

# Verbal Semantics and the Diachronic Development of Differential Object Marking in Spanish

KLAUS VON HEUSINGER

to appear in *Probus* 20 (1) 2008

## Abstract

*The use of the marker *a* with a direct object in Spanish is an instance of Differential Object Marking (DOM), which is cross linguistically a well-documented phenomenon. In Spanish, there are two main dimensions that determine *a*-marking of direct objects: (i) the semantic and (discourse) pragmatic properties of the direct object, such as animacy, definiteness, specificity and topicality; and (ii) the lexical semantics of the verb, such as aktionsart properties and selectional restrictions with respect to the position of the direct object. Diachronically, *a*-marking spreads along the Referentiality Hierarchy from personal pronouns and proper names to definite and finally indefinite noun phrases, a process that has been well examined. In this study, I focus on the influence of the lexical semantics of verb classes on this process. I present original findings from two corpus searches from the 12<sup>th</sup> to the 20<sup>th</sup> century. The data show that the diachronic evolution of *a*-marking crucially depends on the verb semantics of the governing predicate, and that the change proceeds according to the Constant Rate Hypothesis of Kroch. These findings suggest that *a*-marking in Spanish, and DOM in general, can only be described in a multi-dimensional space consisting of the semantic properties of the direct object and the lexical semantics of the verb.*

## 1. Introduction\*

Spanish exhibits Differential Object Marking (DOM) by the marker *a*, which shows an interesting range of synchronic variation and a well documented diachronic evolution. DOM is a widespread instantiation of case alternation among the languages of the world (see Butt

---

\* Georg Kaiser and I have presented previous versions of this paper at the International Workshop *Definiteness, Specificity and Animacy in Ibero-Romance Languages (Nereus III)* in Alcalá de Henares (Madrid) in October 2006, at the International Workshop *New Perspectives on Morphological Change* in Berlin in October 2006, at the International Workshop *Differential Object Marking in Romance* in Stuttgart in June 2007, at the *Workshop on Empirical Approaches to Morphological Case* at the LSA Summer Institute, Stanford in July 2007 and at the *Coloquio Internacional sobre Corpus Diacrónicos en Lenguas Ibero-romances* in Palma de Mallorca in October 2007. I would like to thank the organizers and the audiences for their constructive comments. I am especially indebted to Miriam Butt, Andres Enrique, Victoria Escandell-Vidal, Udo Klein, Brenda Laca, Manuel Leonetti, Edgar Onea and Elisabeth Stark for long discussions and valuable comments. Special thanks go to Georg Kaiser who not only has edited this volume but who was the co-author of previous versions and who had given enough intellectual input into the final paper to be a proper co-author. Finally I would like to express my gratitude to the two reviewers who have given me very detailed and valuable comments which have helped to restructure this paper and develop a hopefully clearer argument. A substantial part of the corpus searches were undertaken by my student assistants Annika Deichsel and Helga Szilagyi. The research was supported by the German Science Foundation (project C2 “Case and Referential Context” in the SFB 732 “Incremental Specification in Context”).

2006 for an overview on case alternation). Bossong (1985) mentions that at least 300 known languages exhibit DOM in one way or another. DOM or *a*-marking of the direct object in Spanish is a well studied subject (e.g. Brugè and Brugger 1996, Torrego 1999, Leonetti 2004, Pensado 1995 with an annotated bibliography), but there is no overall account of the different parameters that determine DOM.

DOM in Spanish, as in other languages, can only be explained by the interaction of two main types of parameters for DOM: (i) the properties of the direct object, and (ii) the lexical semantics of the verb. The properties of the argument include animacy, definiteness, specificity and topicality, often summarized in terms of degrees of “individuation”. In European Spanish, human direct objects can be *a*-marked if they have certain referential properties, while inanimate direct objects are generally unmarked.<sup>1</sup> The lexical semantics of the verb interacts with the mentioned nominal semantics: certain verbs require *a*-marking (with human direct objects), while others allow for variation. Even though the literature has always acknowledged a certain influence of the verbal semantics, it was never systematically investigated or accounted for.

This study presents the results of two corpus searches from the 12<sup>th</sup> to the 20<sup>th</sup> century. I distinguish three verbal classes that differ in their selectional restrictions with respect to animacy: class 1 strongly requires a human direct object (*matar* ‘to kill’), class 2 has no preference for human direct objects (*ver* ‘to see’), while class 3 has a preference for inanimate direct objects (*tomar* ‘to take’). The corpus searches show that DOM is found first with verbs of class 1, and about three centuries later with verbs of class 2 for definite as well as for indefinite human NPs.

The structure of the paper is as follows: in section 2, I discuss the synchronic variation of *a*-marking in Spanish, the relevant nominal and verbal parameters. Section 3 provides a summary of diachronic studies which show that DOM in Spanish spread from personal pronouns and proper names to definite and finally indefinite noun phrases (all human or animate). In section 4, I summarize the detailed corpus study reported in von Heusinger and Kaiser (2007). Their findings suggest that besides this general picture, the lexical semantics of the verb is an additional driving force in the diachronic evolution of DOM. They compared two chapters of the Bible from different periods (14th, 16th and 20th century) and regions (including an American Bible translation from the 20th century). However, this search was

---

<sup>1</sup> There are exceptions to this rule, see example (8) below for inanimate direct objects that can be marked if the subject is also inanimate (Isenberg 1968, Roegiest 1989, Garcia 1997 for a comprehensive discussion). Company (2002) discusses examples from American Spanish, where *a*-marking of inanimate direct objects is more common.

based on very few items from each class and it was coarse-grained since it had only three time slices. Therefore, I undertook a broader corpus search with the results presented in section 5. This second search covered the time period from the 12<sup>th</sup> to the 19<sup>th</sup> century and collected significantly more occurrences. The results of this suggest that the selectional restrictions of verbs are one factor in accounting for the diachronic evolution of *a*-marking (and its synchronic variation) in Spanish. In section 6, I discuss and evaluate the findings with respect to the main theories accounting for the underlying forces of DOM and its diachronic evolution.

## 2. Differential Object Marking in Spanish

Two main approaches to Differential Object Marking (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) proposes that languages that do not formally 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 and on their contrast to the subject. The Transitivity Thesis (Hopper and Thompson 1980, Naess 2004, 2007), in contrast, assumes that a direct object is overtly marked if it is a “good” argument in a transitive sentence and the sentence expresses a “salient event”. Transitivity investigates a broad range of semantic properties, in particular the semantic features of the verb such as telicity, aspectuality and thematic information of the argument roles such as volitionality and agency. De Hoop and Narasimhan (2005) modify the Transitivity Theory and use the concept of “Strength of an Argument” which stands for different nominal properties. According to them, DOM-languages mark strong arguments in direct object positions.

Differential Object Marking (DOM) in Spanish is expressed by the marker *a*, which is a homophone to the preposition *a* ‘to’ and the dative marker *a* of the indirect object.<sup>2</sup> DOM or *a*-marking in Spanish is determined by two main types of parameters: (i) properties of the direct object, and (ii) transitivity properties of the verb, i.e. the lexical semantics of the verb.

---

<sup>2</sup> For want of space I cannot discuss the role of clitic doubling for DOM, but see Suñer (1988), Brugè and Brugger (1996) and Parodi (1998) for clitic doubling, and Fontana (1993) for the diachronic development of clitic doubling; 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 [...]”. See Leonetti (this volume) for a comprehensive analysis of the interaction of clitic doubling and DOM in Spanish.

## 2.1 Nominal properties

It is commonly assumed that the main factors for DOM are semantic and (discourse) pragmatic categories such as animacy, referentiality (definiteness and specificity), and topicality (see Comrie 1975, Bosson 1985, Croft 1988, Aissen 2003, and others).<sup>3</sup> 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) or Butt (2006) for an account of case alternation based on the notion “control”. I assume that animacy is a lexicalized conceptual category, i.e. speakers categorize objects they speak about according to different values of animacy. The Animacy Scale (1) distinguishes three values.

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

There is no agreement in the literature whether Spanish distinguishes between +human vs. –human or +animate vs. –animate. I will therefore suggest a distinction between +human and –animate and assign -human and +animate objects to either category depending on the context. Thus a human direct object can take DOM, as in (2a), while DOM with inanimate direct objects is ungrammatical, as in (2b) (for European Spanish).

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

---

<sup>3</sup> There are additional referential properties such as number, collectivity, concreteness, etc., and discourse pragmatic ones, such as prominence, which influence the “individuation” or the “referential status” of an argument.

The main parameter, however, for the individuation of a noun is referentiality as expressed in the Referentiality Scale<sup>4</sup>, which combines definiteness and specificity. This scale ranks personal pronouns highest, 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, I have added the non-argumental noun slot, which is crucial for the description of DOM in Spanish.

- (3) Referentiality Scale: personal pronoun > proper noun > definite NP > indefinite specific NP > indefinite non-specific NP > non-argumental

For Modern (European) Spanish, I assume that the cut is somewhere in the slot for non-specific indefinites (-Spec), since they can optionally be marked by *a*, while non-argumental (-Arg) can never be marked. In other words, DOM in Spanish (for human direct objects) indicates that the noun is an argument and cannot be a predicative expression that might be incorporated. Bleam (1999) formulates this distinction in terms of the semantic type of noun: argument type *e* vs. non-argumental or predicative type  $\langle e, t \rangle$ . The definite noun phrase in (4a) and the indefinite (specific) noun phrase in (4b) must be marked by *a*. The non-specific indefinite noun phrase in (4c) may optionally be marked with *a*. The non-specificity is clearly indicated by the subjunctive *sepa* in the relative clause. Even the indefinite pronoun *alguien* in (4d) takes *a* in its non-specific reading. Only the non-specific reading of (4e) does not allow *a*. Note that in the specific reading ('to need a certain assistant) *a*-marking is adequate:<sup>5</sup>

- (4) (a) Vi        *\*(a)*        **la** **mujer**.  
          saw-1SG DOM the woman  
          'I saw the woman.'  
       (b) Vi        *\*(a)*        **una** **mujer**.  
          saw-1SG DOM a woman  
          'I saw a woman.'  
       (c) Necesitan (*a*)    **un** **ayudante** **que sepa**                **inglés**.  
          need-3PL DOM a assistant that speak-SUBJ.3SG English  
          'They need an assistant who knows English.'  
       (d) Está buscando *a*    **alguien**.  
          is looking DOM someone  
          '(S)he is looking for someone.'

<sup>4</sup> Contrary to the use of the "Definiteness Scale" in Aissen (2003) and others, I follow Croft (2003:130) and name the scale under discussion "Referentiality Scale".

<sup>5</sup> As noted by the first reviewer the situation is probably not as clear as suggested by Bleam (1999). Non-specific indefinite direct objects of type *e* can appear with or without *a*.

- (e) El dentista necesita *\*a* **un ayudante**.  
 the dentist needs DOM a assistant  
 Intended reading: ‘The dentist needs some assistant.’

Topicality is a notion of information structure and therefore somewhat vague and difficult to test in corpora. I assume two values for topicality, as in (5).

- (5) Topicality: topical > non topical

Escandell-Vidal (2007) suggests in her discussion of DOM in Balearic Catalan that topicality might have more than two values. However, I assume that there are only topical and non-topical direct objects. I further assume that topicality is expressed (or at least detectable in texts) by left-dislocation. The indefinite direct object right of the verb may optionally take *a*, while the left-dislocated one in (6b) must take it (cf. Leonetti 2004:86).

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

Table 1 summarizes the conditions in Modern (European) Spanish: the direct object is marked if it is human and specific, DOM is optional for non-specific indefinites and ungrammatical for non-argumental indefinites, or inanimates NPs.

+ human direct objects	personal pronoun	> proper noun	> definite NP	> indefinite spec. NP	> indefinite non spec. NP	> non- argumental
contemporary European Spanish	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a/Ø</i>	<i>Ø</i>

Table 1: Conditions for *a*-marking in Contemporary European Spanish

There are additional conditions that can overwrite the general picture given in Table 1: *a*-marking can be suspended if there is an indirect *a*-marked object in the same sentence. This avoidance of *a* is only a “stylistic rule” and is, according to the Real Academia Española (1973:374f), mostly applied in cases when both objects are full nouns. Both examples (7a) and (7b) are grammatical. In contrast, *a*-marking can optionally be used even with inanimate

objects if they contrast with an inanimate subject. (Isenberg 1968; see also Torrego 1999, García 2007), as in (8) and (9).

- (7) (a) Ha sido forzoso dejar **al conde** en rehenes **al enemigo**.  
 has been compelling leave DOM.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 was necessary to leave the count as a hostage for the enemy.’
- (8) ...que los gerundios modifican **al sujeto**.  
 that the gerunds modify DOM.the subject  
 ‘... that the gerunds modify the subject’
- (9) En esta receta, la leche puede sustituir **al huevo**.  
 In this recipe the milk can replace DOM.the egg  
 ‘In this recipe egg can replace the milk.’

## 2.2 Transitivity and Affectedness

DOM in Spanish also depends on the lexical properties of the verb. This has long been noted in descriptive grammars of Spanish (Bello 1847:567-570, Fernández Ramírez 1951:151-190 and others, quoted after Torrego 1999:1783). Particular approaches to describe DOM depending on properties of certain 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). But there has been no systematic account of DOM in terms of verbal semantics and there is no study that investigates the diachronic evolution of DOM from the perspective of semantic properties of verbs. The correlations of nominal and verbal parameters in the expression of DOM are described in the more general Theory of Transitivity of Hopper and Thompson (1980). They maintain that the categories in Table 2 are ordered or aligned in a particular way: languages prefer to mark high transitivity values formally, rather than the lower values. They account for the particular alignment of the categories by assuming that all high transitive values contribute to the discourse salience of the event described by the verb and its arguments. A prototypical salient event has two participants, expresses action, is telic and has a totally affected and highly individuated direct object. A prototypical non-salient event has only one participant, expresses no action, etc. Note that not all of these parameters must be instantiated at the same time.

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

Table 2: Parameters of Transitivity (Hopper and Thompson 1980)

The last parameter in the table, namely *Individuation*, summarizes the semantic and pragmatic factors of the direct object discussed in the last section: animacy, definiteness, specificity and topicality. I will now focus on one of the other parameters, namely *affectedness*. See for other parameters of Transitivity, such as telicity, volitionality, type of causation, mode, agency and affectedness Torrego (1999), Leonetti (2004) and others. 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 following examples with eventive or stative verbs. The affectedness can relate to physical circumstances, as in (10), or to psychological ones, as in (11). The (b)-examples without the marker are ungrammatical.

- (10) (a) Golpearon ***a*** **un** **extranjero**.  
beat-3PL DOM a stranger  
(b) \*Golpearon **un** **extranjero**.  
beat-3PL a stranger  
‘They have beaten a stranger.’

- (11) (a) Odia ***a*** **un** **vecino**.  
hate-3SG DOM a neighbor  
(b) \*Odia **un** **vecino**.  
hate-3SG a neighbor  
‘(S)he hates a neighbor.’

Modern Spanish has lexicalized this contrast: a whole class of verbs obligatorily take *a* with human objects, such as *saludar* (‘greet’), *odiar* (‘hate’), *insultar* (‘insult’), *castigar* (‘punish’), *sobornar* (‘bribe’) or *atacar* (‘attack’), while other verbs like *encontrar* (‘find’), *buscar* (‘look



Affectedness itself – like individuation – seems to be a complex category that comprises different aspects of involvement of the direct object in the event expressed by the verb. Pottier (1968:87) was the first to propose a Scale of Affectedness, with different verb classes ranked according to the degree of the affectedness of their direct object.

$\leftarrow +$ 
*matar*
*ver*
*considerar*
*tener*
 $\rightarrow -$

‘kill’
‘see’
‘consider’
‘have’

To simplify matters, let us assume that one underlying parameter of affectedness is the animacy restriction imposed on the direct object by the verb. *Matar* ‘to kill’ always requires a human direct object, while *ver* ‘to see’ does not. Note that it is not the animacy of the direct object itself, it is rather the selectional restriction of the verb. Therefore, von Heusinger and Kaiser (2007) modify or simplify the Scale of Affectedness into a version that depends only on this parameter. They assume that the particular ranking depends on the animacy requirement imposed by the verb on the object. The verb *matar* ‘to kill’ has a strong tendency to take human objects and is high in affectedness, while *ver* ‘see’ has no restriction with

<sup>7</sup> See von Stechow and Klein (2007) for a discussion of these verbal properties in different languages including Hindi, Mongolian and Uzbek. A preliminary result of this study is that verbal properties of this kind often play a role of preference in otherwise “optional” cases. See Guntsetseg (2007) for a contrast between *read* and *write* for DOM in Mongolian.

respect to animacy. However, it is difficult to assign some preference restriction to the direct object of *considerar* ‘consider’ or *tener* ‘have’.

(13) Scale of Affectedness and expected animacy of the direct object

[+ human] >	[± human] >	[±/- animate] >	[(±)/– animate]
<i>matar</i>	<i>ver</i>	<i>considerar</i>	<i>tener</i>
‘kill’	‘see’	‘consider’	‘have’

### 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), DOM in Old Spanish is less frequent and used in different conditions than in Modern Spanish. The main results of these studies 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 (14). Human proper names acting as direct object are obligatorily *a*-marked, as in (15). Human definite direct objects are optionally *a*-marked, as in (16). Animate indefinite direct objects are never *a*-marked, as in (17) (cf. Laca 2006:444):<sup>8</sup>

- (14) e ssi fuéredes vençidos, non rebtedes                      *a*    **nós**                      (Cid, 3566)  
 and if would-2PL defeated not blame-IMP.2PL DOM us  
 ‘but if you are defeated you are not to blame us.’

- (15) Matastes *a*    **Bucar** & arrancamos el campo                      (Cid, 2458)  
 killed-2SG DOM Búcar and rupture-1PL the field  
 ‘you killed Búcar and and we have won the battle.’

<sup>8</sup> There was no clear case of a human indefinite direct object in the text. The animate indefinite direct object *muchos gañados deovejás e de vacas* in (17) is the closest that we come to an instance of it.

- (16) (a) Reçiba *a* **mios yernos** commo elle pudier mejor (Cid, 2637)  
 receive-IMP.2SG DOM 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 (Cid, 2956)  
 for I married.1SG. his daughters with Infantes of Carrion  
 ‘for I married his daughters to the Infantes of Carrion.’
- (17) Tanto traen las grandes ganancias, **muchos gañados de ovejas e de vacas**  
 very brought.3PL the big wealths many herds of sheep and of cows  
 ‘They brought such great wealth, many herds of sheep and cows.’ (Cid, 480-481)

Comparing these facts in Old Spanish to the situation in Modern Spanish, we see that there is a crucial difference in the marking of definite objects and (specific) indefinite NPs, both animate and human ones. According to Laca’s (2006) research, 36 percent of all animate definite objects are marked with DOM in Old Spanish. In Modern Spanish, as already shown, these objects always appear with *a*. This difference is illustrated in (18) and (19), 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):

- (18) Old Spanish:  
 (a) En braços tenedes **mis fijas** tan blancas commo el sol. (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**. (Cid, 2551)  
 will-humiliate-1PL the daughters-of-the Battler  
 ‘We shall humiliate the Battler’s daughters’
- (19) Modern Spanish:  
 (a) tenéis *a* **mis hijas**, tan blancas como el sol, en vuestros brazos  
 have.2.PL DOM 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 humiliate DOM 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 observe that DOM has expanded considerably towards the right of the Referentiality Scale (while the Animacy Scale has so far not been affected from this extension in European Spanish): the development for human direct objects 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, as illustrated in table 3:

+ human direct objects	personal pronoun	> proper noun	> definite NP	> indefinite spec. NP	> indefinite non spec. NP	> non- argumental
Old Spanish (Cid)	a	a	a/Ø (36%)	Ø	Ø	Ø
Modern Spanish	a	a	a	a	a/Ø	Ø

Table 3: Diachronic evolution of the DOM in Spanish along the Referentiality Scale  
for human direct objects (based on Laca 2006, Melis 1995)

### 3.2 Transitional 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 of *a* in these cases are structures with topicalization. In a study on *El Cantar de Mio Cid*, Melis (1995:134) observes that direct objects occurring in postverbal position, are in general not employed with DOM, while preposed direct objects 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 objects used without *a* appear in postverbal position (see the examples in (18)). On the other hand, Laca (2006:455) notes that 73 percent of the definite objects used with *a* are either preposed, doubled by a co-referential clitic, or both preposed and doubled, as illustrated below:<sup>9</sup>

<sup>9</sup> One reviewer pointed out that the syntax of Old Spanish is quite different from the syntax of Modern Spanish (see Fontana 1993 for a detailed study). In particular, the word order was not as fixed as in Modern Spanish. Old Spanish was much more flexible with accepting preverbal direct objects, such as in (i) (quoted from Fontana 1993, 63, ex (16b), source: *General Estoria* of Alfonso X of Castille, 13<sup>th</sup> century). We find even a postverbal position for both the subject and the object, such as in (ii) (quoted from Fontana 1993, 261, ex. (44), the source is *Historia del gran Tamerlán* from the 15<sup>th</sup> century). Both syntactic constructions are contexts in which subject and object might be confused and therefore such constructions might have triggered DOM.

(i) este logar mostro dios a abraam  
this place show-PER.3SG God to Abraham  
'God showed Abraham this place.' (GE-I.62v)  
(ii) & vencio= lo al turco el senor tamurbeque  
& defeat-PER.3SG-him DOM-the Turk the lord Tamurbeque  
'And Tamurbeque defeated the Turk.'

Fontana (1993,240) also reports that the rate of topicalized direct objects drops from the 12<sup>th</sup> to the 15<sup>th</sup> century indicating a change of basic word order.

- (19) (a) Assi **las** escarniremos **a las fijas del Campeador** (Cid, 2555)  
 so them humiliate.FUT-1PL DOM-the daughters of-the Battler  
 ‘So, we shall humiliate the Battler’s daughters’
- (b) **A las sus fijas** en braço **las** prendia (Cid, 275)  
 DOM 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 objects, i.e. with full definite or indefinite nouns, in Old Spanish, showing that topicality plays 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 definite human direct objects. In a further step, DOM marking extends to indefinite NPs. An additional feature here could have been [ $\pm$ specific]: specific indefinite objects share some features with definite ones. However, it is difficult to find enough material to give life to this hypothesis.<sup>10</sup> Still, von Heusinger and Kaiser (2005) conclude that the evolution of DOM is driven 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. Topical definite NPs are more similar to proper names than non-topical definite NPs, since they are not only assumed to be known to speaker and listener but also that the sentence is about them, and they are interpreted independently of the material in the sentence. Specific indefinite NPs behave more like definite NPs since they have wide scope and are referential expressions. In addition, we have assumed a further distinction, indicated by the feature [ $\pm$ Arg(umental)] that facilitates the development into the nonspecific domain, as illustrated in table 4:

---

<sup>10</sup> See von Heusinger and Onea (this volume) for the transitional function of specificity for the diachronic development of DOM in Romanian.

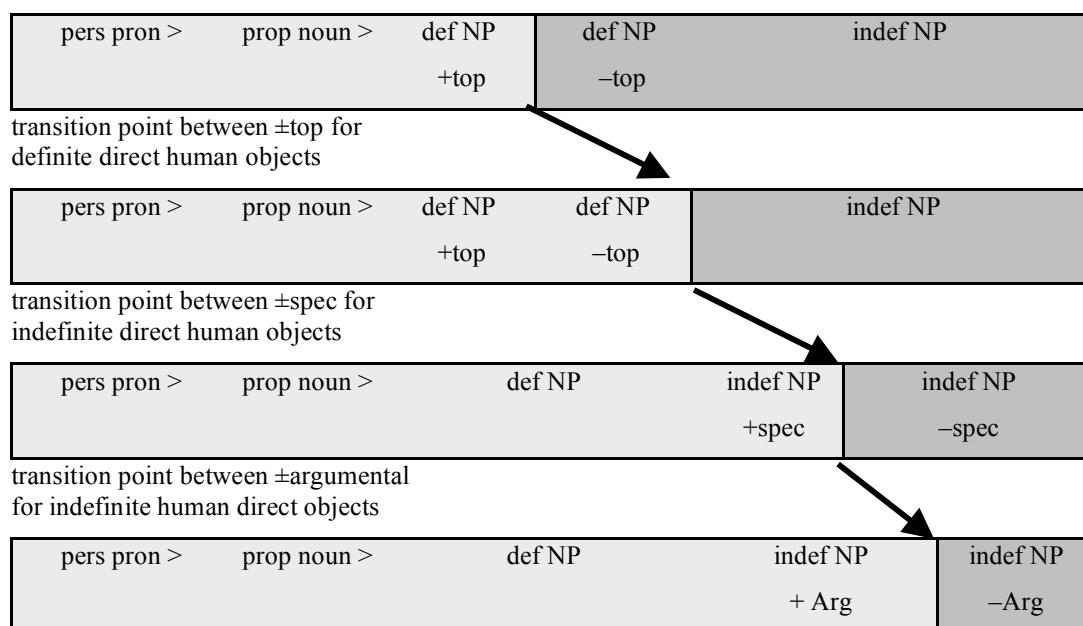


Table 4: Evolution of DOM from Old Spanish to Modern Spanish for human direct objects

#### 4. Diachronic development and lexical classes

The diachronic development of DOM in Spanish as described in table 4 is based on referentiality together with transitional conditions such as topicality and specificity. While this dimension has been well studied, von Heusinger and Kaiser (2007) undertake the first diachronic study that investigates the dimension expressed by the lexical semantics of the verb, which are summarized in this section. They modify the Scale of Affectedness (13) above in section 2.2 to a scale which ranks verb classes according to their selectional restrictions: class 1 has a high preference for human direct objects, class 2 has no preference, and class 3 has a preference for inanimate direct objects. They do not include existential verbs since they have even today a very strong tendency not to take objects that are *a*-marked.<sup>11</sup>

<sup>11</sup> See Bolinger (1953), Brugè and Brugger (1996:38, fn. 40) for the definition of “existential verb”. 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).

- (i) tiene **a** su mujer enferma  
has-3.SG DOM his wife sick  
‘His wife is sick’
- (ii) tiene una mujer muy inteligente  
has-1.SG a wife very intelligent  
‘He has a very intelligent wife’

## (20) Scale of verb classes

[+ human] >	[± human] >	[(±)/- animate]
Class 1	Class 2	Class 3
<i>matar / herir</i>	<i>ver / hallar</i>	<i>tomar / poner</i>
‘kill’ / ‘hurt’	‘see’ / ‘find’	‘take’ / ‘put’

Furthermore, only instances of full definite or indefinite human direct objects are investigated. Personal pronouns and proper names of human objects are always *a*-marked in the 12<sup>th</sup> century such that the evolution of DOM can only be observed with full NPs.

#### 4.1. Comparing bible translations

The first corpus consisted of the two books of Samuel and the two books of Kings in four Bible translations, abbreviated as A-D: translation A is from the 14<sup>th</sup> and was only available in a printed version. All other translations were electronically available at Biblegate: B, *Reina Valera Antigua* from 16<sup>th</sup>/17<sup>th</sup>; its contemporary version C from 1995 (*Reina Valera*); and a modern American translation D 1971 (*La Biblia de las Américas*).<sup>12</sup> The English translation is from *The 21st Century King James Version*. 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 and Wälchli 2007). The assumption is that Bible translations serve this requirement best (cf. Kaiser 2005, Enrique-Arias 2007). They constitute a very archaic text and often have quite a specialized register which differs substantially from that of the spoken language, but they also contain a considerable amount of natural-sounding direct speech.

#### 4.2 Definite NPs

DOM continuously spreads from one referential category to the next in the Referentiality Scale. Therefore, we first investigate the situation with definite human direct objects. In (21) 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 human ones. In the translation from the 14<sup>th</sup> century, the direct object is left-dislocated, an indication for being topicalized. Topicalization being a

<sup>12</sup> According to Enrique-Arias (pc), all these versions are translations from a Hebrew text. The A version remains very faithful to the original phrasing of the Hebrew text, so it often uses DOM, copying the Hebrew marker *et* even for inanimate direct objects as in Genesis 1,1 *Crio dios a los cielos & a la tierra* (God created DOM the heavens and DOM the earth). However, the translators of the later versions were aware of the necessity of producing a natural sounding text, so we do not find interference with the source text.

transitional condition, it is not surprising that the direct object is *a*-marked. In the B version from the 16<sup>th</sup> century, the direct object is neither moved nor marked. Both contemporary texts *a*-mark the object as expected.

(21) 1 Samuel 8, 13:

A (14<sup>th</sup>) E **a vuestras hijas tomará** por espeçieras e cosineras e panaderas.

B (16<sup>th</sup>) **Tomará** también **vuestras hijas** para que sean perfumadoras, cocineras, y amasadoras.

C (20<sup>th</sup> E) **Tomará** también **a vuestras hijas** para perfumistas, cocineras y amasadoras.

D (20<sup>th</sup> A) **Tomará** también **a vuestras hijas** para perfumistas, cocineras y panaderas.

English He will **take your daughters** to be perfumers, cooks and bakers.

This example illustrates the advantages of comparing the same text at different historical stages. Assuming that the translators had a similar understanding of the meaning of the text, we could say that *vuestras hijas* is topical in all texts with respect to thematic structure. However, it makes a difference whether the topicalized direct object is left-dislocated or not. In translation A from the 14<sup>th</sup> century the left-dislocated object is marked, while in translation B from two centuries later the object is not left-dislocated and not marked. This particular example cannot be generalized, but further investigation into the difference of topicalized direct objects and left-dislocated ones may reveal interesting contrasts (see also Escandell-Vidal 2007).

At the other end of the scale of verbal classes is the verb *matar* ‘to kill’ of class 1. We therefore would expect an early appearance of DOM, which is confirmed by the corpus, as illustrated by (22). Only translation A from the 14<sup>th</sup> century does not mark the direct object, while all others do.<sup>13</sup>

(22) 1 Reyes 19, 1:

A (14<sup>th</sup>) ... e como **mató todos los profetas** a espada.

B (16<sup>th</sup>) ... de como **había muerto** á cuchillo **á todos los profetas**.

C (20<sup>th</sup> E) ... y de cómo **había matado** a espada **a todos los profetas**.

D (20<sup>th</sup> A) ... y cómo **había matado** a espada **a todos los profetas**.

English ... how he had **killed all the prophets** with the sword.

<sup>13</sup> Again, one could speculate about the particular circumstances in this sentence: first, all versions have a second expression with *a*, i.e. a preposition without any blocking effect. Second, only in the A version is the direct object adjacent to the verb, while in the other versions the idiomatic expression *a cuchillo* or *a espada* separates verb from object, which might express the topicality of the direct object, thus triggering DOM.



Example (23) is of special interest since it provides two kinds of variation: first, translation A differs in its use of DOM in the same environment. In verse 3, the construction *mató jonatán el capitán* does not have *a*-marking, while the same structure *matara saul al capitán* in the next verse does have it. Here one can only speculate that DOM indicates discourse prominence. Second, we find an interesting variation between the verbs *matar* and *herir*. The verb *matar* in A seems to prefer DOM, while the verb *herir* in B does not, even though both express the same action in this example (see also the differences in table 5, below):

(23) 1 Samuel 13, 3-4

- A (14<sup>th</sup>) E **mató** jonatán **el capitán** de los filisteos que estaban en gueba  
todo ysraael oyeron desir que **matara** saul **al capitán** de los filisteos.
- B (16<sup>th</sup>) Y Jonathán **hirió la guarnición** de los Filisteos que había en el collado  
Y todo Israel oyó lo que se decía: Saúl ha **herido la guarnición** de los Filisteos
- C (20<sup>th</sup> E) Jonatán **atacó a la guarnición** de los filisteos que había en el collado.  
Cuando todo Israel supo que se decía: «Saúl ha **atacado a la guarnición** de los filisteos».
- D (20<sup>th</sup> A) Y Jonatán **hirió la guarnición** de los filisteos que estaba en Geba,  
Y todo Israel oyó decir que Saúl había **herido la guarnición** de los filisteos.
- English Jonathan **attacked the Philistine** outpost at Geba.  
So all Israel heard the news: “Saul **has attacked the Philistine outpost**”.

If we continue to discuss examples in such detail we would find for each example particular circumstances that might explain the *a*-marking in that case, but not the general principles. Therefore, the following presentation will give the statistics for all instances of DOM with the six verbs in the four different Bible translations in table 5. The number in brackets gives the instances of all definite human NPs for that verb. So for *poner* in A “25% (4)” means that 1 instance out of 4 of definite human direct objects is *a*-marked, while three are not, which correspond to the 25% of *a*-marking.

class		A: 14 <sup>th</sup> cent.	B: 16 <sup>th</sup> /17 <sup>th</sup> cent.	C: 20 <sup>th</sup> cent. (Euro)	D: 20 <sup>th</sup> cent. (Am)
3	<i>poner</i>	25% (4)	50% (6)	83% (6)	100% (6)
	<i>tomar</i>	31% (19)	23% (17)	62% (24)	68% (25)
	<b>sum</b>	<b>30% (23)</b>	<b>30% (23)</b>	<b>67% (30)</b>	<b>74% (31)</b>
2	<i>ver</i>	35% (20)	41% (22)	83% (29)	75% (20)
	<i>hallar</i>	50% (4)	80% (5)	66% (3)	75% (4)
	<b>sum</b>	<b>38% (24)</b>	<b>48% (27)</b>	<b>81% (32)</b>	<b>75% (24)</b>
1	<i>matar</i>	59% (32)	85% (27)	92% (27)	100% (27)
	<i>herir</i>	62% (8)	48% (29)	83% (12)	81% (16)
	<b>sum</b>	<b>60% (40)</b>	<b>66% (56)</b>	<b>92% (39)</b>	<b>93% (43)</b>

Table 5: Percentage of DOM with definite human direct objects (number of all definite human objects in brackets; Bible translations of 1+2 Samuel and 1+2 Kings)

The table clearly shows the development of DOM in Spanish for definite noun phrases along the timeline and by verb class. In translation A from the 14<sup>th</sup> century, class 3 provides 30% DOM, class 2 – 38%, and class 1 – 60%. These numbers increase through the centuries. In the American translation D class 3 shows 74%, class 2 exactly 75%, and class 1 more than 90% for *a*-marking. Keeping in mind that Bible translations are somewhat archaic and might represent the state of the language a half century or a century earlier, one could say that for (nearly) all definite human direct objects DOM has become obligatory in Modern Spanish.

### 4.3 Indefinite noun phrases

In a second step von Heusinger and Kaiser (2007) searched for instances of indefinite human direct objects. In general, there are less indefinite human direct objects than direct objects and *a*-marking can be found less frequently and only some centuries later. This is illustrated in example (24), where all European translations have *hallarás dos hombres* ('to find two men') without the marker and only the American translation D expresses the marker in *hallarás a dos hombres*.

(24) 1 Samuel 10, 2

- A (14<sup>th</sup>) En yéndote oy de mí **fallarás dos omnes** cerca la sepultura de rachel  
 B (16<sup>th</sup>) Hoy, después que te hayas apartado de mí, **hallarás dos hombres** junto al sepulcro de Rachêl, ...  
 C (20<sup>th</sup> E) Hoy, después que te hayas apartado de mí, **hallarás dos hombres** junto al sepulcro de Raquel  
 D (20<sup>th</sup> A) Cuando te apartes hoy de mí, **hallarás a dos hombres** cerca del sepulcro de Raquel, ...  
 English When you leave me today, you will **meet two men** near Rachel's tomb, ...

The discussion of example (25) below again reveals the variety of factors that might influence DOM, including different constructions and different lexical items. There are three instances of killing described in one and the same verse. Translation A uses the lexical item *matar* and a partitive-like construction *mató de los omnes* for the first instance, and indefinite direct objects for the other two instances. Translation B uses *herir*, DOM for the definite direct object *hirió Dios á los de Bethsemes*, and no marking for the indefinite one *hirió en el pueblo cincuenta mil y setenta hombres*. Translation C employs the periphrastic causative construction *hacer morir* (‘make die’) with DOM for the definite as well as the indefinite human direct object, and version D combines *herir* with DOM in all instances.

(25) 1 Samuel 6, 19

- A (14<sup>th</sup>) E **mató de los omnes** de betsemes, porque vieron el arca del señor, e **mató** en el pueblo **çinquenta mill e setenta omnes**. E pusieron luyto el pueblo, ca **mató** el señor en el pueblo **grant matanza**.
- B (16<sup>th</sup>) Entonces **hirió** Dios **á los de Bethsemes**, porque habían mirado en el arca de Jehová; **hirió** en el pueblo **cincuenta mil y setenta hombres**. Y el pueblo puso luto, porque Jehová **le había herido** de tan gran plaga.
- C (20th E) Entonces Dios **hizo morir a los hombres** de Bet-semes, 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.
- D (20th A) El Señor **hirió a los hombres** de Bet-semes 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.
- English ‘But God **struck down some of the men** of Beth Shemesh, **putting seventy 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.’

Table 6 summarizes the percentages for *a*-marking and the absolute numbers of all indefinite human direct objects for each verb in brackets. We can observe that there are fewer instances of indefinite human direct objects, and that *a*-marking starts some centuries later than with definite human direct objects. Therefore, we cannot see a significant difference for the three verb classes in the two older translations A and B. However, the contemporary translations C and D significantly indicate that *a*-marking depends on the verb class: class 3 shows about 13% DOM, class 2 – 67%, and class 1 exactly 100%.

class		A: 14 <sup>th</sup> cent.	B: 16 <sup>th</sup> /17 <sup>th</sup> cent.	C: 20 <sup>th</sup> cent. (Euro)	D: 20 <sup>th</sup> cent. (Am)
3	<i>poner</i>	0% (7)	0%(14)	14% (7)	0% (9)
	<i>tomar</i>	0% (8)	0% (14)	20% (5)	28% (7)
	<b>sum</b>	<b>0% (15)</b>	<b>0% (28)</b>	<b>17% (12)</b>	<b>13% (16)</b>
2	<i>ver</i>	0% (7)	20% (10)	50% (8)	56% (9)
	<i>hallar</i>	0% (4)	0% (3)	33% (3)	100% (3)
	<b>sum</b>	<b>0% (11)</b>	<b>15% (13)</b>	<b>45% (11)</b>	<b>67% (12)</b>
1	<i>matar</i>	7% (14)	14% (7)	87% (8)	100% (9)
	<i>herir</i>	-- (0)	0% (7)	100% (3)	100% (4)
	<b>sum</b>	<b>7% (14)</b>	<b>7% (14)</b>	<b>90% (11)</b>	<b>100% (13)</b>

Table 6: Percentage of DOM with indefinite human direct objects (number of all indefinite human objects in brackets; Bible translations of 1+2 Samuel and 1+2 Kings)

#### 4.4 Summary and further questions

This detailed but restricted corpus search has given clear evidence that the evolution of DOM in Spanish correlates with the verb class. The following two tables summarize the findings and provide the percentage for each class (based on the sum of the instances of both verbs per class) for definite human direct objects in table 7, and indefinite human direct objects in table 8.

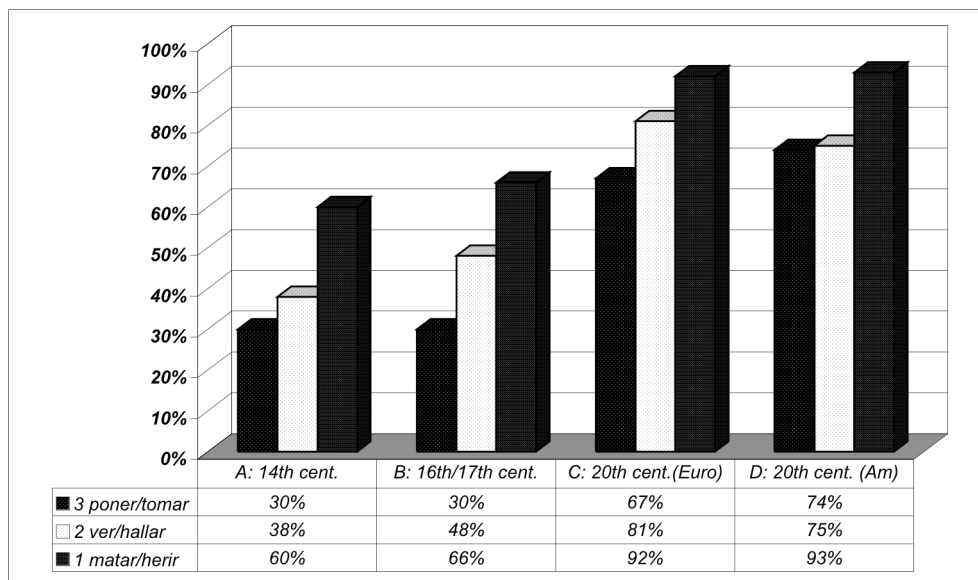


Table 7: Percentage of DOM with definite human direct objects  
(Bible translations of 1+2 Samuel and 1+2 Kings)

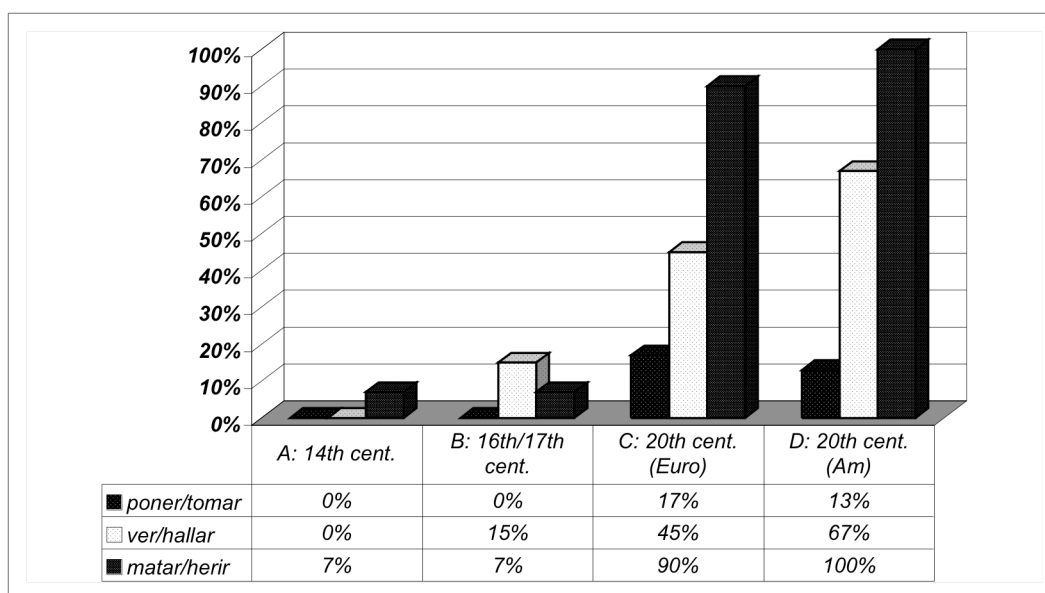


Table 8: Percentage of DOM with indefinite human direct objects  
(Bible translations of 1+2 Samuel and 1+2 Kings)

Table 9 summarizes both charts and shows that the distribution of DOM depends on (i) the timeline, (ii) the position of the Referentiality Scale, and (iii) the verb class. It is very interesting to note that for the latter two contextual factors the “distance” stays constant through time. Take, for example, the contrast between definite human and indefinite human direct objects for class 2 (*ver / hallar*). In translations A-C, the distance in percentages of DOM between these two contexts stays around 20-25%. Take the distance in percentage for definite human direct objects between the three verb classes (dotted lines). The difference stays between 10% to 15% in all four translations. This observation would fit the Constant Rate Hypothesis of Kroch (1989). However, it is hard to decide for this data whether the curve for each context (verb class and position of the Referentiality Scale) is S-shaped or not. See section 6 for a more detailed discussion of these issues.

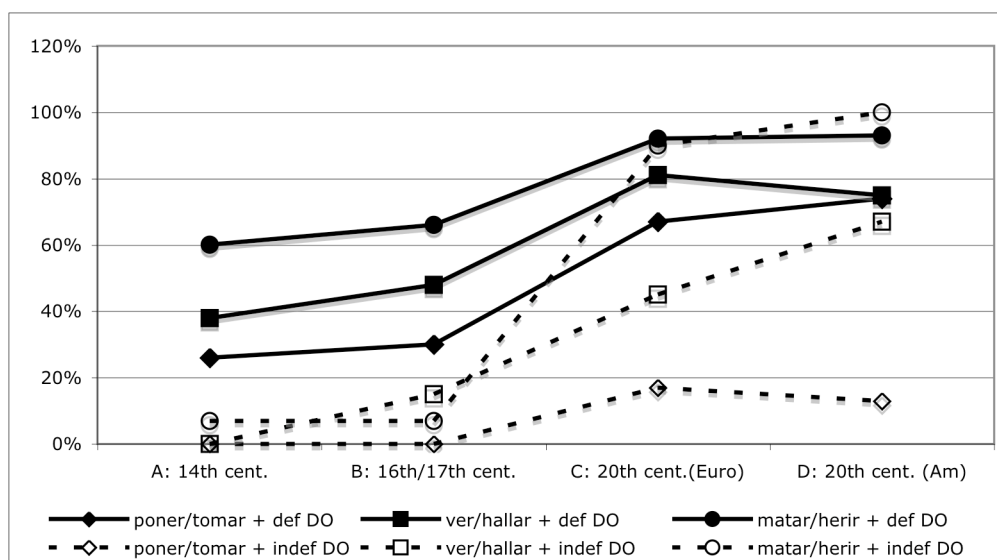


Table 9: Percentage of DOM comparing verb class 1-3 with definite and indefinite human direct objects  
(Bible translations of 1+2 Samuel and 1+2 Kings)

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 a larger corpus, as well? (ii) What has happened between early stages of the language and later stages? The Bible corpus is somewhat coarse-grained and only provides data from the fourteenth, sixteenth, and twentieth centuries. A more fine grained analysis would help us to see the development in more detail. (iii) What kind of curves do we find for each context, once we have more data through time?

### 5 *Corpus del Español*: 12<sup>th</sup> to 19<sup>th</sup> century

For confirming the analysis, I extended the corpus search, this time using the *Corpus del Español* of Mark Davies (<http://www.corpusdelespanol.org>). The corpus comprises 100 million words of Spanish texts from the twelfth to the nineteenth century. The corpus interface allows one to search for lemmas, rather than for word forms (as in simple text files as the Bible texts). However, my searches were still very time-consuming since I had to select the human definite or indefinite direct full NP objects by hand. In the case of *tomar* only about 1-7% of all hits for *tomar* were human definite or indefinite full NPs.<sup>14</sup> The others were

<sup>14</sup> For *tomar* I not only compared all hits with the human definite or indefinite full pronouns we also counted the relevant instances for the past tense *tomaron*, present tense *toman*, future *tomarán*. However there was no significant result between tense forms, in particular due to the very few instances for present and future tense. Moreover, the numbers for the 15<sup>th</sup> and 16<sup>th</sup> century are very low and therefore less reliable, see table below:

<i>tomaran</i>	<i>toman</i>	<i>tomaran</i>	all verb forms
----------------	--------------	----------------	----------------

either inanimate, 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, I searched all instances of a lemma if it produced fewer than 1,000 tokens. However, I had to confine the corpus study, so I searched for instances of *matar* (verb class 1), until we got 20-25 instances each of definite human direct objects, and indefinite human direct objects, I did the same for *tomar*, however I did not always obtain the same range of instances. The expectation was to find the early development of class 1 (*matar*) and a very late development in class 3 (*tomar*).

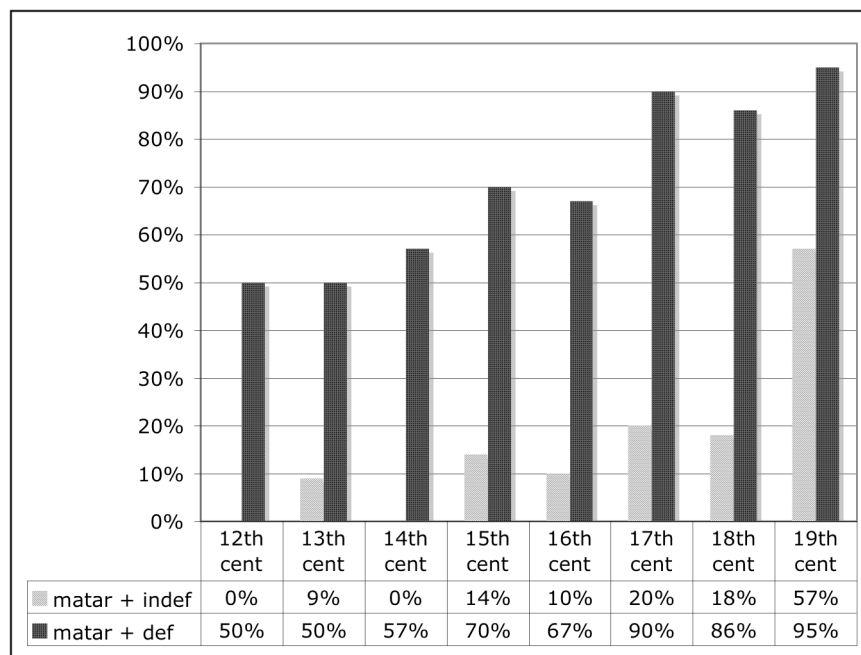
### 5.1 *Matar*

The verb *matar* ('to kill') from class 1 shows a high percentage of DOM in the earliest available texts. Table 10 provides the absolute figures and shows an increase of instances of DOM for definite and indefinite human direct objects, and table 11 gives the percentages for *a*-marking. In the twelfth century, 50% of definite human direct objects are marked with *a*. This number continually increases and reaches about 90% by the 17<sup>th</sup> century. As expected, indefinite direct objects are much less often *a*-marked. The development does not start before the seventeenth century and develops to about 50% by the 19<sup>th</sup> century. This confirms the finding from section 4.1.

	def	/	all		def	/	all		def	/	all		def	/	all	
	indef		forms	%	indef		forms	%	indef		forms	%	indef		forms	%
12th cent	66		1308	5,01	7		453	1,55	5		146	3,42	78		1907	4,09
13th cent	47		482	9,75	6		256	2,34	1		30	3,33	54		768	7,03
14th cent	32		762	4,20	8		582	1,37	1		84	1,19	41		1428	2,87
15th cent	39		1253	3,11	21		823	2,55	1		25	4	61		2101	2,90
16th cent	7		439	1,59	2		219	0,91	1		17	5,88	10		675	1,48
17th cent	3		397	0,76	6		362	1,66	0		4	0	9		763	1,18
18th cent	12		483	2,48	5		494	1,01	1		27	3,70	18		1004	1,79

Table (i) Absolute numbers of all human definite and indefinite full nouns, of all hits and the percentages with different tense forms of *tomar* (Corpus de Español)

<i>matar</i>	12th cent	13th cent	14th cent	15th cent	16th cent	17th cent	18thcent	19th cent
def Ø	13	12	10	6	7	2	3	1
def +a	13	12	13	14	14	18	19	20
sum	26	24	23	20	21	20	22	21
indef Ø	20	20	19	18	18	16	18	9
indef +a	0	2	0	3	2	4	4	12
sum	20	22	19	21	20	20	22	21

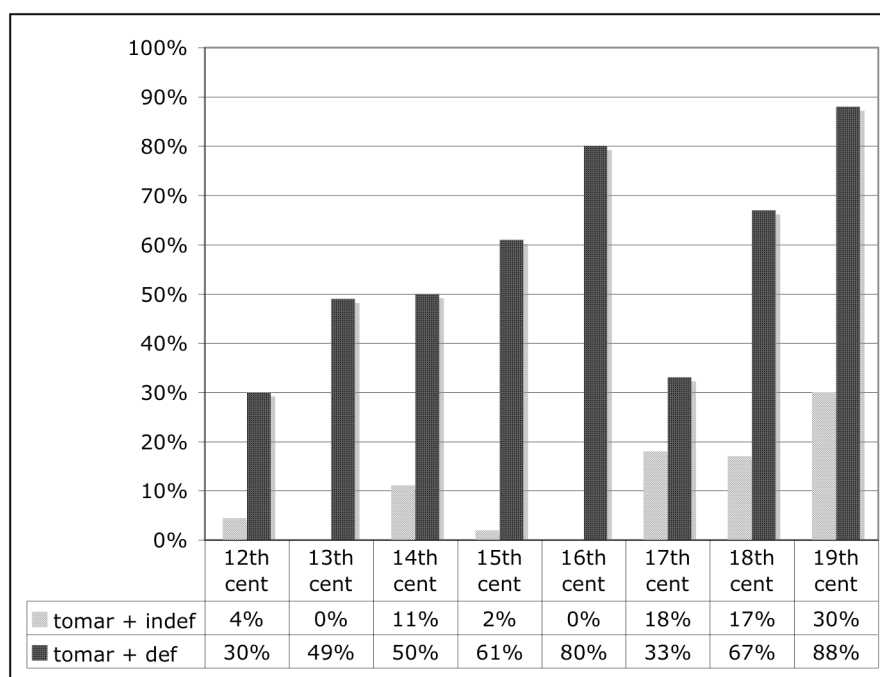
Table 10: Instances of definite and indefinite human direct objects with *matar* (Corpus del Español)Table 11: Percentages of DOM for *matar* with definite and indefinite human direct objects (Corpus del Español)

## 5.2 Tomar

The verb *tomar* ('to take') from class 3 also shows an increasing use of DOM with definite human direct objects from the twelfth to the nineteenth century. However, the development is two centuries later than for *matar* described in the last subsection. *Tomar* shows 50% DOM in the fourteenth century (*matar* in the twelfth century) and 90 percent in the nineteenth century (*matar* in the seventeenth century). DOM appears with indefinite human direct objects some centuries later. One finds not much marking before the nineteenth century, and even there the percentage is not very high. The absolute instances are summarized in table 12, the percentages of DOM in table 13:



tomar	12th cent	13th cent	14th cent	15th cent	16th cent	17th cent	18thcent	19th cent
def Ø	32	25	16	9	1	2	2	2
def +a	14	24	16	14	6	1	5	15
sum	46	49	32	23	7	3	7	17
indef Ø	28	5	8	37	3	5	15	9
indef +a	1	0	1	1	0	1	3	4
sum	29	5	9	38	3	6	18	13

Table 12: Instances of definite and indefinite human direct objects with *tomar* (Corpus del Español)Table 13: Percentages of DOM for *tomar* with definite and indefinite human direct objects (Corpus del Español)

### 5.3 Comparing verb classes through time

Table 14 compares the development of DOM for definite and indefinite human direct objects for *matar* and *tomar*. It shows three points: (i) DOM in Spanish increases through time; (ii) DOM depends on the Referentiality Scale as indefinite human direct objects show less preference for DOM than definite ones; (iii) there is a clear tendency that DOM depends on the verb class, i.e. on the selectional restrictions of the verb towards its direct object. This tendency is weaker than the clear correlation in the last section, but see the discussion below. (iv) Table 14 also shows a very constant correlation between the contexts created by definite and indefinite human direct objects, and the curves for the particular contexts are not S-

shaped indicating a confirmation of Kroch's (1989) Constant Rate Hypothesis, which will be discussed in the next section.

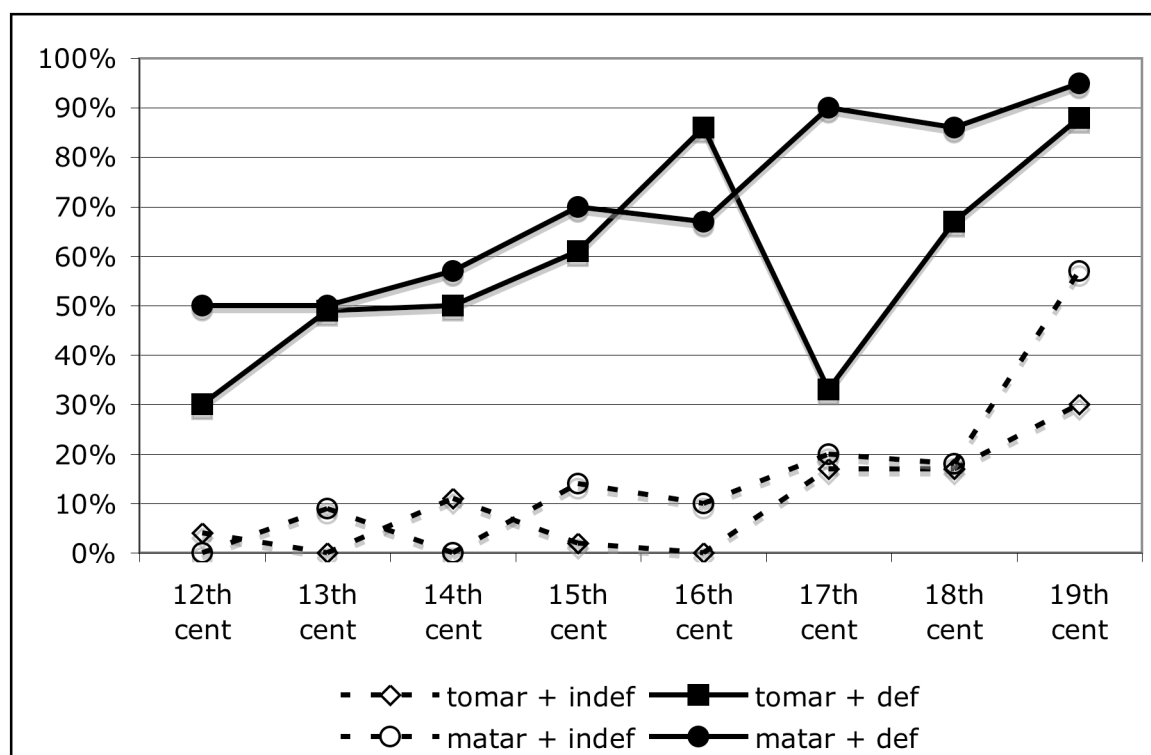


Table 14: Percentage of DOM comparing verb class 1 and 3 with definite and indefinite human direct objects (Corpus del Español)

The extended corpus search in the Corpus de Español has confirmed the results of the first search in four parallel Bible texts, in particular with respect to the dependency of the diachronic evolution of DOM on the verb class. The contrast between verb classes was very evident for the first corpus search (see table 9 above), while the second corpus search did not give such a clear result (see table 14). However, the sum of all instances of one context type across time significantly demonstrates the contrast between verb classes. The absolute numbers are listed in table 15 and the percentages in table 16. The absolute numbers show that I recorded considerably more occurrences of definite and indefinite human direct objects in the second search, which had 9 time slices from the 12<sup>th</sup> to the 19<sup>th</sup> century. It also shows that the Bible translations (with three time slices in the 14<sup>th</sup>, 16/17<sup>th</sup> and 20<sup>th</sup> century) did not provide many instances of indefinite human direct objects.

absolute numbers of all <i>a</i> -marking	indef. DO in Corpus Español	indef. DO in Bible translation	def. DO in Corpus Español	def. DO in Bible translation
<i>tomar</i> + <i>a</i>	11 (121)	3 (34)	95 (184)	42 (85)
<i>matar</i> + <i>a</i>	27 (165)	18 (38)	123 (177)	95 (113)

Table 15: Instances of DOM with definite and indefinite human direct objects for *tomar* and *matar* at all time points (number of all instances in brackets; Corpus de Español (12<sup>th</sup> - 19<sup>th</sup> cent) and Bible translations of 1+2 Samuel and 1+2 Kings 14<sup>th</sup>-20<sup>th</sup> cent.))

The percentages given in table 16 provide evidence that verb class is a relevant factor for determining DOM in Spanish. The contrast between verbs like *tomar* and verbs like *matar* is consistently shown by both corpus searches, even though they include quite different material and time points. The only deviance is the high percentage (47%) for indefinite human direct objects in the Bible translation. This might be an artifact of the summation of the same instances of the four translations of the same verses.

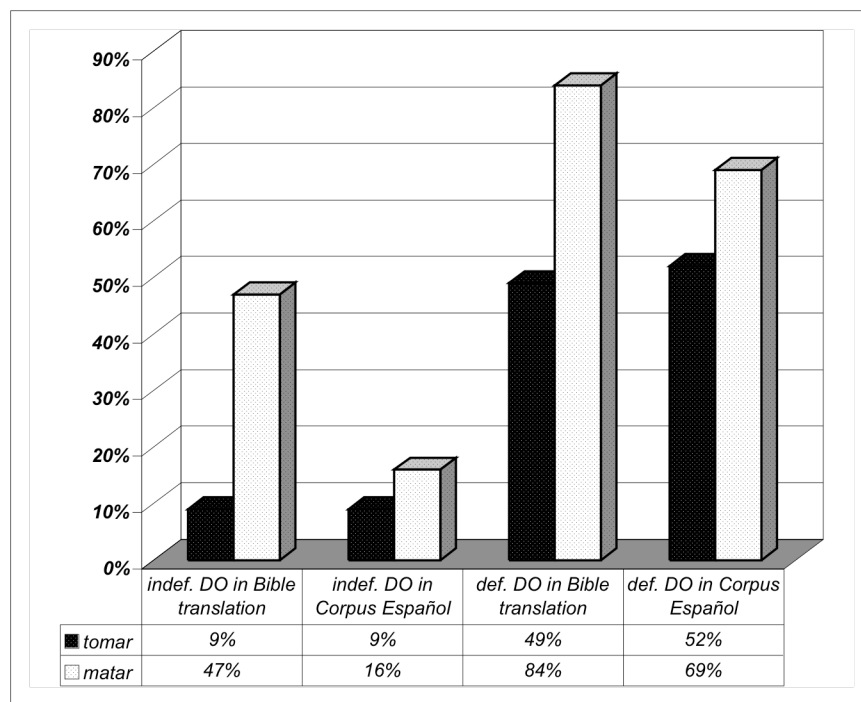


Table 15: Percentages of DOM with definite and indefinite human direct objects for *tomar* and *matar* at all time points (Corpus de Español (12<sup>th</sup> - 19<sup>th</sup> cent) and Bible translations of 1+2 Samuel and 1+2 Kings (14<sup>th</sup>-20<sup>th</sup> cent.))

## 6. Discussion and further questions

The two corpus studies have provided three observations with respect to the diachronic development of DOM in Spanish: one, the dependency on the position of the Referentiality Scale; two, the dependency on the verb class; and three, the type of change. However, some more questions have also been raised, which will be briefly addressed below. Furthermore, I evaluate the two main theories of DOM. While the paper investigates the factors that determine DOM diachronically, the question of why it develops at all is outside the scope of this paper.

The analysis presented in the last two sections has confirmed the well-known observation that the evolution of DOM in Spanish depends on the Referentiality Scale. From the early sources in the 12<sup>th</sup> century to Modern Spanish, the evolution concerns the definite and indefinite slot on that scale. We find not only clear evidence for that contrast, but also that the contrast stays stable through time (see discussion below). This paper does not investigate the transitional conditions for the evolution into the definite slot and then into the indefinite slot of the scale. In table 4 in section 3.2, I have suggested that topicality is a transitional condition for definite human NPs and specificity for indefinite ones. However, another study is necessary to confirm this. An additional question is whether these two transitional conditions only work for one category each and whether they might interact.

The second main result of this study is that verb classes differ with respect to their preference to take *a*-marking with definite and indefinite human direct objects. The first question that arises is how we can measure such a “preference”. I assume that frequency qualifies as being a clear indicator of preference. To show this I have categorized the first 100 hits for *mataron* and *tomaron* for the 19<sup>th</sup> century in the Corpus de Español according to NP-type, DOM and animacy. I list the following NP-types: clitics (Cl), personal pronouns (PerPr), proper names (PN), definite NPs, indefinite NPs, bare nouns (BN), and other types of complements (ot). I further distinguish between direct objects without marker ( $\emptyset$ ), with marker (a), and with marker and clitic doubling (a+cl). I have not listed all logical possibilities, since not all NP-types distribute over these subclasses, e.g. clitics and bare nouns never have *a*-marking, while personal pronouns always trigger *a*-marking and clitic doubling. Only definite and indefinite distribute over these subclasses. I finally distinguish between human, animate and non-animate direct objects. Even though most of these categories do not qualify for an interesting variation of *a*-marking, I think it is noteworthy to compare the absolute number of occurrences with that of definite and indefinite direct objects.

The data confirm the intuition about the preferences: *mataron* takes 95% human direct objects, while *tomaron* only 16%. If we compare the figures for definite and indefinite NPs, we find 37 human and 4 animate NP for *mataron*, and 3 human and 49 inanimate NPs for *tomaron*, which also shows frequent combinations with bare nouns and other constructions. This result, namely only 3% of human definite or indefinite NPs, confirms the findings in section 5.

<i>mataron</i>	Cl	PerPr	PN		def NP			indef NP		BN	ot	sum			
	∅	a	a	a+cl	∅	a	a+	∅	a	∅	∅	∅	a	a+cl	all
							cl								
human	39	1	9	1	1	10	6	1	19	4	4	49	39	7	95
animate	1	0	0	0	2	1	0	1	0	0	1	4	1	0	5
inanimate	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sum	40	1	9	1	3	11	6	2	19	4	5	53	40	7	100

Table 16: Instances of DOM according to NP-type and animacy for the first 100 hits for *mataron* (19<sup>th</sup> cent.)

<i>tomaron</i>	Cl	PerPr	PN		def NP			indef NP		BN	ot	sum			
	∅	a	a	a+cl	∅	a	a+cl	∅	a	∅	∅	∅	a	a+cl	all
human	11	1	0	0	1	0	0	2	0	0	1	15	1	0	16
animate	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
inanimate	4	0	7	0	38	0	0	11	0	9	15	84			84
sum	15	1	7	0	39	0	0	13	0	9	16	99	1	0	100

Table 17: Instances of DOM according to NP-type and animacy for the first 100 hits for *tomaron* (19<sup>th</sup> cent.)

Frequency shows the preference of one type of verb for a certain type of animacy: the higher the preference for human direct objects in general, the higher the preference to mark such a human direct object. Note that for this contrast it is not the animacy of the direct object, but the preference of the verb for a human direct object that matters. This observation needs more investigation, in particular we need to know what property of the verb might correlate with its preference towards human objects. We also need further studies with respect to other verbal properties or subproperties, such as the rather vague concept “Affectedness”.

The third result concerns the type of change and contributes to the general discussion of the way language change progresses (see Kroch 1989 and Pintzuk 2003).<sup>15</sup> One of the crucial questions of language change is how it proceeds. According to Baily (1973) the change slowly starts at one category (e.g. definite human direct objects), accelerates in using the new form up to a certain threshold above which it becomes grammaticalized, and then slows down since there are only very few old forms left. It then can invade into the next category in the same way, e.g. into the human indefinite slot. Such a change results in an S-shaped curve (for each category) and expresses the idea that contexts favoring a change allow for a higher rate of change than contexts that do not favor the change. This picture underlines the general assumption about diachronic DOM (Aissen 2003, see also section 3.2 above). In contrast to Bailey, Kroch (1989) proposes a different model of linguistic change, the Constant Rate Hypothesis (CRH), according to which the change in contexts that differ in favoring the change proceed at the same rate. The results in table 9 for the Bible translations and in table 14 for the data from the Corpus de Español show exactly this behavior. While definite human NPs strongly favor DOM, indefinite weakly favor DOM. Still, for both contexts the development proceeds at a similar rate. If more data confirms this kind of change, it would not only confirm Kroch's CRH, but also require a new theoretical model for diachronic (and probably also for synchronic) DOM.

These three observations provide an interesting testing ground for the two main approaches to DOM, the Ambiguity Theory and the Transitivity Theory (see section 2). The Ambiguity Theory assumes that DOM is used if the direct object becomes too similar to the subject in order to disambiguate the sentence. We find two versions of this assumption: first, the actual direct object is too similar to the actual subject; and second, the actual direct object is too similar to a typical subject. While languages usually follow the latter version, the *a*-marking of inanimate direct objects in Spanish (see examples (8) and (9)) can only be explained according to the first version. In contrast, the Transitivity Theory assumes that DOM is used to indicate that the direct object is a "salient" object such that the whole event becomes a salient event. "Salient" properties are high positions on the Animacy Scale and Referentiality Scale, thus the typical properties of the subject. Therefore, both theories make similar predictions for regular cases of DOM, as it can be shown for the observation that diachronic DOM has a preference for definite human NPs over indefinite NPs. Definite humans NPs are more like subjects and they are salient objects.

---

<sup>15</sup> I would like to acknowledge my gratitude to one reviewer, who brought this very important aspect to my attention.

The second finding is that verb classes constitute different contexts for DOM. Verbs of class 1 (type: ‘the man kills the boy’) have a higher preference for DOM than verbs of class 3 (type: ‘the man takes the boy’). Both classical approaches to DOM have difficulties explaining these data: the Ambiguity Theory cannot explain the difference between the two constructions since they are very similar, if not identical, with respect to the similarity between subject and direct object. The Transitivity Theory also falls short of explaining the difference between *kill* and *take* in terms of affectedness or salience of the direct object or the event as a whole. There might be a slight difference in Affectedness, however, the concept itself was left unexplained (see discussion in section 2.2).

A new kind of explanation is necessary. The observation of this paper is that DOM signals that the verb generally has a high preference for human objects, while no marking indicates that the verb has a low preference for human object (in contexts where all other parameters are constant). It is still unclear to me why we need a marker that indicates what we already know, namely that the verb has a high preference. In a certain way we would expect the opposite: marking the unexpected. However, we may interpret our observations that DOM is facilitated by those contexts in which the marker is coherent with the verb semantics. If this is correct, the marker not only expresses nominal properties of the object, but also properties of the relation between verb and object. Finally, the data from the diachronic development are crucial for models of diachronic DOM, as discussed above.

In summary, Differential Object Marking in Spanish can be explained by the interaction of nominal properties of the direct object and verbal properties of the governing predicate, including verb classes that differ in their preference for taking human direct objects. The diachronic data have confirmed this picture, since not only animacy and the position of the Referentiality Scale determine the diachronic development, but also the verb class. Additionally, we found that the change of DOM through time does not show an S-curve, but rather curves of similar distances between different contexts, confirming the Constant Rate Hypothesis of Kroch. These findings suggest that *a*-marking in Spanish, and DOM in general, can only be described in a multi-dimensional space consisting of the semantic properties of the direct object and the lexical semantics of the verb.

## Sources

Bible A: 14<sup>th</sup> century: *Biblias medievales romanceadas. Biblia medieval romanceada judio-cristiana. Versión del Antiguo Testamento en el siglo XIV, sobre los textos hebreo y latino*. Vol. I: Genesis-Reyes. Edición y estudio introductorio por el P. José Llamas. Madrid: Instituto «Francisco Suarez» (= E4 or Escorial I.I.4).

Bible B: 16<sup>th</sup>/17<sup>th</sup> century: *Reina Valera Antigua* (1569/1602). Source: <http://www.biblegateway.com>  
 Bible C: 20<sup>th</sup> century: *Reina Valera* (1995) (United Bible Societies). (<http://www.biblegateway.com/>)  
 Bible D: 20<sup>th</sup> century: *La Biblia de las Américas* (1971) (The Lockman Foundation). (<http://www.biblegateway.com/>)  
 Holy Bible: The 21st Century King James Version of the Holy Bible (KJ21®) is an updating of the 1611 King James Version (KJV)(source: <http://www.biblegateway.com/>)  
 Corpus del Español from the 12<sup>th</sup> to the 19<sup>th</sup> century (<http://www.corpusdelespanol.org>).

## References [please adopt bibliography to the *Probus* style sheet]

- Aissen, Judith (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21: 435-483. [→correct it in the rest of the bibliography]
- Baily, Charles-James (1973). *Variation and linguistic theory*. Arlington, Va.: Center for Applied Linguistics.
- Bello, Andrés (1847). *Gramática de la lengua castellana destinada al uso de los americanos*. Con las notas de Rufino José Cuervo. Vol. 2. Madrid: Arco Libros, estudio y edición de Ramón Trujillo 1988.
- Bolinger, Dwight L. (1953). Verbs of being. *Hispania* 36: 343-345.
- Bleam, Tonia (1999). *Leista Spanish and the Syntax of Clitic Doubling*, PhD. Dissertation, University of Delaware.
- Bossong, Georg (1985). Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen. Tübingen: Narr.
- Brugè, Laura and Gerhard Brugger (1996). On the acusative *a* in Spanish. *Probus* 8: 1-51.
- Butt, Miriam (2006). The Dative-Ergative Connection. In Empirical Issues in Syntax and Semantics 6, Olivier Bonami and Patricia Cabredo Hofherr (eds.), 69-92. (<http://www.cssp.cnrs.fr/eiss6>)
- Company, Concepción (2002). El avance diacrónico de la marcación prepositiva en objetos directos inanimados". In *Actas del II Congreso de la Sociedad Española de Lingüística*. Madrid: CSIC, 146-154.
- Comrie, Bernard (1975). Definite and animate direct objects: a natural class. *Linguistica Silesiana* 3: 13-21.
- Croft, William (1988). Agreement vs. case marking and direct objects. In *Agreement in Natural Language. Approaches, Theories, Descriptions*, Michael Barlow and Charles A. Ferguson (eds.), 159-179. Stanford: Center for the Study of Language and Information.
- Croft, William (2003). *Typology and Universals*. Cambridge: Cambridge University Press, 2<sup>nd</sup> edition.
- Cysouw, Michael and Bernhard Wälchli (2007). Parallel texts: using translational equivalents in linguistic typology. *Sprachtypologie und Universalienforschung* 60: 95-99.
- Enrique-Arias, Andrés (2007). Biblias romanceadas e historia de la lengua. To appear in: *Actas del congreso de Historia de la Lengua de Mérida*.
- Escandell-Vidal, Victoria (2007). Topics from Ibiza: differential object marking and clitic-dislocation. In *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*, Georg A. Kaiser and Manuel Leonetti (eds.), 23-43. Universität Konstanz: Fachbereich Sprachwissenschaft (Arbeitspapier 122).
- Fernández Ramírez, Salvador (1951). *Gramática española. Vol. 4: El verbo y la oración (volumen ordenado y completado por Ignacio Bosque)*. Madrid: Arco/Libros, segunda edición 1986.
- Fish, Gordon A. (1967). 'A' with Spanish direct object. *Hispania* 50: 80-85.



- García García, Marco (2007) Differential object marking with inanimate objects. In *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*, Georg A. Kaiser and Manuel Leonetti (eds.), 63-84. Universität Konstanz: Fachbereich Sprachwissenschaft (Arbeitspapier 122).
- Guntsetseg, Dolgor (2007). *Differential Object Marking in Mongolian*. Ms. Universität Stuttgart.
- von Heusinger, Klaus and Georg A. Kaiser (2005). The evolution of differential object marking in Spanish. In *Proceedings of the Workshop "Specificity and the Evolution / Emergence of Nominal Determination in Romance"*, Klaus von Heusinger, Georg A. Kaiser and Elisabeth Stark (eds.), 33-69. Universität Konstanz: Fachbereich Sprachwissenschaft (Arbeitspapier 119).
- von Heusinger, Klaus and Georg A. Kaiser (2007). Differential Object Marking and the Lexical Semantics of Verbs in Spanish. In *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*, Georg A. Kaiser and Manuel Leonetti (eds.), 83-109. Universität Konstanz: Fachbereich Sprachwissenschaft (Arbeitspapier 122).
- von Heusinger, Klaus and Edgar Onea (2008). *Triggering and blocking effects in the diachronic development of DOM in Romanian* (in this volume).
- de Hoop, Helen and Bhuvana Narasimhan (2005). Differential case-marking in Hindi. In *Competition and Variation in Natural languages the Case for Case*, Mengistu Amberber and Helen de Hoop (eds.), 321-345. Amsterdam Oxford: Elsevier.
- Hopper, Paul and Sandra Thompson (1980). Transitivity in grammar and discourse. *Language* 56: 251-299.
- Isenberg, Horst (1968). *Das direkte Objekt im Spanischen*. Berlin: Akademie Verlag.
- Fontana, Josep M. (1993). Phrase structure and the syntax of clitics in the history of Spanish. PhD thesis, University of Pennsylvania.
- Kaiser, Georg A. (2005). Bibelübersetzungen als Grundlage für empirische Sprachwandeluntersuchungen. In *Romanistische Korpuslinguistik II. Romance Corpus Linguistics II. Korpora und diachrone Sprachwissenschaft. Corpora and Diachronic Linguistics*, Claus D. Pusch, Johannes Kabatek and Wolfgang Raible (eds.), 71-85. Tübingen: Narr.
- Kroch, Anthony S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199-244.
- Laca, Brenda (2002). Gramaticalización y variabilidad - propiedades inherentes y factores contextuales en la evolución del acusativo preposicional en español. In *Sprachgeschichte als Varietätengeschichte: Beiträge zur diachronen Varietätenlinguistik des Spanischen und anderer romanischer Sprachen. Festschrift für Jens Lüdtko zum 60. Geburtstag*, Andreas Wesch (ed.), 195-303. Tübingen: Stauffenburg.
- Laca, Brenda (2006). El objeto directo. La marcación preposicional. In *Sintaxis histórica del español. Primera parte: La frase verbal. Volumen 1*, C. Company (dir.), 423-475. Mexico: Fondo de cultura económica / Universidad Nacional Autónoma de México.
- Leonetti, Manuel (2004). Specificity and differential object marking in Spanish. *Catalan Journal of Linguistics* 3: 75-114.
- Leonetti, Manuel (2008). *Clitics do not encode specificity* (in this volume).
- Melis, Chantal (1995). El objeto directo personal en *El Cantar de Mio Cid*. Estudio sintáctico-pragmático. In *El complemento directo preposicional*, Carmen Pensado (ed), 133-163. Madrid: Visor.
- Moravcsik, Edit (1978). On the case marking of objects. In *Universals of Human Language IV*, Joseph H. Greenberg (ed.), 249-290. Stanford: Stanford University Press.

- Naess, Ashild (2004). What markedness marks. The markedness problem with direct objects. *Lingua* 114: 1186-1212.
- Naess, Ashild (2007). Prototypical Transitivity. Amsterdam: Benjamins.
- Parodi, Teresa (1998). Aspects of clitic doubling and clitic clusters in Spanish. In *Models of Inflection*, R. Fabri, O. Albert and Teresa Parodi (eds.), 85-102. Tübingen: Niemeyer.
- Pensado, Carmen (1995). El complemento directo preposicional. Estado de la cuestión y bibliografía comentada. In *El complemento directo preposicional*, Carmen Pensado (ed), 11-60. Madrid: Visor.
- Pintzuk, Susan (2003). Variationist approaches to syntactic change. In *The Handbook of Historical Linguistics*, Brian D. Joseph and Richard D. Janda (eds.), 509-528. Oxford: Blackwell Publishing.
- Pottier, Bernard (1968). L'emploi de la préposition *a* devant l'objet en espagnol. *Bulletin de la Société de Linguistique* 1: 83-95.
- Real Academia Española (1973). *Esbozo de una gramática de la lengua española*. Madrid: Espasa Calpe.
- Roegiest, Eugeen (1979). A propos de l'accusatif prépositionnel dans quelques langues romanes. *Vox Romanica* 38: 37-54.
- Silverstein, Michael (1976). Hierarchy of features and ergativity. In *Grammatical Categories in Australian Languages*, Robert M. W. Dixon (ed.), 112-171. Canberra: Australian Institute of Aboriginal Studies.
- Suñer, Margarita (1988). The Role of Agreement in Clitic-Doubled Constructions. *Natural Languages and Linguistic Theory* 6: 391-434.
- Torrego, Esther (1999). El complemento directo preposicional. In *Gramática descriptiva de la lengua española Vol. 2: Las construcciones sintácticas fundamentales*. Relaciones temporales, aspectuales y modales, Ignacio Bosque and Violeta Demonte (eds.), 1779-1805. Madrid: Espasa Calpe.