

## Entrenchment and discourse traditions in Spanish auxiliary selection

Malte Rosemeyer

Albert-Ludwigs-Universität Freiburg

### 1. Introduction

Usage-based linguistics (UBL) assumes that frequency of use has a crucial influence on the development of linguistic elements. Among other frequency effects, Bybee (2006) argues for a *conserving effect* of frequency. Thus, a linguistic element with a high token frequency is subject to *entrenchment*: the repeated activation of the mental representation of that element leads to a strengthening of its mental representation. In language change, entrenchment is said to counter processes of language change. In situations where a linguistic element is in the process of being replaced by a second, more productive, linguistic element, high-frequency instantiations of the replaced element will resist the change longer than low-frequency instantiations.

As a cognitivist approach to language, UBL is mostly concerned with the cognitive foundations causing the frequency effects observed in language use and language change. However, when applying the notion of entrenchment to the study of language change, a problem with this cognitivist perspective immediately arises: the context-dependency of entrenchment. It is unclear whether entrenchment has to be modelled as a function of the global token frequency of an instantiation of a construction, or as the token frequency of that instantiation in specific linguistic contexts. From a methodological perspective, diachronic studies of entrenchment thus always face the problem of which token frequencies to count. This problem could potentially cause huge biases in these studies.

In a recent paper, Kabatek (2013) discusses the question of how to conduct quantitative studies in diachronic corpora and puts special emphasis on the problem of *discourse traditions* (Schlieben-Lange 1983; Koch 1997). A language as represented in historical texts is not a system but a conglomerate of several systems, each corresponding to a historically determined way of text organisation, i.e., a discourse tradition. Diachronic studies of language change tend to neglect this influence of text organisation on the use of language in these texts and consequently assume an ideal, generalised language that does not exist. Kabatek notes that diachronic studies of entrenchment commit the same error, since they typically consider the product of a language (i.e., the texts) the language itself (Kabatek 2013: 20). If the rules of use of a language are specifically bound to discourse traditions, entrenchment should also be determined by discourse traditions.

In this paper, I develop a new perspective on the question of frequency effects in language change. I propose that entrenchment not only varies according to discourse traditions, but also argue that the very notion of entrenchment is inseparable from discourse traditions. I suggest that entrenchment plays a pivotal role in the creation of discourse traditions and that in turn, discourse traditions enable entrenchment. In order to demonstrate the relevance of these assumptions, I conduct a quantitative analysis of the development of Spanish auxiliary selection in three different discourse traditions: Historiographical texts, administrative documents, and private letters. The findings from the study support earlier studies which claim that the disappearing *ser* + PtcP construction was conserved with verbs that have a high token frequency – in particular, verbs that express change-of-location predicates. In addition, they show that the strength of this conserving effect varies according to the discourse tradition. As a result, the analysis demonstrates that it is indeed possible to study entrenchment in historical texts if close attention is paid to the question of discourse traditions.

This paper is structured as follows. In section 2, I elaborate the notion of the context-dependency of entrenchment processes and consequently, conservation in language change. In section 3, I give a brief overview on recent findings on the development of auxiliary selection in Spanish. After an introduction to the data used in this study (section 4), I conduct a quantitative analysis of auxiliary selection in these data (section 5). The final section 6 discusses the relevance of these findings for studies of frequency effects and discourse traditions.

## 2. Discourse traditions and entrenchment

The term *discourse tradition* was coined in Romance linguistics and is usually taken refer to historical conventionalised patterns of text production (Schlieben-Lange 1983; Koch 1997; Oesterreicher 1997; Jacob and Kabatek 2001; Wilhelm 2001: 467; Aschenberg 2003: 5; Kabatek 2005b; a; Schrott and Völker 2005; Kabatek 2013). As an example, consider the following quote from Jacob and Kabatek (2001: viii):

[Discourse traditions] are historical-normative patterns that have been established in a society and which speakers follow in the production of discourse. The existence of these categories means that each discourse, and consequently, each historical text, is not only part of a specific language (of several languages) but is rather situated in an intertextual succession constituted from a series of repeating elements. These elements refer to both the level of the “context” (situational, medial or institutional constellations) and the level of forms on the text surface (e.g., specific text passages, formulaic use of constructions and language). (Jacob and Kabatek 2001: viii, translation MR).

Studies on discourse traditions take a genuinely pragmatic approach to the description of language: they claim that the use of a linguistic element is to a great extent determined by its usage context. One and the same linguistic element can have a different function in different discourse traditions. This consideration has important implications, as differences in discourse traditions can govern linguistic phenomena as distinct as polysemy (Blank 2005), the function of grammatical elements such as verbal periphrases (Jacob 1994; 2001), or the use of politeness markers (Held 2005).

The discourse tradition approach to the description of language can be related to the increasing recognition in historical linguistics that language change is intricately related to discourse contexts. Many historical linguists claim that the discourse context has an instrumental role in linguistic innovation processes. For instance, recent studies on grammaticalisation processes emphasise the importance of the discourse context for the reanalysis of a linguistic element and consequently, the emergence of a new grammatical meaning (Diewald 2002; Heine 2002; Diewald 2006). If a grammatical element is used in a novel discourse context, its grammatical meaning changes to fit the discourse context. This idea basically corresponds to the notion that the reanalysis of a grammatical element in a specific context is caused by a pragmatic implicature (Dahl 1985: 10-11). Grammaticalisation can then be described as the conventionalisation of this conversational implicature (Traugott 1989; Traugott and König 1991; Hopper and Traugott 2003: 81-93).

As suggested by this parenthesis on the importance of discourse traditions for the description of mechanisms of language change such as grammaticalisation, the notion of discourse traditions has a direct impact on the idea of how language change works. Kabatek (2013: 15-16) argues that if we accept the idea of discourse traditions, we must also take seriously that there are distinctions between (a) a language system and a text, and (b)

language systems that serve as the basis for the generation of texts, and completed texts. Kabatek's first point emphasises that a language is a much less homogeneous entity than a text, given that a language can be characterised as a conglomerate of several systems characterised by factors such as diatopic, diastratic and diaphasic variation. However, historical linguists often neglect this problem and study a text as if it were representative of "the language" used in that time period. The second point addressed by Kabatek is the core assumption of the study of discourse traditions: the inherent intertextuality of language use. Thus, all levels of the use of a language system – grammatical, phonetic and semantic rules – are intrinsically historically determined. In Kabatek's words, "every text, from the salutation to the novel, evokes anterior texts, even if it does not repeat elements belonging to these texts" (Kabatek 2013: 16, translation mine). It follows that it is impossible to argue for an abstract development of a language distinct from the development within discourse traditions: discourse traditions or text genres do not merely modify language use like a sociolinguistic variable, for instance, but directly determine language use.

There are interesting parallels between the perspective on language use drawn by researchers on discourse traditions and researchers working in the paradigm of usage-based linguistics. In the following, I give a brief introduction to UBL and in particular, its model of the *conserving effect of frequency*, in order to demonstrate these similarities.

UBL claims that the speakers' experience with language has a direct influence on their language use (Bybee and Hopper 2001; Bybee 2006; Diessel 2007; 2011; Pfänder et al. 2013). The frequency with which speakers perceive a linguistic element alters the form of the linguistic element, causing frequency effects such as *chunking* (Newell 1990: 185-193; Ellis 1996; Bybee 2010: 33-56).

When modelling chunking, many UBL approaches make use of *exemplar theory* (Medin and Schaffer 1978; Kuhl 1994; Lacerda 1995; Johnson 1997; Pierrehumbert 2001; 2002). In essence, exemplar theory is a theory of categorisation. It proposes that when a person perceives a novel stimulus, s/he will categorise this stimulus according to its similarity to other stimuli that s/he has previously encountered. This similarity relationship determines the way these exemplars are stored in the mind: in the storage process, similar exemplars are stored proximally to each other, whereas dissimilar exemplars are stored less proximally to each other. From a neurological point of view, this locationist metaphor can be understood as specifying the strength of activation of the links between the stored exemplars. As a result, exemplar theory proposes that categories be represented as clusters of specific tokens perceived by a person that in one way or another resemble each other. The categories formed by exemplars are thus gradient.

Similarity between exemplars as modelled in UBL can also be a function of context. In particular, a repeated co-occurrence of two linguistic elements corresponds to a similarity between these elements in terms of situational contexts (trivially, these elements are perceived at two points in time that are close to one another). The repeated simultaneous activation of two linguistic elements will therefore lead to a strengthening of the link between the exemplars. This process results in chunking: Due to the repeated simultaneous activation of two linguistic elements, the speaker will start conceptualising these two elements as forming a chunk, i.e. a new category. Chunking results in *entrenchment* (Langacker 1987: 59), i.e. an improvement of both accessibility and production of the chunked element. Important evidence for the psychological reality of chunking can be found in studies of L1 acquisition. These studies suggest that chunking may be a first step in the creation of grammatical rules, given that chunking leads to the creation of a complex hierarchical structure between the chunked elements and the resulting chunk (Tomasello 1992; Lieven, Pine and Baldwin 1997; Goldberg 2006; Behrens 2009).

The notions of chunking and entrenchment are crucial for the explanation of the *conserving effect of frequency* (Bybee 2006). Due to chunking, a morphologically complex

form will typically decrease in analysability. In Bybee's words, "the more a sequence of morphemes or words is used together, the stronger the sequence will become as a unit and the less associated it will be to its component parts" (Bybee 2010: 48). As an example, consider Poplack's (2001) study of subjunctive mood in Canadian French. Although in Canadian French the subjunctive is currently being replaced with the indicative, the raw numbers of use would appear to indicate that the subjunctive is still very frequent. Within the envelope of variation isolated by Poplack, the subjunctive is used in 77 per cent of the cases (Poplack 2001: 411). Poplack, however, demonstrates that this numerical dominance of the use of subjunctive is misleading, since the productivity of the subjunctive is restricted to a very small but frequent group of both matrix verbs and embedded verbs (Poplack 2001: 412). Although in theory any verb can be used in subjunctive mood, in Poplack's data only four verbs occur regularly with the subjunctive: *être* ('be'), *aller* ('go'), *avoir* ('have') and *faire* ('do'). It is evident that these verbs have a very high overall token frequency. Consequently, Poplack claims that

although the token frequency of the standard subjunctive variant is elevated, virtually all its uses are concentrated among a handful of highly favoring matrix verbs collocated with a small cohort of frequent and irregular embedded verbs. Outside of these few contexts, in which its use has become ritualised, selection of the subjunctive is very rare. (Poplack 2001: 414)

Due to the high token frequency of specific instantiations of the complex subordinating construction (for instance, *falloir* + *que* + subjunctive), these syntagms are subject to an entrenchment process. This entrenchment process leads to a loss of internal analysability. In other words, the high-frequency instantiations of the construction start forming a category on their own. This means that over time, their productivity is less bound to the productivity of the construction of which they are instantiations. As a result, the high-frequency syntagms are affected less by the replacement process operating on the Canadian French subjunctive. This is the *conserving effect of frequency*. Sections 3 and 5 of this paper will demonstrate a scenario very similar to the one suggested by Poplack for the development of Spanish auxiliary selection.

When comparing the descriptions of the discourse-traditional and the usage-based approach to language change given in this section, striking similarities between the two approaches arise. In particular, both approaches assume that the experience of language users with language has a direct influence on the way linguistic elements are employed. This overlap is evident in the importance of the notions of *conventionalisation* and *intertextuality* for the two research approaches.

First, both approaches assume that the co-occurrence of linguistic elements is conventionalised in language use over time. The crucial factor for this conventionalisation process is repeated exposure. Indeed, the emergence of discourse traditions can be conceptualised as a type of entrenchment on a higher level of linguistic complexity: the repeated use of a specific constellation of linguistic elements in a specific discourse situation will lead to the strengthening of the connections between the mental representations of these linguistic elements.

Second, the UBL and discourse tradition approach share a focus on intertextuality as developed in poststructuralism (Barthes 1967; Foucault 1988 [1969]; 2002 [1969]). These

studies put into perspective the role of the author in text production. If any text “is made of multiple writings, drawn from many cultures and entering into mutual relations of dialogue, parody, contestation” (Barthes 1967), the author plays a less innovative role than commonly assumed. Rather, texts are the product of preceding texts. As already noted in Kabatek’s (2013) discussion of the discourse tradition approach, intertextuality is a core assumption of the discourse tradition approach. Intertextuality is also crucial for UBL. The exemplar model of language production assumes that in the production of a linguistic element, the language user necessarily evokes earlier language events involving this linguistic element. Entrenchment only differs from this “regular” model of language production in that it cancels out the abstraction processes involved in linguistic productivity and thus leads to the explicit copying of the entrenched linguistic element. This means that entrenchment can be conceptualised as a process by which the use of a linguistic element comes to increasingly depend on intertextuality.

This discussion of the notions of conventionalisation and intertextuality has demonstrated broad similarities between the discourse tradition approach to language and UBL. I have suggested that discourse traditions can be understood as a kind of “macro-entrenchment”: the repeated co-occurrence of a constellation of linguistic elements leads to the conventionalisation of this constellation and consequently, a discourse tradition. In turn, the UBL notion of entrenchment is intrinsically intertextual in nature and can be understood as a “micro- discourse tradition”: just as the use of a linguistic element within a certain discourse tradition depends on the previous use of that element in that discourse tradition, the use of an entrenched linguistic element depends on the repeated previous use of that linguistic element.

The identification of the similarities in the approaches described in this section of the paper offers a potential solution to a problem that is highly relevant to historical linguistics, i.e. the problem of the “accumulative episteme” (Jacob 2001) of medieval writing. Medieval authors typically take their material from several sources, often copying entire text passages. With regard to Spanish, this problem can be illustrated clearly for the discourse tradition of historiographical texts. The historiographical texts attributed to Alfonso X are composed of passages taken from works by classical Latin authors such as Lucan, Ovid and Pliny the Elder, religious texts like the bible or the texts written by Eusebius of Caesarea, as well as French and Arabian literature and epic poems in Romance (Fernández-Ordóñez 2004: 390).

This method of text production is indeed a defining property of medieval literacy (Burke 2004: 97). In the case of Spanish historiographical texts, however, it continues well into the 17th century: Many of the later historiographical texts both explicitly adopt the model of the Alfonsine chronicles and copy from them (Von Hoegen 2000: 30). Indeed, this indicates that these Spanish texts belong to the same discourse tradition. It also constitutes a problem because many of the tokens from these texts are copies from earlier tokens. For this reason, Jacob (2001: 156) argues that these tokens are not representative of the language of the writer and consequently, that quantitative analyses of medieval texts are impossible to conduct.

From the perspective developed in this section, however, the phenomenon of explicit copying is not qualitatively different from other mechanisms that lead to the constitution of a discourse tradition. In other words, within the Spanish discourse tradition of historiographical texts, these texts refer to previous texts both implicitly (e.g. by topic and narrative style) and explicitly (explicit copies). Effectively, copied tokens mirror the principle of intertextuality.

Likewise, an interaction between entrenchment and copying can be assumed. During the process of copying, old-fashioned linguistic forms are sometimes modernised. Evidence for this phenomenon is found in Rodríguez Molina (2006b). He demonstrates that when copying tokens belonging to the *ser* + PtcP construction in medieval texts, the copiers often replace these tokens with the more modern *haber* + PtcP (cf. section 3). The discussion of

entrenchment in this chapter suggests that these replacement mechanisms might not be aleatory. Rather, it can be assumed that entrenchment has an important effect on whether or not a linguistic element is modernised when copied: more entrenched elements may be less susceptible to modernisation than less entrenched elements. This is due to (a) the strengthened mental representation of these entrenched elements and (b) the fact that arguably, entrenched linguistic elements are more likely to be perceived as typical for the discourse tradition.

The theoretical perspective developed in this section thus suggests that there is no principled distinction between the mechanism of entrenchment in spoken and written discourse traditions. Just as speakers use entrenched linguistic elements because of repeated exposure to these elements in spoken language, writers use entrenched linguistic elements due to repeated exposure to these elements in specific discourse traditions in written language. Intertextuality follows similar rules in written language as in spoken language.

If there is a difference regarding entrenchment in written and spoken language, it is quantitative in nature. In particular, this is due to differences in the way discourse traditions in written and spoken language are constituted. Spoken discourse traditions are mainly determined by factors regarding the situational context. For instance, due to the use of stereotypical and recurrent patterns, it is licit to characterise the dialogues between customer and clerk at grocery store counters as a discourse tradition. This discourse tradition is basically determined by the situational context: The speakers do not situate their speech in the discourse tradition in order to explicitly refer to earlier language usage events. Rather, entrenchment mainly takes place on a subconscious level. In contrast, references to earlier usage events in written discourse traditions are caused by more deliberate processes: Writers reflect on the fact that within their discourse tradition, a certain linguistic element is entrenched, and often deliberately use this element in order to follow the stylistic prescripts of that discourse tradition. Due to the fact that written texts persist longer in time than spoken texts, this also means that archaisms are more likely to occur in written language than in spoken language. As a result, written discourse traditions are typically expected to be more conservative than spoken discourse traditions.

Crucially, this assumption also leads to a sub-division within written texts. Thus, some written texts share more characteristics with spoken texts than others. Following Koch and Oesterreicher's terminology, some texts are characterised by *communicative proximity*, whereas other texts are characterised by *communicative distance* (Koch and Oesterreicher 1985; 1990; 2001). As illustrated by the grocery store example, it is this parameter which governs the conservatism of some discourse traditions in relation to other discourse traditions.

Summarising these considerations, I establish two hypotheses:

1. The cognitive mechanism of entrenchment plays a pivotal role in the emergence of discourse traditions both in the written and the spoken modality. By repeatedly using elements that are typical of a certain discourse situation, a writer or speaker signals that her text belongs to a specific discourse tradition. Conversely, entrenchment can be conceptualised as a process by which a linguistic element increasingly comes to depend on intertextuality. This means that discourse traditions have an enabling function for entrenchment: Entrenchment operates within specific discourse traditions.
2. Discourse traditions characterised by communicative distance display a more conservative language use than discourse traditions characterised by communicative proximity.

The next section of this paper will illustrate the relevance of these hypotheses for the example of Spanish auxiliary selection.

### 3. Auxiliary Selection in Spanish

In contrast to Modern Spanish, Old Spanish possessed two auxiliaries for the formation of the verbal periphrases commonly termed *compound tenses* in intransitive sentences: *ser* ('be') and *haber* ('have'). The two construction types are illustrated in (1) and (2). In contrast, *haber* is used invariably in transitive sentences (3).<sup>1</sup>

- (1) *perdonamus a don Marcus e a los sobreditos cavalleros el*  
 pardon.PRS.1PL to Don Marcus and to the abovementioned knights the  
*daño que nos es venido por la mala voz*  
 damage that to.us be.PRS.3SG come.PTCP.M.SG because.of the bad voice  
*que nos pusieron*  
 that to.us put.PST.PFV.3SG

'We forgive Marcus and the abovementioned knights the damage caused by their attacks on our reputation' (*Carta mediante la cual don Marco de Artieda y otros muchos piden perdon al Monasterio de San Salvador y este se lo concede*, March 1262, apud CODEA)

- (2) *porque Sancho Pérez avía venido e entrado en*  
 because Sancho Pérez have.PST.IPFV.3SG come.PTCP.M.SG and enter.PTCP.M.SG in  
*la iglesia*  
 the church

'Because Sancho Pérez had come and entered the church' (*Carta mediante la cual Joán Gómez encomienda a Pascual Martínez la iglesia de Alhama y le entrega las llaves de dicha iglesia*, September 13 1417, apud CODEA)

- (3) *todas las mejorías, e aprovechamientos e edificaciones que*  
 all the improvements and improvements and constructions that  
*en las dichas casas [...] oviéredes fecho, e*  
 in the say.PTCP.F.PL houses have.PST.IPFV.SBJ.2PL make.PTCP.M.SG and  
*labrado e mejorado*  
 work.PTCP.M.SG and improve.PTCP.M.SG

'All the improvements that you have made in the said houses' (*Carta de venta de unas casas de Catalina Rodríguez a Juan Sánchez y Elvira Sánchez*, August 29 1447, apud CODEA)

The variation between *ser* and *haber* is to a great degree determined by verb semantics (Benzing 1931; Yllera 1980). *Ser* + PtcP is typically used with intransitive verbs that express a change of state or a change of location (4-5), whereas *haber* + PtcP is typically used with intransitive verbs that express states or activities. (6-7). This means that in Old Spanish texts, examples such as (2) are rather atypical.

<sup>1</sup> Sources marked in capital letters correspond to the following corpora: CODEA = *Corpus de Documentos Españoles Anteriores a 1700* (Grupo de Investigación de Textos para la Historia del Español 2012), BIBEL = *Biblia Medieval* (Enrique-Arias 2008), LETTERS = *Cartas de Particulares en Indias de Siglo XVI* (Fernández Alcaide 2009), HIST = *Corpus de la selección de auxiliares en textos historiográficos* (Rosemeyer 2014). The HIST corpus mainly consists of tokens extracted from the *Corpus Diacrónico del Español* (Real Academia Española 2012). For a more detailed description of the corpora, cf. section 4.

- (4) *cuerpo de ombre | que maten o que sea por si*  
 body of man that kill.PRS.SBJ.3PL or that be.PRS.SBJ.3SG because.of REFL

**muerto**

die.PTCP.M.SG

‘Corpse of a man who has been killed or who has died on his own’ (*Bible E8/6*, 13th c., apud BIBLE)

- (5) *Et eillos aun, no heran llegados, quando vino*  
 and they yet not be.PST.IPFV.3PL arrive.PTCP.M.PL when come.PST.PFV.3SG  
*al rey mandado que absalon matara a todos los hijos*  
 to.the king message that Absalon kill.PST.IPFV.SBJ.3SG to all the sons

‘And they had not yet arrived, when the King received a message which said that Absalon was killing all of the sons’ (*Bible E8/6*, 13th c., apud BIBLE)

- (6) *assaz auedes estado en esta montanna*  
 enough have.PRS.2PL stay.PTCP.M.SG in this mountain

‘You have stayed long enough on this mountain’ (*Bible E8/6*, 13th c., apud BIBLE)

- (7) *quando el ouo fablado folgo el espiritu deillo*  
 when he have.PST.PFV.3SG speak.PTCP.M.SG rejoice.PST.PFV.3SG the spirit of.him

‘When he had spoken, he was happy’ (*Bible E8/6*, 13th c., apud BIBLE)

This distinction in the distribution of the auxiliaries in older stages of Spanish has been explained as (a) a universal contrast between unaccusative and unergative verbs (Elvira González 2001), (b) an interaction between argument structure and verbal aspect (Aranovich 2003; Mateu 2009) and (c) a functional contrast between the resultative *ser* + PtcP construction and the perfective *haber* + PtcP construction (Rodríguez Molina 2006a; Rosemeyer 2012; 2014; Mateu and Massanell this volume). Since this paper is concerned with the development of Spanish auxiliary selection, I will not discuss this issue here but refer the reader to the studies mentioned above.

After the first half of the 15th century writers start replacing the *ser* + PtcP construction with the *haber* + PtcP construction in active intransitive sentences. In texts dating back to the end of the 17th century, *ser* + PtcP has mostly disappeared from these usage contexts. Especially since the beginning of the 16th century there are many examples of the use of *haber* + PtcP in intransitive contexts expressing a change of state or change of location:

- (8) *pues no me e muerto do muchas graçias*  
 since not REFL have.PRS.1SG die.PTCP.M.SG give.PRS.1SG many thanks  
*a mi dios*  
 to my God

‘I thank God because I have not died’ (*De Alonso de Herrojo a su mujer, Teresa González, vecina de Reina (Badajoz)*, March 10 1583, apud LETTERS)

- (9) *auia nueue meses que se auia ydo*  
 have.PST.IPFV.3SG nine months that REFL have.PST.IPFV.3SG go.PTCP.M.SG

‘Nine months ago he had gone away’ (*De Celedon Favalis a su padre, Simón Favalis, en*

Madrid, March 20 1587, apud LETTERS)

- (10) *Vos, señor, habéis crecido mucho*  
 you master have.PRS.2PL grow.PTCP.M.SG much

‘You, master, have grown much’ (*Historia de la vida y hechos del Emperador Carlos V, 1604-1611, apud HIST*)

- (11) *Andrea Gasparo Gorço, que por el Rey Católico había*  
 Andrea Gasparo Gorço that because.of the king catholic have.PST.IPFV.3SG  
*vuelto a Fez*  
 return.PTCP.M.SG to Fez

‘Andrea Gasparo Gorço, who had returned to Fez because of the Catholic King’ (*Historia de Felipe II, rey de España, 1619, apud HIST*)

In Aranovich (2003) and Rosemeyer (2014) this replacement process is assumed to result from *actualisation* (Andersen 2001; De Smet 2012). Following usage-based approaches to language change such as Bybee (2003) and Barðdal (2008), it can be suggested that the grammaticalisation of the *haber* + PtcP construction causes an increase in the productivity of the construction and particularly, a gradual expansion of its type frequency. As a result, the writers start using verbs of change of state and change of location in the *haber* + PtcP construction. The actualisation process is determined by the similarity of these usage contexts to the original usage context of *haber* + PtcP, i.e. transitive sentences. In Rosemeyer (2014), I model this process as illustrated in Fig. 1. Thus, in contexts classified as syntactically and semantically intransitive, predicates that do not contain a transition in their event template [-DIR] and are used in atelic configurations [-TE] are first to be affected by the actualisation process, followed in a first step by atelic predicates of change of state and change of location [+DIR,-TE], and in a second step by telic predicates of change of state and change of location [+DIR,+TE].<sup>2</sup>

Syntactically transitive, semantically transitive	Mixed forms	Syntactically intransitive, semantically intransitive
	Locative adverbials in object position	
Transitive sentences	Sentences with dative arguments	[-DIR] [+DIR] [+DIR]
	“Absolute” transitives	[-TE] [-TE] [+TE]
	Reflexives and reciprocals	

Fig. 1. The actualisation of *haber* + PtcP (Rosemeyer 2014: 265)

In addition to this actualisation process, frequency effects have had an important influence on the development of the replacement of *ser* + PtcP with *haber* + PtcP. In Rosemeyer (2013; 2014), evidence is given for the existence of conserving effects in the development of Spanish auxiliary selection. In line with the notion of entrenchment introduced in section 2, these studies demonstrate that in a corpus of historiographical texts, the use of *ser* + PtcP persists longer with high-frequency verbs than low-frequency verbs, since *ser* + PtcP syntagms formed with these verbs have a high token frequency and are therefore susceptible to entrenchment.

<sup>2</sup> Fig. 1 also illustrates the predominance of the criterion of transitivity. For instance, compound tense of reflexive verbs, which can be characterised as syntactically transitive, are typically formed with *haber* already in Old Spanish. This observation illustrates the need to take into account both syntactic and semantic factors in the analysis of the development of auxiliary selection systems, as suggested by Loporcaro (2011) and Loporcaro (this volume).

In Rosemeyer (2014), it is also shown that there are systematic asymmetries in the frequency of use of the verbs investigated in the study (cf. section 4 for a list). In particular, the semantic parameter of a change-of-location meaning is positively correlated with the usage frequency of verbs. In other words, many of the high-frequency verbs express a change-of-location situation. Examples are verbs such as *ir* ('go'), *venir* ('come'), *pasar* ('pass'), *partir* ('go away') and *volver* ('return'). The majority of late *ser* + PtcP tokens involve participles formed from these verbs. While the semantic parameter of change of location does not significantly influence Old Spanish auxiliary selection, this parameter starts determining auxiliary selection in Early Modern Spanish, favouring *ser* + PtcP over *haber* + PtcP. Synchronic evidence for this rule in the 16th century is found in Rosemeyer (2013). This observation suggests a similar situation in Modern French, where the selection of *être* ('be') instead of *avoir* ('have') is often motivated by the parameter of change-of-location semantics (Kailuweit 2011). Interestingly, Heidinger's (this volume) case study of the French verb *monter* ('go up'), suggests similar developments in French and Spanish. Between the 16th and 20th century, *être*-selection becomes more probable, relatively speaking, according to whether or not *monter* denotes a change of location. It seems reasonable to interpret this development as the result of a conserving effect: During the slow process of replacement of *être* with *avoir*, *être*-selection persists longer with change-of-location verbs due to their high token frequency. This development leads to the creation of a synchronic grammatical rule: "when expressing a change-of-location event, use *être* over *avoir*". Similar observations apply for German (Gillmann this volume).

In Rosemeyer (2014) I restrict my analysis to a sole discourse tradition, i.e. historiographical texts. This decision was taken (a) in order to avoid skewing of the data through the simultaneous analysis of several discourse traditions and (b) because that discourse tradition is most stable across the investigated time period and offers enough data to license a quantitative analysis. In accordance with the discourse tradition approach to language history introduced in section 2, one can assume that both the use of auxiliary selection and its development differ from discourse tradition to discourse tradition. In particular, it is to be expected that the conserving effect demonstrated in Rosemeyer (2014) is dependent on discourse tradition. Stolova (2009) provides preliminary evidence for this assumption. Stolova analyses 372 *ser* + PtcP tokens from texts after 1700 in the *Corpus del Español* (Davies 2002-). These tokens are manifestations of a fossilised use of *ser* + PtcP. They are formed from the verbs *bajar* ('do down'), *correr* ('run'), *entrar* ('enter'), *ir* ('go'), *llegar* ('come'), *partir* ('go away'), *pasar* ('pass'), *salir* ('leave'), *subir* ('go up'), *tornar* ('return') and *volver* ('return'). All of these verbs typically express a change-of-location event, which can be considered evidence for the conserving effect demonstrated in Rosemeyer (2014). Stolova also demonstrates how the use of these fossilised items perpetuates discourse traditions (Hypothesis 1). For instance, in the examples (12) and (13) taken from Stolova's data a *ser* + PtcP token from the 16th century (12) is copied into a 19th century poem (13).

- (12) *Como el ciervo huiste, habiéndome herido. / Salí*  
 like the deer flee.PST.PFV.2SG have.prog.REFL wound.PTCP.M.SG leave.PST.PFV.1SG  
*tras ti clamando y eras ido*  
 after you shout.PROG and be.PST.IPFV.2SG go.PTCP.M.SG

'Like a deer you fled after wounding me / I followed you shouting and you were gone'  
 (*Cántico spiritual*, 1542-1591, apud Stolova [2009: 395])

- (13) *La vida transpusiste! .../ Hermanito querido; Salí tras*  
 the life change.PST.PFV.2SG little.brother love.PTCP.M.SG leave.PST.PFV.1SG after  
*ti clamando ... y eras ido!*

you shout.PROG and be.PST.IPFV.2SG go.PTCP.M.SG

‘You turned my life upside down! / beloved little brother; I followed you shouting ... and you were gone’ (*En la muerte de un hermano niño*, 1811-1863, apud Stolova [2009: 395])

Stolova furthermore suggests that the use of these fossilised *ser* + PtcP syntagms differs according to discourse tradition. In the legal texts in her corpus, *ser* + PtcP typically appears in the fossilised form *era llegado el caso* (‘the case had arisen’) (14). In narrative texts, *ser* + PtcP is typically used to indicate time, as in *era llegado septiembre* (‘September had come’) (15). While the latter two uses of *ser* + PtcP are metaphorical, *ser* + PtcP typically expresses the literal event of a change of location in historiographical texts, like in (16).

- (14) *y no era llegado el caso de examinarse*  
 and not be.PST.IPFV.3SG arrive.PTCP.M.SG the case of examine.INF.REFL  
*si ésta incurrió realmente en las contravenciones*  
 if she commit.PST.PFV.3SG really in the violations

‘And it had not been examined whether she had indeed committed these violations of the law’ (*Revista de derecho y jurisprudencia*, 20 January 1927, apud Stolova 2009: 396)

- (15) *había pasado el verano y era llegado*  
 have.PST.IPFV.3SG pass.PTCP.M.SG the summer and be.PST.IPFV.3SG arrive.PTCP.M.SG  
*septiembre*  
 September

‘Summer had passed and September had come’ (*La gaviota*, 1849, apud Stolova 2009: 396)

- (16) *El capitán le respondió que había poco tiempo*  
 the captain to.him answer.PST.PFV.3SG that have.PST.IPFV.3SG little time  
*que era llegado á la isla*  
 that be.PST.IPFV.3SG arrive.PTCP.M.SG to the island

‘The captain answered him that he had only recently come to the island’ (*Viajes al Maluco de Fray García de Loáisya y Álvaro de Saavedra*, 1837, apud Stolova 2009: 397)

In summary, Stolova’s results suggest both similarities and differences between entrenchment and conservation mechanisms between different discourse traditions. On the one hand, the distribution of verbs in the *ser* + PtcP construction appears to be relatively homogeneous across discourse traditions: *ser* + PtcP is typically used with verbs expressing a change of location. On the other hand, the function of the conserved construction appears to differ according to discourse tradition. An explanation for this difference may reside in whether the construction expresses a metaphorical or literal meaning.

#### 4. Data

The analyses conducted in this study are based on 4680 *haber* + PtcP and *ser* + PtcP tokens from Spanish texts dated between 1260 and 1688. According to the criteria developed in Rosemeyer (2014), only tokens that fall into the envelope of variation between the two constructions were included in the corpus, leading to the exclusion of, for example, syntactically transitive tokens. Likewise, the study is based on the 43 verbs selected in Rosemeyer (2014) for the investigation of the development of Spanish auxiliary selection. These verbs vary systematically according to the semantic features “change of state” [DIR],

“telicity” [TE], and “change of location” [MO], and control by the subject referent [CON], as well as overall usage frequency of the verb. Table 1 summarises these verbs and their typical semantic values regarding the features [DIR] and [TE] which have special relevance for the description.

**Table 1. Verbs investigated in this study**

Predicate class	Description	Verbs and their typical translation
[+DIR,+TE]	Telic change of state or change of location	<i>acaecer</i> (‘happen’), <i>ahogar</i> (‘drown’), <i>alzar</i> (‘rise’), <i>aparecer</i> (‘appear’), <i>avenir</i> (‘happen’), <i>ayuntar</i> (‘gather’), <i>cabalgar</i> (‘mount’), <i>caer</i> (‘fall down’), <i>cuntir</i> (‘happen’), <i>espantar</i> (‘get frightened’), <i>finar</i> (‘end’, ‘die’), <i>hundir</i> (‘sink’), <i>maravillar</i> (‘get astonished’), <i>morir</i> (‘die’), <i>nacer</i> (‘be born’), <i>pasar</i> (‘pass’), <i>perecer</i> (‘perish’), <i>quebrar</i> (‘break’), <i>sentar</i> (‘sit down’), <i>tornar</i> (‘return’), <i>venir</i> (‘come’), <i>volver</i> (‘return’)
[+DIR,-TE]	Atelic change of state or change of location	<i>andar</i> (‘walk’), <i>caminar</i> (‘walk’), <i>correr</i> (‘run’), <i>corromper</i> (‘corrode’), <i>crecer</i> (‘grow’), <i>descender</i> (‘go down’), <i>escapar</i> (‘escape’), <i>exir</i> (‘leave’), <i>huir</i> (‘flee’), <i>ir</i> (‘go’), <i>menguar</i> (‘diminish’), <i>partir</i> (‘leave’), <i>subir</i> (‘go up’, ‘rise’)
[-DIR,-TE]	State or persistence of pre-existing state	<i>arrepentir</i> (‘repent’), <i>durar</i> (‘last’), <i>fincar</i> (‘stay’), <i>holgar</i> (‘stay’), <i>morar</i> (‘stay’, ‘live’), <i>quedar</i> (‘stay’), <i>sobrar</i> (‘remain’), <i>yacer</i> (‘lie’)

The corpus is divided into three smaller subcorpora, each representing a specific discourse tradition. The first and biggest subcorpus – HIST – is taken from the data collected in Rosemeyer (2014). HIST represents the auxiliary selection in 44 Spanish historiographical texts from between 1270 and 1698 and consists of 4158 tokens. These texts mainly stem from the *Corpus Diacrónico del Español* (Real Academia Española 2012). The selection of the texts was guided by the criteria established in Fernández-Ordóñez (2006), thus guaranteeing a relative authenticity of the manuscripts used in the editions of the texts.

The second subcorpus – CODEA – represents auxiliary selection in Spanish documents from the royal chancellery, ecclesiastical institutions and municipal administrations, as well as less formal letters between dignitaries from these institutions. It consists of 100 tokens taken from the *Corpus de Documentos Españoles Anteriores a 1700* (Grupo de Investigación de Textos para la Historia del Español 2012). The small number of occurrences is a result of (a) the small overall size of the corpus and (b) the fact that the use of *haber* + PtcP and *ser* + PtcP is less frequent in that corpus. However, these tokens can be regarded as more representative because the texts in the CODEA corpus are based on original manuscripts. Thus, from a philological point of view these data are more reliable than the data collected in HIST.

Likewise, the third subcorpus – LETTERS – offers more reliable data than the HIST corpus. LETTERS shows auxiliary selection in private letters from Spanish immigrants in the American colonies between 1500 and 1600. It consists of 422 tokens. The corpus is based on the edition of these letters from original manuscripts by Marta Fernández Alcaide (2009). As pointed out in Fernández Alcaide’s study, the corpus provides important data, since private letters constitute a discourse tradition characterised by a high *communicative proximity* (in terms of Koch and Oesterreicher 1985; 1990; 2001). Thus,

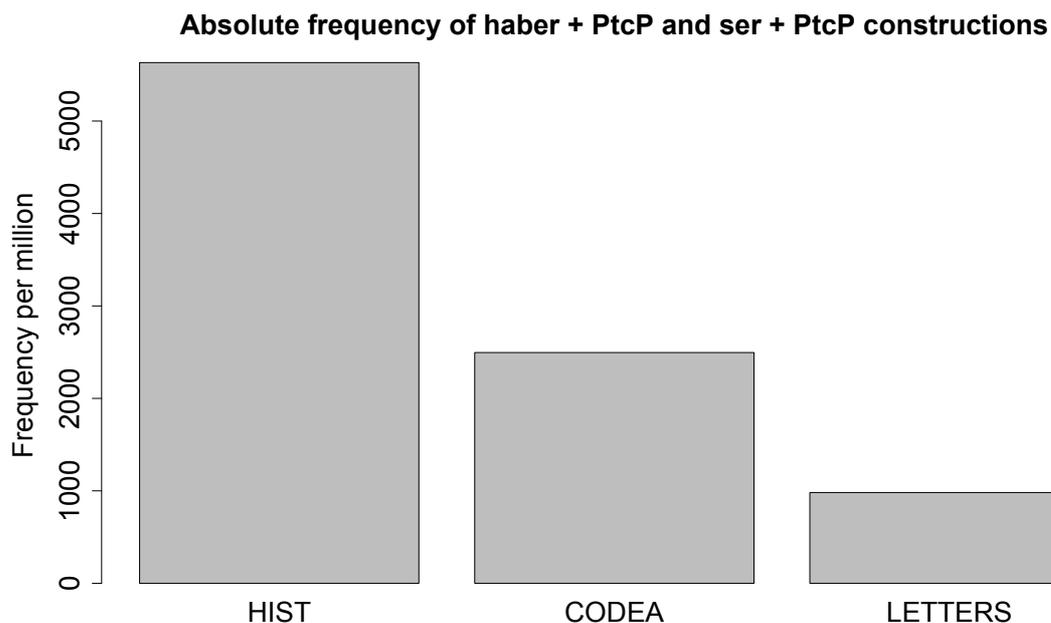
[the corpus is characterised] by a form of discourse organisation that is essentially based on the fact that the utterances are produced with less planning than in official or more formal texts, i.e. a text characterised by communicative distance; because the linguistic



with the corresponding verb class. This finding explains why in the LETTERS corpus, *ser*-selection seems to be no less frequent than in the other corpora.

### 5.1 Usage frequency of *haber* + PtcP and *ser* + PtcP

Figure 3 illustrates the approximate joint absolute usage frequency of *haber/ser* + PtcP- construction per million words in the three corpora. The frequency measure was normalised using the formula  $freq = \frac{n \text{ of tokens}}{n \text{ of words in the corpus}} * 1.000.000$ .



**Fig. 3 Joint usage frequency of *haber* + PtcP and *ser* + PtcP in HIST, CODEA and LETTERS<sup>3</sup>**

Within the envelope of variation selected for this study, the use of *haber* + PtcP and *ser* + PtcP constructions is most frequent in historiographical texts (HIST), followed by administrative documents (CODEA), and private letters (LETTERS). Arguably, this variation results from the differences in text function between the three discourse traditions. The verbs investigated in this study (cf. Table 1) are typical for situations described in historiographical texts. These texts often elaborately describe successions of events. In contrast, administrative texts and private letters usually refer to events in isolation and tend to describe the emotional or legal consequences of these events.

Theoretically, these differences in the joint usage frequency of the two constructions should not lead to differences in the distribution of *haber* + PtcP and *ser* + PtcP between the three discourse traditions. As elaborated in section 3, the most important predictors of auxiliary selection in Spanish are the semantic parameters “change of state” [DIR] and

<sup>3</sup> The application of the formula is straightforward for the LETTERS and CODEA corpora because no randomisation procedure was applied. However, due to the great token number in the HIST corpus, frequent *haber* + PtcP and *ser* + PtcP types were randomised in that corpus. As a result, an approximation method had to be introduced in order to calculate the overall number of *haber* + PtcP and *ser* + PtcP tokens in that corpus. This approximation method has the following formula:

$$n \text{ of tokens} = \text{overall } n \text{ of tokens encountered using a search string} / \left( \frac{n \text{ of reviewed tokens}}{n \text{ of randomly selected tokens}} \right).$$

“telicity” [TE]. Tables 2-4 illustrate the influence of these semantic parameters on auxiliary selection over time in the three subcorpora. In these tables, “H” refers to *haber* + PtcP tokens, whereas “S” refers to *ser* + PtcP tokens.

**Table 2. Development of auxiliary selection over time according to [DIR] and [TE] in HIST**

	1200-1299			1300-1399			1400-1499			1500-1599			1600-1699		
	H	S	% S	H	S	% S	H	S	% S	H	S	% S	H	S	% S
[+DIR,+TE]	26	405	94.0	41	397	90.6	144	551	79.3	468	189	28.8	426	75	15.0
[+DIR,-TE]	15	75	83.3	33	128	79.5	92	150	62.0	187	46	19.7	251	15	5.6
[-DIR,-TE]	56	16	22.2	43	28	39.4	98	19	16.2	88	3	3.3	93	0	0.0
<b>TOTAL</b>	97	496	83.6	117	553	82.5	334	720	68.3	743	238	24.3	770	90	10.5

**Table 3. Development of auxiliary selection over time according to [DIR] and [TE] in CODEA**

	1200-1299			1300-1399			1400-1499			1500-1599			1600-1699		
	H	S	% S	H	S	% S	H	S	% S	H	S	% S	H	S	% S
[+DIR,+TE]	0	4	100.0	0	5	100.0	1	5	83.3	19	15	44.1	10	1	9.1
[+DIR,-TE]	0	1	100.0	(NA)			1	0	0.0	18	3	14.3	2	0	0.0
[-DIR,-TE]	0	1	100.0	(NA)			(NA)			5	0	0.0	9	0	0.0
<b>TOTAL</b>	0	6	100.0	0	5	100.0	2	5	71.4	42	18	30.0	21	1	4.8

**Table 4. Development of auxiliary selection over time according to [DIR] and [TE] in LETTERS**

	1200-1299			1300-1399			1400-1499			1500-1599			1600-1699		
	H	S	% S	H	S	% S	H	S	% S	H	S	% S	H	S	% S
[+DIR,+TE]	(NA)			(NA)			(NA)			205	113	35.5	(NA)		
[+DIR,-TE]	(NA)			(NA)			(NA)			61	10	14.1	(NA)		
[-DIR,-TE]	(NA)			(NA)			(NA)			33	0	0.0	(NA)		
<b>TOTAL</b>	(NA)			(NA)			(NA)			299	123	29.2	(NA)		

First, the data summarised in Tables 2-4 demonstrates that only in the 16th century there are enough tokens to license a quantitative comparison of the use of auxiliary selection in the three discourse traditions.

Second, the distribution of *haber* + PtcP and *ser* + PtcP according to [DIR] and [TE] is similar in all of the corpora. Thus, *ser*-selection is most frequent with telic predicates of change of state or change of location [+DIR,+TE], followed by atelic predicates of change of state or change of location [+DIR,-TE] and lastly predicates of state or continuation of a pre-existing state which are always atelic [-DIR,-TE]. However, there also seem to be slight quantitative differences. In particular, it seems that in the 16th century, *ser*-selection is relatively more frequent for the predicate class [+DIR,+TE] in CODEA (44.1 per cent) and LETTERS (35.5 per cent) than in HIST (18.8 per cent). In contrast, *ser*-selection in the predicate classes [+DIR,-TE] and [-DIR,-TE] is relatively more frequent in HIST than in CODEA and LETTERS. This suggests that in CODEA and LETTERS, the grammatical productivity of the *ser* + PtcP construction is more restricted to a specific predicate class than in HIST. The analyses in the next section will demonstrate that this finding is actually an effect of differences in the frequency with which change of location predicates are expressed in the three discourse traditions.

Third, a comparison of auxiliary selection in the 16th century section of the three corpora (marked by a grey shade in Tables 2-4) seems to suggest that *ser*-selection is relatively less frequent in historiographical texts than in the other two discourse traditions. For sake of clarity, the distribution of the two constructions in the 16th century section of the three corpora is summarised in Table 5.

**Table 5. Auxiliary selection in the 16th century section of HIST, CODEA, and LETTERS**

Corpus	H	S	% S
HIST	743	238	24.3
CODEA	42	18	30.0
LETTERS	299	123	29.2

Although there seem to indeed be differences in the distribution of auxiliary selection in the three corpora, these differences do not reach statistical significance ( $\chi^2 = 4.216$  (2),  $p = 0.121$ ). Consequently, the data from the corpus do not warrant the conclusion that there is a difference in the distribution of *haber* + PtcP and *ser* + PtcP in the three discourse traditions.

This result is surprising because it contradicts Hypothesis 2. The discourse tradition represented by the LETTERS corpus, i.e. private letters, is characterised by a high degree of communicative proximity. Thus, it would be expected that the historical process by which *ser* + PtcP came to be replaced with *haber* + PtcP is most advanced in this discourse tradition. Although in line with this hypothesis, the grammatical productivity of *ser* + PtcP is more restricted in the LETTERS corpus, the overall frequency of *ser*-selection does not appear to reflect this difference. The analysis of the functional differences between the two constructions in the three discourse traditions will, however, offer an explanation for this finding that allows Hypothesis 2 to be maintained.

## 5.2 Functional differences in auxiliary selection

In the last section, two observations regarding the 16th century data were made that have no straightforward explanation. First, the raw data in Tables 2-4 suggests that *ser*-selection with telic change of state or change of location predicates [+DIR,+TE] is relatively more frequent in CODEA and LETTERS than in HIST. Second, it was shown that in terms of the overall usage frequency of *haber* + PtcP and *ser* + PtcP, the distribution of the two constructions does not appear to differ. This finding contradicts the hypothesis that discourse traditions characterised by communicative proximity display a more modern language use than discourse traditions characterised by communicative distance.

In this section, I demonstrate how the assumption of conserving effects in language change can provide an explanation for these two findings. In addition, the discussion illustrates the context-dependency of entrenchment postulated in section 2.

A more detailed analysis of the distribution of auxiliary selection according to predicate class suggests that the semantic parameter “change of location” [MO] can explain the contradictory results from the last section. Predicates of change of location are a subclass of predicates of change of state (cf. Table 1). Consequently, variation in auxiliary selection according to change of location predicates can explain the apparent variation in auxiliary selection according to change of state predicates. In Table 6, I restrict the data to 16th century *haber* + PtcP and *ser* + PtcP tokens that express a change of state [+DIR,-MO] or change of location [+DIR,+MO] and illustrate the percentage with which a token from this predicate class in the three corpora expresses a change of location [+DIR,+MO].

**Table 6. Usage frequency of change of location predicates [+DIR, +MO] relative to the usage frequency of change of state predicates [+DIR,-MO] in HIST, CODEA and LETTERS in the 16th century**

	[+DIR,+MO]	[+DIR,-MO]	% [+DIR,+MO]
HIST	503	387	56.5
CODEA	37	18	67.3
LETTERS	313	76	80.5

Compared to other situations of change of state, historiographical texts describe fewer events that can be characterised as change of location events than legal texts and private letters. While the difference between HIST and CODEA does not reach statistical significance ( $\chi^2 = 2.028$  (1),  $p = 0.155$ ), the difference between HIST and LETTERS does ( $\chi^2 = 4.264$  (1),  $p = 0.039^*$ ). I summarise this observation in (17).

- (17) Observation 1: In comparison to historiographical texts in the 16th century, private letters from the 16th century describe significantly more events of change of location than other events of change of state.

In the light of observation 1, it is worth investigating whether the differences between the three corpora in terms of the relative usage frequency of *ser* + PtcP with predicates marked as [+DIR,+TE] are epiphenomenal and actually derive from the differences in the usage frequency of predicates marked as [+DIR,+MO]. Table 7 illustrates the 16th century auxiliary selection in the three corpora according to whether the predicate refers to a change of location [+DIR,+MO], a change of state [+DIR,-MO] or a state viz. continuation of a pre-existing state [-DIR,-MO].

**Table 7. Auxiliary selection in the 16th century according to subcorpus and the predicate classes [+DIR,+MO], [+DIR,-MO] and [-DIR,-MO].**

	HIST			CODEA			LETTERS		
	H	S	%S	H	S	%S	H	S	%S
[+DIR,+MO]	378	125	24.9	27	10	27.0	192	121	38.7
[+DIR,-MO]	277	110	28.4	10	8	44.4	74	2	2.6
[-DIR,-MO]	88	3	3.3	5	0	0.0	33	0	0.0
<b>TOTAL</b>	743	238	24.3	42	18	30.0	299	123	29.2

In the 16th century section of the HIST and CODEA corpora, the difference between change of state and change of location predicates does not have a significant influence on auxiliary selection. Neither in the HIST corpus nor in the CODEA corpus does the difference between the conditions [+DIR,+MO] and [+DIR,-MO] reach statistical significance (HIST:  $\chi^2 = 1.259$  (1),  $p = 0.262$ , CODEA:  $\chi^2 = 0.971$  (1),  $p = 0.324$ ). In contrast, the difference between these two types of predicates has a crucial influence on auxiliary selection in the LETTERS corpus. First, only in two of 76 cases in the LETTERS corpus in which the predicate refers to a change of state, *ser* is used as the auxiliary (2.6 per cent). Second, in 121 of 313 cases in which the predicate expresses a change of location in the LETTERS corpus, *ser* is used as the auxiliary (38.7 per cent). This effect reaches a very high statistical significance ( $\chi^2 = 35.061$  (1),  $p < 0.001^{***}$ ). I summarise this observation in (18):

- (18) Observation 2: When writing about events of change of location, the authors of the texts in the LETTERS corpus use *ser* + PtcP significantly more often than when writing about other events of change of state. No similar statistically significant effect can be observed in the 16th century sections of the HIST corpus and the CODEA corpus.

A last observation drawn from Table 7 refers to the differences between the three discourse traditions with respect to the usage frequency of *ser* + PtcP with predicates

expressing a mere change of state [+DIR,-MO], as well as a state or a continuation of a pre-existing state [-DIR,-MO]. With these predicates, *ser*-selection is least frequent in the private letters. Only the difference in auxiliary selection in the domain of predicates marked as [+DIR,-MO] between the HIST corpus and the LETTERS corpus reaches statistical significance ( $\chi^2 = 22.547$  (1),  $p < 0.001$ \*\*\*). I summarise this third observation in (19).

- (19) Observation 3: When describing situations of change of state [+DIR,-MO], *ser*-selection is significantly less frequent in LETTERS than in the 16th century section of HIST.

The three observations in (17-19) explain the contradictory findings from the last section, namely the facts that apparently, *ser*-selection is as frequent in the LETTERS corpus as in the other corpora and that *ser*-selection is more frequent in predicates marked as [+DIR,+TE] in the LETTERS corpus than in the HIST corpus. Observation 3 suggests that the parameter [DIR], i.e. the fact that the situation refers to a change of state, is not relevant for auxiliary selection in the LETTERS corpus. This means that the high frequency of *ser*-selection in the predicate classes [+DIR,+TE] and [+DIR,-TE] in the LETTERS corpus results from the fact that the use of *ser*-selection is very frequent with change of location predicates in the LETTERS corpus (observation 2) and that in the LETTERS corpus, more than 80 per cent of the predicates marked as [+DIR] are change of location predicates (observation 1).

The interplay between entrenchment and discourse traditions as postulated in Hypothesis 1 has explanatory potential for this result. Crucially, the observations 1-3 suggest a positive correlation between the frequency with which a predicate of change of location is used in a discourse tradition and the frequency of *ser*-selection with change of location predicates in that discourse tradition. In the data investigated in this study, the conserving effect caused by entrenchment is strongest in the discourse tradition in which change of location predicates are most frequent, i.e. private letters. Although in Rosemeyer (2013; 2014) I have demonstrated a similar trend for historiographical texts, from a synchronic perspective this trend is not strong enough to result in a significant difference in auxiliary selection between change-of-state and change-of-location predicates.

The detailed analysis of the data in this section has also demonstrated that Hypothesis 2 can be maintained when describing the development of auxiliary selection in the three discourse traditions. In the LETTERS corpus, the grammatical productivity of *ser* + PtcP has apparently receded to the degree that it is exclusively used with a row of entrenched syntagms formed from change of location verbs. In contrast, the type frequency of *ser* + PtcP is higher in the HIST and CODEA corpora. This suggests that in comparison to the discourse traditions represented by these latter corpora, the replacement of *ser* with *haber* has progressed faster in discourse traditions characterised by communicative proximity – in this case, private letters. The high absolute frequency of *ser*-selection in the LETTERS corpus results from the fact that in that discourse tradition, change of location predicates are very frequent.

## 6. Summary and discussion

In this paper, I have analysed the differences in auxiliary selection between three discourse traditions in Spanish – historiographical texts, administrative documents, and private letters. An analysis of these differences in the 16th century section of the corresponding corpora suggests a positive correlation between the frequency with which a

predicate of change of location is used in a discourse tradition and the frequency of *ser*-selection with change of location predicates in that discourse tradition.

On the one hand, these results confirm the evidence proposed in Rosemeyer (2014): The conserving effect of frequency specifically affects change-of-location predicates, since verbs expressing this meaning are frequent in all three investigated discourse traditions.

On the other hand, the results also suggest that entrenchment in the ensuing conserving effect in language change differs quantitatively across the three examined discourse traditions. The analysis has revealed a difference in the distribution of *ser* + PtcP and *haber* + PtcP in 16th century historiographical texts and private letters. While the semantic parameter “change of state” is an important predictor of auxiliary selection in historiographical texts, the authors of the private letters appear to regard the semantic parameter “change of location” as the crucial parameter of auxiliary selection. Extremely few *ser* + PtcP tokens in the LETTERS corpus do not display a change-of-location semantics. I have argued that this finding needs to be explained by assuming context-dependent entrenchment: Given that change-of-location events are more frequent in private letters and that due to the communicative proximity of these texts the change has progressed further, change of location has become the most important predictor of auxiliary selection. This assumption also has explanatory power for Stolova’s (2009) finding that there are slight differences in the function of fossilised *ser* + PtcP tokens across different discourse traditions. As the discussion of the data from the LETTERS corpus has shown, entrenchment or fossilisation can lead to the reanalysis of the function of the entrenched syntagms. This reanalysis, however, can be claimed to depend on the involved discourse function.

Consequently, the study at hand proposes that it is necessary to reconcile the cognitivist approach taken in usage-based linguistics (UBL) and the pragmatic approach taken in studies of discourse traditions. This reconciliation is advantageous for either theoretical perspective.

Regarding the usage-based approach to language change, my proposal leads to a more precise definition of how entrenchment works in language change. In order to refine the notion of the frequency effect of entrenchment in language change, it is crucial to take into account the fact that entrenchment operates within specific experiential domains. The notion of discourse traditions offers a possible model for these experiential domains. Thus, accounting for discourse traditions in the study of frequency effects allows us to account for differences in usage frequency between different types of texts which cannot be explained if we assume that entrenchment effects operate globally.

Regarding the discourse tradition approach to language change, my proposal offers a way of getting clearer idea of how discourse traditions emerge. In order to study their emergence, it is not only necessary to study “text-internal factors” such as recurrent topics, but also phenomena on the “text surface”. Thus, writers recur to linguistic elements entrenched within a certain discourse tradition in order to situate their texts within that discourse tradition. As a result, the notion of entrenchment could be exploited in order to study how discourse traditions emerge and are consolidated over time. In addition, these assumptions are highly relevant for the methodology of the study of historical texts, since they suggest that in addition to the concept of discourse traditions, the concept of entrenchment is crucial for the assessment of the representativity of tokens from historical texts.

## References

- Andersen, Henning (ed.) 2001 *Actualization: Linguistic change in progress*. Amsterdam, Philadelphia: John Benjamins.
- Aranovich, Raúl 2003 The semantics of auxiliary selection in Old Spanish. *Studies in Language* 27 (1): 1-37.
- Aschenberg, Heidi 2003 Diskurstraditionen - Orientierungen und Fragestellungen. In Heidi Aschenberg and Raymund Wilhelm (eds.), *Romanische Sprachgeschichte und Diskurstraditionen*, 1-18. Tübingen: Narr.
- Barðdal, Jóhanna 2008 *Productivity: evidence from case and argument structure in Icelandic*. Amsterdam, Philadelphia: John Benjamins.
- Barthes, Roland 1967 The Death of the Author. *Aspen* 5-6. Available online at <http://www.ubu.com/asp/asp5and6/threeEssays.html-barthes>. Last access: 16 January 2012.
- Behrens, Heike 2009 Usage-based and emergentist approaches to language acquisition. *Linguistics* 47 (2): 383-411.
- Benzing, Joseph 1931 Zur Geschichte von *ser* als Hilfszeitwort bei den intransitiven Verben im Spanischen. *Zeitschrift für romanische Philologie* LI: 385-460.
- Blank, Andreas 2005 Polysemie: Sprachbewußtsein, semantische Relationen, Diskurstraditionen. In Daniel Jacob, Thomas Krefeld and Wulf Oesterreicher (eds.), *Sprache - Bewusstsein - Stil. Theoretische und historische Perspektiven*, 31-46. Tübingen: Narr.
- Burke, James 2004 Medieval Spanish prose. In David T. Gies (ed.), *The Cambridge History of Spanish Literature*, 95-104. Cambridge: Cambridge University Press.
- Bybee, Joan L. 2003 Mechanisms of change in grammaticization: the role of frequency. In Richard Janda and Brian Joseph (eds.), *The Handbook of Historical Linguistics*, 624-647. Oxford: Blackwell.
- Bybee, Joan L. 2006 From usage to grammar: the mind's response to repetition. *Language* 82 (4): 711-733.
- Bybee, Joan L. 2010 *Language, Usage, and Cognition*. Cambridge, New York: Cambridge University Press.
- Bybee, Joan L. and Paul J. Hopper (eds.) 2001 *Frequency and the emergence of linguistic structure*. Amsterdam, Philadelphia: John Benjamins.
- Dahl, Östen 1985 *Tense and aspect systems*. Oxford: Blackwell.
- Davies, Mark 2002- *Corpus del español (100 million words, 1200s-1900s)*. Available online at <http://www.corpusdelespanol.org>. Last access: 23 March 2013.
- De Smet, Hendrik 2012 The course of actualization. *Language* 88: 601-633.

- Diessel, Holger 2007 Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25: 108-127.
- Diessel, Holger 2011 Review article of 'Language, usage and cognition' by Joan Bybee. *Language* 87: 830-844.
- Diewald, Gabriele 2002 A Model for Relevant Types of Contexts in Grammaticalization. In Ilse Wischer and Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 103-120. Amsterdam, Philadelphia: John Benjamins.
- Diewald, Gabriele 2006 Context types in grammaticalization as constructions. *Constructions* 1 (9).
- Ellis, Nick 1996 Sequencing in SLA: phonological memory, chunking and points of order. *Studies in Second Language Acquisition* 18: 91-126.
- Elvira González, Javier 2001 Intransitividad escindida en español: El uso auxiliar de *ser* en español medieval. *EdLing* 15: 201-245.
- Enrique-Arias, Andrés (dir.) 2008 *Biblia Medieval*. Available online at <http://www.bibliamedieval.es>. Last access 26 May 2011.
- Fernández Alcaide, Marta 2009 *Cartas de particulares en Indias del siglo XVI. Edición y estudio discursivo*. Madrid, Frankfurt am Main: Iberoamericana, Vervuert.
- Fernández-Ordóñez, Inés 2004 Alfonso X en la historia del español. In Rafael Cano Aguilar (ed.), *Historia de la lengua española*, 381-422. Barcelona: Ariel.
- Fernández-Ordóñez, Inés 2006 La Historiografía medieval como fuente de datos lingüísticos. Tradiciones consolidadas y rupturas necesarias. In José Jesus Bustos Tovar and José Luis Girón Alchonchel (eds.), *Actas del VI Congreso Internacional de Historia de la Lengua Española*, 11, 1779-1807. Madrid: Arco Libros.
- Foucault, Michel 1988 [1969] Was ist ein Autor? In Michel Foucault (ed.), *Schriften zur Literatur*, 7-31. Frankfurt am Main: Fischer.
- Foucault, Michel 2002 [1969] *Archeology of Knowledge*. London and New York: Routledge.
- Gillmann, Melitta This volume Auxiliary selection in closely related languages: the case of German and Dutch.
- Goldberg, Adele E. 2006 *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Grupo de Investigación de Textos para la Historia del Español 2012 *Corpus de Documentos Españoles Anteriores a 1700*, Available online at <http://demos.bitext.com/codea/>. Last access 10 December 2012.
- Heidinger, Steffen This volume The auxiliary selection of French *monter* 'move upward' from the 16th to the 20th century.

Heine, Bernd 2002 On the Role of Context in Grammaticalization. In Ilse Wischer and Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 83-101. Amsterdam, Philadelphia: John Benjamins.

Held, Gudrun 2005 Der Einfluss von Höflichkeit auf die mittelalterliche Briefkunst – exemplarische Überlegungen zur Entwicklung von Textstruktur und Syntax vom *dictamen* zur freien Briefpraxis. In Angela Schrott and Harald Völker (eds.), *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*, 45-62. Göttingen: Universitätsverlag Göttingen.

Hopper, Paul J. and Elizabeth Closs Traugott 2003 *Grammaticalization*. 2. Ed. Cambridge: Cambridge University Press.

Jacob, Daniel 1994 Die Auxiliarisierung von *habere*. Habilitationsschrift, Albert-Ludwigs-Universität Freiburg.

Jacob, Daniel 2001 Representatividad lingüística o autonomía pragmática del texto antiguo? El ejemplo del pasado compuesto. In Daniel Jacob and Johannes Kabatek (eds.), *Lengua medieval y tradiciones discursivas en la Península Ibérica*, 153-176. Frankfurt am Main: Vervuert.

Jacob, Daniel and Johannes Kabatek (eds.) 2001 *Lengua medieval y tradiciones discursivas en la Península Ibérica: Descripción gramatical, pragmática histórica, metodología*. Frankfurt am Main: Vervuert.

Johnson, Keith 1997 Speech perception without speaker normalization. In Keith Johnson and John W. Mullennix (eds.), *Talker variability in speech processing*, 145–165. San Diego, Cal.: Academic Press.

Kabatek, Johannes 2005a Las tradiciones discursivas del español medieval: historia de textos e historia de la lengua. *Iberoromania* 62: 28-44.

Kabatek, Johannes 2005b Tradiciones discursivas y cambio lingüístico. *Lexis* 29 (2): 151-177.

Kabatek, Johannes 2013 ¿Es posible una lingüística histórica basada en un corpus representativo? *Iberoromania* 77 (1): 8-28.

Kailuweit, Rolf 2011 Le choix de l'auxiliaire: *Être* ou *avoir* en français standard contemporain. In Renata Enghels, Machteld Meullemann and Clara Vanderschueren (eds.), *Peregrinatio in Romania: Artículos en homenaje a Eugen Roegiest con motivo de su 65 cumpleaños*, 397–420. Gent: Academia Press.

Koch, Peter 1997 Diskurstraditionen: zu ihrem sprachtheoretischen Status und ihrer Dynamik. In Barbara Frank, Thomas Haye and Doris Tophinke (eds.), *Gattungen mittelalterlicher Schriftlichkeit*, 43-79. Tübingen: Narr.

Koch, Peter and Wulf Oesterreicher 1985 Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgebrauch. *Romanistisches Jahrbuch* 36: 15-43.

Koch, Peter and Wulf Oesterreicher 1990 *Gesprochene Sprache in der Romania : Französisch, Italienisch, Spanisch*. Tübingen: Niemeyer.

Koch, Peter and Wulf Oesterreicher 2001 Gesprochene Sprache und geschriebene Sprache. Langage parlé et langage écrit. In Günter Holtus, Michael Metzelin and Christian Schmitt (eds.), *Lexikon der Romanistischen Linguistik*, I,2, 584-627. Tübingen: Niemeyer.

Kuhl, Patricia K. 1994 Learning and representation in speech and language. *Current opinion in Neurobiology* 4: 812-822.

Lacerda, Francisco 1995 The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In Kjell Elenius and Peter Branderup (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, II, 140-147. Stockholm: KTH, Stockholm University.

Langacker, Ronald W. 1987 *Foundations of cognitive grammar. Theoretical prerequisites*. Stanford: Stanford University Press.

Lieven, Elena V. M., Julian M. Pine and Gillian Baldwin 1997 Lexically-based learning and early grammatical development. *Journal of Child Language* 24: 187-219.

Loporcaro, Michele 2011 A euroversal in a global perspective: auxiliation and alignment. In Peter Siemund (ed.), *Linguistic Universals and Language Variation*, 55-91. Berlin, New York: Mouton de Gruyter.

Loporcaro this volume Perfective auxiliation with reflexives in Medieval Romance: syntactic vs. semantic gradients

Mateu, Jaume 2009 Gradience and auxiliary selection in Old Catalan and Old Spanish. In Paola Crisma and Giuseppe Longobardi (eds.), *Historical Syntax and Linguistic Theory*, 176-193. Oxford: Oxford University Press.

Mateu, Jaume and Mar Massanell This volume A constructional approach to auxiliary selection: Evidence from existential constructions.

Medin, Douglas L. and Marguerite M. Schaffer 1978 Context theory of classification learning. *Psychological Review* 85 (3): 207-238.

Newell, Allen 1990 *Unified theories of cognition*. Cambridge: MIT Press.

Oesterreicher, Wulf 1997 Zur Fundierung von Diskurstraditionen. In Barbara Frank, Thomas Hays and Doris Tophinke (eds.), *Gattungen mittelalterlicher Schriftlichkeit*, 19-42. Tübingen: Narr.

Pfänder, Stefan, Heike Behrens, Peter Auer, Daniel Jacob, Rolf Kailuweit, Lars Konieczny, Bernd Kortmann, Christian Mair and Gerhard Strube 2013 Erfahrung zählt. Frequenzeffekte in der Sprache - ein Werkstattbericht. *Zeitschrift für Literaturwissenschaft und Linguistik* 43 (169): 7-32.

- Pierrehumbert, Janet B. 2001 Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee and Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 123-136. Amsterdam, Philadelphia: John Benjamins.
- Pierrehumbert, Janet B. 2002 Word-specific phonetics. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory phonology 7*, 101-139. Berlin: Mouton de Gruyter.
- Poplack, Shana 2001 Variability, frequency, and productivity in the irrealis domain of French. In Joan L. Bybee and Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 405-428. Amsterdam, Philadelphia: John Benjamins.
- Real Academia Española 2012 *Banco de datos (CORDE) [en línea]. Corpus diacrónico del español*. Available online at <http://www.rae.es>. Last access 20 April 2012.
- Rodríguez Molina, Javier 2006a *Ser + participio en español antiguo: perífrasis resultativa, no tiempo compuesto*. In José Jesus Bustos Tovar and José Luis Girón Alchonchel (eds.), *Actas del VI Congreso Internacional de Historia de la Lengua Española*, II, 1059-1072. Madrid: Arco Libros.
- Rodríguez Molina, Javier 2006b Tradición manuscrita y gramática histórica: los tiempos compuestos en los textos medievales. In Lola Pons Rodríguez (ed.), *Historia de la lengua y crítica textual*, 19-67. Madrid, Frankfurt: Iberoamericana Vervuert Verlag.
- Rosemeyer, Malte 2012 How to measure replacement: Auxiliary selection in Old Spanish bibles. *Folia Linguistica Historica* 33 (1): 135-174.
- Rosemeyer, Malte 2013 *Tornar and volver: The interplay of frequency and semantics in compound tense auxiliary selection in Medieval and Classical Spanish*. In Elly Van Gelderen, Jóhanna Barðdal and Michela Cennamo (eds.), *Argument Structure in Flux. The Naples-Capri Papers*, 435-458. Amsterdam, Philadelphia: John Benjamins.
- Rosemeyer, Malte 2014 *Auxiliary Selection in Spanish. Gradience, Gradualness, and Conservation*. Amsterdam, Philadelphia: John Benjamins.
- Schlieben-Lange, Brigitte 1983 *Traditionen des Sprechens: Elemente einer pragmatischen Sprachgeschichtsschreibung*. Stuttgart: Kohlhammer.
- Schrott, Angela and Harald Völker 2005 Historische Pragmatik und historische Varietätenlinguistik. Traditionen, Methoden und Modelle in der Romanistik. In Angela Schrott and Harald Völker (eds.), *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*, 1-24. Göttingen: Universitätsverlag Göttingen.
- Stolova, Natalya I. 2009 Los corpus diacrónicos al servicio del estudio de los arcaísmos gramaticales. In Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde a lingüística de corpus*, 385-400. Madrid, Frankfurt am Main: Iberoamericana, Vervuert.
- Tomasello, Michael 1992 *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.

Traugott, Elizabeth C. 1989 On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65: 31-55.

Traugott, Elizabeth C. and Ekkehard König 1991 The semantics-pragmatics of grammaticalization revisited. In Elizabeth C. Traugott and Bernd Heine (eds.), *Approaches to Grammaticalization*, 1, 189-218. Amsterdam, Philadelphia: Benjamins.

Von Hoegen, Saskia 2000 *Entwicklung der spanischen Historiographie im ausgehenden Mittelalter. Am Beispiel der Crónicas de los Reyes de Castilla Don Pedro I, Don Enrique II, Don Juan I y Don Enrique III von Pero López de Ayala, der Generaciones y semblanzas von Fernán Pérez de Guzmán und der Crónica de los Reyes Católicos von Fernando del Pulgar.* Frankfurt am Main: Peter Lang.

Wilhelm, Raymund 2001 Diskurstraditionen. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible (eds.), *Language Typology and Language Universals / La typologie des langues et les universaux linguistiques / Sprachtypologie und sprachliche Universalien. An International Handbook / Manuel international / Ein internationales Handbuch*, I, 467-477. Berlin, New York: De Gruyter.

Yllera, Alicia 1980 *Sintaxis histórica del verbo español: Las perífrasis medievales.* Zaragoza: Universidad de Zaragoza.