic, or both preposed and doubled, as illustrated by the examples in (30):

(30) (a) Assi las escarniremos alas hijas del Campeador (Cid, 2555)
so them humiliare.FUT-1PL to-the daughters of the Battler
‘So, we shall humiliate the Battler's daughters’
(b) A las sus hijas en braço las prendia  (Cid, 275)
to the his daughters in arm them hold-3.SG
‘He gathered his daughters in his arms’

These findings provide one explanation for the variability found in the use of DOM with direct object NPs in Old Spanish, showing that topicality played a crucial role for DOM
marking in the earlier periods of Spanish. Note that in a later period topicality ceases to be a relevant factor for the use of DOM with (animate and definite) direct object NPs. As a consequence, almost all of these object types are used with a, which leads to a significant increase of DOM marking in Spanish. In a further step, DOM marking extends to indefinite NPs. The crucial feature here is [+specific]: while specific indefinite object NPs tend to be marked with DOM, unspecific ones are not. This observation has led us to conclude that the evolution of DOM is facilitated by intervening “transitional” categories, such as topicality and specificity. These categories are only active for the category to which DOM is developing: topicality for definite NPs, specificity for indefinite NPs. We can only speculate why we find such pairs: Topicality expresses a prominent contrast that (most often) affects definite NPs, while specificity expresses a contrast that (most) often affects indefinite NPs. In addition, a further distinction, indicated by the feature [+Arg(umental)] can be made with respect to unspecific indefinite NPs.

Extending von Heusinger & Kaiser (2005), we propose the following model of diachronic change, according to which the diachronic evolution of a-marking is triggered by additional parameters:

(31) Evolution of DOM from Old Spanish to Modern Spanish for animate objects:

<table>
<thead>
<tr>
<th>Strong Pro &gt; PN &gt; Definite +top</th>
<th>Definite – top</th>
<th>Indefinite</th>
</tr>
</thead>
<tbody>
<tr>
<td>transition point between ±top for definite direct objects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strong Pro &gt; PN &gt; Definite +top</th>
<th>Definite – top</th>
<th>Indefinite</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutralization for ±top</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strong Pro &gt; PN &gt; Definite +spec</th>
<th>Indefinite –spec</th>
</tr>
</thead>
<tbody>
<tr>
<td>transition point between ±spec for indefinite direct objects</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strong Pro &gt; PN &gt; Definite +spec</th>
<th>Indefinite –spec + Arg</th>
<th>Indefinite –spec –Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td>transition point between an unknown feature for non-specific indefinite direct objects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Diachronic development and lexical classes

The commonly assumed summary (31) of the diachronic development of DOM in Spanish rests on two main parameters, referentiality and animacy, together with triggering conditions such as topicality (from proper names to definite NPs) and specificity (from definite to indefinite NPs). While these parameters have been well studied, they have not sufficed to account for the observable diachronic and synchronic variation. Our detailed analyses of three different kinds of accessible corpora show that the diachronic development of DOM is crucially dependent on the lexical class of the governing verb. Our corpus searches have confirmed our original hypothesis that the verb class is a main parameter for DOM in Spanish and that DOM-spreading depends on time, referential properties of nouns, and verbal class. In
general, this diachronic corpus study strongly suggests that case-marking is determined by parameters in a multi-dimensional space (see von Heusinger & Klein 2007 for an account to DOM as “strength of case relation”).

We start with the classification developed in section 2.3, which is based on Pottier (1968:87). While Pottier (1968) takes affectedness as the main dimension according to which we arrange our verbs, we simplified this in assuming that the semantic restriction on the type of animacy for the object is the decisive factor, as in (22), repeated as (32). Our view is motivated by similar observations for Hindi (Mohanan 1994).

(32) Scale of Affectedness:

\[
\begin{array}{c|c|c|c}
+ & matar & ver & considerar & -
\hline
'kill' & 'see' & 'consider' & tener & 'have'
\end{array}
\]

Since purely existential verbs such as tener 'have' have a very strong tendency, even today, not to take objects that are a-marked, we excluded existential verbs.\footnote{Various authors note that there are certain conditions under which even tener requires (or allows) a-marking of the direct object. Pensado (1995:32) mentions the contrast between (i) and (ii). See also Bolinger (1953).}

That left us with three verbal classes, and we selected particular verbs for each class depending on availability in the original material. Class 1: herir, matar all have a clear preference for animate direct objects, class 2: ver, hallar do not have any preference with respect to the animacy of the direct object, and class 3: poner and tomar have a slight preference for inanimate objects. We can locate them on the scale in the following way.

(33) Scale of verbal classes in Spanish according to animacy preferences

\[
\begin{array}{c|c|c}
[+ animate] & [\pm animate] & [\pm/- animate]
\hline
Class 1 & Class 2 & Class 3
\hline
herir / matar & ver / hallar & poner / tomar
'hurt / kill' & 'see / find' & 'put / take'
\end{array}
\]

For a selection of these verbs (depending on the corpus) we categorized the direct objects according to the Referentiality Hierarchy. We used three types of corpora: Firstly, we compared the two books of Samuel and the two books of Kings in four Spanish bible translations: A from the 14th, B from 16th/17th, C from an European Spanish translation of the 20th century and D from an American translation of the 20th century, on the assumption that American Spanish has developed further. Secondly, we extended the search to the whole Bible, as the last three versions are electronically available (bible-gateway). Thirdly, we used the broader Corpus del Español from the 12th to the 19th century (http://www.corpusdelespanol.org). These corpus searches confirmed our original hypothesis that the verb class is a main parameter for DOM in Spanish and that DOM-spreading depends on time, referential properties of nouns and verbal class.

4.1. Comparing sentences in the same environment across time

Using parallel texts in general provides the great advantage of allowing one to compare the very same kind of construction, expression or lexical unit in texts from different languages or
from different periods of the same language (cf. Cysouw & Wälchli 2007). Our assumption is that Bible translations serve this requirement best (cf. Kaiser 2005, Enrique-Arias 2007). Bible translations constitute a quite archaic text and often have a quite specialized register that differs substantially from that of spoken language, but they also contain a considerable amount of natural-sounding direct speech. In addition, unlike many other older texts, most parts of the Bible are written (and translated) in prose, which eliminates the possibility that rime constraints influence the morphosyntax. Furthermore, for many languages, including Spanish, Bible translations (as well as other religious texts) provide the very earliest written documentation, and many Bible translations are easily and freely available for download from the Internet. We chose the following four Bible translations from different times and from different regions, taking the book of Samuel as our data base:

(34) Bible translations used in the corpus search

Version A: 14th century

Version B: 16th / 17th century
Reina Valera Antigua (1569/1602). Source: http://www.biblegateway.com
(The Reina-Valera Antigua was first translated and published in 1569 by Casiodoro de Reina, after twelve years of intensive work, and later put out in 1602 in revised form by Cipriano de Valera, who gave more than twenty years of his life to its revision and improvement.)

Version C: 20th century (Standard European Spanish)
Source: http://www.biblegateway.com/

Version D: 20th century (American Spanish)
La Biblia de las Américas (1971) (The Lockman Foundation).
Source: http://www.biblegateway.com/
This is a new translation of the Scriptures from the original languages. Completed in 1986 by a team of Latin American evangelical Bible scholars, La Biblia de las Américas is an original work translated from the Hebrew-Aramaic and Greek directly into modern Spanish.

4.1.1 Definite NPs

Our first corpus search concerned the presence or absence of DOM with definite human direct objects in the first ten chapters of the first book of Samuel. As expected, we found that pronouns and proper names are always marked with DOM in all four Bible translations. For definite object NPs, the results are summarized in table (35):

<table>
<thead>
<tr>
<th>1 Sam 1-10</th>
<th>Definite direct object noun ([+human])</th>
<th>+ top (preverbal /clitic doubling)</th>
<th>− top (postverbal / no clitic doubling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>± a</td>
<td>− a</td>
<td>+ a</td>
<td>− a</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 1: DOM-marking in 1 Samuel 1-10 for definite human NPs (Bibles A-D)
As predicted, DOM increases through time, i.e. while only half of the definite NPs are marked in the earliest Bible (12th century), nearly all are DOM-marked in the two versions from the 20th century. However, another prediction could not be tested: In (31) we said – following Laca and others – that topicalized definite NPs have a strong tendency or preference to be DOM-marked. Unfortunately, we found very few cases of direct object NPs occurring in a syntactic topicalization structure (i.e. left moved), six of them in version A, one in B and none in the modern versions C and D. As predicted, direct objects are in most cases DOM-marked in these structures, with only two topicalized objects appearing without DOM. An example for each case is given in (35):

(35) (a) E **a vuestras hijas** tomará por espejeras e cosineras
and to your **daughters** will-take-3SG for perfumers and cooks
and bakers  (A: 1 Sam 8,13)

‘He will take your daughters to be perfumers and cooks and bakers’

(b) [...] ca del señor **lo** demandó el varón.
because of-the **Lord** him asked-1SG the man  (A: 1 Sam 1,20)

‘because I asked the man from the Lord’

In other words, the comparison of Bible translations from different centuries does not provide evidence for the hypothesis that an increase of topicalization structures causes DOM marking to become more frequent and to finally extend to non-topicalized direct objects.⁷

Note that the distribution of non-topicalized object NPs is more complex. Although there is a clear development of DOM-marking through time, the triggering factors are not clearly identifiable. Topicalization cannot be an important factor, given that fact the already in A a great number of non-topicalized object NPs is DOM-marked. Obviously, not just position in the sentence (±topic) plays a role, but additional factors, as well. Our restricted but detailed search strongly suggested that the verb class determines the probability of DOM-marking. We therefore categorized DOM-marking according to our verb classes defined in 4.1. and extended our search to the two books of Samuel and the two books of Kings.

Table 2 lists the instances of DOM-marking with particular verbs in the four different Bible translations. The number in brackets gives the instances of all definite NPs for that verb, i.e. the unmarked instances are the difference between the two numbers. For convenience we have transferred the absolute instances into percentages in table 3, which clearly shows that DOM marking is increasing along the time dimension (left to right) and along the Scale of verbal classes (top to bottom).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>poner</td>
<td>1 (4)</td>
<td>3 (6)</td>
<td>5 (6)</td>
<td>6 (6)</td>
</tr>
<tr>
<td></td>
<td>tomar</td>
<td>6 (19)</td>
<td>4 (17)</td>
<td>15 (24)</td>
<td>17 (25)</td>
</tr>
<tr>
<td>2</td>
<td>ver</td>
<td>7 (20)</td>
<td>9 (22)</td>
<td>24 (29)</td>
<td>15 (20)</td>
</tr>
<tr>
<td></td>
<td>hallar</td>
<td>2 (4)</td>
<td>4 (5)</td>
<td>2 (3)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>1</td>
<td>matar</td>
<td>19 (32)</td>
<td>23 (27)</td>
<td>26 (27)</td>
<td>27 (27)</td>
</tr>
<tr>
<td></td>
<td>herir</td>
<td>5 (8)</td>
<td>14 (29)</td>
<td>10 (12)</td>
<td>13 (16)</td>
</tr>
</tbody>
</table>

Table 2: Bible translation of 1+2 Samuel and 1+2 Kings, instances of DOM with definite direct object (number of all definite NPs in brackets)

---

⁷ See Gabriel & Rinke (in preparation) for a more sophisticated approach to detecting topicalised direct objects that are not left located.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>poner</td>
<td>25%</td>
<td>50%</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td>tomar</td>
<td>31%</td>
<td>23%</td>
<td>62%</td>
</tr>
<tr>
<td>2</td>
<td>ver</td>
<td>35%</td>
<td>41%</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td>hallar</td>
<td>50%</td>
<td>80%</td>
<td>66%</td>
</tr>
<tr>
<td>1</td>
<td>matar</td>
<td>59%</td>
<td>85%</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>herir</td>
<td>62%</td>
<td>48%</td>
<td>83%</td>
</tr>
</tbody>
</table>

Table 3: Bible translations of 1+2 Samuel and 1+2 Kings, percentage of DOM with definite direct object

The tendencies are clearly visible, even though there is still a certain amount of variation. In the following we discuss examples that show the detailed comparison between the different translations. In (36) the verb *tomar* (‘take’) is of class 3, i.e. it prefers to take inanimate direct objects. However, as in the example given, it can also take animate ones. In the translation from the 14th century, the direct object is left moved, which we interpreted as being topicalized. Accordingly, it is DOM-marked. In the B-version from the 16th century, it is not moved and not marked. Both contemporary texts DOM-mark the object, as expected.8

(36) 1 Samuel 8, 13:

<table>
<thead>
<tr>
<th>A (14th century)</th>
<th>B (16th century)</th>
<th>C (20th century) (Spain)</th>
<th>D (20thc.) (Am)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E a vuestras hijas <strong>tomará</strong> por especieras e cosineras e panaderas.</td>
<td><strong>Tomará</strong> también a vuestras hijas para que sean perfumadoras, cocineras, y amasadoras.</td>
<td><strong>Tomará</strong> también a vuestras hijas para perfumistas, cocineras y amasadoras.</td>
<td><strong>Tomará</strong> también a vuestras hijas para perfumistas, cocineras y panaderas.</td>
</tr>
</tbody>
</table>

‘He will take your daughters to be perfumers and cooks and bakers.’

At the other end of the Scale of verbal classes is the verb *matar* ‘to kill’ of class 1, i.e. it only has a strong preference to take animate direct objects. As shown in table 3, objects of *matar* are highly affected by the verb and are almost always animate or human. There we would expect a high degree of DOM marking during all periods of time, as is evidenced in (37).

(37) 1 Reyes 19, 1:

<table>
<thead>
<tr>
<th>A (14th century)</th>
<th>B (16th century)</th>
<th>C (20th century) (Spain)</th>
<th>D (20thc.) (Am)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E notificó acab a ysebel todo lo que fiso elias e como <strong>mató</strong> <em>todos los profetas</em> a espada.</td>
<td>Y Acháb dió la nueva á Jezabel de todo lo que Elias había hecho, de como <strong>había muerto</strong> á cuchillo á <em>todos los profetas</em>.</td>
<td>Acab dio a Jezabel la noticia de todo lo que Elias había hecho y de cómo <strong>había matado</strong> a espada a <em>todos los profetas</em>.</td>
<td>Y Acab le contó a Jezabel todo lo que Elias había hecho y cómo <strong>había matado</strong> a espada a <em>todos los profetas</em>.</td>
</tr>
</tbody>
</table>

‘Now Ahab told Jezebel everything Elijah had done and how he had killed all the prophets with the sword.’

In (38), however, we see the oldest translation varies between *mató* (‘he killed’) ... *el capitán* in verse 3 and *matara* (‘he had killed’) ... *al capitán* in the next verse. The B version uses different lexical items, *herir* (‘to hurt’) *la guarnición* without marker, while version C uses *atacar* (‘to attack’) with marker.

---

8 This example is a nice case for close comparison: One could argue that in all versions the object must be topicalized, since all texts have the same underlying text structure. In that case, however, the later translation (16th century) would be more conservative than the earlier version (12th century).
(38) Samuel 13, 3-4:

<table>
<thead>
<tr>
<th>Class</th>
<th>A (14th century)</th>
<th>B (16th century)</th>
<th>C (20th century)</th>
<th>D (20thc.) (Am)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>matar</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>herir</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><em>tomar</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>hallo</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><em>poner</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>hallo</em></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Jonathan attacked the Philistine outpost at Geba, and the Philistines heard about it. Then Saul had the trumpet blown throughout the land and said, "Let the Hebrews hear!". So all Israel heard the news: "Saul has attacked the Philistine outpost, \ldots"} \]

4.1.2 Indefinite noun phrases

While the development of the DOM-marker with definite object NPs started early, marking of indefinite object NPs starts several centuries later. Even though in our sample text we find very few indefinite direct objects, it becomes obvious that there has been some diachronic development, at least for the verbs from class 1 (i.e. that verbs that take only animate objects). Table 4 gives the absolute figures and table 5 the percentages.

![Table 4: Bible translation of 1+2 Samuel and 1+2 Kings, instances of DOM with indefinite direct object (number of all definite NPs in brackets)](image)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td><em>poner</em></td>
<td>0 (7)</td>
<td>0 (14)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td><em>tomar</em></td>
<td>0 (8)</td>
<td>0 (14)</td>
<td>1(5)</td>
</tr>
<tr>
<td>2</td>
<td><em>ver</em></td>
<td>0 (7)</td>
<td>2 (10)</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td><em>hallo</em></td>
<td>0 (4)</td>
<td>0 (3)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>1</td>
<td><em>matar</em></td>
<td>1 (14)</td>
<td>1 (7)</td>
<td>7 (8)</td>
</tr>
<tr>
<td></td>
<td><em>herir</em></td>
<td>0 (0)</td>
<td>0 (7)</td>
<td>3 (3)</td>
</tr>
</tbody>
</table>

Table 5: Bible translation of 1+2 Samuel and 1+2 Kings, percentage of DOM with indefinite direct object

For verbs of class 3 (*tomar, poner*) we find a significant development only in the twentieth century. For verbs of class 2, *hallo* in (39) shows the different stages of development: Only in the American version of the Bible from the twentieth century, do we find the marker with the indefinite object NP.
(39) 1 Samuel 10, 2

<table>
<thead>
<tr>
<th>A (14th century)</th>
<th>B (16th century)</th>
<th>C (20th century) (Spain)</th>
<th>D (20th c.) (Latin America)</th>
</tr>
</thead>
<tbody>
<tr>
<td>En yéndote oy de mí fallarás o dos omnes cerca la sepultura de Rachel …</td>
<td>Hoy, después que te hayas apartado de mí, hallarás o dos hombres junto al sepulcro de Raquel, …</td>
<td>Hoy, después que te hayas apartado de mí, hallarás o dos hombres junto al sepulcro de Raquel, …</td>
<td>Cuando te apartes hoy de mí, hallarás a dos hombres cerca del sepulcro de Raquel, …</td>
</tr>
</tbody>
</table>

‘When you leave me today, you will meet two men near Rachel's tomb …’

For verbs of class 1 (only animate objects) we find an interesting situation. The case for herir and matar is complex since both verbs can substitute for each other, which results in the somewhat unexpected distribution in table 5. Example (40) reveals the full complexity of our data. We find three instances of the verb matar or herir. In one and the same verse, the first instance is a partitive for A with matar and definite NPs with herir (B, D) and hizo morir ‘make die’ (C). The second instance is an indefinite NP-object (cinquenta mil e setenta omnes ’fifty thousand and seventy men’). A and B do not mark, while C and D do. Finally, the last instance in that verse has in B the pronoun le, while in C the pronoun lo and in D the full NP al pueblo.

(40) 1 Samuel 6, 19

<table>
<thead>
<tr>
<th>A (14th century)</th>
<th>B (16th century)</th>
<th>C (20th century) (Spain)</th>
<th>D (20th c.) (Latin America)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E mató de los omnes de betsemes, porque vieron el arca del señor, e mató en el pueblo o cincuenta mill e setenta omnes. E pusieron luyo el pueblo, ca mató el señor en el pueblo grant matanza.</td>
<td>Entonces hirió Dios á los de Beth-semester, porque habían mirado en el arca de Jehová; hirió en el pueblo o cincuenta mil y setenta hombres. Y el pueblo puso luto, porque Jehová le había herido de tan gran plaga.</td>
<td>Entonces Dios hizo morir a los hombres de Beth-semester, porque habían mirado dentro del Arca de Jehová. Hizo morir a cincuenta mil setenta hombres del pueblo. Y lloró el pueblo, porque Jehová lo había herido con una mortandad tan grande.</td>
<td>El Señor hirió a los hombres de Beth-semester porque habían mirado dentro del arca del SEÑOR. De todo el pueblo hirió a cincuenta mil setenta hombres, y el pueblo lloró porque el SEÑOR había herido al pueblo con gran mortandad.</td>
</tr>
</tbody>
</table>

‘But God struck down some of the men of Beth Shemesh, putting seventy [c] of them to death because they had looked into the ark of the LORD. The people mourned because of the heavy blow the LORD had dealt them.’

### 4.1.3 Summary and further questions

Our very detailed and restricted corpus search in four Bible translations of the same chapters from the fourteenth to the twentieth century has shown that there is clear evidence for the evolution of DOM in Spanish. By examining identical sentences in identical contextual settings, we could show that the marker a becomes more frequent in later Bible versions. However, we were not able to corroborate the claim that topicality is a trigger for DOM-spreading into the definiteness slot or specificity for moving into the indefinite slot. We found rather that the evolution and the pace with which the evolution takes place depends strongly on the verb class. Definite NPs differ from indefinite in that they start to become marked at an earlier time – approximately three to four centuries earlier, as illustrated in the two charts below:
Both charts clearly indicate that DOM-marking develops along the axes of time and verbal class. The third main factor is the position of the object on the Referentiality Scale: Indefinite objects come to show the same pattern as definite ones, but three to four centuries later. The questions that arise from these observations are: (i) The generalization for marking definite...
and indefinite objects was based on selected chapters of the Bible. Is the generalization valid for the whole Bible? (ii) What has happened between early stages of the language and later stages? Our corpora were somewhat coarse grained since we had only access to data from the fourteenth, sixteenth, and twentieth centuries. A more fine grained analysis would help us to see the development in more detail.

We will address both questions in the next two sections. Firstly, we extend a sample search to the whole Bible, and then we make some searches in the *Corpus del Español* with texts from the twelfth to the nineteenth century.

4.2 The whole Bible as corpus

Three of the four versions of the Bible translations are available on-line (see (34) above for a more detailed description).

(41) Bible translations used in the in the on-line corpus search

Version B: 16\textsuperscript{th} / 17\textsuperscript{th} century


Version C: 20\textsuperscript{th} (European Spanish)


Source: http://www.biblegateway.com/

Version D: 20\textsuperscript{th} century (American Spanish)


Source: http://www.biblegateway.com/

The on-line versions only provided the text, but no kind of morphological or syntactic information. Therefore we had to search for word forms, rather than for lemmata. We decided to check whether more instances from the whole Bible would confirm the tendencies we found in the last section by analyzing all instances in 1 Samuel 1-10. We decided to search for more instances of *ver* in the whole Bible, but to avoid the complication of searching different word forms (*ver, ve, vi, veia, visto* etc.) we restricted the search to forms beginning in *ve*- and to the participle *visto*. The result is summarized in table 8, where we have listed instances of human direct definite and indefinite direct objects. The total numbers are 43 for the B-version, 65 for C, and 68 for D. These are numbers that allow for stabler generalizations. However, the number of indefinite objects is only around one fifth of the number of definite objects.

<table>
<thead>
<tr>
<th>ver (instances)</th>
<th>B: 16\textsuperscript{th}/17\textsuperscript{th} cent.</th>
<th>C: 20\textsuperscript{th} cent. (Euro)</th>
<th>D: 20\textsuperscript{th} cent. (Am)</th>
</tr>
</thead>
<tbody>
<tr>
<td>def Ø</td>
<td>12</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>def +a</td>
<td>24</td>
<td>51</td>
<td>55</td>
</tr>
<tr>
<td>indef Ø</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>indef +a</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>all instances</td>
<td>43</td>
<td>65</td>
<td>68</td>
</tr>
</tbody>
</table>

*Table 8: Extended search for direct objects ([+human], [±definite]) selected word forms of *ver* in three electronically available Bible translations*

Tables 9 and 10 compare the results from the last section, i.e. from the investigation of the two books of Samuel and the two books of Kings with the result from our extended search. Both tables unequivocally confirm the earlier results. Table 9 for the definite direct objects demonstrates that the tendencies of DOM-marking are even stronger in the broader corpus. The numbers for indirect objects in table 10 are also very similar in the two searches.\footnote{While the second extended search did not result in many more instances, it gives different records and thus...}
Table 9: Comparison between search result in Samuel / Kings and an extended search for DOM-marking with definite object NPs

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Samuel / Kings</td>
<td>7 (20)</td>
<td>9 (22)</td>
<td>24 (29)</td>
<td>15 (20)</td>
</tr>
<tr>
<td></td>
<td>35%</td>
<td>41%</td>
<td>83%</td>
<td>73%</td>
</tr>
<tr>
<td>extended search</td>
<td>--</td>
<td>24 (36)</td>
<td>51 (55)</td>
<td>55 (58)</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>67%</td>
<td>93%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Table 10: Comparison between search result in Samuel / Kings and an extended search for DOM-marking with indefinite object NPs

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Samuel / Kings</td>
<td>0 (7)</td>
<td>2 (10)</td>
<td>4 (8)</td>
<td>5 (9)</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>20%</td>
<td>50%</td>
<td>56%</td>
</tr>
<tr>
<td>extended search</td>
<td>--</td>
<td>1 (7)</td>
<td>3 (10)</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>14%</td>
<td>30%</td>
<td>50%</td>
</tr>
</tbody>
</table>

The detailed comparisons among the different translations reveal interesting facts. As expected, the oldest translation from the sixteenth century exhibits DOM-marked direct objects less frequently. We discuss below some instances of DOM-marking (or the lack of) for indefinite direct objects. In (42) all three versions mark the indefinite direct object, which is a very long and very descriptive noun phrase: *a un hijo de Isai de Belén que sabe tocar* (‘a son of Jesse of Bethlehem who knows how to play the harp.’). In (43) from Proverbs 26, 12 the B-version from the sixteenth century uses the bare noun *hombre* modified by a complex adjective *sabio en su opinion*, while the contemporary C and D-versions use an indefinite noun with a relative clause *a un hombre que se tiene por sabio*, both in the sense of ‘a man wise in his own eyes’.

(42) 1 Samuel 16:18

<table>
<thead>
<tr>
<th>B (16th century)</th>
<th>C (20th century) (Spain)</th>
<th>D (20th c.) (Latin America)</th>
</tr>
</thead>
<tbody>
<tr>
<td>He aquí yo he <em>visto a un hijo</em> de Isai de Beth-lehem, que sabe tocar.</td>
<td>He <em>visto a un hijo</em> de Isai de Belén que sabe tocar.</td>
<td>he <em>visto a un hijo</em> de Isai, el de Belén, que sabe tocar.</td>
</tr>
</tbody>
</table>

‘I have seen a son of Jesse of Bethlehem who knows how to play the harp.’

(43) Proverbs 26, 12

<table>
<thead>
<tr>
<th>B (16th century)</th>
<th>C (20th century) (Spain)</th>
<th>D (20th c.) (Latin America)</th>
</tr>
</thead>
<tbody>
<tr>
<td>¿Has <em>visto hombre</em> sabio en su opinión?</td>
<td>¿Has <em>visto a un hombre</em> que se tiene por sabio?</td>
<td>¿Has <em>visto a un hombre</em> que se tiene por sabio?</td>
</tr>
</tbody>
</table>

‘Do you see a man wise in his own eyes? There is more hope for a fool than for him.’

To summarize, the extended corpus search in the whole Bible translation (in the electronically available versions) confirmed our findings in the last sections. The diachronic development of DOM-marking for verb class 2 (*ver, hallar*) for definite object NPs started in the fourteenth century and increased to nearly 100 percent in the twentieth century. The marking of indefinite direct object NPs started around four centuries later, in the sixteenth century, and has arrived at not more than 50 percent in the twentieth century, which is comparable to the situation for definite objects in the sixteenth century. There seems to be an interesting contrast between the European and the American contemporary Spanish. The latter seems to be more advanced with respect to DOM-marking of indefinite objects. In this context, we like to add the following observation for a contrast between the European translations and the American

confirms the first search.
one, as illustrated in (44) from Psalm 50:18. For the generically used ‘a thief’, both the older and the contemporary European Spanish versions (B and C) use the definite form with the DOM-marker *al ladrón* ‘DOM-the thief’, while the contemporary American Spanish version uses the indefinite *a un ladrón* ‘DOM a thief’. Note that neither form is referential nor specific. This interesting case might motivate the extension of DOM-marking from specific indefinite to non-specific indefinites by analogy to non-referential definite cases. This observation needs further investigation.

(44) Psalm 50:18

<table>
<thead>
<tr>
<th>B (16th century)</th>
<th>C (20th) (Spain)</th>
<th>D (20th c.) (Latin America)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Si veías al ladrón, tú corrías con él; Y con los adulteros era tu parte</em></td>
<td><em>Si veías al ladrón, tú corrías con él, y con los adulteros era tu parte</em></td>
<td><em>Cuando ves a un ladrón, te complaces con él, y con adulteros te asocias</em></td>
</tr>
</tbody>
</table>

‘When you see a thief, you join with him; you throw in your lot with adulterers.’

4.3 Corpus del Español, twelfth to nineteenth century

Useful as they are, our Bible searches are somewhat coarse grained, and additionally we are not quite sure how archaic Bible translations are. Therefore we extended our corpus search again, this time to the *Corpus del Español* of Mark Davies (http://www.corpusdelespanol.org). This corpus comprises 100 million words of Spanish texts from the twelfth to the nineteenth century. The corpus interface allows one to search for lemmata, rather than for word forms (as in simple text files as the Bible texts). However, our searches were still very time-consuming since we had to select the human definite or indefinite direct objects by hand. Only 5 to 8 percent of all direct objects were identifiable as human and definite or indefinite. The others were either nonhuman or human and of a different type on the Referentiality Scale, such as clitics, personal pronouns, proper names and different types of quantifiers. Depending on the availability of texts, we searched all instances of a lemma if it produced fewer than 1000 tokens.

Due to time limits we restricted our searches to class 1 and class 3. We expected to find the early development of class 1 (*matar, herir*) and a very late development in class 3 (*tomar, poner*). We first present the particular results for *matar*, then for *tomar*, and give finally a comparison of all the verbs.

4.4.1 Matar

The verb *matar* (‘to kill’) from Class 1 (that takes only animate direct objects) is the starting point of the development of DOM in the history of Spanish. In the twelfth century we already find exactly 50 percent of definite direct objects marked with *a*, as shown in table 12. Table 11 shows the absolute token count we found in each corpus search.

<table>
<thead>
<tr>
<th>m</th>
<th>12th cent</th>
<th>13th cent</th>
<th>14th cent</th>
<th>15th cent</th>
<th>16th cent</th>
<th>17th cent</th>
<th>18thcent</th>
<th>19th cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>def Ø</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>def +a</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 11: Instances of DOM-marking of direct human definite object NPs according to the *Corpus del Español*
Table 12: Percentage of DOM-marking of direct human definite object NPs according to the Corpus del Español

As expected, for indefinite direct objects we find that the DOM-marking starts much later, in the Davies corpus not before the seventeenth century. This confirms the finding from section 4.1, where we compared Bible translations. Table 13 provides the absolute number, and table 14, the percentages:

<table>
<thead>
<tr>
<th>matar</th>
<th>12th cent</th>
<th>13th cent</th>
<th>14th cent</th>
<th>15th cent</th>
<th>16th cent</th>
<th>17th cent</th>
<th>18th cent</th>
<th>19th cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>indef Ø</td>
<td>20</td>
<td>20</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>16</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>indef +a</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 13: Instances of DOM-marking direct of human indefinite object NPs according to the Corpus del Español

Table 14: Percentage of DOM-marking of direct human indefinite object NPs according to the Corpus del Español

4.4.2 Tomar

The verb *tomar* (‘to take’) from class 3 (i.e. the class that prefers inanimate direct objects) shows an increasing use of DOM-marking with definite objects-NPs from the twelfth to the nineteenth century. However, it seems that the development is two centuries later than the one for *matar* described in the last subsection. *Tomar* has 50 percent DOM-marking in the fourteenth century (*matar* in the twelfth century) and 90 percent in the nineteenth century (*matar* in the seventeenth century.)
Table 15: Instances of DOM-marking of direct human definite object NPs according to the *Corpus del Español*

<table>
<thead>
<tr>
<th>tomar</th>
<th>12th cent</th>
<th>13th cent</th>
<th>14th cent</th>
<th>15th cent</th>
<th>16th cent</th>
<th>17th cent</th>
<th>18th cent</th>
<th>19th cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>def Ø</td>
<td>32</td>
<td>25</td>
<td>16</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>def +a</td>
<td>14</td>
<td>24</td>
<td>16</td>
<td>14</td>
<td>15</td>
<td>2</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 16: Percentage of DOM-marking of direct human definite object NPs according to the *Corpus del Español*

As expected, DOM-marking of indefinite direct objects is even more delayed. We find not much marking before the nineteenth century, and even there the absolute numbers are not very high, as summarized in tables 17 and table 18:

<table>
<thead>
<tr>
<th>tomar</th>
<th>12th cent</th>
<th>13th cent</th>
<th>14th cent</th>
<th>15th cent</th>
<th>16th cent</th>
<th>17th cent</th>
<th>18th cent</th>
<th>19th cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>indef Ø</td>
<td>28</td>
<td>5</td>
<td>8</td>
<td>37</td>
<td>9</td>
<td>9</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>indef +a</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 17: Instances of DOM-marking of direct human indefinite object NPs according to the *Corpus del Español*

Table 18: Percentage of DOM-marking of indirect human indefinite object NPs according to the *Corpus del Español*
4.4.3 Comparing verb classes through time

We can now combine the development in the different verb classes: We take two verbs from class 3 (tomar, poner) and two from class 1 (herir, matar). Table 19 shows that we have a continuously increasing probability of DOM-marking for definite direct object.

![Graph showing DOM-marking percentage for different verbs and time periods.](image)

Table 19: Percentage of DOM-marking of indirect human definite object NPs according to the Corpus del Español depending on verb-class and time

The same is the case for indefinite direct objects, but with a delay of 3 to 4 centuries. We expect a further development of the DOM-marking in Spanish, until all direct objects are marked.

![Graph showing DOM-marking percentage for different verbs and time periods.](image)

Table 20: Percentage of DOM-marking of indirect human indefinite object NPs according to the Corpus del Español depending on verb-class and time
5. Summary and further research

Our original findings from this first corpus search show for definite human direct objects a diachronic development of a-marking depending on verb class. The same development can be observed for indefinite human direct objects, but with some delay. In a second step, we extended our corpus search to the whole Bible, and in a third one we used an even broader text corpus (the electronically available Corpus del Español from the twelfth to the nineteenth century). These corpus searches confirmed our original hypothesis that the verb class is a main parameter for DOM in Spanish and that DOM-spreading depends on time, referential properties of nouns, and verbal class. In general, this diachronic corpus search strongly suggests that Differential Object Marking is determined by parameters in a multi-dimensional space.

6. References


García García, Marco (this volume). Differential object marking with inanimate objets, 63-84.


