

## Affectedness and Differential Object Marking in Spanish

Klaus von Heusinger · Georg A. Kaiser

**Abstract** In this study we investigate the impact of affectedness on the diachronic development of Differential Object Marking (DOM) in Spanish. DOM in Spanish synchronically depends on (i) the referential features of the direct object, such as animacy and referentiality, and (ii) the semantics of the verb. Several studies have also shown that the diachronic development of DOM proceeds along the Animacy Scale and the Referentiality Scale, and some recent corpus studies have indicated an influence of the verb's semantics on this diachronic process. This study presents new findings from a detailed analysis of extensive corpus research on the distribution of DOM with respect to affectedness, understood as "the persistent change of an event participant". We use Tsunoda's Affectedness Scale to order the verb classes under investigation. Our findings provide evidence that this scale can be partly correlated with the diachronic spread of DOM in Spanish which would tend to confirm the influence of verbal semantics on DOM in Spanish.

**Keywords** Case marking · Affectedness · Differential Object Marking · Spanish · Diachronic change

### 1 Introduction

Affectedness is one of the key parameters of Hopper and Thompson's (1980) Transitivity Scale and a central notion for defining direct objecthood (Fillmore 1968; Anderson 1971; Jackendoff 1990; Dowty 1991; Beavers 2010). It is generally understood as the change in the direct object that is imposed by the main predicate.

---

K. von Heusinger (✉)  
Universität Stuttgart, Stuttgart, Germany  
e mail: Klaus.vonHeusinger@ling.uni-stuttgart.de

G. A. Kaiser  
Universität Konstanz, Konstanz, Germany  
e mail: Georg.Kaiser@uni-konstanz.de

Apart from many other applications, it plays a central role in selecting case frames for predicates. Tsunoda (1985) proposed the Affectedness Scale that is based on cross-linguistic generalizations over transitive case frames. The hierarchy predicts that a transitive case frame that is found with verbs of one class is also found with verbs of higher classes. In this paper we apply this Affectedness Scale to one particular well-studied phenomenon: Differential Object Marking (DOM) in Spanish. DOM in Spanish depends on the referential properties (often summarized as “individuation”) of the direct object and the lexical semantics of the main predicate (Bossong 1985; Pensado 1995; Torrego 1999; Leonetti 2004). Melis (1995) and Laca (2002, 2006) have investigated the diachronic development of DOM and relate it to referential parameters of the arguments, such as animacy, referentiality (definiteness and specificity) and topicality. There have, however, been no systematic diachronic investigations into the conditions imposed by the verb on the argument. Von Heusinger (2008) presents a corpus study from the 14th to the 20th century for three verb classes that differ in their selectional restrictions on the direct object in terms of animacy. In the first class (type ‘to kill’), the predicate obligatorily requires an animate object, the second class (‘to see’) does not show any restriction, and the third class (‘to put’) has a strong preference for inanimate objects. The study provides evidence that the development of DOM in Spanish is determined not only by the referential properties of the direct object, but also by the verb semantics governing the direct object. The present study of the diachronic investigation of Spanish verb classes focuses on the concept of affectedness, generally understood as “persistent change in an event participant”, as an additional parameter for DOM. Following the Transitivity Hypothesis that aligns individuation with affectedness, we propose that direct objects high on the Affectedness Scale tend to get DOM earlier and more often than objects low on that scale. We compare the increase of DOM from the 15th to the 19th century for definite and indefinite human direct objects of verbs in five different classes according to Tsunoda’s (1985) Affectedness Scale ranked in the following way: ACTION > PERCEPTION > PURSUIT > KNOWLEDGE > FEELING. Our findings indicate that the distribution of DOM in Spanish is partly aligned with this scale, but the findings also indicate that affectedness is only a secondary parameter and that the current concept of affectedness needs more clarification.

In Sect. 2 we discuss affectedness, its relation to semantic transitivity and different hierarchies. Section 3 presents a brief overview of the referential parameters of the direct object that trigger DOM in Spanish. Section 4 reports on earlier studies on the evolution of DOM in Spanish. We compare approaches that focus on the individuation of the direct object with studies on the influence of verb classes. Section 5 presents new and extensive diachronic data that arranges the verb classes according to Tsunoda’s Affectedness Hierarchy. The analysis of the findings reveals evidence for a partial correlation of the Affectedness Scale with the diachronic evolution of DOM.

## 2 Affectedness

Affectedness is a central notion in the discussions of different phenomena of argument realization. Affectedness generally understood as the “persistent change in

an event participant” has been a key concept in analyzing argument realization and defining direct objecthood (Fillmore 1968; Anderson 1979, 2006; Jackendoff 1990; Dowty 1991; Beavers 2006, 2010). It is one parameter of Transitivity and aligned with high transitive case frames (Hopper and Thompson 1980; Tsunoda 1985; Malchukov 2005). Affectedness is known to determine telicity (Tenny 1987; Krifka 1989) and it is also correlated with other syntactic operations such as passivisation and reflexivization (cf. Jaeggli 1986). However, affectedness is used as a rather vague concept and there is no agreement on a clear definition. In syntactically oriented literature it is analysed as a feature having a  $\pm$  value (cf. Anderson 2006). In the typological and functional literature it is rather understood as gradable on a scale, similar to individuation. The literature proposes different hierarchies of affectedness determined by various criteria: grades of affectedness can depend on the change of state of the patient or on the movement of a theme along some path; these grades can also depend on the type of change or the domain to which the change applies, such as existence, location or sensation. In the following we first present the Transitivity Hypothesis and then discuss the question whether affectedness is a property only of the argument (Naess 2004) or of a relation between the predicate expressing the event and the argument expressing one participant. We then show that affectedness has at least two dimensions, the domain it applies to (existence, location, sensation) and the degree it applies to an object (total, partial, minimal).

Affectedness is one of ten semantic transitivity parameters suggested by Hopper and Thompson (1980). They maintain that the categories in Table 1 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, or has a less affected and less individuated object. Note that not all of these parameters must be instantiated at the same time.

**Table 1** Parameters of transitivity (Hopper and Thompson 1980, p. 252)

	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	Agent high in potency	Agent low in potency
9. Affectedness of O	Object totally affected	Object not affected
10. Individuation of O	Object highly individuated	Object not individuated

These properties can be grouped into three main classes (see Givón 1985, p. 90 and the discussion in Malchukov 2005, p. 79): (i) *Agent-related properties*: The prototypical transitive clause has a visible, salient, volitional, controlling agent-cause which initiates the event. (ii) *Patient-related properties*: The prototypical transitive clause has a visible, salient, non-volitional, non-controlling patient-effect which registers the bulk of change associated with the event. (iii) *Verb-related*: The prototypical transitive clause has a compact, perfective, realis verb or verbal tense-aspect-modality. The last parameter in the table, namely *individuation*, summarizes the semantic and pragmatic factors (the referential features) which contribute towards the referential strength of the direct object: animacy, definiteness, specificity and topicality. Affectedness, on the other hand, comprises the property of induced change of the direct object and thus is object-related, but rooted in the lexical semantics of the verb.

Naess (2004) develops a modified view of affectedness. She argues that affectedness is a property of the argument comprising referential features like animacy, definiteness and more importantly “saliency”. She assumes that actions with animate objects are more saliently affected than inanimate ones. According to her, *kill* entails higher affectedness than *break* because it has more dramatic consequences for humans. Thus affectedness is closely related (and often aligned) with animacy and definiteness, but she shows that in some cases saliency can override animacy and definiteness (see example (i) in footnote 1 below). While we agree with Naess that the referential features of the argument have to include discourse information and notions like saliency, we think that the term ‘affectedness’ should be reserved for the interaction between the verb semantics and the argument features.

Lehmann (1991, p. 217) illustrates the traditional notion of affectedness with the contrast between *affected* and *effected* objects in (1) and (2):

- (1) Paul corrected the letter.
- (2) Paul wrote the letter.

The direct object *the letter* in (1) is an *affected* object, i.e., an object that is changed by the event expressed by the predicate. This change can also happen at other domains such as the physical form, the mental representation etc. The direct object *the letter* in (2) is an *effected* object, i.e., its existence is caused by the event or situation expressed by the predicate. It is also called *object of result*. Lehmann (1991, p. 218) notes that “while affected objects may be affected in different ways and to different degrees, effected objects cannot be said to be affected by the situation in any way or degree.” They are either created or not. The degree of affectedness is illustrated by (3a-d) from Beavers (2010, p. 2):

- (3) a. John ate the apple up. (Apple is completely gone)
- b. John cut the apple. (Apple cut, not necessarily to a particular degree)
- c. John kicked the apple. (Apple impinged upon, not necessarily affected)
- d. John touched the apple. (Apple manipulated, not necessarily impinged upon)

The apple in (3a-d) is increasingly less affected by the situation or event expressed by the predicate. Lehmann (1991, p. 221) proposes a two-dimensional affectedness space by quality and quantity as in Fig. 1. Effected objects are created and therefore do not show grades of existence (we will see below that other approaches include effected objects in the affectedness hierarchies). Affected objects vary on two dimensions: quality or the domain in which the object is affected: motion, existence, mental, affection, non-attainment; and the quantity or the grade to which an object is affected: total, partial or minimal.

Figure 1 indicates a problem with measuring the degree of affectedness: Is the degree a function of the quantitative dimension, of the qualitative dimension or of both? We find different positions to this question: In the generative tradition starting with Tenny (1987) affectedness is a grammatical primitive that may consist of different aspectual properties, but that shows one categorical value (plus or minus affected). Beavers (2006, 2010) provides a four step categorical distinction in quantity, while functional approaches (Tsunoda 1985; Malchukov 2005) provide an integrated scale of quality and quantity.

Beavers (2010, p. 5) summarizes the different domains of affectedness i.e., different types of change discussed in the literature, as in (4). According to Beavers, the verb classes in (4) are related, they express prototypical properties of direct objects and they trigger certain syntactic constructions such as DP-passive formation and middles.

- (4) a. x changes in some observable property. (*clean/paint/delouse/fix/break x*)  
 b. x transforms into something else. (*turn/carve/change/transform x into y*)  
 c. x moves to and stays at some location. (*move/push/angle/roll x into y*)  
 d. x is physically impinged. (*hit/kick/punch/rub/slap/wipe/scrub/sweep x*)  
 e. x goes out of existence. (*delete/eat/consume/reduce/devour x*)  
 f. x comes into existence. (*build/design/construct/create/fashion x*)

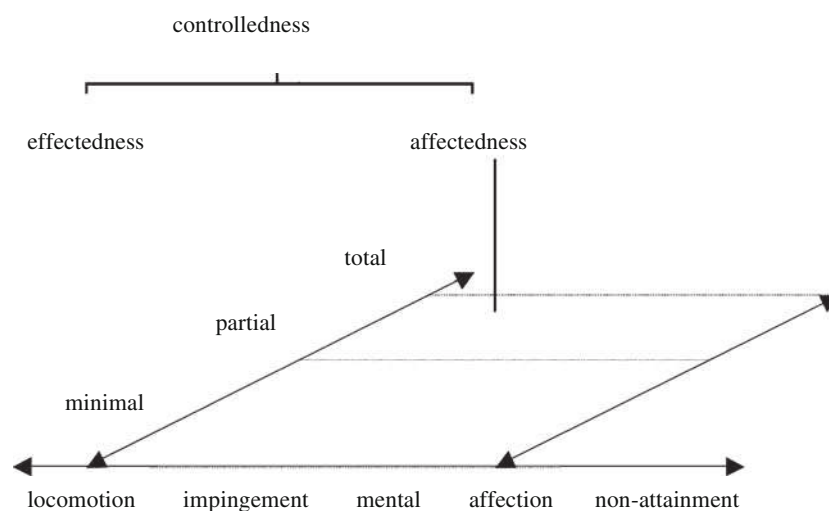


Fig. 1 Two dimensional affectedness space (Lehmann 1991, p. 221)

**Table 2** The Affectedness Hierarchy (Beavers 2010, p. 24, ex (64))

Quantized change	Non quantified change	Potential for change	Unspecified for change
<i>kill, break</i>	<i>widen</i>	<i>hit, kick</i>	<i>wait, search for</i>

Beavers develops a semantic model with semantic properties that allows us to account for this common behaviour of the verb classes in (4). He proposes a scalar analysis of the quantity aspect or the grade of imposed change where every degree of affectedness can be separated from other degrees by linguistic tests. He conceives affectedness as a transition of a theme along a path or scale that defines the change. He defines the Affectedness Hierarchy in Table 2 with four degrees of affectedness, which correspond to the degree of specificity in the verb about the endpoint of the theme's movement on the path or scale. Values range from highly specific on the left side to unspecified on the right side.

Like Lehmann, Beavers suggests a two-dimensional space of affectedness. One dimension represents the type of change and the other the degree of change. He gives a clear definition of the latter in terms of implicational properties. However, he does not provide a ranking of the first dimension, the type of change. Such a ranking is proposed by Tsunoda (1985, p. 388). He starts with a ranking of different types of verb classes as in (5) based on cross-linguistic generalizations of case pattern.

(5) Verb type hierarchy (Tsunoda 1985, p. 388)

EFFECTIVE ACTION >> PERCEPTION >> PURSUIT >> KNOWLEDGE >> FEELING >>  
RELATION >> ABILITY

This hierarchy has the more transitive verbs on the left and the less transitive ones on the right and it predicts that if there is a transitive case frame (nominative-accusative in accusative languages, and ergative-absolutive in ergative languages), then the verb types to the left have this case frame as well. The verbs in higher classes correspond more to the prototype of semantic transitivity of Hopper and Thompson (1980), i.e., they express a salient event or situation. Verbs of direct effect like *kill* or *hit* affect the object much more than verbs of pursuit like *search for* or *wait for*. The functional principle behind this order is the grade of affectedness of the object. Tsunoda assumes that different types of change also express different grades of affectedness, thus combining the two-dimensional space of Lehmann, Beavers and others into a one-dimensional scale, which is also known as Tsunoda's Affectedness Scale, as in Table 3. He admits further sub-classes for direct effect and perception, which correspond to Beavers' quantity analysis in Table 2. *Kill* belongs to class 1a since it has a direct result on the object, while *hit* belongs to 1b since it does not need to change the object as such (see discussion in Malchukov 2005).

Table 3 Affectedness Scale (Tsunoda 1985, p. 388, first 5 classes)

1		2		3	4	5
Direct effect on patient ( Effective Action)		Perception		Pursuit	Knowledge	Feeling
1a +result	1b result	2a +attained	2b attained			
<i>kill break</i>	<i>hit, shoot</i>	<i>see, hear</i>	<i>look, listen</i>	<i>search for, wait for</i>	<i>know, understand</i>	<i>like, fear, fond of</i>

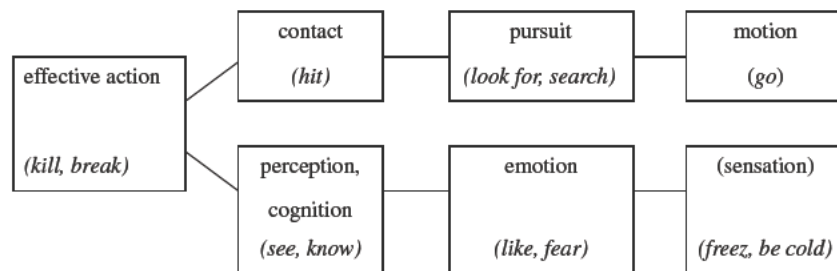


Fig. 2 Partial Affectedness Scale (Malchukov 2005, p. 83)

Malchukov (2005, p. 79ff) decomposes this hierarchy into a partially ordered hierarchy in Fig. 2. He argues, following other functionalist approaches (Hopper and Thompson 1980; Comrie 1989; Croft 1988; Givón 1985; Lehmann 1991) that semantic transitivity is composed of different properties (see above): (i) agent-related, (ii) patient-related and (iii) verb-related. The upper sub-scale is patient-related: *kill* > *hit* > *look for* form a decreasing scale with respect to the change in the patient. *Go* is an intransitive verb, i.e., does not have a direct object, so it would be at the end of this sub-scale. The lower sub-scale expresses a decreasing prototypicality of the subject as a salient, volitional controlling agent along the cline of *kill* > *see* > *like* > *be cold*.

Summarizing this section, we can say that affectedness can be expressed in degrees in a two-dimensional space, a one-dimensional scale or in a scale with partial orderings of verb classes. It is widely used for explaining different grammatical contrasts, structural properties and universal implications of ranked verb classes. As a first approximation we use Tsunoda's Affectedness Scale in Table 3 for our analysis of verbal properties triggering the development of DOM in Spanish. Since we are investigating one language in depth, we hope not only to learn more about DOM in Spanish, but also to contribute to a better understanding of affectedness as a linguistic category and its interaction with grammatical constructions. We are aware that we are still working with a rather informal concept and not with a well-defined and clearly detectable grammatical notion.



### 3 Diachronic DOM and nominal parameters in Spanish

Differential Object Marking in Spanish is expressed by the marker *a*, which is a homophone of the preposition *a* ‘to’ and of the dative marker *a* of the indirect object. We will confine the investigation to European Spanish throughout this paper, but see Company (2002) for American Spanish. DOM or *a*-marking in Spanish is determined by two main parameters: (i) properties of the direct object, and (ii) transitivity properties of the verb, including the lexical semantics of the verb. It is commonly assumed that the main factor favouring DOM in the languages of the world is the referential status of the direct object, i.e., the combination of semantic and (discourse) pragmatic features such as animacy, referentiality (definiteness and specificity), and topicality (see Comrie 1975; Bossong 1985; Croft 1988; Aissen 2003; de Hoop and Narasimhan 2005; Butt 2006; de Swart 2007). There are additional semantic features such as number and collectivity, and discourse pragmatic ones, such as prominence or saliency (see above), which influence the referential status of an argument, but which we cannot consider here. Each particular parameter can be expressed by a scale of two or more values. A language locates its DOM cut-off point at one particular point on the scale – the language-specific transition point. We confine our presentation to the Animacy Scale (6) and the Referentiality Scale (7), 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 bottom.

- (6) Animacy Scale:  
human > animate > inanimate
- (7) Referentiality Scale:  
personal pronoun > proper noun > definite NP > indefinite specific NP >  
indefinite non-specific NP > non-argumental

In Modern Spanish, a human direct object receives *a*-marking, i.e., DOM, as in (8a), while *a*-marking with inanimate direct objects is in general ungrammatical, as in (8b) (Pensado 1995; Torrego 1999; Delbecque 1998; Leonetti 2004).<sup>1</sup>

- (8) a. Conozco \*(a) este actor.  
know-1.SG DOM this actor  
‘I know this actor.’

<sup>1</sup> Inanimate direct objects can take *a* marking under certain conditions, as in (i) where the subject is inanimate as well. See Weissenrieder (1991), Delbecque (2002) and García (2007) for an extensive discussion of other examples and other conditions. In the remainder of this article we focus on the distribution of *a* marking with human direct objects.

(i) Un adjetivo acompaña/califica a un sustantivo.  
an adjective accompanies/qualifies DOM a noun  
‘An adjective accompanies/qualifies a noun.’



- b. Conozco (\*a) esta película.  
 know-1.SG DOM this film  
 'I know this film.'

We further assume that the cut-off point for DOM on the Referentiality Scale is somewhere in the slot for non-specific indefinites, since they can optionally be marked by *a*, while non-arguments can never be marked (see Leonetti 2004, who convincingly argues that *a*-marking does not express specificity). In other words, *a*-marking in Spanish (for human direct objects) indicates that the noun is an argument and introduces a discourse referent, but cannot be a predicative expression that might be incorporated. The definite noun phrase in (9a) and the indefinite (specific) noun phrase in (9b) must be marked by *a*. The non-specific indefinite noun phrase in (9c) may optionally be marked with *a*. The non-specificity is clearly indicated by the subjunctive form *sepa* in the relative clause. Even the indefinite pronoun *alguien* in (9d) takes *a* in its non-specific reading. Only the non-specific reading of (9e) does not allow *a*. Note that in the specific reading 'to need a certain assistant' *a*-marking is appropriate.

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

Like Modern Spanish, Old Spanish (10th century to 16th century) exhibits DOM. However DOM in Old Spanish is less frequent than in Modern Spanish and is used in different conditions, as has been shown in several diachronic studies (Melis 1995; Laca 2002, 2006). The main results of these studies are repeated here briefly and illustrated with some examples from the *Cantar de mio Cid* from the 14th century (following Melis 1995; Laca 2006). Object personal pronouns, strong or weak, carry obligatory DOM in Old Spanish, as in (10). Human proper names acting as direct object are obligatorily *a*-marked, as in (11).

- (10) 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.'

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

Human definite direct objects are optionally *a*-marked, as the two examples in (12) illustrate. Animate indefinite direct objects are never *a*-marked, as in (13) (cf. Laca 2006, p. 444).

- (12) a. Reçiba *a* **mios yernos** commo elle pudier  
 receive-IMP.2SG DOM my sons-in-law as he could-3.SG  
 mejor (Cid, 2637)  
 better  
 ‘Let him give to my sons-in-law the finest possible welcome.’  
 b. Ca yo case **sus hijas** con yfantes de Carrion  
 for I married.1SG. his daughters with Infantes of Carrion  
 ‘for I married his daughters to the Infantes of Carrion.’ (Cid, 2956)

- (13) Tanto traen las grandes ganancias, **muchos gañados**  
 very brought.3PL the big wealths many herds  
**de ovejas e de vacas**  
 of sheep and of cows (Cid, 480 481)  
 ‘They brought such great wealth, many herds of sheep and cows.’

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. Laca (2006) describes the diachronic development of *a*-marking in Spanish based on a qualitative evaluation of selected texts from the 12th century to the 19th century. She collects about 100–150 direct objects for each period and categorizes them according to their animacy and referentiality. Table 4 provides a selection of Laca’s data with the total number of instances in brackets (see for similar results Company 2002, p. 149).

Figure 3, based on Laca’s data, compares the development of *a*-marking for four main lexical classes of the Referentiality Scale: proper nouns, definite NPs, indefinite

**Table 4** Selection from Table 3 of Laca (2006, p. 442)

	XII	XIV	XV	XVI	XVII	XVIII	XIX
NPrHum	96% (26)	100% (8)	100% (35)	95% (44)	100% (65)	79% (29)	89% (27)
HumDef Pro	36% (36)	55% (66)	58% (65)	70% (122)	86% (136)	85% (53)	96% (76)
HumInd Pro	0% (6)	6% (31)	0% (11)	12% (59)	39% (53)	62% (32)	41% (29)
Hum0	0% (12)	0% (7)	16% (12)	5% (40)	2% (39)	9% (22)	6% (17)

*NPrHum*: human proper name, *HumDef Pro*: human definite NP, *HumInd Pro*: human indefinite NP, *Hum0*: human bare noun

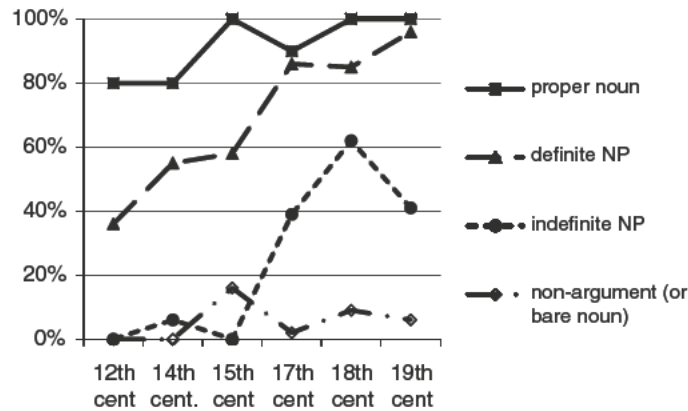


Fig. 3 Percentage of *a* marking depending on lexical type (referential type) and time (based on Laca 2006, p. 442, Table 3)

NPs, and bare nouns (or non-argumental nouns) as in (7). Since personal pronouns are always *a*-marked in the oldest texts, they are not included in the overview. All categories show an increasing preference for *a*-marking. Compare the percentages for human definite direct objects with those for human indefinite direct objects: 36% of definite direct objects in Old Spanish are marked, while 96% are marked in the 19th century. There is no instances of marked indefinite objects in the 12th century, while 41% are marked in the 19th century.

In this section we have concentrated on the referential properties of the direct object in isolation. Delbecque (1999, 2002) systematizes the observation that *a*-marking depends on the relation between the subject and the direct object (also Pensado 1995; Torrego 1999; Company 2002; García 2007), rather than on the referential status of the direct object in isolation. While this relation is often understood in terms of disambiguation, Delbecque gives an overall account in terms of construction grammar (Croft 1998). Her central claim is that the contrast between direct objects with and without *a*-marking is captured by two transitive paradigms. The *a*-paradigm provides two roles to the direct object – the typical goal-role of the event and an attributor-role, which has properties of a second subject. In other words, the *a*/ $\emptyset$  contrast does not rely on the properties of the noun itself, but on the whole sentence or construction. The *a*-marked direct object receives a much more independent role in the event structure of the sentence, while the  $\emptyset$ -marked direct object tends to be integrated into the verb meaning. Delbecque (2002, p. 84) not only accounts for examples with inanimate direct objects as in (14), but also for alternations with animate direct objects as in (15).

- (14) El presidente contesta (a) la pregunta.  
‘The president answers (DOM) the question’
- (15) Los franceses eligen el/al Presidente para cinco años.  
‘the French elect the/DOM-the President for five years’

Delbecque (2002, p. 115) argues that the “effect of the *a* construal on the S[subject] entity is that it is conceived of as a particular subset, viz. the French people involved in the elections, as opposed to the generic S[subject] entity of the  $\emptyset$  construal which ranges over the whole set, viz. the French people in general and over time”.

Our own proposal is neutral with respect to the assumption of two different verb frames as assumed by Delbecque. Still we think we have to extend the analysis: *a*-marking in Spanish does not only depend on the referential status of the direct object or the relation between the subject and the direct object, but also on the verbal semantics, i.e., on the restriction from the verb on the direct object.

#### 4 Diachronic DOM and verbal parameters

The diachronic development of DOM in Spanish as described in the last section is based on the referential properties of the direct object and its spread along the Referentiality Scale (see Melis 1995; Laca 2006; for a detailed discussion of the vast literature see Pensado 1995). Ever since the spread of DOM in Spanish and other Romance languages has been described, both traditional grammarians and contemporary researchers have noted the influence of the verbal semantics of the main predicate on the distribution of DOM. Bello (1847, pp. 567–570) and Fernández Ramírez (1951, pp. 151–190) present rich material on the variation according to different verb types in Spanish. See also the observations in Reichenkron (1951, pp. 367–368) and Meier (1945, p. 142) for Spanish, Stimm (1986, p. 443) for a Swiss Raeto-Romance dialect, Roegiest (1979, pp. 41–43) and Delille (1970) for Portuguese, and Puscariu (1937, pp. 449–456) for Romanian, to name only a few. However, there has been no quantitative investigation of Spanish diachronic corpora that allows us to evaluate the influence of verb classes or verbal semantics on *a*-marking in Spanish. Delille (1970) is alone in giving a comprehensive overview of verb classes in Portuguese following Pottier’s (1968) categorization. In von Heusinger and Kaiser (2007) we use the same verb classes for a diachronic survey of four chapters of the Bible from 12th century to the 19th century and in von Heusinger (2008) the corpus search was extended to two large diachronic corpora of Spanish.

Pottier (1968, p. 87) proposes a two dimensional space for *a*-marking in Spanish. One dimension corresponds to the Referentiality Scale and the other to a verbal scale (“un axe sémantique verbal”) with four different verb classes ranked according to the degree of activity that the predicate attributes to the object.

- (16) Verbal Scale (Pottier 1968, p. 87 “un axe sémantique verbal”)  
*matar* ‘kill’ > *ver* ‘see’ > *considerar* ‘consider’ > *tener* ‘have’

Pottier argues that the spread of *a*-marking in Spanish develops in this two dimensional space. He gives some examples, but no quantitative analysis. Delille (1970) assumes that DOM in Portuguese (in the 17th century) follows individuation (referential status) and the verb class according to Pottier. Delille (1970, p. 100) reduces the four classes of Pottier to three classes defined as: (a) verbs that express

**Table 5** Verbal scale on preferred animacy of the direct object (von Heusinger and Kaiser 2007, p. 94)

Class 1 [+human]	Class 2 [ $\pm$ human]	Class 3 [( $\pm$ )/ animate]
<i>matar</i> / <i>herir</i> 'kill' / 'hurt'	<i>ver</i> / <i>hallar</i> 'see' / 'find'	<i>tomar</i> / <i>poner</i> 'take' / 'put'

an intended action or that are applied to a human theme; (b) verbs that take an inanimate direct object, and (c) verbs that allow for both (a) and (b). Delille gives lists and numbers of verbs that fall under these classes, but he does not provide a quantitative overview of the distribution of DOM across these classes.

In von Heusinger and Kaiser (2007) we used the verbal scale of Pottier in the interpretation of Delille. We assume there that the particular ranking depends on the animacy requirement imposed by the verb on the direct object. The verb *matar* 'to kill' has a strong tendency to take human objects, while *ver* 'to see' has no restriction with respect to animacy. *Considerar* 'to consider' would prefer an abstract object and *tener* 'to have' an inanimate one. It is important to note that this requirement of the verb on the direct object is different from assuming that DOM depends on the animacy of the direct object. As mentioned above, the study only considers human direct objects. This means that even with verbs like *tener* 'to have' which strongly prefer an inanimate direct object, DOM can only appear with human direct objects (in general). We used a verbal scale of three verb classes as in Table 5 that differ with respect to the preference for animacy in the direct object, as in (16). We did not include existential verbs since they have a very strong tendency not to allow *a*-marking with human direct objects even today.

Furthermore, only instances of full definite or indefinite human direct objects are investigated in this study. Personal pronouns and proper names of human direct objects have always been *a*-marked as far back as in the 14th century, which means that the evolution of DOM can be much better observed with full human (definite and indefinite) direct objects.

In von Heusinger and Kaiser (2007) we used as a small corpus the two books of Samuel and the two books of Kings in three Bible translations, abbreviated as A C: translation A is from the 14th century and is only available in a printed version. All the other translations are available in electronic form: B refers to the *Reina Valera Antigua* from 16th/17th century, and C to the version from 1995 (*Reina Valera*). We used for the translation of the examples into English *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. The assumption is that Bible translations serve this requirement best (cf. Harris and Campbell 1995; Kaiser 2005; Enrique-Arias 2008). Still, they constitute a very archaic text and often have quite a specialized register, which differs substantially from that of the spoken language, although they contain a considerable amount of natural-sounding direct speech.

DOM continuously spreads from one lexical class of the Referentiality Scale to the next. Therefore, we first investigated the situation with human definite direct

objects. The verb *tomar* ‘take’ is of class 3, i.e., it prefers to take inanimate direct objects. As shown in example (17), it can also take human ones. In the translation from the 14th century, the direct object is left-dislocated, an indication of it being topicalized. In the B version from the 16th century, the direct object is neither moved nor marked. The contemporary text *a*-marks the object as expected.

- (17) 1 Samuel 8, 13:  
 A (14th) E **a vuestras hijas tomará** por espeçieras e cosineras e panaderas.  
 B (16th) **Tomará** también **vuestras hijas** para que sean perfumadoras,  
 cocineras, y amasadoras.  
 C (20th) **Tomará** también **a vuestras hijas** para perfumistas, cocineras y  
 amasadoras.  
 English He will **take your daughters** to be perfumers, cooks and bakers.

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 (18). Only translation A from the 14th century does not mark the direct object, while all others do.

- (18) 1 Kings 19, 1:  
 A (14th) ... e como **mató todos los profetas** a espada.  
 B (16th) ... de como **había muerto** á cuchillo **á todos los profetas**.  
 C (20th) ... y de cómo **había matado** a espada **a todos los profetas**.  
 English ... how he had **killed all the prophets** with the sword.

Tables 6 and 7 provide the percentages of *a*-marking of definite and indefinite direct objects (and the absolute numbers of *a*-marked instances and all instances in brackets) for the three verb classes and three different Bible translations. While there are a considerable number of instances of human definite direct objects and a distribution clearly dependent on verb class, the text provides fewer instances of human indefinite direct objects. We cannot therefore see a substantial effect of verb class in the *a*-marking of indefinite direct objects in the two older translations. However, the contemporary translation would suggest that *a*-marking depends on the verb class: 90% (10/11) of indefinite direct objects are marked for class 3 verbs, 45% (5/11) for class 2 verbs and only 17% (1/12) for class 3 verbs.

**Table 6** Percentage of *a* marking of human definite direct objects (three Bible translations of 1 + 2 Samuel and 1 + 2 Kings)

Class	14th cent.	16th/17th cent.	20th cent.
1. <i>matar, herir</i>	60% (24/40)	66% (37/56)	92% (36/39)
2. <i>ver, hallar</i>	38% (9/24)	48% (13/27)	81% (26/32)
3. <i>poner, tomar</i>	30% (7/23)	30% (7/23)	67% (20/30)

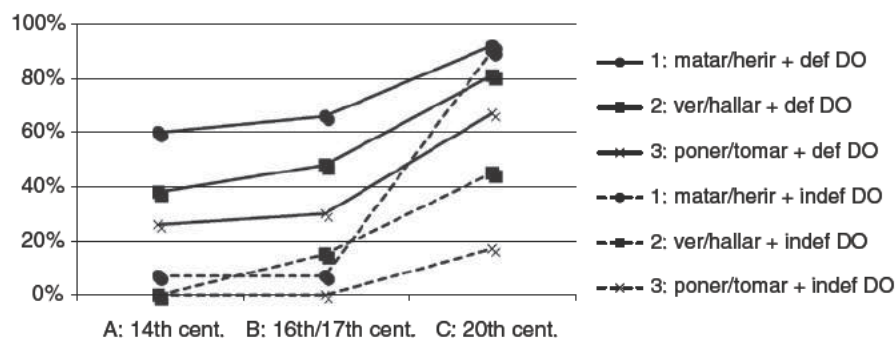
**Table 7** Percentage of *a* marking of human indefinite direct objects (three Bible translations of 1 + 2 Samuel and 1 + 2 Kings)

Class	14th cent.	16th/17th cent.	20th cent.
1. <i>matar, herir</i>	7% (1/14)	7% (1/14)	90% (10/11)
2. <i>ver, hallar</i>	0% (0/11)	15% (2/13)	45% (5/11)
3. <i>poner, tomar</i>	0% (0/15)	0% (0/28)	17% (2/12)

Figure 4 provides a comparison of the percentages for definite and indefinite human direct objects for all three classes and three Bible translations. It shows that the distribution of *a*-marking depends on (i) the historical period, (ii) the position on the Referentiality Scale, and (iii) the verb class. Thus this detailed quantitative corpus search has given significant evidence that the evolution of DOM in Spanish proceeds differentially by verb class. The significance of the corpus result was tested with a generalized linear model. The relevant Wald chi-square tests for difference between the verb classes shows clear statistical significance for indefinites ( $\chi^2 = 7.828$ ,  $df = 2$ ,  $p < 0.05$ ) and for definites ( $\chi^2 = 169.884$ ,  $df = 2$ ,  $p < 0.01$ ). For indefinites, class 2 did not significantly differ from class 1 or class 3.

The questions that arise from these findings are: (i) The generalization for marking definite and indefinite objects is based on selected chapters of the Bible. Is the generalization valid for a larger corpus as well? (ii) The study only uses data from three historical periods, from the fourteenth, sixteenth, and twentieth centuries. The question is whether a comparison with more time periods would give similar results or not.

Von Heusinger (2008) addressed these two questions and extended the corpus search to more precise historical periods, using Mark Davies' *Corpus del Español*. The corpus comprises 100 million words of Spanish texts from the 12th to the 19th century. The corpus interface allows one to search for lemmas, rather than for word forms (as in simple text files of the Bible texts). However, such searches are still very time-consuming since one has to select the definite or indefinite human direct objects by hand. In the case of *tomar* only about 17% of all hits for *tomar* were



**Fig. 4** Percentage of *a* marking depending on verb class, definiteness and time (three Bible translations of 1 + 2 Samuel and 1 + 2 Kings)



**Table 8** Percentage of *a* marking of human definite direct objects (Corpus del Español)

Class	12th + 13th cent.	14th + 15th cent.	16th + 17th cent.	18th + 19 cent.
1. <i>matar</i>	50% (25/50)	63% (27/43)	78% (32/41)	91% (39/43)
3. <i>tomar</i>	40% (38/95)	55% (30/55)	70% (7/10)	83% (20/24)

**Table 9** Percentage of *a* marking of human indefinite direct objects (Corpus del Español)

Class	12th + 13th cent.	14th + 15th cent.	16th + 17th cent.	18th + 19th cent.
1. <i>matar</i>	5% (2/42)	8% (3/40)	15% (6/40)	37% (16/43)
3. <i>tomar</i>	3% (1/34)	4% (2/47)	11% (1/9)	23% (7/31)

human definite or indefinite full NPs. The others were either inanimate, or human and of a different type on the Referentiality Scale, such as clitics, personal pronouns, proper names or different types of quantifiers. The study originally differentiates between eight time periods from the 12th to the 19th century, which we have reduced to four time periods. On the other hand the search was restricted to two verb classes, and one verb for each class: *matar* 'to kill' for class 1 and *tomar* 'to take' for class 3 (see von Heusinger 2008 for the details). We summarize the findings in Tables 8 and 9, where we have combined two centuries to each time period since this provides higher and therefore more reliable numbers. Table 8 shows that in the 12th and 13th century, 50% of human definite direct objects of *matar* are marked with *a*. This number continually increases and reaches about 90% by the 18th and 19th century. The marking of the definite direct object of *tomar* is less preferred. Only about 40% in the 12th and 13th century are marked, a number that continuously increases to about 80% in the 18th and 19th century.

Table 9 provides the numbers for human indefinite direct objects. As expected, *a*-marking is less preferred, but we can see a clear increase over time and some difference between the two verb classes.

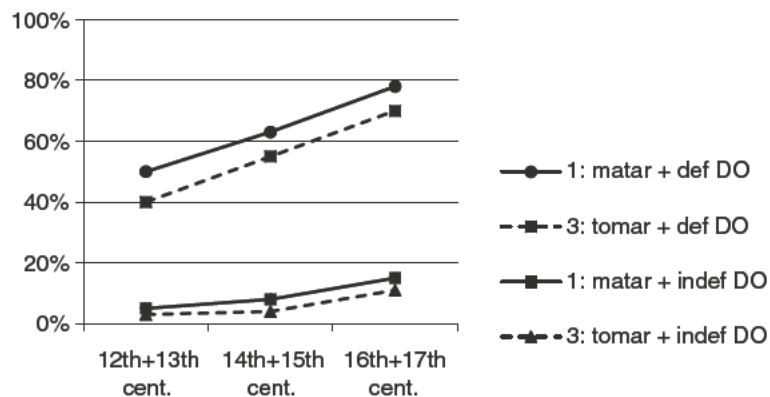
**Fig. 5** Percentage of *a* marking depending on verb class, definiteness and time (Corpus del Español)

Figure 5 compares the development of *a*-marking for definite and indefinite human direct objects for the two verbs. It shows three points: (i) *a*-marking in Spanish increases over time; (ii) it depends on the Referentiality Scale as human indefinite direct objects show less preference for DOM than definite ones; (iii) there is a tendency for *a*-marking to depend on the verb class, i.e., on the preference of the verb for the animacy of the direct object. The significance of the corpus result was tested with a generalized linear model. The relevant Wald chi-square tests for difference between the verb classes is close to statistical significance for indefinites ( $\chi^2 = 72.554$ ,  $df = 1$ ,  $p < 0.15$ ) and for definites ( $\chi^2 = 2.884$ ,  $df = 1$ ,  $Sign < 0.1$ ) Note that only human direct objects were counted, which means that we have two independent parameters: first the actual animacy of the direct object and second the preference of the verb for the animacy of the direct object. The question is now whether this preference depends on a more general property of the verb, such as affectedness.

### 5 Tsunoda's Affectedness Scale and diachronic DOM

In order to understand the findings presented in the last section better, we designed a broader corpus investigation that takes the first five verb classes of Tsunoda's (1985) Affectedness Scale of Table 3 (repeated here as Table 10) as the relevant ordering principle. Each class, including the sub-classes of class 1 and 2, is represented by two Spanish verbs, so there are 12 verbs in total, as listed in Table 10.

We searched in the two largest electronically available corpora: (i) the *Corpus del Español* of Mark Davies mentioned above and the *Corpus diacrónico del Español* of the Real Academia Española ("Corde" in the following). The latter corpus comprises more than 250 million entries from all times and text types. It is the largest existing corpus of historical Spanish texts. The search was confined to human definite or indefinite (full) noun phrases. Since the corpora are not tagged for the referential type of the object, the search required many steps to be undertaken manually, such as selecting the relevant cases from among the full set of hits. Only about 0.5–3% of hits for indefinite noun phrases and 3–8% of hits for definite noun phrases were human full noun phrases. The other hits had either no direct object or they had clitics, pronouns, proper names or inanimate full noun phrases.

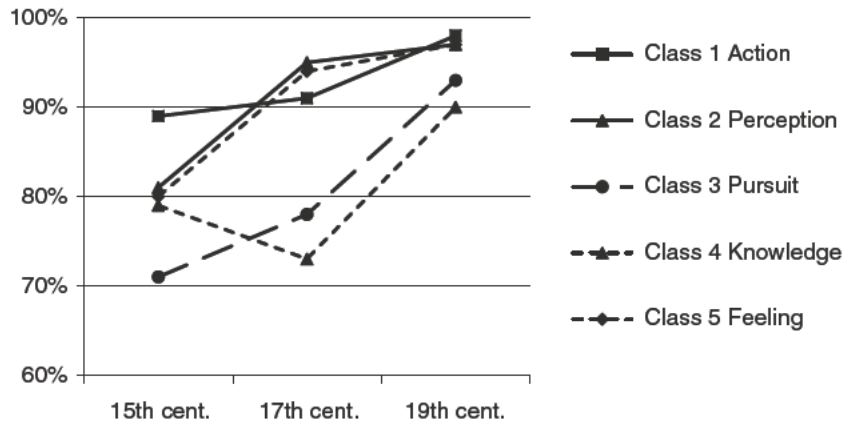
**Table 10** Affectedness Scale of Tsunoda (1985, p. 388, first 5 classes) with Spanish verbs

1		2		3	4	5
Direct effect on patient (effective action)		Perception		Pursuit	Knowledge	Feeling
1a +result	1b result	2a +attained	2b attained		1a +result	1b result
<i>matar</i> 'kill'	<i>golpear</i> 'hit'	<i>ver</i> 'see', <i>oir</i> 'hear'	<i>escuchar</i> 'listen', <i>mirar</i> 'look at'	<i>buscar</i> 'search for', <i>esperar</i> 'wait for'	<i>conocer</i> 'know', <i>entender</i> 'understand'	<i>querer</i> 'like', <i>temer</i> 'fear'
<i>herir</i> 'violate'	<i>tirar</i> 'shoot'					

The selection process therefore entailed checking 500 to 1,500 hits to obtain 10–30 relevant cases. In general, there were far fewer hits for indefinite noun phrases.

**Table 11** Percentage of *a* marking of human definite direct objects for five verb classes (Corpus de Español and Corde)

Class	15th cent.	17th cent.	19th cent.
1a + 1b EFFECTIVE ACTION: <i>matar, herir, golpear, tirar</i>	89% (136/153)	91% (112/123)	98% (132/135)
2a + 2b PERCEPTION: <i>oir, ver, escuchar, mirar</i>	81% (81/100)	95% (142/149)	97% (224/231)
3 PURSUIT: <i>buscar, esperar</i>	71% (73/103)	78% (82/105)	93% (102/110)
4 KNOWLEDGE: <i>conocer, entender</i>	79% (33/42)	73% (48/66)	90% (61/68)
5 FEELING: <i>querer, temer</i>	80% (32/40)	94% (58/62)	97% (71/73)

**Fig. 6** Percentage of *a* marking of human definite direct objects depending on verb class and time (Corpus de Español and Corde)

Therefore adding together the results of the two or four verbs of each class resulted in a reasonable number of relevant instances.

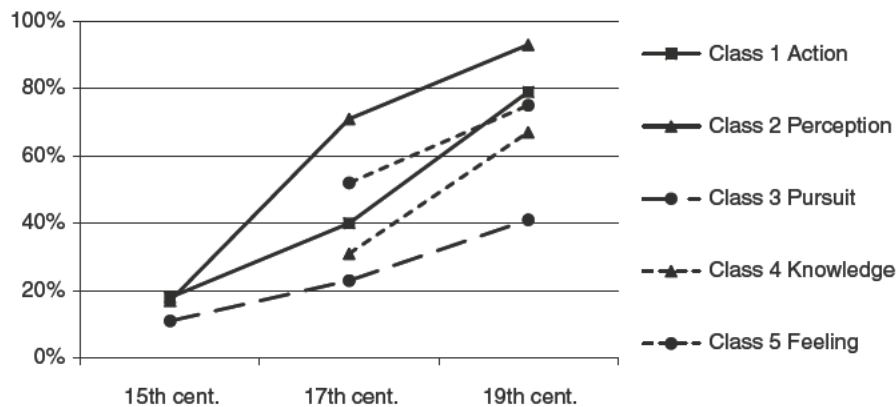
Table 11 provides the figures for human definite direct objects for the five verb classes.<sup>2</sup> There were about 2,000 relevant hits, and for each century around 500 for definite noun phrases and between 70 and 170 for indefinite noun phrases. For definite direct objects there is in general a high preference for *a*-marking from the 15th century onwards, with a clear increase towards the 19th century. Since the percentage of *a*-marking is already quite high we do not see a very dramatic difference between the verb classes, even though we can detect some preference, as presented in Fig. 6.

Table 12 provides the figures for human indefinite direct objects. As expected, they are much lower than those for the definites, but they show a clear increase over time. Since the difference between the frequency of *a*-marking in the 15th century and in the 19th century is larger, we can see a more dramatic development for the different classes, as illustrated in Fig. 7.

<sup>2</sup> We also compared the figures between the search in Davies and in Corde. There was an average of 5.12% difference for particular verbs, which can be expected in such low numbers.

**Table 12** Percentage of *a* marking of human indefinite direct objects for five verb classes (Corpus de Español and Corde)

Class	15th cent.	17th cent.	19th cent.
1a + 1b EFFECTIVE ACTION: <i>matar, herir, golpear, tirar</i>	18% (9/51)	40% (21/53)	79% (46/58)
2a + 2b PERCEPTION: <i>oir, ver, escuchar, mirar</i>	17% (1/6)	71% (22/31)	93% (27/29)
3 PURSUIT: <i>buscar, esperar</i>	11% (1/9)	23% (8/35)	41% (17/41)
4 KNOWLEDGE: <i>conocer, entender</i>	(0/0)	31% (5/16)	67% (14/21)
5 FEELING: <i>querer, temer</i>	(0/0)	52% (11/21)	75% (15/20)

**Fig. 7** Percentage of *a* marking of human indefinite direct objects depending on verb class and time (Corpus de Español and Corde)

The significance of the corpus results was tested with a generalized linear model. The relevant Wald chi-square tests for the difference between the verb classes shows clear statistical significance for indefinites ( $\chi^2 = 52.609$ ,  $df = 4$ ,  $Sign < 0.01$ ) and for definites ( $\chi^2 = 34.970$ ,  $df = 4$ ,  $Sign < 0.01$ ). For indefinites all verb classes show significant difference to each other, while for definites we find that the five verb classes cluster into two main groups: Class 1 ACTION, Class 2 PURSUIT, Class 5 FEELING > Class 3 PURSUIT, Class 4 KNOWLEDGE.

As in the earlier studies, we see that (i) *a*-marking changes over time and (ii) increases along the Referentiality Scale. Figure 8 compares the development of *a*-marking of definite and indefinite direct objects for *matar* 'to kill' from class 1 and *buscar* 'to search' from class 3 to illustrate a clear difference between two classes.

Finally we observe that Tsunoda's Hierarchy is not totally reflected in the ranking of the percentages of *a*-marking for definite and for indefinite noun phrases. We repeat the verb class hierarchy in (19) and indicate the relative order of *a*-marking for definite objects in (20) and for indefinite objects in (21).

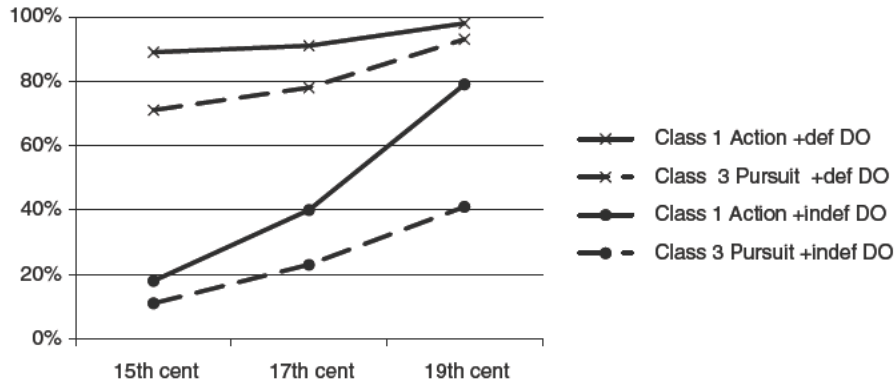


Fig. 8 Percentage of *a* marking of human direct objects depending on definiteness, verb class and time (Corpus de Español and Corde)

- (19) Tsunoda's Hierarchy:  
ACTION > PERCEPTION > PURSUIT > KNOWLEDGE > FEELING
- (20) definite noun phrases:  
PERCEPTION, FEELING, ACTION >> PURSUIT, KNOWLEDGE
- (21) indefinite noun phrases:  
PERCEPTION > FEELING, ACTION > KNOWLEDGE > PURSUIT

The ranking of *a*-marking for definites can be grouped into two main groups: PERCEPTION, FEELING and ACTION versus PURSUIT and KNOWLEDGE, and for indefinites in a similar order. Thus, the ordering of verb classes still does not correspond to the original scale of Tsunoda. It seems that the scale is too general or comprises too many verbs in each class to capture the fine-grained conditions for *a*-marking in Spanish.

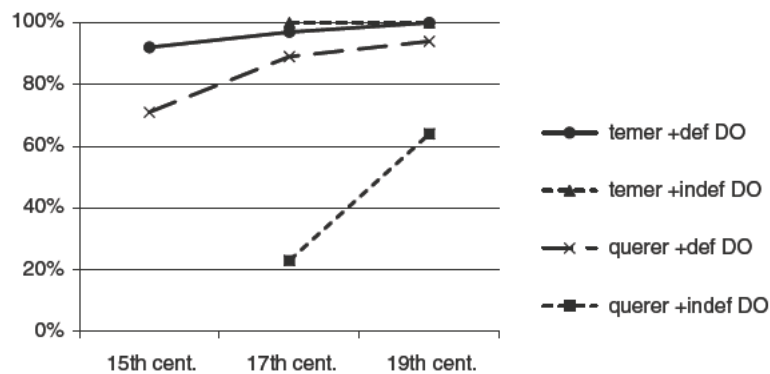
The discrepancy between Tsunoda's Affectedness Scale and our findings are particularly obvious with the class 5 FEELING, which appears at the end of Tsunoda's scale but takes a middle position in our results. Taking a closer look at this class, we can identify an interesting contrast. Following Tsunoda (and others) we have assigned the verbs *querer* 'to like' and *temer* 'to fear' to this class. However, there are clear differences between the two verbs, as summarized in Tables 13 and 14 and compared in Fig. 9.

Table 13 Percentages of *a* marking of human definite direct objects for *querer* and *temer* (Corpus de Español and Corde)

	15th cent.	17th cent.	19th cent.
<i>temer</i>	92% (22/24)	97% (34/35)	100% (38/38)
<i>querer</i>	71% (10/14)	89% (24/27)	94% (33/35)

**Table 14** Percentages of *a* marking of human indefinite direct objects for *querer* and *temer* (Corpus de Español and Corde)

	15th cent.	17th cent.	19th cent.
<i>temer</i>	(0/0)	100% (8/8)	100% (6/6)
<i>querer</i>	(0/0)	23% (3/13)	64% (9/14)



**Fig. 9** Percentages of *a* marking of direct objects for *querer* and *temer* depending on definiteness and time (Corpus de Español and Corde)

The two verbs behave very differently: while *temer* ‘to fear’ has a high preference for *a*-marking already in the 15th century, *querer* ‘to like’ has a low preference and only a slow increase over time. If we take *temer* out of the verb class FEELING, the result, i.e., the verb *querer*, would show lower figures than any other verb class—following the prediction of Tsunoda’s Affectedness Hierarchy. We then have to explain the unexpected behavior of *temer*. As a first attempt we would suggest that the direct object of ‘fear’ has more typical properties of a subject than a prototypical object of ‘like’ (see Kirsner and Thompson 1976). This might be the cause of *temer*’s high scores. This behaviour, however, has nothing to do with affectedness, but rather with the competition between the agentivity of the participants involved in the event. This corresponds to the well-known observation that verbs like ‘fear’ and ‘frighten’ express a different perspective from other transitive verbs. They show a closer interaction between the two participants, since the stimulus causes a change in the agent. This is only possible if the direct object has more properties of a subject than of a direct object. Thus the direct object of verbs like ‘fear’ can also act as a subject in verbs like ‘frighten’. However, the situation of the verb pairs like ‘like’ and ‘please’ is different since the stimulus does not have the same causative effect on the agent (see Croft 1993; Barödal 2001; Delbecque 2002).<sup>3</sup> We see a similar contrast

<sup>3</sup> We are very grateful to one of the reviewers who brought this point to our attention.

**Table 15** Percentages of a marking of human definite direct objects for *oir*, *escuchar*, *ver* and *mirar* (Corpus de Español and Corde)

	15th cent.	17th cent.	19th cent.
<i>escuchar</i>	100% (17/17)	100% (51/51)	98% (84/86)
<i>oir</i>	87% (27/31)	96% (55/57)	96% (65/68)
<i>mirar</i>	75% (40/53)	91% (61/67)	97% (95/98)
<i>ver</i>	76% (39/51)	78% (21/27)	89% (32/36)

**Table 16** Percentages of a marking of human indefinite direct objects for *oir*, *escuchar*, *ver* and *mirar* (Corpus de Español and Corde)

	15th cent.	17th cent.	19th cent.
<i>escuchar</i>	100% (3/3)	100% (6/6)	100% (9/9)
<i>oir</i>	67% (4/6)	81% (13/16)	100% (12/12)
<i>mirar</i>	(0/0)	33% (3/9)	82% (9/11)
<i>ver</i>	(0/0)	8% (2/25)	55% (6/11)

between *oir* ‘hear’ and *escuchar* ‘listen’ on the one hand, and *ver* ‘see’ and *mirar* ‘look at’ at the other. The objects of ‘hear’ must actively produce a noise to be heard, while the object of ‘see’ do not have to be active to be seen. Therefore the objects of ‘hear’ have more properties of a prototypical subject than the objects of ‘hear’, which is also reflected in the distribution of *a*-marking, as listed in Tables 15 and 16.

While Tsunoda classifies these verbs according to  $\pm$ attainment (see above), i.e., *oir* and *ver* versus *escuchar* and *mirar*, the actual preference for *a*-marking of these verbs show that they rather pattern according to the agentivity properties of the direct object assigned by the verb. Note that this is different from the actual agentivity properties of the direct object, which is always high since we only allowed for human direct objects.

It is also not easy to adapt the Partial Affectedness Scale (Malchukov 2005, p. 83), introduced in Sect. 2, to this behaviour. Malchukov (2005) explains affectedness, or more appropriately, transitivity effects by two sub-scales: (i) one scale that ranks verb classes by salient features of the direct object, and (ii) on prototypicality of the subject as a salient, volitional controlling agent. What is not included is that the direct object is attributed by the verb some features that are prototypical for subjects. In order to represent this, we have to develop a relational approach that covers the relation between the properties of (proto-)theta roles of both arguments in the sense of Dowty (1991) and Primus (1999).

## 6 Summary

Differential Object Marking in Spanish depends on the referential properties of the direct object and the lexical semantics of the verb that governs the direct object.



Earlier studies have shown that the preference of the verb for the animacy of the human direct object correlates with the distribution of *a*-marking through time. In this study we investigated whether *a*-marking also correlates with affectedness understood as “the persistent change in an event participant”. For our analysis we used the Affectedness Scale of Tsunoda, which is a generalization of cross-linguistic research on alternations in case frames. We undertook a corpus search for 12 verbs of 5 verb classes and about 2,000 transitive sentences with definite or indefinite human direct objects for three times periods. Our extensive corpus search demonstrated that there is a correlation between verb classes high on the Affectedness Scale and higher frequency of *a*-marking. However, we also detected some mismatches that provide evidence that the another important factor is the competition of agentivity between the participants in the event. This effect is found with other examples as well, and indicates that we need further studies that investigates this additional parameter besides individuation and transitivity, namely the relative ranking of the two arguments.

**Acknowledgements** An earlier version of this paper was presented at the *Workshop on Empirical Approaches to Morphological Case* at the LSA Summer Institute, Stanford in July 2007. We would like to thank the organizers and the audience for their constructive comments. We are especially indebted to Miriam Butt, Andres Enrique, Victoria Escandell Vidal, Sam Featherston, Udo Klein, Brenda Laca, Beth Levin, Manuel Leonetti, Edgar Onea, Elisabeth Stark and Fernando Zuñiga for long discussions and valuable comments. Special thanks go to Johanna Barðdal and Cathryn Donohue for organizing the workshop and editing this volume, and for many very helpful comments and suggestions. Finally we would like to express our gratitude to the three reviewers who gave us very detailed and valuable comments, which have helped us to restructure the paper and make our argument more coherent. The usual disclaimer applies. A substantial part of the corpus research was undertaken by our student assistant Annika Deichsel. The first author acknowledges the support of the German Science Foundation (project C2 “Case and Referential Context” in the SFB 732 “Incremental Specification in Context”) and of the VolkswagenStiftung and the Fritz Thyssen Stiftung (opus magnum).

## References

- Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21, 435–483.
- Anderson, S. R. (1971). On the role of deep structure in semantic interpretation. *Foundations of Language*, 7, 387–396.
- Anderson, M. (1979). *Noun phrase structure*. Ph.D. dissertation, University of Connecticut.
- Anderson, M. (2006). Affectedness. In M. Everaert, H. van Riemsdijk, R. Goedemans, & B. Hollebrandse (Eds.), *The Blackwell companion to syntax* (pp. 121–140). Oxford: Blackwell.
- Barðdal, J. (2001). The perplexity of Dat Nom verbs in Icelandic. *Nordic Journal of Linguistics*, 24, 47–70.
- Beavers, J. (2006). *Argument/oblique alternations and the structure of lexical meaning*. Ph.D. dissertation, Stanford University.
- Beavers, J. (2010). On affectedness. *Natural Language and Linguistic Theory* (to appear).
- Bello, A. (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.
- Bosson, G. (1985). *Empirische Universalienforschung. Differentielle Objektmarkierung in den neu iranischen Sprachen*. Tübingen: Narr.
- Butt, M. (2006). The dative ergative connection. In O. Bonami & P. Cabredo Hofherr (Eds.), *Empirical issues in syntax and semantics 6* (pp. 69–92). <http://www.cssp.cnrs.fr/eiss6>.
- Company, C. (2002). El avance diacrónico de la marcación prepositiva en objetos directos inanimados. In *Actas del II Congreso de la Sociedad Española de Lingüística* (pp. 146–154). Madrid: CSIC.

- Comrie, B. (1975). Definite and animate direct objects: A natural class. *Linguistica Silesiana*, 3, 13-21.
- Comrie, B. (1989). *Language universals and linguistic typology. Syntax and morphology* (2nd ed.). Oxford: Blackwell.
- Croft, W. (1988). Agreement vs. case marking and direct objects. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural language. Approaches, theories, descriptions* (pp. 159-179). Stanford: Center for the Study of Language and Information.
- Croft, W. (1993). Case marking and the semantics of mental verbs. In J. Pustejovsky (Ed.), *Semantics and the lexicon* (pp. 55-72). Dordrecht: Kluwer.
- Croft, W. (1998). Event structure in argument linking. In M. Butt & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 21-63). Stanford: CSLI.
- de Hoop, H., & Narasimhan, B. (2005). Differential case marking in Hindi. In M. Amberber & H. de Hoop (Eds.), *Competition and variation in natural languages: The case for case* (pp. 321-345). Amsterdam: Elsevier.
- Delbecq, N. (1998). Why Spanish has two transitive construction frames. *Leuvense Bijdragen*, 87, 387-415.
- Delbecq, N. (1999). Two transitive construction frames in Spanish: The prepositional and the non-prepositional accusative. In L. De Stadler & C. Eyrich (Eds.), *Issues in cognitive linguistics* (pp. 407-424). Berlin: De Gruyter.
- Delbecq, N. (2002). A construction grammar approach to transitivity in Spanish. In K. Davidse & B. Lamiroy (Eds.), *The nominative & accusative and their counterparts* (pp. 81-130). Amsterdam: Benjamins.
- Delille, K. H. (1970). *Die geschichtliche Entwicklung des präpositionalen Akkusativs im Portugiesischen*. Bonn: Romanisches Seminar.
- de Swart, P. (2007). Cross Linguistic variation in object marking. Ph.D. dissertation, Radboud University Nijmegen, LOT Publication.
- Dowty, D. (1991). Thematic proto roles and argument selection. *Language*, 67, 547-619.
- Enrique Arias, A. (2008). Biblias romanceadas e historia de la lengua. In C. Company & J. Moreno de Alba (Eds.), *Actas del VII Congreso Internacional de Historia de la Lengua Española, Mérida (Yucatán), 4-8 de septiembre de 2006*, (pp. 1-4) 1781-1794 Madrid: Arco/Libros.
- Fernández Ramírez, S. (1951). *Gramática española. Vol. 4: El verbo y la oración (volumen ordenado y completado por Ignacio Bosque)*. Madrid: Arco/Libros, segunda edición 1986.
- Fillmore, C. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in language* (pp. 1-88). New York: Holt, Rinehart & Winston.
- García, M. (2007). Differential object marking with inanimate objects. In G. A. Kaiser & M. Leonetti (Eds.), *Proceedings of the workshop "Definiteness, specificity and animacy in Ibero Romance languages"*. *Arbeitspapier 122. Fachbereich Sprachwissenschaft* (pp. 63-84). Konstanz: Universität Konstanz.
- Givón, T. (1985). Ergative morphology and transitivity gradients in Newari. In F. Plank (Ed.), *Relational Typology* (pp. 89-107). Berlin: Mouton.
- Harris, A., & Campbell, L. (1995). *Historical syntax in cross linguistic perspective*. Cambridge: Cambridge University Press.
- Hopper, P., & Thompson, S. (1980). Transitivity in grammar and discourse. *Language*, 56, 251-299.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jaeggli, O. A. (1986). Passive. *Linguistic Inquiry*, 17, 587-622.
- Kaiser, G. A. (2005). Bibelübersetzungen als Grundlage für empirische Sprachwandeluntersuchungen. In C. D. Pusch, J. Kabatek, & W. Raible (Eds.), *Romanistische Korpuslinguistik II. Romance Corpus Linguistics II. Korpora und diachrone Sprachwissenschaft. Corpora and Diachronic Linguistics* (pp. 71-85). Tübingen: Narr.
- Kirsner, R. S., & Thompson, S. A. (1976). The role of pragmatic inference in semantics: A study of sensory verb complements in English. *Glossa*, 10, 200-240.
- Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In R. Bartsch, J. van Benthem, & P. van Emde Boas (Eds.), *Semantics and contextual expressions* (pp. 75-115). Foris: Dordrecht.
- Laca, B. (2002). Gramaticalización y variabilidad – propiedades inherentes y factores contextuales en la evolución del acusativo preposicional en español. In A. Wesch (Ed.), *Sprachgeschichte als Varietätengeschichte: Beiträge zur diachronen Varietätenlinguistik des Spanischen und anderer romanischer Sprachen. Festschrift für Jens Lüdtke zum 60. Geburtstag* (pp. 195-303). Tübingen: Stauffenburg.

- Laca, B. (2006). El objeto directo. La marcación preposicional. In C. Company (Ed.), *Sintaxis histórica del español. Primera parte: La frase verbal. Volumen 1* (pp. 423-475). Mexico: Fondo de cultura económica/Universidad Nacional Autónoma de México.
- Lehmann, C. (1991). Predicate classes and PARTICIPATION. In H. Seiler & W. Premper (Eds.), *Participation: Das sprachliche Erfassen von Sachverhalten* (pp. 183-239). Tübingen: Narr.
- Leonetti, M. (2004). Specificity and differential object marking in Spanish. *Catalan Journal of Linguistics*, 3, 75-114.
- Malchukov, A. L. (2005). Case pattern splits, verb types and construction competition. In M. Amberber & H. de Hoop (Eds.), *Competition and variation in natural languages: The case for case* (pp. 73-117). Amsterdam: Elsevier.
- Meier, H. (1945). O problema do acusativo preposicional no catalão. *Boletim de Filologia*, 8, 237-260.
- Melis, C. (1995). El objeto directo personal en *El Cantar de Mio Cid*. Estudio sintáctico pragmático. In C. Pensado (Ed.), *El complemento directo preposicional* (pp. 133-163). Madrid: Visor.
- Naess, A. (2004). What markedness marks. The markedness problem with direct objects. *Lingua*, 114, 1186-1212.
- Pensado, C. (1995). El complemento directo preposicional. Estado de la cuestión y bibliografía comentada. In C. Pensado (Ed.), *El complemento directo preposicional* (pp. 11-60). Madrid: Visor.
- Pottier, B. (1968). L'emploi de la préposition *a* devant l'objet en espagnol. *Bulletin de la Société de Linguistique*, 1, 83-95.
- Primus, B. (1999). *Cases and thematic roles*. Tübingen: Niemeyer.
- Puscariu, S. (1937). Au sujet de p(r)e avec l'accusatif. In S. Puscariu (Ed.), *Etudes de linguistique roumaine* (pp. 439-457). Bukarest: Imprimeria national.
- Reichenkron, G. (1951). Das präpositionale Akkusativ Objekt im ältesten Spanisch. *Romanistische Forschungen*, 63, 342-97.
- Roegiest, E. (1979). A propos de l'accusatif prépositionnel dans quelques langues romanes. *Vox Romanica*, 38, 37-54.
- Stimm, H. (1986). Die Markierung des direkten Objektes durch *a* im Unterengadinischen. In G. Holtus & K. Ringger (Eds.), *Raetia antiqua et moderna: W. Theodor Elwert zum 80. Geburtstag* (pp. 407-448). Tübingen: Niemeyer.
- Tenny, C. (1987). *Grammaticalizing aspect and affectedness*. Ph.D. thesis, MIT.
- Torrego, E. (1999). El complemento directo preposicional. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española Vol. 2: Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales* (pp. 1779-1805). Madrid: Espasa Calpe.
- Tsunoda, T. (1985). Remarks on transitivity. *Journal of Linguistics*, 21, 385-396.
- von Heusinger, K. (2008). Verbal semantics and the diachronic development of differential object marking in Spanish. *Probus*, 20, 1-31.
- von Heusinger, K., & Kaiser, G. A. (2007). Differential object marking and the lexical semantics of verbs in Spanish. In G. A. Kaiser & M. Leonetti (Eds.), *Proceedings of the workshop "Definiteness, specificity and animacy in Ibero Romance languages"* (pp. 83-109). Universität Konstanz: Fachbereich Sprachwissenschaft (Arbeitspapier 122).
- Weissenrieder, M. (1991). A functional approach to the accusative 'a'. *Hispania*, 74, 146-156.

## Sources

- Bible A: 14th 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: 16th/17th century: *Reina Valera Antigua* (1569/1602). <http://www.biblegateway.com>
- Bible C: 20th century: *Reina Valera* (1995) (United Bible Societies). <http://www.biblegateway.com>
- Holy Bible: The 21st Century King James Version of the Holy Bible (KJ21®) is an update of the 1611 King James Version (KJV) source: <http://www.biblegateway.com>
- Corpus del Español from the 12th to the 19th century. <http://www.corpusdelespanol.org>
- Real Academia Española Corpus Diacrónico del Español (CORDE). <http://corpus.rae.es/cordenet.html>