



Københavns Universitet



## The Spanish impersonal se-construction

Pedersen, Johan

*Published in:*

Deixis and Pronouns in Romance Languages

*Publication date:*

2013

*Document Version*

Peer reviewed version

*Citation for published version (APA):*

Pedersen, J. (2013). The Spanish impersonal se-construction: Constructional motivation for case-marking - quantitative evidence. In K. Kragh, & J. Lindschow (Eds.), *Deixis and Pronouns in Romance Languages* (pp. 109-123). Amsterdam: John Benjamins Publishing Company.

## The Spanish impersonal *se*-construction

### Constructional motivation for case-marking - quantitative evidence

#### Abstract

According to the new edition of the Spanish Academy Grammar (Bosque 2010), it remains unexplained why the clitic object in the Spanish impersonal transitive *se*-construction tends to be in dative (*le*): *se le ve* 'you can see him'. The aim of this paper is to analyze this usage quantitatively, with focus on its constructional motivation. To achieve solid empirical evidence, I extracted large amounts of data from Corpus del Español (20.4 mill. words) and analyzed the data statistically. I found that the constructional motivation for the dative is substantial and that it is due to characteristics of the specific impersonal *se*-construction.

#### 1. Introduction

In Spanish, only pronouns show *per se* case distinctions. The original pronominal case-system is outlined in Table 1.<sup>1</sup> Forms that are relevant for this study are in bold, that is, the clitic forms in accusative (*lo/la/los/las*) and the clitic forms in dative (*le/les*):

---

<sup>1</sup> Frequently, Spanish linguists make mention of the etymological use with reference to the use of the forms derived from Latin: *lo(s)*, *la(s)*, *lo* (accusative) and *le(s)* (dative) for direct and indirect object respectively. See e.g. Cuervo (1895); Fernández-Ordóñez (1993, 1999); Fernández Ramírez (1964); García (1975, 1986, 1990, 1992); Klein-Andreu (1981, 1998); Lapesa (1968); Roldán (1975). In this paper I use the notion *pronominal case-system* with reference to this etymological use.

	Unbound forms		Clitics	
	Nominative	Prep. case	Accusative	Dative
Singular:				
1.p.	yo	mí	me	me
2.p.	tú	ti	te	te
3.p.	él/ella/ello	él/ella	<b>lo/la/lo</b>	<b>le</b>
Plural:				
1.p.	nosotros/as	nosotros/as	nos	nos
2.p.	vosotros/as	vosotros/as	os	os
3.p.	ellos/as	ellos/as	<b>los/las</b>	<b>les</b>

Table 1 The pronominal case system in Spanish.

The aim of this paper is to offer a quantitative analysis of patterns of case marking in the Spanish impersonal *se*-construction. The prototypical personal construction and the corresponding impersonal *se*-construction are exemplified in (1) and (2):<sup>2</sup>

- (1) Juan (**lo**) ve (al niño)  
 Juan ACC see.3s ACC the boy  
 'Juan is seeing (him) (the boy)'

- (2) se (**le**) ve (al niño)  
 MM DATsee.3s ACC the boy  
 'you can see (him) (the boy)'

The impersonal *se*-construction (2) has no subject, instead there is a middle marker (MM = *se*), in which the impersonal meaning originates, that is, a reference to an unspecified subject role. It is

<sup>2</sup> I analyze *se* as a middle marker (MM) following Kemmer (1993) and Maldonado (1992).

*impersonal* in the sense that the agent is only implicitly present. In contrast, (1) is *personal* since the agent is explicitly constructed as a grammatical subject. In both constructions, the verb is transitive and it has a direct object. In (2) the unbound direct object is obligatorily marked by the preposition *a*, whereas in (1) only human, or animate, definite objects are *a*-marked (cf.... *al niño*). Inanimate objects are typically not *a*-marked.<sup>3</sup> The direct object may be replaced, or doubled, by a clitic object. When comparing the two constructions, it is particularly interesting to observe that in old Spanish (and still present in modern Spanish) a clitic object in the impersonal *se*-construction had to be in dative case (*le(s)*), which appears from example (2) and (3):

- (3) *se le encerr-ó en Monzón junto a Palencia ...* (Corde, Jerónimo Zurita, 1562)  
 MM DAT jail-pst.3sg in Monzón next to Palencia  
 ‘they jailed him in Monzón next to Palencia’

This was (and is) not so in the prototypical personal construction, as shown in example (1) and (4):

- (4) *Leones lo empuxa-ron; y el primero lo com-io* (CORDE, anónimo, 1553)  
 Lions ACC push-pst.3pl and the first ACC eat-pst.3sg  
 ‘lions pushed him; and the first ... ate him’

The paper is structured as follows: In the next section (Section 2), some influential proposals that have specifically addressed the question of case marking in this construction will briefly be discussed. This will lead to a specification of the research questions in this paper (Section 3).

---

<sup>3</sup> In some languages, e.g. Spanish, case marking of the NP as a direct object is due to more specific conditions for the NP, structured by the dimensions of animacy and definiteness. This phenomenon has been termed Differential Object Marking (see e.g. Aissen 2003; Bossong 1985, 1991, 1998; Comrie 1989; Haspelmath 2001).

Sections 4 and 5 are outlines of the theoretical framework and the applied methodology. In Section 6 and 7, the results will be presented and discussed, and finally a conclusion will be drawn.

## 2. Previous studies

The impersonal *se*-constructions have been thoroughly studied in Spanish linguistics (see e.g. Maldonado 1992, 2000; Martín Zorraquino 1979; Mendikoetxea 1999; Pedersen 2005; Santiago 1975; and references cited there). Traditionally Spanish grammarians have maintained that the noun phrase in impersonal *se*-constructions was not a direct object, but an indirect object (e.g. Bello 1847). This claim supposedly explains the use of the dative straightforwardly. However, the analysis of the noun phrase as an indirect object primarily relies on the distributional condition that the NP could be substituted by a clitic in dative and is therefore implausible due to the circularity of the argument. It has also been suggested by Bello (1847), and by those grammarians who followed his original proposals, that the impersonal *se*-construction was closely related to and derived from the ditransitive variant of the passive *se*-construction as exemplified in (5):<sup>4</sup>

(5) se le da admiración (a Juan) passive with indirect object →  
MM DAT give.prs.3sg admiration (to Juan)  
'he is given admiration'

(5') se le admir-a Impersonal construction  
MM DAT admire-prs.3sg

---

<sup>4</sup> See e.g. Bello 1847, Martín Zorraquino 1979, Mendikoetxea 1992.

‘he is admired/someone admires him’ (Bello 1847)

The argument is, without going into detail, that the pronominal variant of the ditransitive *se*-passive, cf. (5), develops into an impersonal *se*-construction, whose NP thereby is given indirect object status. This proposal is problematic for the simple reason that the prototypical impersonal *se*-construction implies transitivity, cf. (2), and that the noun phrase therefore must be assigned direct object status.

Various scholars have claimed that the dative clitic is preferred over the accusative to avoid ambiguity (e.g. Fernández Ramírez 1964; Santiago 1975). The argument goes as follows: if the accusative were grammatical in the impersonal *se*-construction, example (6) would have two interpretations:

(6) se lo quita

It would have a personal interpretation:

(6) a. (él) se lo quit-a  
he DAT ACC remove-prs.3sg  
‘he removes it from himself/another person’

and an impersonal interpretation:

(6) b. se lo quit-a  
MM ACC remove.prs.3sg

'someone (unspecified agent) removes him'

The ambiguous interpretation is due to the fact that the *se*-marker is also used as a dative when it is followed by another 3. person clitic. Defenders of this point of view argue that the ambiguity is avoided by having dative *le* as an object marker. Against this position, it might be objected that this argument rests on the questionable assumption that language is basically unambiguous (see e.g. Newmeyer 2003).

Another line of argument is that internal constructional motivation may be decisive for case marking. The basic idea is that the degree of transitivity (see e.g. Hopper & Thompson 1980; Langacker 1991) to some extent must affect the choice of case. In the impersonal *se*-construction, as compared with the prototypical personal construction, the implicitly construed agent implies a transitivity reduction (Maldonado 1992; Pedersen 2005). Transitivity is conceived of as reduced in users' grammar, because the existence of a merely schematic, and implicitly construed, agent implies that the object participant is no longer directly affected by a specific subject participant. Moreover, the object was originally always a human, or animate, participant in the impersonal *se*-construction – in its unbound (non-clitic) form iconically marked by the DOM-marker *a*; see ex (2). It was therefore a marked object that clearly resembled, and in its unbound form was identical with, the prototypical indirect object (e.g. Aissen 2003). Now, basically, the idea is that there are structural motivations for the use of the dative: Transitivity reduction along with object markedness and similarity with the indirect object motivated the original choice of the dative, instead of the accusative, from users' paradigm of pronominal clitics, which only included these two options.

There is a strong tendency in the literature to highlight an extensive variation in form. According to many linguists (e.g. Alarcos Llorach 1994; Mendikoetxea 1999; Santiago 1975), the question of case marking in impersonal *se*-constructions is closely related to the existence of dialectal case

marking patterns that diverge from the original pronominal case-system, cf. Table 1. Particularly, it is often claimed that the phenomenon in question is a matter of *leísmo*, according to which the dative *le(s)* is used systematically as a marker of animate objects. However, it is a fact that the dative has been the preferred form in these constructions in various regions of *non-leísmo* as well (Bosque 2009). Such evidence is a serious problem for explanations that exclusively focus on this specific kind of variation. In addition, given that *leísmo* is a factor to be taken into account, we should equally take into account that accusative marking of the direct object in prototypical personal constructions, cf. (1), motivates the use of the accusative in the impersonal *se*-construction. This may be an even stronger motivational factor for selecting the accusative case than *leísmo* is for selecting the dative, particularly because occurrences of the construction with inanimate object are relatively frequent in modern Spanish, and by no means ungrammatical.

Not only diatopic variation can be observed. Also elements of diaphasic variation, including personal preferences, seem to play a role, as exemplified in (7) and (8):

- (7) Es verdad que yo **lo** (ACC) había encontrado tan triste, pero me quedé muy sorprendida  
 ‘It’s true that I found **him** sad, but I was surprised’  
 cuando fui a buscar**le** (DAT),..., hasta a veces **le** (DAT) he oído gritar (Matute 1969: 119).  
 when I went to look for **him**,..., sometimes I have even heard **him** shouting...’
- (8) a esa persona se **le** (DAT) saca del grupo, se **la** (ACC) advierte; si hay niños, por  
 ‘They take **this person** out of the group, they warn **him/her**; if there are children, for  
 ejemplo, pues se **les** (DAT) advierte también (Corpus del Español (CE)).  
 ‘instance, they will also be warned’.

In (7), the author is switching from etymological case marking (*lo* for direct object) to *leísmo* style (*le* for direct object) within the same text sequence of prototypical personal clauses.<sup>5</sup> The same kind of case switching is being practiced in (8), in a syntactic environment of impersonal *se*-constructions. Notice that the two variants [*se le(s) V*] versus [*se lo(s)/la(s) V*] are basically associated with the same meaning.

To sum up, according to the new edition of the Spanish Academy Grammar (Bosque 2009), the dative case marking in the impersonal *se*-construction remains a controversial topic. Studies of the internal structure of the construction and its influence on case marking are not conclusive, and they are not based on quantitative evidence. Moreover, dative case marking cannot be explained exclusively by purely formal variation (*leísmo*), and it cannot be explained by functional differences either; the use of the dative or the accusative does not imply a different reading of argument structure.

### 3. Research questions

The aim of this study is to assess case marking in the impersonal *se*-construction from a quantitative perspective, with focus on its possible structural motivation. I will determine the constructional motivation for case marking that can be ascribed to the characteristic *se*-marker (MM) taking into account the conventionalized use in transitive constructions that are inherited by the impersonal *se*-construction. I will analyze case marking patterns with particular focus on quantification: Is the observed tendency to use the dative supported by large corpus data? Is the constructional motivation

---

<sup>5</sup> See note 1 for a clarification of the notion *etymological case*.

substantial, or is it only marginal? It is not the objective of this study to testify the usage of specific dialects, or other user segments.

#### **4. Theoretical framework**

This study has been carried out within the framework of construction grammar (e.g. Croft 2001; Goldberg 2006), which is a usage-based approach. In usage-based approaches to grammar, grammatical structure emerges from language use in the sense that linguistic units are seen as being abstracted from usage events (e.g. Langacker 1987). Grammatical development and change is thus grounded on abstractions made upon actual usage events. The fundamental implication of the usage-based model is that the existence of constructions (pairings of form and meaning) in grammatical representation is a function of frequency and similarity in form and meaning. Constructions form a network and are linked by inheritance relations, which motivate many of the properties of particular constructions. Inheritance allows us to capture generalizations across constructions and particularly the fact that two constructions may be in some ways the same and in other ways distinct (Goldberg 1995: 72).

My principal concern is to argue that according to Construction Grammar, the syntactic-semantic context conditions the status of grammatical relations, e.g. the object relation, and hence it is a motivational environment for case-marking (Pedersen 2005). Grammatical categories and relations are construction specific and undergo constant abstraction and (re)analysis by the users.

As language users and language learners we face the task of categorizing utterances into construction types. There are clearly discontinuities between the prototypical personal constructions, and the impersonal *se*-construction, see (1) and (2). The difference is basically due to

the middle marker *se*, the impersonal meaning and in particular the lack of an explicitly encoded agent in the impersonal *se*-construction. It is therefore plausible to assume that (1) and (2) are categorized and treated as different construction types in users' grammar even though some properties clearly are inherited, e.g. the transitive structure: in both constructions there is a DOM (= *a*) marked NP and/or a coreferential pronominal clitic case marker with (construction specific) direct object status.

A clausal expression very often contains in its internal structure a range of different construction types. In the internal structure of the impersonal *se*-construction, the basic skeletal construction is an inherited transitive pattern. This transitive pattern, itself, transmits the influence from similar transitive construction types, including structural priming effects (Gries 2005). What makes the assessment of case marking complicated, particularly, is that its motivation depends on different conventions about case marking (accusative or dative) in these basic transitive constructions.

## 5. Methodology

The empiric goal is to quantify case-marking in the impersonal *se*-construction by means of a distributional analysis of corpus data. The data sample was extracted from searches in Corpus del Español, which is a large monolingual corpus available on the Internet.<sup>6</sup> Corpus del Español is an annotated corpus, tagged for lemma and parts of speech. It is therefore a suitable data source for our purpose, which is to conduct a quantitative corpus study. The corpus consists of around 100 million words in more than 20,000 Spanish texts from the 12th to the 20th centuries. The present study is concerned only with modern Spanish usage. Searches therefore were carried out exclusively in texts

---

<sup>6</sup> Corpus del Español is available on this web-page: <http://www.corpusdelespanol.org/x.asp>

from the 20th century (approx. 20.4 mill. words). Regarding the composition of the corpus, it contains oral as well as written language (interviews and transcripts, newspaper and magazine texts, fiction and academic texts) and Iberian as well as Latinamerican Spanish is represented.

I decided not to look at the role of the *a*-marker in the impersonal *se*-construction, mainly because we already know from other studies (e.g. Aissen 2003) that it is a motivational factor for the use of the dative, but also because I wanted to focus on the role of the *se*-marker *per se*. This focus was implemented by restricting the study to include only the most simple variant of the construction: [*se* case-marker V] / ‘impersonal transitive relation’, in which there is a case marked pronominal clitic object, but no coreferential DOM (=a)-marked nounphrase constituent:

- (9) *se le ve*  
MM DAT see-prs.3sg  
‘you (unspecified agent) can see him’

The search-strings for the corpus study were consequently: [*se + lo/los/la/las/le/les + V*]. *Se*-constructions is a very complex area of Spanish linguistics and one may easily confuse the specific construction studied in this paper with other *se*-constructions, for instance, reflexive *se*-constructions. It is therefore necessary to go through and analyze all the extracted data manually to be sure that in each case we have got the right *se*-construction.

The principal methodological challenge is how to assess the motivation for case marking quantitatively as it is determined by various factors. On the one hand, it seems to be constructionally motivated by the characteristic impersonal *se*-marker. On the other hand, it is reasonable to assume that it is influenced by case marking patterns in prototypical constructions of transitivity since the impersonal *se*-construction has adopted the transitive feature by means of

constructional inheritance; see Section 2. In that respect, notice that the data sample does not consist exclusively of impersonal *se*-constructions with an animate direct object. In fact, many instances of the construction with an inanimate direct object can be observed, see, e.g., examples (6), (12), (18), (19), (23), (26), (27), (28) and (30) in Table 3. This is an important observation because it implies that the assumed influence from the prototypical construction of transitivity is not restricted to constructions with an animate direct object, but includes constructions with all kinds of objects.

Firstly, the distributional analysis should provide a measure of case that takes into account a possible conventional bias in the corpus, i.e. predominance of regional variation that prescribes the use of either the accusative (*lo(s)/la(s)*) or the dative (*le(s)*) for direct object in transitive constructions. Using methodology from collocational analysis (Biber 1993), we will therefore measure the association in the corpus of clitic case-marking (dative or accusative) with the impersonal *se*-construction, taking into account its general distribution in the corpus. Occurrences in other constructions may include prototypical transitive constructions whose case marking patterns may interfere via inheritance in the impersonal *se*-construction, but also other construction types whose case marking patterns may interfere more indirectly, for instance via structural priming. The basic methodological idea is that if a specific case is relatively low-frequent in general in case-marked (transitive) argument structure constructions, it is also, statistically, expected to be relatively low-frequent in one specific transitive argument structure construction. If, however, the corpus data show a higher frequency than the expected one, this is an indication of a relatively strong association of this case-marker with the specific constructional environment. In addition, the measure of association should be an absolute value in order to be directly comparable, for instance, if we in a future study want to see how case-marking in this construction has changed over time.

Like other measures of collocation strength – typically between two words - the Fisher exact test can be performed on a two-by-two table, see Table 2, representing the single and joint frequencies

of, in our case, the case marker and the impersonal *se*-construction in the corpus. To calculate the association strength of the case marker for the impersonal *se*-construction, we need four frequencies: the frequency of dative case in the construction (A), the frequency of dative case in all other constructions (C), the frequency of the construction with the alternative case, the accusative (B), and the frequency of all other constructions with the accusative (D). These frequencies can then be entered in a 4-by-4 table and submitted to the Fisher exact test.

	Dative case	Accusative case
Impersonal <i>se</i> -construction	<b>A</b>	<b>B</b>
Other case-marked constructions	<b>C (= E-A)</b>	<b>D (= F-B)</b>
All case-marked constructions	E	F

Table 2 Input for Fischer exact test

The frequencies of all dative-marked (E) and all accusative-marked (F) argument structure constructions are extracted from the annotated corpus. The frequencies of case marked “other constructions” (C and D respectively) are then calculated by subtraction. The use of Fisher exact tests is based on the following rationale: given the frequencies in the table (A, B, C, D) and assuming that a specific case option is not associated with the impersonal *se*-construction ( $H_0$  = the null hypothesis of independence), how likely is it to have frequencies as the ones that were observed, or more extreme?  $H_0$  (the null hypothesis of independence) should be understood as the assumption that the general usage patterns for pronominal case (the frequency of accusative versus dative) in the corpus, is reflected in the use of case-marking in the impersonal *se*-construction; and that case-marking in this construction is not motivated by characteristics of this specific construction, but rather by general case-marking patterns in transitive constructions, and related constructions. Thus,  $H_0$  captures the idea that case marking in the impersonal *se*-constructions

simply is reflecting the dialectal conventions for case marking in transitive constructions, that is, the use of the accusative, or the dative, for direct object, see Section 2.

The Fisher exact test computes, thus, the exact probability ( $p$ ) under the null hypothesis of obtaining the current distribution of frequencies, or one that is more extreme. The  $p$ -level may be chosen to be, e.g.,  $p < 0.01$  or  $0.001$ , which indicates a positive relative association between the case marker and the impersonal *se*-construction for  $p$ -values lower than  $0.01$  or  $0.001$ , thus rejecting the null hypothesis. The lower the  $p$ -value, the stronger the association of case-marking with the impersonal *se*-construction – either in direction of dative case or in direction of accusative case.

## 6. Results

Table 3 is an arbitrary extraction of examples from the data sample that shows dative and accusative case marking in various forms of gender and number:

1	CR:PrLibre:98Jun1	<b>se le llamó</b> dos veces más, pero no respondió.
2	CR:PrLibre:98Jun27	cuando <b>se le consultó</b> sobre las razones que le motivaron a hablar de intenciones de expansionismo de parte
3	Cien años de soledad	En los días siguientes <b>se le vio</b> con una malla y una canastilla cazando mariposas en los alrededores del pueblo.
4	Habla Culta: Caracas: M7	, y por otra parte, <b>se le ha adiestrado</b> , entiendo yo, en una forma bastante aceptable también
5	Habla Culta: Madrid: M13	hay un profesor que es el que me ha dicho que usted, que <b>se le ha nombrado</b> tres veces,
6	Habla Culta: Sevilla: M6	Pero también a Sevilla se le ha hecho como una especie de mito, <b>se le ha querido encasillar</b> ,
7	Habla Culta: San Juan (PR):...	solamente quizás <b>se le ha preparado</b> para bregar con tubos de ensayos, o para analizar obras de teatro
8	España Oral: ADEP017C	, no <b>se le ha entendido</b> - Tiene que salir Muiños. Vuelve a organizar su ataque...
9	Entrevista (ABC)	¿ Quiere decir que <b>se les ayuda</b> activamente?
10	Habla Culta: Santiago: M41	así que jamás <b>se les entiende</b> bien; yo no sé si será el micrófono o el parlante,
11	España:ABC	<b>Se les oye</b> en todas partes.
12	España Oral: CDEB033A	el espionaje electrónico y espacial no <b>se le podía mitificar</b> de la forma que <b>se le ha mitificado</b> porque en la Guerra del Golfo se ha descubierto que ...

13	España:ABC	<b>se les obliga</b> al trabajo en equipo.
14	España:ABC	si <b>se les ayuda</b> bajando el tipo de interés y facilitando el crédito
15	Habla Culta: Buenos Aires: ...	<b>se lo entretenía</b> , se enriquecía su imaginación,
16	Habla Culta: Buenos Aires: ...	hay corrales donde <b>se lo alimenta</b> durante varios días
17	Habla Culta: Buenos Aires: ...	un novillo desde que <b>se lo compra</b> hasta que llega a Chile pueden pasar diez días,
18	Habla Culta: Havana: M49	¿ Cómo <b>se lo baila</b> ?
19	Habla Culta: La Paz: M24	... <b>se lo estudia</b> dentro de la psicología general,
20	Las Vértebras de Pan	<b>se lo bañó</b> con agua traída del Ganges sagrado.
21	El destino, el barro y...	y entonces ya no <b>se lo ayudó</b> ,
22	Habla Culta: Buenos Aires: ...	<b>Se los lleva</b> en camioneta hasta... Mendoza.
23	Enc: Cachalote	escaso en las costas cantábricas y atlánticas de la península Ibérica, y nunca <b>se le ha visto</b> en el Mediterráneo. Es frecuente en Canarias.
24	Habla Culta: Havana: M15	¿ cómo <b>se los llama</b> ? Inf.
25	Habla Culta: La Paz: M1	no <b>se los castiga, se los lleva</b> a un hogar,
26	Habla Culta: La Paz: M10	manteniendo estos valores culturales que <b>se los puede tipificar</b>
27	Habla Culta: La Paz: M29	Pero que esa fórmula <b>se la ha</b> conocido antes del voto.
28	Habla Culta: Madrid: M1	creo que no es la situación de angustia que algunas veces <b>se la ha</b> calificado,
29	Habla Culta: La Paz: M21	A esas masas no <b>se las ha educado. Se las ha explotado.</b>
30	Habla Culta: La Paz: M26	pero esas cosas, ya he dicho, <b>se las acepta.</b>

Table 3 Extraction of data

The results of the quantitative inquiry are presented in Table 4:

	Dative case	Accusative case
Impersonal <i>se</i> -construction	<b>747</b>	<b>366</b>
Other case-marked constructions	<b>75,188</b>	<b>113,961</b>
All case-marked constructions	75,935	114,327

Table 4 Distribution of case in [*se* case-marker V] / ‘impersonal transitive relation’

A right-tailed Fischer exact test returns a *p*-value that indicates a very strong association of dative case to the impersonal *se*-construction in the corpus:  $p = 2.16e-75$  ( $p_{\log_{10}} = 74.66$ ), e.g. at the critic

level of  $p < 0.001$ .<sup>7</sup> Taken into account the inherited transitive pattern that implies transfer of conventionalized case patterns for direct object, and pressure from other case marked argument structure constructions, the dative is strongly associated to the impersonal *se*-construction in Corpus del Español.

## 7. Discussion

Based on the distributional analysis, the dative case seems to be strongly motivated by the impersonal *se*-marker. We have to be careful, though, about how we interpret the calculated  $p$ -value for association strength. The corpus is relatively large and it facilitates the calculation of a robust indication of association strength. However, it should, by no means, be taken as a precise measure. It seems plausible to assume that conventions about case marking in transitive constructions will influence via inheritance case marking patterns in the impersonal *se*-construction. But we cannot know to what extent the availability of other case marked argument structure constructions will interfere. For instance, constructions with dative for indirect object may, via structural priming, put pressure on and promote the use of the dative in the impersonal *se*-construction, but we cannot know more precisely how strong such an effect may be on case marking in the impersonal *se*-construction. Nevertheless, what we can say is that the calculated  $p$ -value ( $=2.16e-75$ ) is a “conservative” measure in the sense that it takes into account all dative marked argument structure constructions, even though some of them may have no effect at all on case marking in the impersonal *se*-construction – and with a lower count of the dative we would get an even more significant  $p$ -value (closer to zero). We may say that the calculation of the association

---

<sup>7</sup>  $E-75 = 1/10^{75}$ , which is very close to zero. The  $p_{\log10}$ -value is a transformation of the  $p$ -value into a positive number that may be a better indicator of high association strength (Stefanowitsch & Gries 2005).

strength is an attempt to capture some reverse pressures and motivational factors for case in the chosen corpus as a very rough estimate. We should also remember that the conclusions are based on a distributional analysis of data extracted from one specific corpus with a specific composition that may not be a perfect reflection of the Spanish language. Moreover, the substantial regional variation – with respect to the use of the accusative versus dative for direct object (+/- *leísmo*) – whose effects on the overall picture of case in the impersonal *se*-construction are taken into account quantitatively by the chosen methodology, has not been brought into focus *per se* in this study. We should also take into consideration the possibility that transitive verbs may have expected or unexpected preferences for case. Among the verbs that occur in the data of this study, none of them are so called *dative governing verbs*. Among the most frequent ones, such as *llamar*, *conocer*, *considerar*, *ver*, *denominar*, *comparar*, *designar*, *calificar*, only the first three of them: *llamar*, *conocer* and *considerer* show a (slightly) more extreme selection of case (the dative-accusative ratio is 78/16, 64/19, 47/17 respectively) when compared to the general proportion in the corpus, see Table 4. It is simply out of the question that this kind of potential verb-biased distribution should explain the overall distribution of case in the corpus sample.

The measurement of relative association strength in absolute (*p*) values facilitates direct comparisons of case marking patterns over time, taking into account changes of regional convention as reflected in diachronic corpora.

## 8. Conclusion

According to the usage-based model, frequency-effects will lead to the expectation that the prototypical transitive construction, in which the accusative as well as the dative is widely used,

may function as a model for clitic case marking in the impersonal *se*-construction. Nevertheless, the original and sustained use of the dative in this construction suggests that case is not only constrained by case conventions in other (inherited) more basic constructions, but also by the specific syntactic-semantic environment of the impersonal *se*-construction. In this study, we have observed a strong motivation for the use of dative case in the impersonal *se*-construction, in spite of the interference of different conventions for case-marking in transitive constructions. The kind of measurements developed in this study facilitates a method to assess how case marking has changed in argument structure constructions.

## References

- Alarcos Llorach, Emilio. 1994. *Gramática de la Lengua Española*, Real Academia Española. Madrid: Espasa Calpe.
- Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory* 21: 435-483.
- Bello, Andrés. 1847. *Gramática de la lengua castellana*, con notas de Rufino José Cuervo & Niceto Alcalá-Zamora. Buenos Aires: Sopena.
- Biber, Douglas. 1993. Co-occurrence patterns among collocations: A tool for corpus-based lexical knowledge acquisition. *Computational Linguistics*, 19(3): 531–538.
- Bosque, Ignacio. 2009. *Nueva Gramática del Español*. Real Academia Española. RAE. Real Academia Española (RAE). Madrid.
- Bossong, Georg. 1985. *Differentielle Objectmarkierung in den Neuiranischen Sprachen*. Tübingen: Gunter Narr Verlag.

- Bossong, Georg. 1991. Differential Object Marking in Romance and Beyond. In *New Analyses in Romance Linguistics: Selected Papers from the XVIII Linguistic Symposium on Romance Languages*, Urbana-Champaign, April 7-9, 1988, Dieter Wanner & Douglas A. Kibbee (eds), 143-170. Amsterdam: John Benjamins.
- Bossong, Georg. 1998. Le marquage différentiel de l'objet dans les langues d'Europe. In *Actance et valence*. EALT EUROTYP 20-2, Jack Feuillet (ed), 193-258. Berlin: Mouton de Gruyter.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.
- CORDE: Internet-based diachronic database: <http://www.rae.es/>. The Royal Spanish Academy (RAE).
- CE (Corpus del Español). <http://www.corpusdelespanol.org/x.asp>
- Croft, Willam. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
- Cuervo, Rufino José. 1895. Los casos enclíticos y proclíticos del pronombre de tercera persona en castellano, *Romania*, 24: 95-113 and 219-263.
- Fernández-Ordóñez, Inés. 1993. Leísmo, laísmo y loísmo: estado de la cuestión. In *Los pronombres átonos*, Olga Fernández Soriano (ed), 63-96. Madrid: Taurus.
- Fernández-Ordóñez, Inés. 1999. Leísmo, laísmo y loísmo. In *Gramática Descriptiva de la Lengua Española*, Ignacio Bosque & Violeta Demonte (eds), 1319-1390.
- Fernández Ramírez, S. 1964. *Un proceso lingüístico en marcha: Presente y futuro de la lengua española II*, 277-285. Madrid: Ofines.
- García, Erica. 1975. *The role of theory in Linguistic Analysis: The Spanish Pronoun System*, North-Holland Linguistic Series, 19. Amsterdam: North-Holland Publishing Company.
- García, Erica. 1986. The case of Spanish Gender: Referential Strategies in Language Change. *Neophilologische Mitteilungen* (87): 165-184.

- García, Erica. 1990. Bilingüismo e interferencia sintáctica. *Lexis* 14(2): 151-195.
- García, Erica. 1992. Sincronización y desfase del leísmo y laísmo, *Neuphilologische Mitteilungen. Bulletin de la Société Néophilologique/Bulletin of the Modern Language Society* 93: 235-256.
- Goldberg, Adele. 1995. *Constructions*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.
- Gries, Stefan. 2005. Syntactic Priming: A Corpus-based Approach. *Journal of Psycholinguistic Research* 34(4): 365-399.
- Haspelmath, Martin. 2001. Non-canonical marking of core arguments in European languages. In *Non-canonical marking of subjects and objects*, Alexandra Y. Aikhenvald, Robert M. W. Dixon & Masayuki Onishi (eds), 53-83. Amsterdam: John Benjamins.
- Hopper, Paul. J. & Sandra. A. Thompson (1980). Transitivity in Grammar and Discourse. *Language* 56, 251-99.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam: John Benjamins.
- Klein-Andreu, Flora. 1981. Distintos sistemas de empleo de *le*, *la*, *lo*: Perspective sincrónica, diacrónica y sociolingüística, *Thesaurus. Bolitín del Instituto Caro y Cuervo* 36, 284-304. Bogotá.
- Klein-Andreu, Flora. 1998. *Le/s, la/s, lo/s: Variación actual y evolución histórica*. Munich: Lincom Europa.
- Langacker, Ronald W. 1987/91. *Foundations of Cognitive Grammar*, Vol. I + II. Stanford, CA: Stanford University Press.
- Lapesa, Rafael. 1968. Sobre los orígenes y evolución del leísmo, laísmo y loísmo. In *Festschrift Walther von Wartburg*, Kurt Baldinger (ed), 523-551. Tübingen: Max Niemeyer.
- Maldonado, Ricardo. 1992. *Middle Voice: the case of Spanish 'se'*. Phd.-dissertation, University of

California, San Diego. UMI, Ann Arbor.

Maldonado, Ricardo. 2000. Conceptual distance and transitivity increase in Spanish reflexives. In *Reflexives: forms and functions*, Zygmunt Frajzyngier & Traci S. Curl (eds), 167-185. Amsterdam: John Benjamins.

Matute, Ana María. 1969. *Paulina*. Barcelona: Lumen.

Martín Zorraquino, María Antonia. 1979. *Las construcciones pronominales en español*. Madrid: Gredos.

Mendikoetxea, Amaya. 1992. *On the nature of Agreement: The Syntax of ARB SE in Spanish*, tesis doctoral, Universidad de York, Great Britain.

Mendikoetxea, Amaya. 1999. Construcciones con *se*: medias, pasivas e impersonales. In *Gramática Descriptiva de la Lengua Española*, Ignacio Bosque & Violeta Demonte (eds), ch. 26. Real Academia Española, Colección Nebrija y Bello, Espasa.

Newmeyer, Frederick. J. 2003. Grammar is grammar and usage is usage. *Language* 79: 4, 682-707.

Pedersen, Johan. 2005. The Spanish impersonal *se*-construction. Constructional variation and change. *Constructions* 2005(1): 1-49.

Roldán, Mercedes. 1975. The great Spanish *le-lo* controversy, *Linguistics* 147: 15-30.

Santiago, Ramón. 1975. Impersonal *se le(s)*, *se lo(s)*, *se la(s)*, *BRAE LV*: 83-107.