Mark Davies (Provo)

On diachronic shifts with Spanish *se*: preliminary evidence from large electronic corpora

Cet article présente le développement historique du pronom espagnol [*se*] d'après le *Corpus del Español* (100 millions de mots). Nos données préliminaires traitent toutes les structures principales (l'inchoatif, l'impersonnel, l'inaccusatif, etc.) et se basent sur le modèle *n-gram*. Grâce à l'architecture connexe d'une base de données, on obtient facilement des chiffres de fréquence et de cooccurrence sur des milliers de formes verbales et leurs pronoms associés, sans devoir les énumérer au préalable. L'examen des changements de ces formes par tranches historiques nous permet de discerner et de comparer les trajectoires historiques du mot [*se*] depuis l'ancien espagnol jusqu'à nos jours.

1. Introduction

One of the most studied aspects of pan-Romance and Spanish syntax are the constructions with pronominal verbs that take [*se*] (*mirarse*, *acostarse*, *sorprenderse*, *irse*, *jactarse*, *venderse*, *hundirse*, *bautizarse*, etc). Bibliographies such as that of the *Modern Language Association* show more than one hundred articles on [*se*] during the past ten to fifteen years (cf. also Martins, in this volume). Yet the one aspect of [*se*] in Spanish that has been relatively neglected during this time – and in fact has never received much attention – is its historical development.

It is true that there are several studies of historical Spanish [*se*], but nearly all of these suffer from certain shortcomings. First, the most data-oriented studies were written more than twenty years ago, and many of them even farther back than that, e.g. Hanssen (1912), Monge (1954), Brown (1928), Karde (1943), Mendeloff (1964), and Hernández Alonso (1966). As a result, these studies – which were written long before there were any large electronic corpora of historical Spanish – deal with the use of [*se*] in just a few texts, or a small number of instances, or a very narrow historical period.

Another problem is that most of the recent studies deal with [*se*] in just a few of its uses – passive, impersonal, or "focusing" [*se*] (Martín Zorraquino 1979; Sepúlveda Barrios 1988; Maldonado 1989; Melis 1995). Many of these are quite valuable in their own right, but still need to be integrated into an overall model of [*se*] during the past 800 years. Perhaps the most complete account to date is that given by Turley (1997, 1998, 1999). Nevertheless, the focus of these three studies is the theory of historical change, rather than new data to improve on previous studies. What is still lacking is a comprehensive

study of [*se*] in all of its uses, in all historical periods, and based on large electronic corpora such as ADMYTE, CORDE or the *Corpus del Español*. The present study attempts to take a preliminary step in that direction.

What do previous studies already tell us about specific shifts with [*se*], and what are some of the more important questions that still need to be answered? The following table summarizes some of the basic conclusions from Turley (1999), Maldonado (1989), and Melis (1995), which in turn are based on a number of the earlier studies mentioned previously. The symbols in the second column from the left column provide a key to their basic conclusions. The numbers 1–5 show the historical trajectory proposed in Turley (1999), which shows how each of these five stages are linked by potentially ambiguous surface structures. The three [↗] symbols indicate the uses of [*se*] that Maldonado (1989) and Melis (1993) suggest have had a large increase during the past 300–400 years.

		+ AGENTIVE	
Turley	1	true reflexive	María se vio en el espejo
		reciprocal	María y Juan se besaron
Turley	2	intransitive (middle)	María se baño / se sentó
Melis	↗	emotional reaction	María se sorprendió / se enojó
Maldonado	↗	body movement	María se fue / ha vuelto a México
	??	change / inchoative	María se durmió
Maldonado	↗	"energetic"	María se cayó en la nieve María se tomó la cerveza
	??	inherent / lexical	María se queja / se arrepiente
Turley	3	decausative	se hundió el barco se derritió el hielo
Turley	4	passive	se vendieron los libros
Turley	5	impersonal	se vive bien en …
	??	causative	María se operó
		− AGENTIVE	

Fig. 1. Overview of recent research on the historical development of [*se*]

As has been mentioned, due to the previous lack of large-scale corpora, the studies just mentioned have all been of necessity somewhat focused, either in terms of the size of the database of examples, the historical period considered, or the range of uses of [*se*]. As a result, there are a number of important questions that are still unanswered. These include the following:

1. Inchoative and change of state (*se enfermó*). This is perhaps the most problematic category, and a use of [*se*] for which we have very little information regarding its historical trajectory.
2. "Energetic" and "unexpected" (*se cayó, se lo comió, se murio*). Maldonado (1989) suggests a significant increase during the past 300–400 years, but his one article is the only one to have examined this construction.

3. Emotional reaction (*Jorge se sorprendió*). Melis (1995) likewise is a preliminary study of this one construction. While it provides useful data, it has not been considered by any other studies.
4. Decausative (*el barco se hundió ayer*). Turley (1999) suggests that this use arose long enough ago that there should be no residual increase during the past 200–300 years. Is this in fact the case?
5. Inherent / lexical (*Manolo se queja*). Have these always obligatorily taken the "reflexive" marker [*se*], or could they take indirect or direct objects in earlier stages?
6. Causatives (*Juan se cortó el pelo*). In some generative studies, these constructions have occupied an important position (e.g. Chomsky 1988: xx). Have they ever had more than a very peripheral status – in any historical period?
7. Impersonal (*se vive bien en España*). When did this construction arise, and with what verbs?

2. Database of examples

As was mentioned previously, it is only during the past five or six years that there have been large, publicly-available corpora that can be searched to provide large amounts of data on syntactic constructions like [*se*]. Previous corpus-based studies of historical syntax had to rely on smaller, proprietary corpora that the researchers had created. We now have access to at least two 100+ million word corpora of historical Spanish, both of which are publicly available, and which provide comprehensive coverage of texts from the 1200s–1900s.

The CORDE and CREA texts from the Real Academia Española were the first to appear (see <http://www.rae.es>). CORDE covers historical Spanish, while CREA deals with the past 30–40 years. While both of these corpora are valuable for many types of studies, there are quite limited for syntactic research. This is because there is no "part of speech" tagging of the texts, nor any attempt at lemmatization. In addition, there is no possibility of creating customized lists of search terms, and only very limited use of wildcards. It would be completely impossible, for example, to search directly for infinitives ending in [-*se*] (*morirse, caerse*), or for all forms of a given pronominal verb (e.g. *irse*: [*se*] *fue, me fui, nos vamos, te irás*). So while CORDE and CREA may be useful for some lexical studies, they were not used in the present study.

The corpus that we have employed is the *Corpus del Español*, a 100 million word corpus of more than 10.000 texts from the 1200s–1900s (see Davies 2002, 2003 and this volume for an overview). This corpus was funded by the National Endowment for the Humanities (US), and was completed in late 2002 (available online at <http://www.corpusdelespanol.org>). Unlike CORDE

and CREA, the *Corpus del Español* allows a wide range of searches. For example, users can search by part of speech and lemma (1–2), wildcards (3), synonyms (4), and customized lists (5):

1	*.pn_obj querer.* *.v_inf	lo quiero hacer, me quería hablar
2	*.n suave.*	voz suave, viento suave, inviernos suaves
3	s_fr_r*	sufrir, sofre, sufrirán
	tan * como	tan bueno / bien / grande como
4	!difícil.* de *.v_inf	difícil de hacer, imposible de evitar
5	[se] poner.* el/la [lópez:ropa].*	[se] puso la chaqueta, [se] pone el sombrero

Fig. 2. Types of searches in the *Corpus del Español*

For the purposes of this study, perhaps the most important aspect of the corpus is its underlying architecture. In addition to the actual linear / textual database, the corpus also contains relational databases containing all of the 1, 2, 3, and 4 word sequences (ngrams) in the entire 100 million word corpus, as well as their frequency in each century from the 1200s–1900s. For example, the following table shows the frequency of the three-word phrase *no puede ser*, and is indicative of the 40+ million other unique three word strings from the corpus. The columns in the database show the three words (w1, w2, w3) as well as the frequency of this phrase in each century (x12=1200s – x19=1900s). Similar tables have been created for all 1, 2, and 4-word sequences in the corpus.

w1	w2	w3	x12	x13	x14	x15	x16	x17	x18	x19
no	puede	ser	105	13	140	518	423	269	892	646

Fig. 3. Ngrams / frequency tables in the *Corpus del Español*

As can be seen in the following tables, there are also databases containing the part of speech and lemma information for hundreds of thousands of forms.

word	*pos*
no	adv
puede	v_pres
ser	v_inf
ser	n

part of speech

word	*lemma*
no	no
puede	poder
ser	ser

lemma

Fig. 4a and 4b. Part of speech and lemma tables in the *Corpus del Español*

Similar databases exist for more than 30.000 synonyms sets, as well as databases that contain customized lists of words that have been created by the end users. All of these databases can be JOINed together via SQL commands to the main ngram / frequency databases. The result is powerful search syntax, allowing queries as complicated as the following:

(1) *.pn_obj !mandar.* que [lópez:movimiento].* 1900s>1 -1800s -1700s

Object pronoun + any form of any synonym of *mandar* + *que* + any form of any word in the [movimiento] list created by [lópez], which occurs at least twice in the 1900s, but not in the 1700s or 1800s

le mandó que saliera, nos pide que corramos, me dice que camine

In terms of studying the historical development of [*se*], this "ngrams approach" yields much useful data. For example, in the 2-grams table, there are more than 50.000 distinct two word sequences in which the first word is [*se*], as in the following table. These data – containing the frequency of a wide range of constructions with [*se*] in each of the centuries from the 1200s–1900s – will be key in helping us to accurately map out the historical trajectory of the different uses of [*se*].

w1	*w2*	*1200*	*1300*	*1400*	*1500*	*1600*	*1700*	*1800*	*1900*
SE	HA	122	184	599	6550	4515	4042	7054	8516
SE	HAN	107	116	237	3023	1650	2007	2825	4451
SE	PUEDE	757	420	2075	4612	1522	2973	1777	4299
SE	HABÍA	0	0	2	2465	1499	1330	4678	3873
SE	VA	48	92	163	933	790	574	1059	2841
SE	TRATA	1	0	55	443	184	545	1526	2316

Fig. 5. Ngram database; frequency for [*se*] sequences in each century

3. Measuring shifts with [*se*]

Using the ngrams tables, we can easily and accurately determine the frequency of [+*se*] with thousands of verbs at one time. This will then allow us to determine which verbs have the largest increase in the use of [*se*] between different historical periods. The following table shows the type of structures that we will look for, as evidence for both [–*se*] and [+*se*] with nonfinite verbs in different periods. In order to calculate the percentage of [+se] with many different verbs at once, we simply calculate the percentage of constructions like 5–6, divided by the frequency of all constructions 1–6.

–[se]	1	PREP INF	de mover
	2	PREP INF CL-[*se*]	de moverlo
	3	PREP CL-[*se*] INF	de lo mover (earlier stages of Spanish)
	4	PREP-CL-[*se*] INF	delo mover (earliest stages)
+[se]	5	PREP INF [*se*]	de moverse
	6	PREP [*se*] INF	de [se] mover (earlier stages of Spanish)

Fig. 6. [–*se*] / [+*se*] constructions with nonfinite forms

Note that we have focused on the presence or absence of [*se*] with [PREP + INF] constructions, in order to avoid problems with clitic climbing, e.g. *tenían que bañarse* vs. *se tenían de bañar* 'they had to take a bath'. The only real complication is the variation between proclisis and enclisis in older stages of the language (*de se mover, de moverse*). Note also that one of the main advantages of looking at nonfinite forms is that we do not have to consider all of the variant forms (present, preterit, future) for a given verb. There is much less variation when we limit ourselves strictly to the infinitival form.

The following table shows the percentage of [+*se*] for a handful of the thousands of verbs that we considered in each of the centuries from the 1200s–1900s. Again, the percentage is a function of the two [+*se*] structures shown in figure 6 above, divided by all six [*se*] constructions. In this table, for example, we find that *casar* has increased from 9% [+*se*] in the 1400s to 81% in the 1900s, while *acercar* has increased much less, from 67% to 71%. Similar figures were calculated for more than 2000 other verbs.

	1400s							*1900s*			
verb	%	de mover	de moverse	de [se] mover	de moverlo	de lo mover	delo mover	%	de mover	de moverse	de moverlo
CASAR	.09	113	2	10		2	6	.81	26	136	5
SENTAR		4						.80	22	110	5
ACERCAR	.67	2	3	1				.71	45	140	13
DESPEDIR	.47	9	2	6				.66	22	110	35
MOVER	.16	15		3	1			.55	102	151	24
QUEDAR	.07	85	2	5			2	.43	236	180	3
UNIR								.42	73	63	14

Fig. 7. Percentage of [+*se*] with nonfinite forms

Now that we have considered in detail the way in which the differences in [+*se*] were calculated, we can use these formulae to discover the shift in [+*se*] with many different verbs, across each pair of centuries from the 1200s–1900s (e.g. 1300s–1400s, 1600s–1700s, etc). An example of this is found in the following table, which provides data for just eight verbs in the period of the 1500s–1600s. The table shows the percentage of [+*se*] in the two centuries and the increase in [+*se*] from one century to the next. For example, *adelantar* 'to advance' was 18% [+*se*] in the 1500s, but 56% [+*se*] in the 1600s, for an increase of 37%.

#	*verb*	*increase*	*1500s*	*1600s*
1	SECAR	0.43	0.43	0.00
2	ADELANTAR	0.37	0.56	0.18
3	CALENTAR	0.36	0.46	0.10
4	ENSAYAR	0.36	0.43	0.07
5	MENEAR	0.35	0.53	0.18
6	ALBOROTAR	0.34	0.39	0.05
7	ESCONDER	0.29	0.51	0.22
8	PRESERVAR	0.29	0.29	0.00

Fig. 8. Percentage change with different verbs, 1500s > 1600s

4. Diachronic shifts, grouped by type of verb

Ideally, then, we could use the type of data that we have introduced up to this point, to show the increase or decrease in [+*se*] with thousands of different verbs between each set of centuries from the 1200s–1900s, and this would then create a detailed map of all of the different historical trajectories. For example, we could see [*se*] increasing with one type of verb (e.g. decausative or inchoative) from the 1500s–1600s, and then spreading to another use (e.g. impersonals) in the 1700s–1800s. Obviously, such a complete mapping could easily produce a book-length study. Because the present study is mainly an introduction to the methodology of how such a study could be carried out, we will only present partial data on a range of uses of [*se*] for a limited period – just the 1800s–1900s. We will leave it to further studies to complete the project.

In the table on the following page we present data showing the increase or decrease in the percentage of [*se*] with different classes of verbs from the 1800s–1900s. With nearly all of the ten types of [*se*], we looked at the increase or decrease with at least ten to fifteen different verbs. The one exception was causative verbs, where there are a limited number of lexical items (*bautizarse, operarse, cortarse el pelo*, etc). As the table indicates, there is a clear increase with impersonal, inchoative, "energetic", and emotional reaction verbs from the 1800s–1900s. The increase in [*se*] with middle, body movement, and decausative verbs occurred at a much lower rate (9–13% increase). The argument might be made that this level of increase is a function of some type of "background noise" in terms of more general syntactic shifts

in Spanish, such as clitic placement or more general shifts in the pronominal structure of Spanish. Surprisingly, there was a decrease in [*se*] with passives and causatives, though the results with the causatives may be spurious, due to the limited range of examples. In the following sections, we will consider a few of these shifts in more detail.

type	*example*	*Change 1800s–1900s*
change / inchoative	se durmió	+ 37%
"energetic"	se comió los tacos	+ 32%
impersonal	se vive bien en ...	+ 27%
emotional reaction	se sorprendió	+ 20%
middle / arreglo personal	se bañó	+ 13%
body movement	se fue	+ 11%
decausative	se derritió el hielo	+ 9%
inherent / lexical	se queja de ...	0%
passive	se vendieron los libros	– 21%
causative (two verbs)	María se operó	– 22%

Fig. 9. Increase or decrease in [*se*] from the 1800s–1900s, by type of verb

4.1 Inchoative verbs

Figure 12 indicates that perhaps the largest increase in [*se*] is with the inchoative / change of state verbs, such as *marearse* 'to get dizzy', *desmayarse* 'to faint', and *enfermarse* 'to get sick', where there was a 37% increase in use from the 1800s–1900s. Perhaps the best evidence for this are the examples without [*se*] in the 1800s, most of which would sound quite odd to a contemporary Spanish speaker:

(2a) en vez de desmayar, *se* sentían excitados de un nuevo vigor
[34 cases of *que* or *y* + *desmayar* in 1800s, only 2 in 1900s]
(2b) y saltaba de peña en peña [y mareaba con su vertiginosa movilidad]
[13 cases of *que* or *y* + *marear* in 1800s, only 4 in 1900s]
(2c) o que enfermaban de otra dolencia
[20 cases of *que* or *y* + *enfermar* in 1800s, only 2 in 1900s]
(2d) si no equivocases con una falsa interpretación
[9 cases of *que* or *y* + *equivocar* in 1800s, only 1 in 1900s]

4.2 "Energetic" reflexives

The [*se*] with verbs such as *comer*, *tomar*, *tragar* and *saber* does not replace the subject or object, as it does with many of the other uses of [*se*]. In these cases, there is already an object, such as *Pedro lo comió* 'Pedro ate it'. As

Maldonado (1989), de Bruyne (1995: 166–67), Butt and Benjamin (2000: 373), Whitley (2002: 176–77) and others have pointed out, its function is to focus more on the dynamicity of the action. Maldonado (1989) is the only one to examine the historical development of this construction, and shows a gradual increase since at least the 1400s. This is supported by our data. The following table shows the number of cases of [+*se*] (per million words) with each verb during the 1800s and 1900s.

	1200s	*1300s*	*1400s*	*1500s*	*1600s*	*1700s*	*1800s*	*1900s*
comer	0.2	2.4	1.5	6.6	5.4	3.4	6.4	6.6
tragar	0	0	0	1.2	1	1.1	1.3	2.2
beber	0	0	0	0.7	1.1	0	0.6	1.2
tomar	0.5	1.2	1	4.9	1.4	0.5	1.6	3.8
saber	0.6	0.4	0.9	3.3	1.5	1	2.3	5
TOTAL	1.3	4	3.4	16	9.3	6	11.6	17.6

Fig. 10. "Energetic" verbs ([*se*] *las comió, me lo tragué*)

As the data indicate, there have been cases with [*se*] since Old Spanish (see 3a below), and during the 1500s it was particularly common (3b). During the 1600s–1800s, however, its use decreased somewhat, but has rebounded to its highest level ever from the 1800s–1900s (3c).

(3a) 1300s: que los toman y se *los*_comen (*Libro de la maravillas*)
(3b) 1500s: como si fuera manojo de yerba se *la*_comían (*Historia de Chile*)
(3c) 1900s: con vidrio y todo se *lo*_comió y lo masticó (HC: Buenos Aires)

The degree that [*se*] has increased with this construction during the past two hundred or so years, and the relatively unique status that this construction occupies in terms of pan-Romance use of [*se*], suggest that it would make a very interesting topic for further investigation, to complement the one previous study in Maldonado (1989).

4.3 Impersonal verbs

The data indicate that the third area of significant increase in [*se*] since the 1800s has been with the impersonal verb construction, which tends to support earlier claims by Barry (1985), Turley (1999) and others. As the following table shows, there has been a roughly 27% increase in impersonal [*se*] during the past 200 years. This table shows the number of cases of [*se*] with the verb or phrase in question in both centuries (e.g. 153 cases of *se vive* 'one lives' in the 1900s), the total number of cases of any word + the verb or phrase in that century (e.g. 1299 cases of [x]+[vive] in the 1900s), and the resulting percentage of [+*se*] (e.g. 11.8% for *se vive* in the 1900s).

	1800s-%	1800s +se	1800s all	1900s-%	1900s +se	1900s all
es	0.05	44	89490	0.1	104	114529
vive	7.0	102	1470	11.8	153	1299
habla de	40.5	159	393	43.3	273	631
espera que	27.9	12	43	52.1	85	163
sabe que	31.1	173	556	26.2	239	912
supone que	57.9	55	95	79.2	236	298
piensa que	21.1	16	76	23.0	66	287
	26.5	> Increase = 27% >		33.7		

Fig. 11. Impersonal verbs

One caveat regarding the impersonal use of [*se*] is that it is quite difficult to find verbs and phrases that are unambiguously impersonal in meaning. Intransitive verbs like *ir*, *dormir*, and *morir* can have inchoative readings with [*se*]: *se va* 'leaves', *se duerme* 'goes to sleep', and *se muere* 'dies'. More importantly, many of these phrases in figure 11 are ambiguous between an impersonal and a passive reading (e.g. *se supone que* 'one supposes that / it is supposed that'), and this is in fact one motivation for the extension of [*se*] from passives to impersonals (cf. Turley 1999). Yet as figure 9 shows, the unambiguous passives (e.g. *se vendió el coche* 'the car was sold') have no increase in use from the 1800s–1900s, and so the increase with these potentially ambiguous passive/impersonal phrases probably is significant.

4.4 Decausatives

As Turley (1999) and others have suggested, the use of [*se*] with decausatives started long enough ago (in at least Old Spanish), and it would be strange to find much increase in Modern Spanish. Yet, there is some evidence that even here, there may have been some recent increase in [*se*]. The evidence for this takes the form of sentences without [*se*] in earlier stages of the language, which are somewhat awkward in Modern Spanish.

(4a) aunque él movió tras ellos llamándolos y asegurándolos
(1500s: *Felixmarte de Hircania*)

(4b) el yelmo fue en dos partes dividido, el pecho abrió
(1500s: *Las lágrimas de Angélica*)

(4c) cuando la noche cerró con más escuridad (1600s: *Don Quijote*)

Nevertheless, it would be useful to carry out a more widespread and systematic study of these verbs during the past two or three hundred years, to see exactly how much increase in [*se*] with decausatives might still be underway.

4.5 "Lexical" reflexives

One final category of [*se*] for which the *Corpus del Español* indicates some shifts during the past three or four hundred years is the "inherent" or "lexical" use of [*se*]. In Modern Spanish these verbs require an accompanying reflexive pronoun. In other words, *Juan quejaba* 'John complained', *María me jactaba* 'Mary was boasting to me', *Pedro les arrepintió* 'Pedro felt bad (towards them)' would all be ungrammatical in Modern Spanish. As the data from the corpus indicate, however, these types of sentences were grammatical as late as the 1500s:

(5a) tan grande que no le quejó cosa ninguna (*Libro de Job*)
(5b) de qué se quexa, a quién le quexa (1500s)
(5c) tan grande que no le quejó cosa ninguna (1500s)

Preliminary data indicates that most of these verbs had probably already acquired obligatory "reflexives" by the 1600s, but it may be useful to examine the data from this and other corpora such as CORDE to determine the precise nature and chronology of the change.

5. Summary

The present study is an attempt to show how a large, annotated corpus of historical texts can provide insight into the historical trajectory of [*se*] in Spanish. Such a study would most likely be impossible with the CORDE corpus, since it is annotated neither for lemma or part of speech. The *Corpus del Español*, however, easily allows this type of research. Most importantly, the ngrams architecture of the data allows us to quickly and easily identify all of the pronominal [*se*] verbs that have undergone a shift from one century to the next, without knowing *a priori* what the specific verbs might be.

The data from the *Corpus del Español* has provided some preliminary data on the trajectory with given uses of [*se*] during the last eight hundred years, but much remains to be done. In this study we have focused more on certain uses of [*se*], and there still remains much to be done on other uses, such as inchoatives / change of state, the "energetic / focusing" uses, and causatives. In addition, the approach that we have applied to the data from the 1800s–1900s could and should be applied to the entire period from the 1200s–1900s. Once this is done, we will have for the first time a comprehensive picture of the development of [*se*] – which is one of the most challenging, yet interesting topics in historical and modern Spanish syntax.

References

Barry, Anita K. 1985: The rise of the impersonal 'se' constructions. *Hispanic Journal* 6, 209–218.

Brown, Charles B. 1928: *The passive and indefinite reflexives in Old Spanish*. Chicago (diss.).

Butt, John / Benjamín, Carmen ^{3}2000: *A new reference grammar of Modern Spanish*. Chicago: McGraw-Hill.

De Bruyne, Jacques 1995: *A comprehensive Spanish grammar*. Oxford: Blackwell.

Chomsky, Noam 1988: *Language and problems of knowledge: the Managua lectures*. Cambridge, MA: MIT Press.

Davies, Mark 2002: Un corpus anotado de 100.000.000 palabras del español histórico y moderno; in: *SEPLN 2002* (Sociedad Española para el Procesamiento del Lenguaje Natural; Valladolid), 21–27.

— 2003: Relational n-gram databases as a basis for unlimited annotation on very large corpora; in: Simov, Kiril (ed.): *Proceedings from the Workshop on Shallow Processing of Large Corpora* (Lancaster, England, March 2003), 23–33.

Dubravcic, Stephanie 1979: *The passive voice in the* Primera Crónica General. Columbus: The Ohio State University (unpublished diss.).

Hanssen, Friedrich 1912: La pasiva castellana. *Anales de la Universidad de Chile* 131, 97–112, 507–514.

Hernández Alonso, C. 1966: Del *se* reflexivo al impersonal. *Archivum* 16, 39–66.

Karde, Sven 1943: *Quelques manières d'exprimer l'idée d'un sujet indéterminé ou général en espagnol*. Upsala: Appelberg.

Kemmer, Suzanne 1993: *The middle voice*. Amsterdam / Philadelphia: Benjamins.

Maldonado, Ricardo 1989: *Se* gramaticalizó: A diachronic account of energetic reflexives in Spanish. *Proceedings of the Pacific Linguistics Conference*, 339–360.

Martín Zorraquino, María Antonia 1979: *Las construcciones pronominales en español (paradigmas y desviaciones)*. Madrid: Gredos.

Melis, Chantal 1995: A diachronic view of prepositional verbs of emotion in Spanish; in: Andersen, Henning (ed.): *Historical Linguistics 1993*. Amsterdam: Benjamins, 309–322.

Mendeloff, Henry 1964: The passive voice in Old Spanish. *Romanistisches Jahrbuch* 15, 269–287.

Monge, Félix 1954: *Las frases pronominales de sentido impersonal en español*. Zaragoza: CSIC (also in *Archivo de Filología Aragonesa* 7, 7–102).

Sepúlveda Barrios, Félix 1988: *La voz pasiva en el español del siglo XVI: contribución a su estudio*. Madrid: Gredos.

Turley, Jeffrey 1997: The renovation of the Romance reflexive constructions. *Romance Philology* 51, 15–34.

— 1998: A prototype analysis of Spanish indeterminate reflexive constructions. *Language Sciences* 20, 137–162.

— 1999: The creation of a grammaticalization chain: The story of Spanish decausative, passive, and indeterminate reflexive constructions. *Southwest Journal of Linguistics* 18, 101–138.

Whitley, M. Stanley ^{2}2002: *Spanish-English contrasts*. Washington, DC: Georgetown University Press.