

Entrenchment and persistence in language change: the Spanish past subjunctive

Abstract

In this paper, we demonstrate that like frequency, morphosyntactic persistence can have a conserving effect on language change. To substantiate this claim, we analyze the alternation between the Spanish past subjunctive forms ending in *-ra* and *-se* (as in *comiera* and *comiese* 'had eaten'). Due to the ongoing replacement of *-se* by *-ra*, persistence and frequency are the best predictors of the alternation in our data. First, the persistence effect of a prior *-se* is significantly greater than the persistence effect of a prior *-ra*. Second, although *-se* is basically restricted to third person singular morphology in contexts without persistence, when primed by *-se* this restriction is drastically reduced. Our results also shed light on the relationship between frequency and persistence in language change. Although both result in conservation, the conserving effect of frequency causes irregularity such as the paradigmatic atrophy of Spanish *-se* forms. In contrast, persistence can temporarily re-establish paradigmatic regularity and consequently strengthen the cognitive representation of obsolescing constructions. However, this resuscitating effect of persistence appears to be restricted to low-frequency *-se* forms; because they are generally more entrenched, the activation of *-se* high-frequency forms relies less on persistence effects.

Key words: Language change, Persistence, Priming, Frequency, Usage-based Linguistics, Past Subjunctive, Spanish

1. Introduction

Usage-based linguistics emphasizes the importance of cognitive processes for the organization of linguistic structures in that "grammar is seen as an emergent system consisting of fluid categories and dynamic constraints that are in principle always changing under the influence of general cognitive and communicative pressures of language use" (Diessel 2011: 830). As a result of these considerations, usage-based approaches assume that language change is also highly dependent on cognitive mechanisms grounded in usage. Much of the research produced in this paradigm has focused on the cognitive process of entrenchment resulting from repeated exposure to a linguistic item, i.e. its usage frequency. Usage frequency has frequently been shown to have a conserving effect in language change. Due to cognitive entrenchment, high-frequency items are less affected by reductive changes such as replacement processes (Bybee 2006).

In this paper, we will give evidence for the influence of a different cognitive mechanism on language change and specifically on subpatterns of maintenance within larger change processes. In particular, we argue that morphosyntactic persistence conserves the use of that element over time. The concept of persistence, sometimes termed *syntactic priming*, refers to the fact that speakers are more likely to use an expression if it has already been used in the preceding context (Gries 2005; Szmrecsanyi 2005; 2006). For instance, English speakers are prone to using *gonna* over *will* as the future form if *gonna* has already been used before (Szmrecsanyi 2005: 114). However, persistence activates not only the specific element but also the entire construction related to the element. We theorize that this activation effect strengthens the overall mental representation of the construction, leading to a conservation effect in situations of language change. Our results also shed light on the relationship between frequency

and persistence as competing factors in language change. Although both can result in conservation, the conserving effect of frequency will always cause irregularity to arise in paradigmatic contrasts. Our study demonstrates that persistence has the opposite effect; it can reestablish regularity in subparts of a paradigm.

To substantiate these claims, we analyze the alternation between the Spanish past subjunctive forms ending in *-ra* and *-se* (as in *comiera* and *comiese* 'had eaten') in a corpus of nearly 4000 past subjunctive forms taken from the *Corpus del Español* (Davies 2002). The older form *-se* has been gradually replaced by the *-ra* form since the 13th century, leading to a low relative frequency of *-se* in contemporary Spanish. This makes the alternation a perfect example for the analysis of the influence of persistence on language change. Previous studies of this variation (Asratan 2007; Day 2011) claim that the two forms are governed by semantic differences such as (non-)realization, emphasis or (im)probability, but our analysis suggests that these factors play a marginal role. Rather, due to the ongoing replacement of *-se* by *-ra*, persistence and frequency effects are the best predictors of the alternation. First, the probability for a prior *-se* to lead to the choice of *-se* over *-ra* in the following context is significantly greater than the probability of a prior *-ra* to lead to *-ra* over *-se*. Second, *-se* is significantly more likely to occur with high-frequency verbs (e.g. *ser* 'to be') than low-frequency verbs (e.g. *entrar* 'to enter'). Third, although *-se* is basically restricted to 3rd person singular morphology in contexts without persistence, thus demonstrating *paradigmatic atrophy* (Leech et al. 2009: 80), when primed by *-se* this restriction is drastically reduced. However, this resuscitating effect of persistence appears to be restricted to low-frequency forms; our results show persistence to have a much weaker effect on high-frequency forms.

The rest of this paper is structured as follows. In Section 2, we discuss the relationship between persistence and language change and argue for a conserving effect of persistence on forms that are otherwise becoming obsolete. In Section 3, we discuss previous research on the alternation between the two Spanish past subjunctive forms *-ra* and *-se*. In Section 4, we introduce the data used for the corpus study. In Section 5, we present the results of the corpus study, followed by a discussion in Section 6. The paper concludes in Section 7 with a discussion of the general theoretical implications of our analysis.

2. Persistence, entrenchment, and obsolescence

Recent years have seen an increase of usage-based approaches to the description of grammatical change. These approaches assume that grammatical changes such as the creation of new constructions and obsolescence processes by which constructions are lost over time are to a great degree shaped by usage. Many usage-based approaches to language make use of constructional approaches to language description (Langacker 1987; Fillmore, Kay and O'Connor 1988; Goldberg 1995; Croft 2001; Goldberg 2006). Constructions in the sense of such construction grammars can be defined as "conventional, learned form-function pairings at varying levels of complexity and abstraction" (Goldberg 2013: 17). This definition includes linguistic patterns as diverse as simple words (e.g., *banana*), idioms (e.g., *going great guns*) and schematic constructions such as the passive (e.g., SUBJECT + AUXILIARY + PARTICIPLE (+ BY NOUN)) Crucially, many construction grammars assume that constructions can have different degrees of schematicity. In English, the passive construction can be formed by using either the auxiliaries *be* or *get*, leading to two subschemas, SUBJECT + BE + PARTICIPLE (+ BY NOUN) and SUBJECT + GET + PARTICIPLE (+ BY NOUN). These subschemas can be instantiated

by further, less schematic subschemas, such as SUBJECT + BE + DONE (as in, e.g., *the deed is done*).

It is because of this notion of schematicity that usage-based approaches to language change often rely on cognitive representations based on constructions. Regarding obsolescence processes, it has been prominently argued by Joan Bybee and her colleagues (see, e.g., Bybee 2006; 2010) that when a construction is replaced by another construction over time, it will be those specific instantiations of the constructions that have a high usage frequency that resist the change longest. For instance, AUTHOR (2014) shows that when the *be*-perfect was replaced with the *have*-perfect in Spanish, the *be*-perfect survived longer with high-frequency instantiations of the *be*-perfect (e.g., *es muerto* 's/he is died', *es venido* 's/he is come') than with low-frequency instantiations. This conserving effect of frequency can be explained by a mechanism created by cumulated activation. In particular, the frequent repetition of a linguistic element such as *es venido* 's/he is come' leads to an *entrenchment* process (Langacker 1987: 59) in the language users' minds, by which this high-frequency element is stored holistically and no longer has to be assembled from its component parts. While such entrenchment processes strengthen the mental representation of the element and consequently lead to greater ease of production and perception, they necessarily weaken the link between this specific instantiation of the *be*-perfect to the grammatical pattern that constitutes it. Consequently, in the situation where the *be*-perfect is affected by a replacement process, this process is slower for high-frequency instantiations of the pattern because these elements are processed as a pattern of their own. Entrenchment processes can also lead to *paradigmatic atrophy* (Leech et al. 2009: 80). In other words, the entrenched elements come to be invariable in their form. For instance, the English modal verb *shall* is almost inexistent with second-person-subjects and in its future use, basically restricted to first-person-subjects, as in, e.g., *Then shall we start?* (Leech et al. 2009: 80).

However, usage frequency is not the only factor related to language use that might influence grammatical change in general, and obsolescence processes in particular. In this paper, we are interested in the effect of *persistence* or *priming* on grammatical change.¹ It is well known from psycholinguistic experiments that in a situation where a speaker can choose between two alternative expressions, s/he will often opt for the expression that has already occurred in the previous discourse. The primed expression is already active in the speaker's mind, which is why access to it is easier and the expression "springs to mind" more readily than the alternative expression. The relevant parameter here is not only the presence of the expression in the previous discourse, but also the temporal distance between the prime and the target expression. Thus, psycholinguists have found that the strength of priming decreases over time (see, e.g., Branigan, Pickering and Cleland 1999), a phenomenon that is usually measured in distance in words between prime and target expression in corpus studies such as Szmrecsanyi (2005).

¹ The terms persistence and priming are often used almost interchangeably in the linguistic literature. However, there are notional differences between them. In particular, in experimental psycholinguistic literature, persistence is used to refer to "shortened reaction times after previous exposure to an identical or related stimulus in experimental settings" (Tamminga 2014: 4). Consequently, the observation that in texts, a certain item is more likely to be used after a previous occurrence of the same item is not persistence in the strict sense because we have no evidence of reaction times in the production of these texts. Many authors (Szmrecsanyi 2005; 2006; Jaeger and Snider 2008; AUTHOR 2014; Tamminga 2014; AUTHOR 2015) therefore prefer to use the term persistence to refer to repetition in corpus data, as will we in this paper. In addition, we use the term *recency* to describe in neutral terms the situation in which a construction has been used in the previous context.

Interestingly, it has been shown that constructions with a very low usage frequency appear to lead to a stronger persistence effect than constructions with a higher usage frequency (Bock 1986; Ferreira 2003). This inverse correlation of strength of persistence and usage frequency makes the prediction that in obsolescence processes, the obsolescing construction will have a stronger persistence effect. A number of recent studies (AUTHOR 2013; AUTHOR 2014; 2015; Schwarz 2016) confirm this prediction. For instance, AUTHOR (2015) finds that in later stages of the historical replacement of the *be*-perfect (e.g., *soy venido* ‘I am come’) with the *have*-perfect (e.g., *he venido* ‘I have come’) in Spanish, the *be*-perfect is more likely to prime a later use of the *be*-perfect than the *have*-perfect is to prime a later use of the *have*-perfect. In addition, AUTHOR demonstrates empirically that the process of obsolescence is slower for tokens of the *be*-perfect primed by a previous *be*-perfect than for unprimed *be*-perfect tokens.

We believe that there are two possible interpretations for these findings. First, it would be possible to characterize the stronger persistence of obsolescing forms as a *result* of the obsolescence process. This is due to the fact that strength of persistence is not only due to the presence of the prime and the temporal distance between prime and target, but actually the degree of *surprisal* caused by the prime. Jaeger and Snider (2008) investigate the voice alternation in English with data from the Penn Treebank corpus. Their results show that persistence effects on passives are sensitive to the degree to which the prime was surprising to the language user. The surprisal measure was operationalized as the probability of the passive structure given the verb. If the previous context contained a passive with a verb that is not likely to be used in the passive (e.g., *was ferried*), the probability of use of a passive in the subsequent discourse increased in comparison to previous contexts containing a passive with a verb that is more likely to be used in the passive (e.g., *was done*). Jaeger and Snider therefore hypothesize that the language processing system maintains probability distributions for syntactic structures such as passives. Perceiving a syntactic structure will affect the probability distribution for that syntactic structure. In this sense, persistence can be defined as a rise in the probability distribution of a syntactic structure due to previous exposure. Crucially, however, less probable syntactic structures have a greater impact on probability distributions than more probable syntactic structures. This effect in turn leads to an increase in the probability of reusing the same structure, i.e. persistence (Jaeger and Snider 2008: 1062).

As a result of these considerations, one could argue that the finding that obsolescing elements have stronger persistence effects is merely a correlate of the necessary decline in frequency of such elements. This first hypothesis can be summarized as follows:

Hypothesis 1. The stronger persistence effect of obsolescing elements is due to the low usage frequency of these structures and, consequently, the high of degree of surprisal caused by them.

In this paper, we however want to argue for the second, stronger hypothesis that persistence actually has a conserving effect in language change. The starting point for this assumption is the fact that usage-based studies have found that both usage frequency and persistence can advance grammatical change. On the one hand, prefabricated expressions or “prefabs” resulting from high usage frequency are at the vanguard of the semantic changes in a construction because the higher degree of structural autonomy is correlated to a higher degree of pragmatic freedom (Bybee 2003;

Bybee and Torres Cacoullos 2009). In the grammaticalization of *estar* + *V-ndo* constructions like *estar mirando* 'be watching' towards a progressive semantics in Spanish, already in the earliest texts high-frequency prefabs such as *estar hablando* 'be talking' no longer display the original locative semantics and achieve a high unithood index (based on the criteria of adjacency, association, and fusion). Bybee and Torres Cacoullos (2009: 190) therefore argue that "prefabs constitute important loci of grammatical development in the diachronic domain."

On the other hand, it has recently been argued that persistence can also advance semantic change, specifically in cases of grammaticalization (Eckardt 2008; Jäger and Rosenbach 2008; Hilpert 2014). For instance, Jäger and Rosenbach (2008) assume that persistence of form or function may explain why grammaticalization typically proceeds unidirectionally. It has been frequently suggested that temporal expressions (such as *from Monday to Friday*) evolve from locative expressions (such as *from London to Paris*) and not vice versa. In addition, psycholinguistic studies have demonstrated that persistence between space and time is also asymmetric, since spatial expressions prime temporal expressions to a much greater degree than vice versa. Jäger and Rosenbach (2008) therefore propose that such asymmetric persistence relations are what drives semantic change; because spatial meanings prime temporal meanings and not vice versa, temporal meanings derive historically from spatial meanings (see Jäger and Rosenbach 2008: 101-106 for details).

If both entrenchment and persistence can advance grammatical change, it seems reasonable to explore the hypothesis that, like entrenchment, persistence can stall grammatical change. A central argument for this assumption again comes from the psycholinguistic literature, namely the *implicit learning* account of language (Hartsuiker and Kolk 1998; Bock and Griffin 2000), as presented in Jäger and Rosenbach (2008: 99-101). The implicit learning account is derived from the observation that (a) persistence effects do not decay at the moment of utterance of the target and (b) cumulative experience with persistence increases the preference for a certain structure (an effect that is also documented empirically in Jaeger and Snider 2008). For instance, Bock and Griffin (2000) tested whether persistence effects could persist over time by presenting test subjects with a prime and then including fillers before presenting them with the target. They found that the recency effects "persist over intervals that are fairly long by the standards of normal limitations on explicit memory for sentence form" (Bock and Griffin 2000: 189). Likewise, in their experimental study on gerund and infinitival complements in L2 English, Gries and Wulff (2009) found evidence of a within-subject cumulative priming effect in that "the larger the proportion of *to*-complements that a subject provided in their completions, the higher the likelihood of an additional *to*-completion became" (Gries and Wulff 2009: 178). These results suggest that strong and repeated persistence effects will increase the persistence effect, or, in Jäger and Rosenbach's words (2008: 100), "the effects of structural priming [i.e. persistence, AUTHOR] may become entrenched in speakers' grammars over time". It is important to emphasize that according to implicit learning accounts it is the *effect* of persistence that becomes entrenched, and not the repetition of the element itself (this would only lead to common entrenchment, i.e. the conserving effect of frequency described above).

Given that, as argued above, obsolescing elements cause a higher degree of surprisal and therefore cause a stronger persistence effect, from the point of view of the notion of implicit learning it seems reasonable to assume that this high degree of persistence of obsolescing elements can perpetuate itself in the speakers' grammar. As a result, persistence can indeed be argued to conserve the use of the obsolescing element.

It can also be hypothesized that persistence not only conserves the *form* of the obsolescing element (as entrenchment does), but also the *function* of that element. To explain this assumption, it is necessary to address the question of item-specificity in morphosyntactic persistence. Crucially, persistence effects do not necessarily rely on identity, but also on abstract grammatical patterns.² In the classic study by Bock (1986), it was shown that when speakers had previously been presented with a passive sentence, they were more likely to use a passive in subsequent discourse even though the lexical elements in the prime and the target sentence were different. For instance, the use of *was written* might be primed by the use of *are shown*, even though the verbs and the morphology (person/number/tense/mode) differ. Similar findings have been made for ditransitives, datives, and locatives (cf. the summary in Goldberg 2006: 120-125).

These findings necessarily imply that the recent mention of a grammatical element not only activates a cognitive representation of that exact element (such as *was written*), but also the construction to which that element belongs (the passive construction formed with *be* + past participle). In other words, persistence does not only lead to the repetition of the exact token, but rather boosts the productivity of the primed construction because the entire paradigm of the construction is activated. This assumption is relevant to our argument because, as described above, in later stages of obsolescence processes entrenchment of the remaining instantiations of the obsolescing construction is common. As a result, we typically observe paradigmatic atrophy in obsolescing constructions. If persistence is a route to the constructional schema, it follows that when obsolescing constructions are primed, the obsolescing construction should receive a boost in productivity that to some degree counteracts the paradigmatic atrophy due to entrenchment. This hypothesis makes the prediction that in comparison to unprimed tokens of obsolescing constructions, primed tokens should display a higher degree of paradigmatic freedom both regarding their morphosyntactic properties (especially person and number morphology) and the lexical items appearing in the construction.

As a last important piece of this hypothesis, we want to add that in line with the implicit learning account of language production, this boost in productivity need not be temporal. Rather, if persistence effects can themselves be entrenched, the increase in productivity can persist over a longer stretch of time and indeed bolster the use of the construction as such.

To conclude this section, we summarize the considerations that led us to propose that persistence has a conserving effect in language change in Hypothesis 2 below.

Hypothesis 2. Although the stronger persistence effect of obsolescing elements is in fact due to the low usage frequency of these structures and, consequently, the high of degree of surprisal caused by them, persistence has an active conserving effect in language change. We consider that persistence activates not only specific instances of a construction, but the construction itself. Therefore, persistence can be argued to increase the productivity of that construction and as a result counteract paradigmatic atrophy.

²This generalization does not contradict the fact that the persistence effect is stronger if the lexical elements used in the prime are identical to the ones used in the target (Pickering and Branigan 1998). Likewise, it has been found that for verb constructions such as the dative, persistence effects can be verb-specific (Gries 2005).

3. The Spanish past subjunctive

In order to test the hypotheses established in the last section, we conducted an analysis of the alternation of the two Spanish morphemes *-ra* and *-se*, which can both be used to express the past subjunctive in Modern Spanish (see 1).

- (1) a. *Nunca le oí decir que una de su-s*
 Never him hear.PST.PFV.1SG say that one.F of his-PL
escultur-a-s fue-ra perfect-a
 sculpture-F-PL be-PST.IPFV.SBJ.3SG perfect-F
 'I never heard him say that one of his sculptures was perfect' (Entrevista, ABC, apud CdE)
- b. *Si Paco de Lucía no fue-se tan conoci-d-o*
 If Paco de Lucía not be-PST.IPFV.SBJ.3SG that.much know-PTCP-M
como guitarrista flamenco, ese concierto no hab-ría
 as guitarist flamenco DET.M concert not have-COND.3SG
sali-d-o a la luz
 leave-PTCP-M to DET.F light
 'If Paco de Lucía had not been so well-known as a Flamenco guitarist, that concert would have never happened' (Entrevista, ABC, apud CdE)

Both past subjunctive forms derive from Latin pluperfect verb forms. The older, *-se* form originated in a subjunctive pluperfect, while the younger, *-ra* form derived from an indicative pluperfect. As noted both by grammars (Real Academia Española 2010: 458) and other scholars (Lunn 1995), the *-ra* form can still be found, especially in journalistic language, with a pluperfect interpretation (Lunn 1995: 433).

- (2) *Y, al final, bes-ó la bandera roj-a y guald-a*
 and TO.DET end kiss-PST.PFV.3SG DET.F flag red-F and golden-F
que hace treinta año-s besa-ra su padre el
 that before thirty year-PL kiss-PST.IPFV.SBJ.3SG his father DET.M
Rey y que un día borda-ra su
 king and that one day embroider-PST.IPFV.SBJ.3SG his
tatarabuela la Reina doña María Cristina
 great-great-grandmother DET.F queen Doña María Cristina
 'And, at the end, he kissed the red and gold flag that his father the King had kissed thirty years ago, and that his great-great-grandmother Queen María Cristina had once embroidered'

This archaic usage survives despite the fact that in the everyday conversational Spanish of any region, the pluperfect is now expressed by a periphrastic construction with an imperfect form of the verb *haber* 'have' plus the past participle. Thus, for the archaic pluperfect *besara* in (2), in present-day Spanish speakers would have said *había besado* '(he) had kissed' instead. While it is often assumed that the *-se* form is not found in contexts like (2), Lunn (1995: 438) provides other examples showing that this is not the case:

- (3) *Al día siguiente de que Isabel Preysler inicia-se*
 To.DET.SG day after of that Isabel Preysler begin-PST.IPFV.SBJ.3SG
su veraneo, según informa-mos en la página 44
 her summer vacation as inform-PST.PFV.1PL in DET.F.SG page 44
de este número,...
 of DET.M.SG issue
 'The day after Isabel Preysler had begun her summer vacation, as we reported on page 44 of this issue...'

The differing sources of the two forms are also held to be responsible for the use of the *-ra* form in main clause contexts, e.g. to make hyperpolite requests such as *Quisiera pedirle un favor* 'I would like to ask you for a favor'. Since the *-ra* form derives from an indicative pluperfect it is stated that such main clauses uses are a remnant of that source, while *-se* is prohibited in main clauses due to its subjunctive source (cf. Haverkate 2002: 30; Kempas 2011: 247). Nevertheless, speakers clearly identify both *-ra* and *-se* as past subjunctive forms (Lunn 1995: 438), and more recent research (e.g., Day 2011) has also found *-se* forms in these hyperpolite contexts, despite the potential interpretation of the *-se* form as "obsequious and pompous" in such contexts (Lunn 1995: 438).

Qualitative studies such as these have thus attempted to isolate differences between the two forms, but have ultimately failed to show that there are contexts in which only one of the two past subjunctive forms is allowed. Other purported explanations for meaning differences between the two forms, such as that of Sussman Goldberg (1995), who states that *-se* signals "emphasis" while *-ra* is the "neutral" form, are essentially untestable in empirical terms. Quantitative analysis would thus seem to be called for, given the broad overlap in the types of contexts in which both *-ra* and *-se* can and do occur, and in the meanings that they contribute. While few such analyses can be found, there has existed a generalized viewpoint dating from some of the earliest studies (e.g. Lemon 1925; Wright 1926) that (1) *-ra* is much more frequent than *-se*, (2) *-se* is found more frequently in Peninsular Spanish than in Latin America, but still much less frequently than *-ra*, and (3) *-se* is found more often in written language than spoken.

In one of the first studies to use naturally-occurring corpus materials to study the variation between *-ra* and *-se*,³ DeMello (1993) compares the usage of the *-ra* and *-se* forms across 10 Spanish-speaking cities, utilizing spoken data collected as part of the Norma Culta project in the 1960s and 1970s (see Lope Blanch 1977). Overall, there is an enormous asymmetry between the two forms in these corpora: 93% (n = 2912) *-ra* versus only 7% (n = 210) *-se*. The frequency of the *-se* form in these data ranges from a low of 0.6% (1/166) in Lima, Peru, to a high of 20% (61/310) in data from San Juan, Puerto Rico, which is also the dialect with the greatest number of *-se* tokens (only Madrid [16%] and Seville [13%] also show more than 10% *-se* use). When comparing peninsular Spain and Latin America as macro-dialect regions, the difference is 15% *-se* in Spain versus only 6% in Latin America. Ultimately, DeMello predicts the complete loss of the *-se* form from the language in favor of the *-ra* form.

In the first published study to look at the variation between *-ra* and *-se* utilizing quantitative multivariate methods (Varbrul), Asratíán (2007) found for Caracas Spanish a low 5.7% (n = 62) rate of *-se* versus a 94.3% (n = 1034) rate for *-ra*. There was a slight increase in the use of *-se* by younger speakers (under 45) versus older speakers, but younger speakers were also more educated and in general belonged to a high social

³ Questionnaire-based surveys of present-day *-ra* vs. *-se* variation in Spain are found in Valeš (2006) and Kempas (2011).

class than older speakers. As regards internal linguistic factors, Asratían actually found an anti-priming effect, whereby *-se* was more likely to occur after a preceding *-ra*. However, given the very low overall rate of *-se* compared to *-ra* the import of this finding is difficult to judge. In terms of lexical distribution of the *-se* form, fully half (31/62) of the *-se* tokens in Asratían’s data came from one verb, *haber*, and of these 31 tokens, 26 of them (84%) were auxiliary uses in the perfect (*haber* + past participle) construction. Thus, there was both a strong lexical bias for *-se* involving one very frequent verb, as well as a constructional bias in the perfect construction containing that same verb. The only other linguistic factor which showed a favoring effect for *-se* was negative polarity, but unfortunately Asratían did not look at interactions between polarity and verb or construction type. As a result it may be that the negated tokens of *-se* were also found with *haber* or in the perfect construction, and not an independent effect of polarity.

In addition, Asratían’s results suggest that complex conditional clauses increase the likelihood of use of *-se* over *-ra*. First, she found that the use of *-se* is more likely in protasis clauses such as (4).

- (4) *Quizá si die-ra el primer paso todo ocurrir-ía*
 Maybe if give-PST.IPFV.SBJ.3SG DET.M.SG first step all happen-COND.3SG
precipitadamente y no hubie-ra tiempo para pensar en
 too.soon and not have-PST.IPFV.SBJ.3SG time for think in
culpa-s
 guilt-PL
 ‘Maybe if he did the first step everything would occur too soon and there would not be time to think about guilt’ (*Sueño de invierno*, apud CdE)

Asratían’s results also suggested that the use of *-se* is more likely in the apodosis position, as in (5) below.

- (5) *Tal vez si yo hubie-ra estudia-do economía,*
 such time if I have-PST.IPFV.SBJ.3SG study-PTCP economics
hubie-se habi-do esa rivalidad
 have-PST.IPFV.SBJ.3SG have-PTCP DET.F.SG rivalry
 ‘Maybe if I had studied economics, we would have become rivals’ (*Habla Culta: La Paz: M28*)

A recent quantitative study on the alternation (Guzmán Naranjo to appear) analyzed 1269 tokens from spoken Peninsular Spanish and found further parameters relevant for the alternation. Specifically, the analysis suggests an influence of the verbal lexeme, whether or not the verb has a modal meaning (with modals favoring *-ra*), sentence type (with temporal and adversative [subordinate] sentences favoring *-se*), and lastly the definiteness of the subject (with definiteness favoring *-ra*) and the object (with definiteness favoring *-se*). However, Guzmán Naranjo does not give an explanation for his findings, and they are not easy to motivate. Given that Asratían found an effect of auxiliatio with *haber*, Guzmán Naranjo’s finding concerning modality may possibly interpreted as a complexity effect; arguably, *-se* might be less probable in periphrastic contexts due to its overall lower productivity than *-ra*.⁴ The fact that *-se* appears to be

⁴ We conducted a descriptive survey of the usage frequencies of some modal and non-modal verbs in the 20th century section of the Corpus del Español that appears to confirm Guzmán Naranjo’s finding. As modal verbs, we took the verbs analyzed in Guzmán Naranjo’s data (*querer* ‘want’, *poder* ‘can’, *deber* ‘must’),

more likely with temporal sentences is actually counterintuitive given that it was *-ra*, and not *-se*, that started out as a temporal construction. Possibly, the finding could be due to the preference of *-ra* over *-se* in main clauses. Lastly, we are not able to find any theoretical reason for why the *-ra/-se* alternation should be sensitive to definiteness. It is because of these doubts that we did not include the contextual factors isolated in Guzmán Naranjo’s study in our analysis.

To conclude, the results of previous studies of the *-ra/-se* alternation in Spanish suggest an ongoing process of replacement of the older past subjunctive *-se* form by the *-ra* form, which gradually took over the subjunctive function. A preliminary look at the development of the two forms in the *Corpus del español* (henceforth CdE), a diachronic corpus that contains about 100 million tokens in total, and 20 million tokens of 20th century Spanish (Davies 2002), confirms this assumption (Figure 1). The *-se* form has experienced a drastic decrease in usage frequency, to the point that it has become far less common than the *-ra* form, whose usage frequency has remained relatively stable. The fact that the *-se* form can be characterized as a nearly obsolescent form that is being replaced by the competing *-ra* form makes the *-ra/-se* alternation in Spanish a perfect testing ground for our hypotheses.

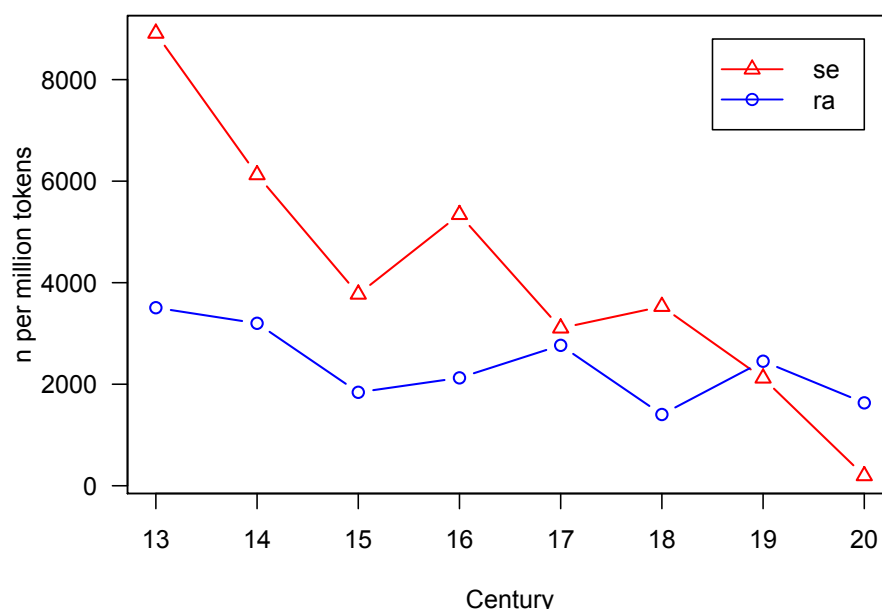


Fig. 1 Development of the relative usage frequencies of *-ra* and *-se* in the CdE

4. Data

In order to investigate the alternation between *-ra* and *-se* forms in Modern Spanish, we extracted 3891 tokens of these forms from the *Corpus del español* (CdE). Importantly, the CdE is tagged and lemmatized, which greatly facilitated the identification and

sober ‘use to’ and *tener* ‘have to’). As non-modal verbs, we took the verbs from our analysis (see Section 4) except, of course, *poder* ‘can’. As illustrated in Table i below, the distribution is statistically significant:

Table i. The past subjunctive in modal and non-modal verbs in the Corpus del Español (20th c.)

	ra	se	Sum
Modal verbs	93.5% (4642)	6.5% (320)	4962
Non-modal verbs	89.1% 3020	10.9% (371)	3391
$\chi^2 = 52.968(1), p < .001^{***}$			

collection of past subjunctive forms for our data. In addition, the CdE can be used to study diatopic and diaphasic variation because (a) it contains texts not only from Spain, but also Latin American countries and (b) these texts differ in genre. Although the CdE distinguishes a rather small number of genres (Academic writing, news texts, oral speech and fiction), the number of tokens from these genres is relatively balanced, with about 5 million tokens for each genre.

Given that we were interested in the cognitive parameters of entrenchment and persistence, we selected the analyzed verbs according to their usage frequency per million words according to the *Corpus del español*. Consequently, we included three verbs of high frequency (*ser* ‘to be’, *haber* ‘to have’, *estar* ‘to be’), three mid-frequency verbs (*poder* ‘to can’, *ver* ‘to see’, *dar* ‘to give’) and six verbs of a comparably low usage frequency (*quedar* ‘to stay’, *salir* ‘to leave’, *sentir* ‘to feel’, *entrar* ‘to enter’, *caer* ‘to fall’ and *acabar* ‘to finish’). Apart from their usage frequency, there was no other criterion governing the selection of these verbs. For *ver*, *dar*, *quedar*, *salir*, *sentir*, *entrar*, *caer* and *acabar*, all past subjunctive tokens from the 20th century section of the CdE were able to be included. In contrast, the number of past subjunctive tokens from *ser*, *haber*, *estar* and *poder* was so high that we opted for a randomization process by which the number of occurrences for each of these verbs was limited to 500.

The resulting 3891 *-ra* and *-se* tokens were manually annotated for a number of morphosyntactic and contextual parameters, which we summarize in Table 1 below. Note that we used two variables to model the effect of negation on the alternation. Thus, we included one variable, NEGATION, to describe the simple fact of whether or not the clause was negated. However, frequently the past subjunctive form was found in a subordinate clause projected from a negated main clause. Whereas in subordinate clauses that are syntactically dependent on negated main clauses (see 6), the action referred to by the past subjunctive is usually truly counterfactual, i.e. not realized, subordinate clauses that depend on non-negated main clauses as in (7) frequently express an intention, i.e. an action that is currently being realized. Since the discourse-pragmatic function of the past subjunctive appears to partially depend on such subordination patterns, we used the variable MAINCLAUSENEGATED to test for an effect of these patterns.

(6) *No sé si ya se pueda ver, porque hasta*
 Not know.PRS.1SG if already REFL can.PRS.SBJ.3SG see because until
hace poco no permit-ía el Instituto de Arqueología
 before little not permit-PST.IPFV.3SG DET.M.SG Institute of Archeology
que se entra-ra todavía
 that REFL enter-PST.IPFV.SBJ.3SG yet
 ‘I don’t know if you can already go see it, because until recently, the Institute for Archeology did not yet permit entry’ (Habla culta Mexico, apud CdE)

(7) *empeza-ndo a mover la cómoda para que el tablón*
 begin-PROG to move DET.F.SG dresser for that DET.M.SG plank
salie-ra poco a poco por la ventana
 leave-PST.IPFV.SBJ.3SG little to little towards DET.F.SG window
 ‘...he began to move the dresser so that the plank moved slowly towards the window’ (Rayuela, apud CdE)

Table 1. Summary of predictor variables

Variable name	Short description	Levels
VERBFREQUENCY	Usage frequency of verb per million tokens in the 20th century	No levels (numerical variable)
TARGET_PERSON	Person morphology	1st, 2nd, 3rd ⁵
TARGET_NUMBER	Number morphology	Singular, plural
RECENT	Occurrence of a past subjunctive form in the previous context	ra (occurrence of <i>-ra</i>) se (occurrence of <i>-se</i>) none (neither <i>-ra</i> nor <i>-se</i>)
DISTANCE_TO_LAST_RA	If RECENT = ra: distance in words to the previous <i>-ra</i> token	No levels (numerical variable)
DISTANCE_TO_LAST_SE	If RECENT = se: distance in words to the previous <i>-se</i> token	No levels (numerical variable)
AUXILIARY	Whether or not the verb serves as an auxiliary for a compound tense form	True, False
NEGATION	Whether or not the sentence is negated	True, False
SUBORDINATED	Whether or not the sentence is a subordinated or a main clause	True, False
MAINCLAUSENEGATED	If SUBORDINATED = True: whether or not the main clause is (also) negated	True, False
PROTASIS	If SUBORDINATED = True: whether or not the clause serves as protasis	True, False
APODOSIS	If SUBORDINATED = False: whether or not the clause serves as apodosis	True, False
REGION	Origin of the author of the source text,	Eur(ope), SouthAm(eric), MidAm(eric)/NorthAm(eric)
GENRE	Genre according to the classification in the CdE	Academic, Oral, News, Fiction

5. Results

In this section, we present the results from an inferential statistical analysis of our corpus data. The data were analyzed using descriptive statistics (5.1), *Classification and Regression trees* (henceforth CART, Breiman et al. 1984) and their recent extension, *random forests* (Breiman 2001) (5.2), as well as logistic regression analysis (5.3).

5.1 Descriptive analysis: the interplay of persistence and frequency effects

In this section, we give a descriptive survey of the variables crucial for the analysis of persistence and frequency effects in the *-ra/-se* alternation. Although such a descriptive analysis can only be preliminary because it does not control for the influence of other contextual factors such as diatopic and diaphasic variation, it serves to illustrate the effects we are interested in here.

A first parameter that is connected to the distribution of the two constructions is the usage frequency of the verbs occurring in them. *-Se* tokens formed from the three most frequent verbs (*ser* 'to be', *haber* 'to have' and *estar* 'to be') make up 52.53 percent (208/396) of all of the *-se* tokens in our data. In contrast, *-ra* tokens formed from these

⁵ The first and third person singular of the Spanish past subjunctive are morphologically identical; depending on the context, a form such as *comiera* can mean either 'I had eaten' or 'he/she/it had eaten'. We therefore annotated the personal morphology of all these ambiguous cases manually on the basis of the context.

three verbs only make up 36.97 percent (1292/3495) of all of the *-ra* tokens in our data. In other words, *-ra* tokens have a higher type/token ratio than *-se* tokens. This finding is also illustrated in Figure 2, which gives the mean verb frequency for *-ra* and *-se* tokens in our data. For *-ra* tokens, the mean verb frequency is significantly lower than for *-se* tokens.

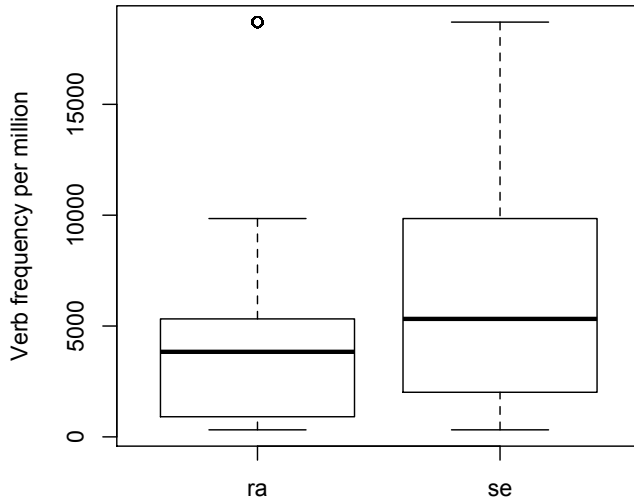


Fig. 2 Mean verb frequency per million by *-ra* and *-se* tokens

Table 2 summarizes the relative usage frequency of *-ra* and *-se* forms in our corpus by the variable RECENT measuring persistence. Like entrenchment, persistence has a strong relationship to the opposition between the two forms. In the neutral condition in which neither *-ra* or *-se* forms are used in the preceding context the usage frequency of *-ra* reaches about 90 percent. If a *-ra* form is used in the preceding context, the usage frequency of *-ra* increases to about 93 percent, a small but nevertheless statistically significant increase ($\chi^2= 9.04$ (1), $p<.01^{**}$). If a *-se* form is used in the preceding context, the usage frequency of *-ra* drops dramatically, to about 51 percent. This effect is significant ($\chi^2= 230.95$ (1), $p<.001^{***}$). It appears, thus, that although persistence is relevant to the use of both *-ra* and *-se* forms, this relationship is much stronger for persistence by *-se* than for persistence by *-ra* forms. Note that we did not encounter cases in which both *-ra* and *-se* forms were used in the previous context.

Table 2. Relative frequency of use of *-ra* vs. *-se* by persistence

	ra	se	Sum
No <i>-ra</i> or <i>-se</i> in previous context	90.64% (2354/2597)	9.35% (243/2597)	2597
<i>-ra</i> in previous context	93.69% (1054/1125)	6.31% (71/1125)	1125
<i>-se</i> in previous context	51.47% (87/169)	48.52% (82/169)	169
Total	3495	396	3891
$\chi^2= 292.11$ (2), $p<.001^{***}$			

Figure 3 illustrates that the difference in the persistence strength of *-ra* and *-se* forms is not merely a result of the presence of these forms in the preceding context, but also of the distance (in words) of the prime to the target. Whereas a greater distance of a *-ra* form in the preceding context corresponds to a moderate decrease of the usage frequency of *-ra* forms compared to *-se* forms (from about 98 percent to about 95 percent, see the left graph in Fig. 3), a greater distance of a *-se* form in the preceding

context corresponds to a drastic decrease of the usage frequency of *-se* forms compared to *-ra* forms (from about 60 percent to about 18 percent), see the right graph in Fig. 3).

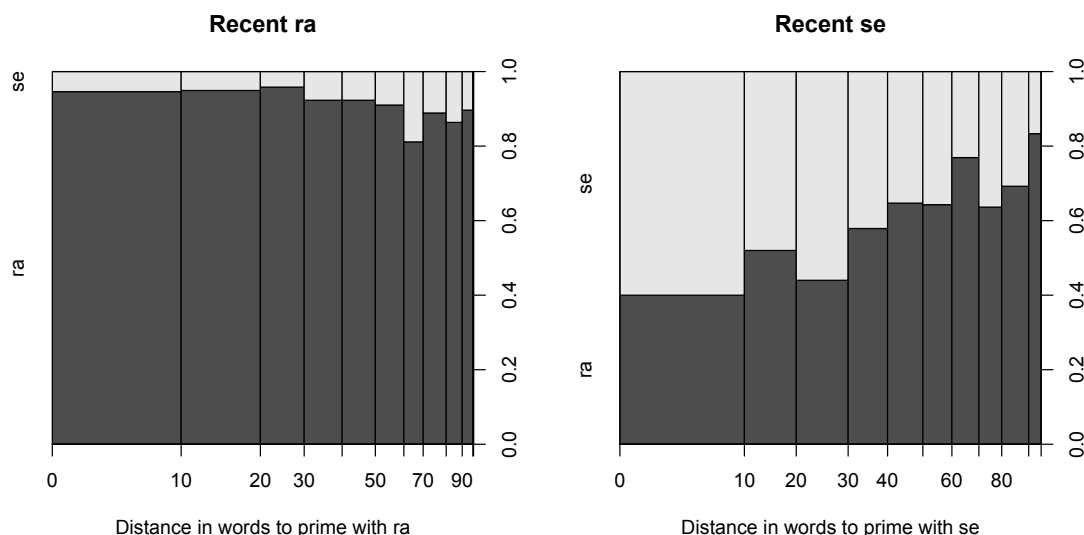


Fig. 3 Relative frequency of use of *-ra* vs. *-se* by persistence and distance to prime

Crucially, we found not only a difference in the strength of persistence by *-ra* and *-se* in our data, but also a relationship between entrenchment and persistence. In line with the assumptions about entrenchment posited in Section 2, we find an effect of *paradigmatic atrophy* for *-se* forms, i.e. a reduction of the grammatical contrasts that constitute the paradigm of *-se* forms due to the general reduction in frequency of the construction. Although both the use of *-ra* and *-se* forms is much more common with third person than with first and second person morphology, this bias is significantly stronger for *-se* forms (Table 3).

Table 3. Relative frequency of use of *-ra* vs. *-se* by personal morphology

	ra	se	Sum
1st	9.84 % (344)	7.58 % (30)	374
2nd	3.09 % (108)	0.51 % (2)	110
3rd	87.07 % (3043)	91.92 % (364)	3407
Total	3495	396	3891

$\chi^2=11.266$ (2), $p<.01^{**}$

Likewise, although both *-ra* and *-se* forms are used more frequently with singular than plural morphology, this bias is significantly stronger for *-se* forms than for *-ra* forms (see Table 4).

Table 4. Relative frequency of use of *-ra* vs. *-se* by number morphology

	ra	se	Sum
Singular	73.62% (2573)	82.32% (326)	2899
Plural	26.38% (922)	17.68% (70)	992
Total	3495	396	3891

$\chi^2=14.2$ (1), $p<.001^{***}$

5.2 CART and random forest analysis

In a second step, we decided to conduct an exploratory CART and random forest analysis on our data. CARTs can be described as nonparametric, data-driven regression analyses. They recursively partition the data such that the resulting subgroups viz. leaves become increasingly homogeneous with respect to the levels of the dependent variable (Tagliamonte and Baayen 2012: 22). For instance, a CART for our data might partition the data according to the predictor variable persistence. If the frequency of *-ra* and *-se* (i.e. the levels of the dependent variable) differs sufficiently in the two resulting subsets to gain statistical significance, the algorithm would then continue with other predictors that further differentiate these subsets. In this way, the CART recursively identifies the predictors that result in significantly different subsets in the data (indicated by *p* values). Because CARTs do not establish a mathematical formula predicting the variation between two (or more) outcomes as in regression analysis but instead proceed in a stepwise fashion, they do not share two major problems of regression analysis. In particular, CARTs can handle collinear predictors, i.e. covariation (Tagliamonte and Baayen 2012: 161), as well as complex interactions (Tagliamonte and Baayen 2012: 171).

Although CARTs are a powerful tool for calculating the influence of predictor variables on a dependent variable, they have a major disadvantage. In particular, the results from CARTs do not appear to be particularly stable because they are sensitive to small changes in the data or to the operationalization of the variables (Strobl, Malley and Tutz 2009: 330). This means that already small changes might lead the CART to create very different subsets of the data. *Random forests* are an extension of the CART methodology that help to alleviate this problem. Basically, random forests differ from CARTs in that they do not rely on only one, but rather on a large number of CARTs. Each of these trees is grown for a different subsample of the data ('in-bag' observations). The results for the tree from this subsample are then used to make predictions about the rest of the data ('out-of-bag' observations). This procedure thus allows for evaluation of the accuracy of a tree's predictions (Tagliamonte and Baayen 2012: 22). The collective predictions by the trees are then used to establish the best overall model. This is achieved by a 'voting' method in which each tree contributes its evaluation of which is the most likely outcome for each data point. The forest will then decide for the outcome voted for by most of the trees.

As also described by Tagliamonte and Baayen (2012: 22), random forests can be used to evaluate the usefulness of a predictor in the model. This is achieved by using a *permutation variable importance measure*. Basically, this procedure works by randomly permuting the values of the predictor variable and using the resulting artificial predictor variable, along with the other predictor variables to grow a random forest. If the permuted version of the original predictor leads to a worse model of the data than the original predictor, we can assume that the original predictor is in fact associated with the response; if the model does not become (significantly) worse by permuting the predictor, the predictor is not important.

We constructed a CART, as well as a random forest, with the dependent variable DEPENDENT (use of *-ra* vs. use of *-se*) and all of the predictor variables listed in Table 1, i.e. RECENT, TARGET_PERSON, TARGET_NUMBER, VERBFREQUENCY, DISTANCE_TO_LAST_RA, DISTANCE_TO_LAST_SE, AUXILIARY, NEGATION, SUBORDINATED, MAINCLAUSENEGATED, PROTASIS, APODOSIS, REGION and GENRE. To do so, we used the package *party* (Hothorn, Hornik and Zeileis 2006) in R (R Development Core Team 2015). We did not include the

title of the source text itself as a predictor variable even though this would have allowed us to estimate the variation due to interpersonal variation. Our reasoning was that interpersonal variation should not be put on a par with the other dependent variables. The appropriate way to control for interpersonal variation would have been to calculate a mixed-effects regression analysis for our data in which control variables are clearly distinguished from predictor variables (Baayen 2008: chapter 7). CART analyses however do not allow for such a feature. We did however calculate a CART with the predictor TITLE, in addition to the predictor variables listed in Table 1, in order to test for interpersonal variation. The inclusion of this predictor turned out to substantially increase the degree of variation explained by the model (c index of concordance = 90.8). TITLE was the most important predictor, followed by RECENT, REGION and VERB. However, the effects of the predictors DISTANCE_TO_LAST_SE, VERBFREQUENCY, TARGET_NUMBER and TARGET_PERSON were lost due to the strength of the effect of TITLE. While the model thus corroborates the most important result from our analysis below (the importance of the influence of persistence on the *-ra/-se* alternation and the relative irrelevance of structural parameters), it also differs from it on some points.

Figure 4 illustrates the CART for the *-ra/-se* alternation resulting from our analysis. It can be interpreted as follows. The nodes of the tree name the variable and give the significance level of the variable. For instance, Node [1] refers to the variable RECENT, which reaches a significance level of $p < .001^{***}$ in the CART analysis. The branches below each node refer to the levels of the variable, according to which the data is partitioned. In the case of the variable RECENT, the analysis distinguishes between persistence by *-se* on the one hand, and persistence by *-ra* and no persistence on the other hand. If we compare the distributions of *-ra* and *-se* on each side of the tree created by this distinction (i.e. Nodes 3 and 4 vs. Nodes 8, 10, 11, 12, 13), we see that in general, *-se* has a much higher relative frequency in the left-hand nodes than in the right-hand nodes. Nodes that are situated below other nodes can be interpreted as interaction effects if they do not appear on both sides of the tree. For instance, Node [2] displays the impact of the variable DISTANCE_TO_LAST_SE on the alternation. Although the variable influences the alternation at $p < .05^*$ (there is a clear difference in the relative frequency of *-se* in Node 3 and 4), its effect is contingent on the effect of RECENT. Whereas there is only a two-way interaction effect on the left-hand side of the tree, the structure of the interaction effects on the right-hand side of the tree is much more complex, with two-, three- and four-way interactions.

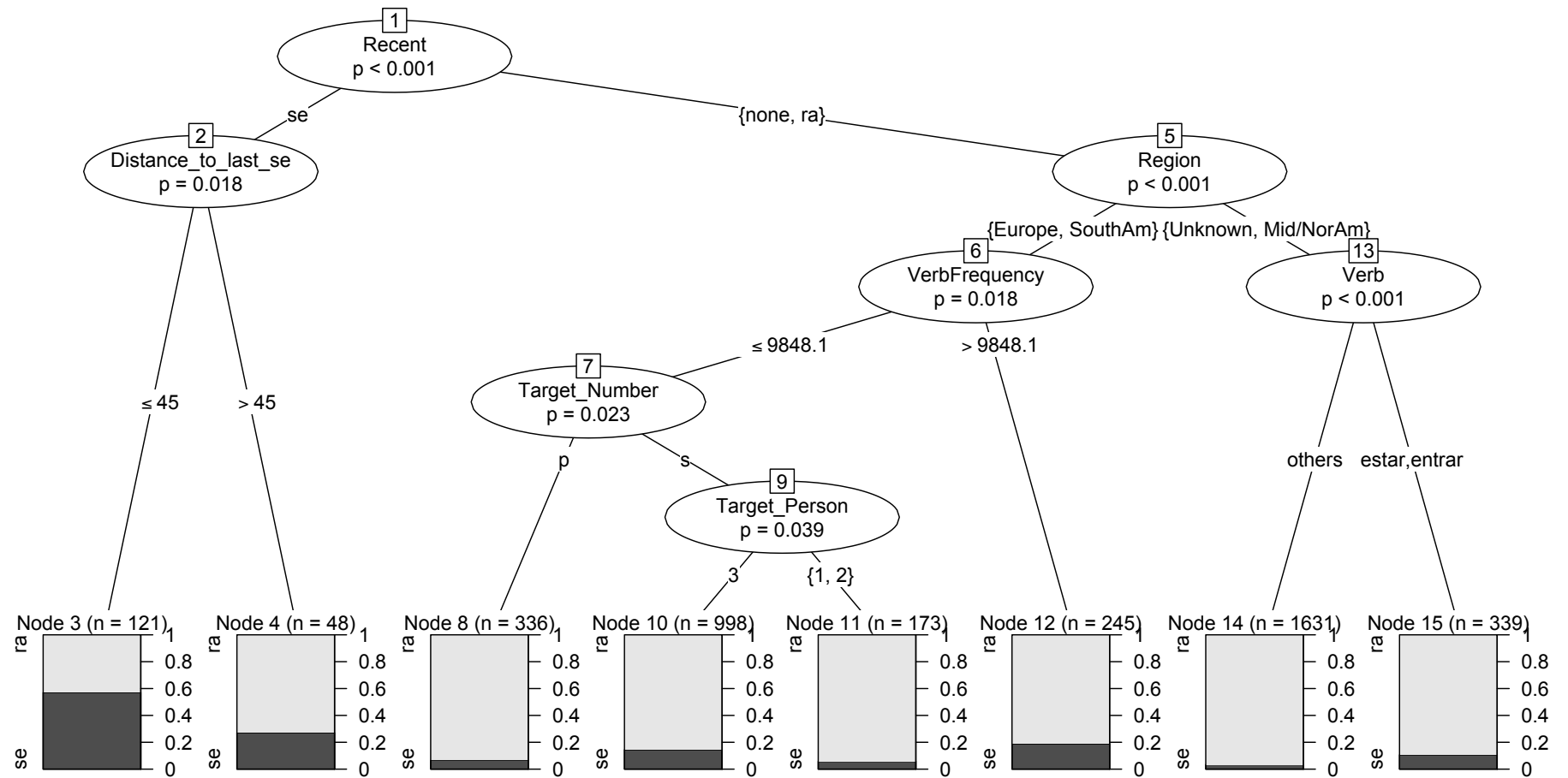


Fig. 4 CART analysis of the *-ra/-se* alternation (c index of concordance = 0.74)

The CART shows at its topmost branch (Node [1]) that there is a statistically significant difference between the presence of a recent *-se* form in the discourse and a recent *-ra* form or no prior form. The relative frequency of *-se* is much higher in the resulting distributions on the left-hand tree than on the right-hand tree. Interestingly, there is no significant difference between the presence of a recent *-ra* form or no prior form at all. In other words, whereas the presence of a recent *-se* form significantly increases the probability of use of *-se* forms, the presence of a recent *-ra* form does not increase the probability of use of *-ra* forms significantly in comparison to contexts without any recent past subjunctive form.

Meanwhile, on the left side of the tree, one can see that the presence of a recent *-se* interacts only with the distance in words to the (last) prior *-se* (see Node [2]): when the most recent prime token of *-se* is found at a distance of less than or equal to 45 words in the prior discourse, then a target *-se* form is significantly more likely (at nearly 60% *-se*) than when the persistence *-se* is found at more than 45 words (only about 25% *-se*). No further effects are found on the left-hand side of the tree, indicating that the presence of a *-se* form in the previous contexts overshadows all other possible effects.

In contrast, in contexts in which either no past subjunctive or a *-ra* form is present, the alternation between the two past subjunctive forms is governed by different factors. First, dialectal variation comes into play (see Node [5]). Thus, as suggested by the descriptive analysis, the use of *-se* is significantly more likely in source texts from the region roughly corresponding to southern America, as well as from Spain, than in source texts from middle or northern America.

Given that in texts from middle and northern America, the *-se* construction is extremely marginal in contexts without a previous *-se* token (see Node 14 and 15), it is unsurprising that in texts from these regions, we only find one further interaction effect. In particular, the model suggests that the verbs *estar* ‘to be’ and *entrar* ‘to enter’ have a significantly higher probability to appear with *-se* than other verbs (see Node [13]).

In texts from Spain and southern America, however, the use of *-se* tokens is more frequent, which is why the model finds more interaction effects in these texts. First, the model identifies verb usage frequency as a significant predictor (Node [6]). Verbs with a usage frequency higher than 9848.1 words per million—precisely the verbs *ser* ‘to be’ and *haber* ‘to have’ and no others—are significantly more likely to appear with *-se* than verbs with a lower frequency. For these high-frequency verbs, no further effects apply, which points to the fact that the frequency effect is due to entrenchment because grammatical distinctions are neutralized for these high-frequency *-se* tokens.

For low-frequency verbs, the model finds additional interaction effects (see Nodes [7] and [9]). In particular, it is here that the paradigmatic atrophy described in Section 4 is relevant. In a nutshell, the use of *-se* is significantly more likely with third person singular morphology than with other person/number morphology. AUTHOR (2013) came to a similar conclusion with data from the CREA, but did not look further at differences between verbs in terms of frequency. In summary, we find that the restriction of *-se* forms to third person singular forms is only relevant in non-primed contexts and with low-frequency forms, a finding that will be replicated in the regression analysis in the next section.

In order to confirm the findings from the CART, we give the conditional permutation variable importance of each predictor from the random forest constructed with the same formula in Figure 5. Although the permutation technique does not allow us to identify the relevance of interaction effects in the same way as the CART analysis, its results for the most part confirm the description of the results from the CART

analysis given above. Note in particular that the predictor variable RECENT has a much higher importance (indicated by its far right placement and distance from the other predictors) than all other predictors in the random forest.

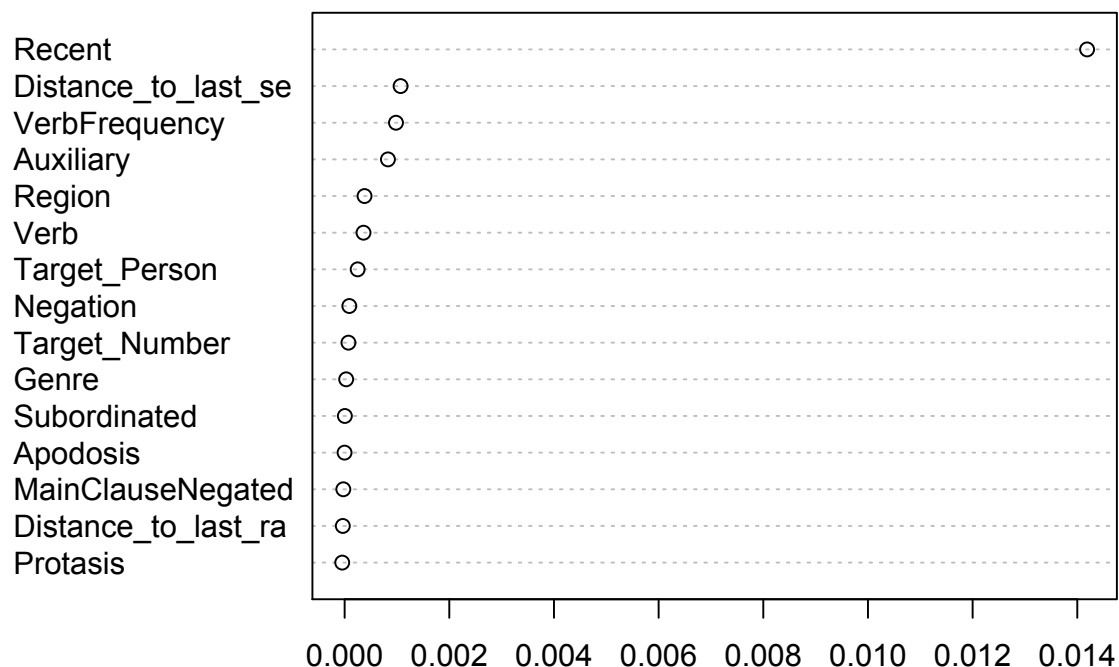


Fig. 5 Conditional permutation variable importance for the random forest for all predictors of the *-ra/-se* alternation

5.3 Logistic regression analysis

Even though CARTs and random forests are ideal tools to explore variation in corpus data, they are not inferential statistical analytics. In other words, since they do not establish a formula of the correlations between the dependent variable and the predictor variables, strictly speaking they do not allow for actual hypothesis testing.

For this reason, in a last step we established a logistic regression model of the *-ra/-se* alternation in our data. We specifically tried to test for the interactions between the recency, frequency and personal morphology variables. The predictor variables included in the model were VERBFREQUENCY, TARGET_PERSONTHIRD, TARGET_NUMBER, RECENT, AUXILIARY, NEGATION, SUBORDINATED, REGION and GENRE, as well as all interactions between the variables VERBFREQUENCY, RECENT, and TARGET_PERSONTHIRD. Note that because the CART analysis clustered first and second person morphology against third person morphology, we decided to collapse the former two levels. The other variables listed in Table 1 could not be used as predictor variables in the model because they involved missing values (e.g., the variable MAINCLAUSENEGATED only received a value if the token in question occurred in a subordinate clause).

The model reproduced the effects of the variables RECENT, TARGET_PERSON, TARGET_NUMBER, VERBFREQUENCY and REGION on the *-ra/-se* alternation from the previous analyses. Figure 4 below illustrates these effects.⁶ Thus, the use of *-se* is much more likely if primed by a previous *-se* token, whereas the use of *-ra* is only moderately more likely if primed by a previous *-ra* token. The use of *-se* is also more likely with third

⁶ The full results from the analysis can be found in the appendix (Table 5).

person and singular number morphology and with high-frequency verbs. In addition, the analysis revealed a diatopic effect (variable REGION). In particular, the European and southern American varieties of Spanish appear to display a more conservative use of the two past subjunctive forms, with a relatively higher likelihood of *-se* usage, whereas in the central and northern American varieties of Spanish the likelihood to use *-se* is lower. Note that because the CdE is not annotated for diatopic variation, the origin of the source texts had to be identified by manual research. For a considerable amount of tokens (n=333), it was not possible to identify the author's origin. These tokens were coded as "Unknown".

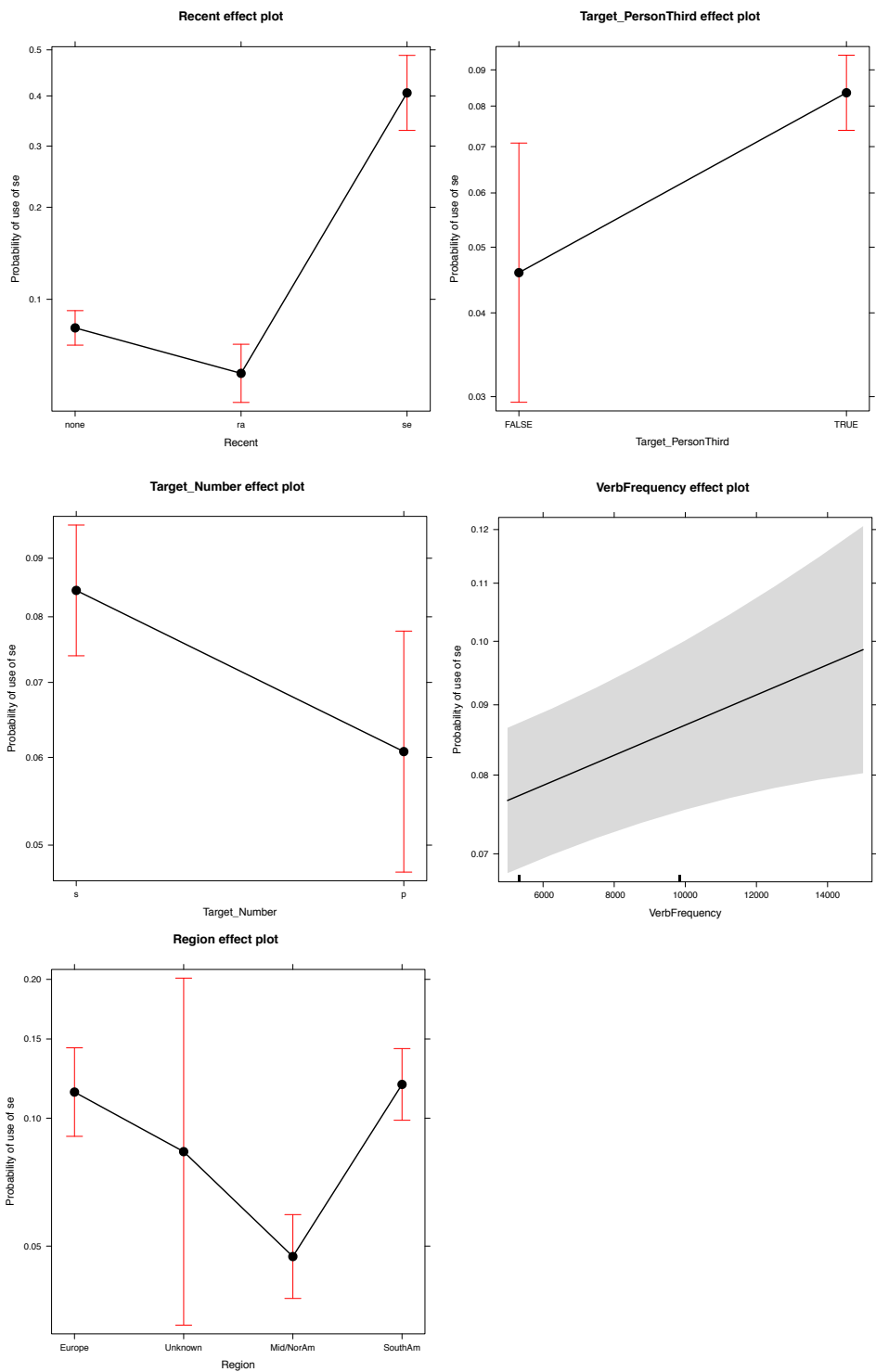


Fig. 4 Effect plots for variables with significant effects (RECENT, TARGET_PERSONTHIRD, TARGET_NUMBER, VERBFREQUENCY, REGION) in the logistic regression model

Crucially, we also found a significant three-way interaction effect between the variables RECENT, TARGETPERSONTHIRD and VERBFREQUENCY, which we illustrate in Figure 5.

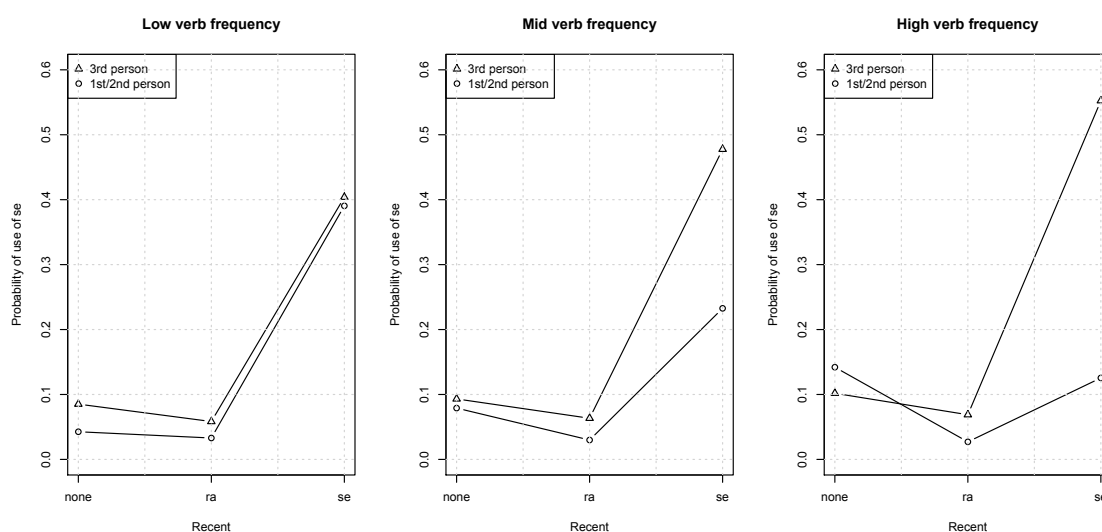


Fig. 5 Effect plot for the interaction effect RECENT:TARGET_PERSONTHIRD:VERBFREQUENCY in the logistic regression model

The interaction effect confirms the existence of an interaction between recency and the personal morphology of the *-ra/-se* tokens in our data. In general, the likelihood of using *-se* in first or second person morphology increases if another *-se* token can be found in the preceding context. In contrast, *-ra*-persistence has almost no effect on personal morphology. In other words, it appears that in general, *-se*-persistence counters paradigmatic atrophy.

Interestingly however, this effect is moderated by usage frequency. The higher the usage frequency of the verb from which the past subjunctive is formed, the less impact *se*-persistence has on the probability of using *-se* in first and second person morphology. For low-frequency verbs (left plot), when primed with *-se*, the likelihood of using *-se* in first and second person is almost identical to the likelihood of using *-se* in third person. In contrast, for high-frequency verbs such as *ser*, *haber*, and *estar* (right plot), if *-se* was used in the preceding context, the likelihood of the past subjunctive appearing in first and second person is even lower than in non-recency contexts. As shown by the results for mid-frequency verbs (middle plot), this effect is gradient and relatively linear. The results suggest that especially for high-frequency verbs, the persistence effect actually bolsters the probability of use of *-se* forms in third person. It is only in these contexts that *-se* actually has a higher overall likelihood of use than *-ra*. This means that whereas for low-frequency verbs *se*-persistence counters paradigmatic atrophy, for high-frequency forms it apparently *creates* paradigmatic atrophy.

Although this result seems paradoxical at first, we believe that it gives crucial clues regarding the interplay of persistence and frequency effects in language use. The reason why, for high-frequency verbs, *-se*-persistence leads to a higher use of third-person morphology resides in the fact that due to their overall higher usage frequency, the primes themselves are usually high-frequency verbs in third person. Thus, the two high-frequency third-person forms *hubiese* 'have.PST.IPFV.SBV.3SG' and *fuese* 'be.PST.IPFV.SBV.3SG' alone make up 43.2 percent of all *-se* primes in our data (73/169). This descriptive result is mirrored by the fact that there is a correlation between the frequency of the target verb (i.e., our variable VERBFREQUENCY) and whether or not the prime and the target are formally identical. A U-test showed that this difference is highly significant ($W=174880$, $p_{\text{one-tailed}} < .001$): formal identity between the prime and the target is more likely for high-frequency than for low-frequency verbs. Consequently, it is

likely that the finding that high-frequency third-person *-se* forms profit most from the persistence effect is due to the well-known fact that lexical identity boosts persistence (see, for instance, Jaeger and Snider 2008). Given that low-frequency verbs are less likely to have been primed by a formally identical prime, they profit to a lesser degree from the lexical identity effect.

Note also that the overall results show that, in general, persistence effects are weaker for high-frequency verbs than for low-frequency verbs; for low-frequency verbs, the increase in the probability of use of *-se* is approximately 0.67 (about 0.35 for first/second person and about 0.32 for third person), for high-frequency verbs the increase is approximately 0.43 (-0.02 for first/second person and 0.45 for third person). The regression model shows this partial interaction effect (RECENT : VERBFREQUENCY) to be statistically significant at $p < .05^*$. Consequently, not only do high-frequency *-se* forms profit less from the canceling out of paradigmatic atrophy attributed to persistence, but persistence has a generally weaker effect on the likelihood of use of high-frequency *-se* forms.

6. Discussion

In Section 2, we established two hypotheses about the influence of persistence on processes of obsolescence in grammatical change that can be roughly summarized as follows. Hypothesis 1 claimed that the higher strength of the persistence effect of obsolescing constructions is due to the stronger surprisal effect of low-frequency constructions and that, consequently, persistence does not necessarily have a conserving effect in language change. Hypothesis 2 established that in line with implicit learning accounts of language production, the stronger persistence of obsolescing constructions to some degree conserves these constructions both in form and function. Note that these two hypotheses do not contradict each other; rather, Hypothesis 2 makes a stronger claim than Hypothesis 1.

The results from the descriptive and inferential analyses presented in this paper allow us not only to maintain the weaker Hypothesis 1, but also the stronger Hypothesis 2. In particular, we assume that persistence can conserve both the form and the function of a disappearing construction, and can therefore have both syntagmatic and paradigmatic effects on that construction. In this section, we explain our reasons for this assumption.

In contrast to previous assumptions from the literature, our analysis found almost no evidence of functional differences that govern the variation between the two past subjunctive forms in Spanish. The morphosyntactic parameters of negation, subordination, protasis, apodosis and negation of main clause that were proposed as indicators of functional differences in the previous literature did not have significant effects on the alternation in our data. We did find a verb bias between the two constructions in that in texts from middle and northern America, the verbs *estar* 'to be' and *entrar* 'to enter' have a higher probability to occur in the *-se* construction than other verbs. The reasons for this bias are unclear to us; one possibility could be phonological factors. Although according to the CART analysis auxiliatioin does not have a statistically significant effect on the distribution of the two variants, the variable importance test did suggest some effect of this variable. Likewise, in the regression analysis, auxiliatioin reached marginal statistical significance. These contradictory findings might be due to the fact that, in our data, all perfect constructions are formed with the auxiliary *haber* 'have', and that *haber* is one of the most frequent verbs in Spanish. Although CARTs and

random forests are unaffected by this multicollinearity, it might lead to considerable variation in the results.

If we based our considerations of the differences between the two constructions on the meaning of these constructions alone, we would thus be forced to conclude that they stand in free variation. Given the relatively low degree of variation explained by our statistical models, it is however likely that our analysis is missing parameters that are of importance for the alternation. Specifically, it is possible that including variables referring to the definiteness of the object and sentence type (cf. Guzmán Naranjo to appear) could add further effects to the model and thus also improve the degree of variance explained by the model. Note however that our findings do not coincide with Guzmán Naranjo's results regarding sentence type; if temporal or adversative subordination were of importance for the opposition in the data, we would have expected some effect of the variable SUBORDINATED. Given that our analysis did not find a statistically significant effect of whether or not the data was from spoken texts (variable GENRE), it appears that this difference should not be due to the fact that Guzmán Naranjo's analysis is solely based on spoken data. Since we only included one modal verb in our data (*poder* 'can'), our analysis could not reproduce the effect of modality tracked in Guzmán Naranjo's paper.

Our analysis has traced important differences between the two past subjunctive forms that are not due to differences in their grammatical meaning. Specifically, there is evidence for considerable diatopic variation. Thus, in European and southern American Spanish, the process of replacement of *-se* by *-ra* appears to have been slower than in middle and northern American Spanish, where the use of *-se* is significantly less likely. This result confirms the findings from previous studies on the alternation such as DeMello (1993).

However, none of these contextual factors reaches the importance of usage-based factors for a description of the *-ra/-se* alternation. The most important predictor governing the selection process is recency, i.e. the presence or absence in the recent context of a prior *-se* token. In particular, the analysis suggests a different strength of the effect of morphosyntactic persistence on the alternation. Whether or not a *-se* token appears in the previous context is by far the most important predictor in that the selection of *-se* becomes much more probable in such a context. In contrast, according to the CART, the effect of a recent appearance of a *-ra* token is statistically indistinguishable from contexts with no recent use of a past subjunctive. Our study thus demonstrates that the use of an obsolescing element, in this case the *-se* form, primes its upcoming use to a greater degree than the use of an element with higher productivity, i.e. the *-ra* form. It therefore suggests that persistence is highly relevant to the distribution of morphosyntactic alternations in which one of the variants is in the process of being replaced by the other variant with a similar discourse-pragmatic function. Our study thus establishes the importance of including persistence as a predictor of morphosyntactic alternations.

As argued in Section 2, this finding does not necessarily mean that persistence actively conserves variation. After all, one could argue that the differential effect of persistence in morphosyntactic alternations is an epiphenomenon of the obsolescence process that affects the less common variant. However, our findings actually suggest that persistence has a conserving effect in language change. Due to the ongoing obsolescence process, *-se* forms display a high degree of entrenchment. *-Se* past subjunctives are overwhelmingly formed from high-frequency verbs, in particular *haber* 'to have' and *ser* 'to be'. The entrenchment process is also visible in the paradigmatic atrophy of *-se* constructions. Thus, the use of *-se* constructions is mostly restricted to third person

singular morphology. In other words, due to their higher absolute usage frequency, forms such as *fue-se* 'ser-PST.IPFV.SBJ.3SG' and *hubie-se* 'haber-PST.IPFV.SBJ.3SG' are much more likely instantiations of the *-se* construction in Modern Spanish than forms such as *sintie-se-n* 'feel-PST.IPFV.SBJ-3PL' or *acaba-se-is* 'finish-PST.IPFV.SBJ-2PL'.

In line with Hypothesis 2, we find that persistence can to some degree counteract this paradigmatic atrophy. The paradigmatic atrophy of *-se* is restricted to contexts without a prime or containing a previous *-ra*. When primed by another token of *-se*, *-se* constructions are more likely to occur with plural number morphology, and especially with first and second person morphology. From a cognitive perspective, this effect demonstrates that persistence not only activates a specific instantiation of the primed construction, but also the constructional schema as a whole. In line with the implicit learning account of persistence (Bock and Griffin 2000), the effect of the global activation of the constructional schema can be cumulative, insofar as the repeated boost in morphosyntactic productivity can itself be entrenched in the speakers' mind. For these reasons, we interpret our findings as showing clear effects of the conserving effect of persistence in language change.

Interestingly, both the CART and the logistic regression analysis showed that this "resuscitating" effect of persistence is modulated by usage frequency. In particular, the probability of using *-se* with first or second person morphology increased in *-se*-persistence contexts only for low- or mid-frequency verbs, but not for high-frequency verbs. We attributed this difference to the fact that *-se* primes are very likely to be formed from high-frequency verbs in third person precisely because these forms are extremely frequent. Consequently, high-frequency verbs in third person are more likely to profit from stronger persistence effects due to lexical identity between prime and target than low-frequency verbs in third person.

This finding has important consequences regarding the influence of persistence on language change. In particular, it suggests that whereas for low-frequency forms, persistence can counter paradigmatic atrophy and thus strengthen their overall productivity, in high-frequency forms it basically serves to reinforce the already-existing frequency effect. In other words, the resuscitating effect of persistence is canceled out by a high degree of entrenchment. In line with recent studies on frequency effects such as Erker and Guy (2012), this finding suggests that usage frequency acts as a "gatekeeper" for other contextual parameters. Above a certain frequency threshold, persistence is a much worse predictor of the alternation because highly entrenched and conserved forms have a stronger representation in the language user's mind and therefore have to rely less on persistence for activation.

7. Conclusion

We have demonstrated in this paper that there is a strong relationship between persistence effects and the conservation of variation: persistence in discourse brings along with it a loss of restrictions (or, alternatively, a recovery of productivity) that obtain when no persistence is present. Persistence in a sense "resuscitates" a form (here the *-se* past subjunctive in Spanish) that otherwise seems well on its way to becoming moribund. While the effects of persistence on syntagmatic aspects of discourse are well-known (Szmrecsanyi 2005), what we have shown is that persistence can also have important effects on certain paradigmatic characteristics of variable phenomena, correlating with greater productivity of individual forms than in contexts without persistence.

More generally, there is an overarching effect of persistence on cases of variation wherein one variant can be shown to be older and obsolescing when compared to a younger, more vibrant, variant with which it is in competition. Persistence effects in such cases always show stronger effects on the obsolescing variant than on the innovative variant. While such asymmetrical persistence effects appear to fall out of research on priming in psycholinguistics (e.g. Jaeger and Snider 2008), since less frequent forms show greater surprisal effects than more frequent ones, it has not, to our knowledge, been tied in explicit fashion in prior research to the variants of a linguistic variable. In diachronic perspective, it is even possible to see the persistence effects on competing variants flip-flop. Torres Cacoullós (2015) shows how this has occurred for the simple present (older) and the present progressive (younger) forms in Spanish: during the period in which the present progressive form is entering into variation with the simple present, its persistence effects are relatively higher, but as it becomes more and more frequent, these effects wane. Correspondingly, the persistence effects on the simple past increase as this form becomes less frequent and its functional territory is partially usurped by the present progressive. Thus, while the simple present does not show signs of becoming obsolete in all of its contexts of use, it is the obsolescing form in the contexts in which it competes with the present progressive, and it is precisely in these contexts where persistence effects exert a strong influence on its occurrence.

Our study moreover demonstrates that the differential strength of persistence effects on variants in morphosyntactic alternations is not only a result of differences in their usage frequencies, as previous research would suggest. Rather, persistence can be said to conserve the use and productivity of an obsolescing variant. Specifically, the fact that persistence influenced the paradigmatic characteristics of the two variants in an asymmetric manner, counteracting the paradigmatic atrophy of *-se* forms due to entrenchment, indicates that persistence leads to an activation of not only specific *-se* forms, but the entire *-se* construction. It is due to this activation effect on the more abstract constructional level that persistence can boost the productivity of the obsolescing variant. This boost in productivity need not be temporal and can actually lead to greater resilience during the obsolescence process.

Lastly, our study has demonstrated an interplay between persistence and frequency effects in language use and language change. The degree to which persistence activates the use of *-se* forms is moderated by usage frequency; high-frequency *-se*-forms are less affected by persistence than low-frequency forms both regarding their overall usage frequency and the countering of paradigmatic atrophy. We attributed this finding to the fact that highly entrenched forms have a stronger representation in the language user's mind and therefore need not rely on persistence for activation. Contrary to expectations, our results demonstrate that for high-frequency *-se*-forms, *-se*-persistence even decreased the likelihood of using those *-se* forms with first and second person morphology because, due to their usage frequency, these forms are more likely to have been primed with lexically identical primes than low-frequency *-se* forms. These findings suggest that the effect of persistence on obsolescing forms differs regarding their usage frequency; whereas for low-frequency forms, persistence leads to a global productivity boost, for high-frequency forms it basically reinforces the conserving effect of usage frequency. Given that entrenchment and persistence are domain-general cognitive processes, we expect that these results can be replicated in the study of obsolescence phenomena in other languages.

From a more general perspective, the analysis has demonstrated the power of using variationist methodology for hypotheses derived from usage-based approaches, as well as the importance of including persistence when modeling morphosyntactic

alternations. Our findings suggest that approaches to morphosyntactic alternations that neglect cognitive dimensions of language use – particularly persistence and usage frequency – would have problems reaching a motivated distinction between *-ra* and *-se*. Our analysis also shows how corpus-based research can inform psycholinguistic research in the study of phenomena such as persistence. Our results could easily be tested in psycholinguistic experiments. For instance, a possible method of analysis would use the masked-priming paradigm (Forster and Davis 1984) to evaluate the difference in the perception of more and less entrenched *-se* forms as a function of the types of primes. Such an experiment would measure the participants' reaction times to more entrenched tokens such as *fuese* 'BE.PST.IPFV.3SG' and less entrenched tokens such as *llega-seis* 'ARRIVE-PST.IPFV2PL' after masked priming with an unrelated form, a *-se* form or a *-ra* form. The results from our study would lead us to expect that (a) *-se* forms are more likely to be primed by masked *-se* forms than *-ra* forms by masked *-ra* forms and that (b) less entrenched *-se* forms should 'profit' more from priming with *-se* than more entrenched *-se* forms. In the light of the complex interaction effect between usage frequency, persistence, and personal morphology found in this study, such an experimental approach should control carefully for the usage frequency of the target constructions, as well as formal identity between the masked prime and the target.

Appendix

Table 5. Full results from the logistic regression model

Variable	Level	n ra	n se	% se	Beta	OR	SE	Z	P
(Intercept)					-4.073	0.017	0.799	-5.1	0.000***
RECENT	None	2354	243	9.35	<i>Reference level</i>				
	ra	1054	71	6.31	0.492	1.636	0.755	0.653	0.514
	se	87	82	48.52	4.073	58.713	0.904	4.507	0.000***
TARGET_PERSONTHIRD	False	452	32	6.61	<i>Reference level</i>				
	True	3043	364	10.68	1.297	3.657	0.45	2.879	0.004**
VERBFREQUENCY	Numerical variable				0	1	0	3.279	0.001**
TARGET_NUMBER	Singular	2573	326	11.25	<i>Reference level</i>				
	Plural	922	70	7.06	-0.354	0.702	0.147	-2.403	0.016*
REGION	Europe	599	103	14.67	<i>Reference level</i>				
	Unknown	312	21	6.31	-0.348	0.706	0.574	-0.606	0.544
	Mid/NorAm	1590	86	5.13	-0.961	0.383	0.164	-5.869	0.000***
	SouthAm	994	186	15.76	0.045	1.046	0.158	0.282	0.778
GENRE	Academic	284	17	5.65	<i>Reference level</i>				
	Fict	1849	255	12.12	0.686	1.985	0.611	1.122	0.262
	News	496	57	10.31	0.656	1.928	0.637	1.031	0.303
	Oral	866	67	7.18	0.247	1.281	0.634	0.39	0.696
NEGATION	False	3102	349	10.11	<i>Reference level</i>				
	True	393	47	10.68	-0.041	0.96	0.177	-0.231	0.817
SUBORDINATED	False	291	32	9.91	<i>Reference level</i>				
	True	3204	364	10.2	0.291	1.338	0.231	1.262	0.207
AUXILIARY	False	3096	329	9.61	<i>Reference level</i>				
	True	399	67	14.38	0.317	1.373	0.176	1.799	0.072
RECENTRA : TARGET_PERSONTHIRDTRUE					-0.889	0.411	0.782	-1.137	0.256
RECENTSE : TARGET_PERSONTHIRDTRUE					-2.289	0.101	0.941	-2.433	0.015*
RECENTRA : VERBFREQUENCY					-0.00015	0.00012	0.000	-1.271	0.204
RECENTSE : VERBFREQUENCY					-0.00028	0.00014	0.000	-2.009	0.044*
TARGET_PERSONTHIRDTRUE : VERBFREQUENCY					-0.00011	0.00004	0.000	-2.69	0.007**
RECENTRA : TARGET_PERSONTHIRDTRUE : VERBFREQUENCY					0.00015	0.00012	0.000	1.229	0.219
RECENTSE : TARGET_PERSONTHIRDTRUE : VERBFREQUENCY					0.00032	0.00014	0.000	2.235	0.025*
Model evaluation	AIC		2291.2						
	c index of concordance		0.740						
	Somers' dxy		0.480						

References

- Asratían, Arucia. 2007. Variación *-ra/-se* en el español hablado en Caracas. *Boletín de lingüística* 19(5). 5-41.
- Baayen, Harald. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Bock, Kathrin J. and Zenzi M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129(2). 177-192.
- Bock, Kathryn J. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18. 355-387.
- Branigan, Holly, Martin Pickering and Alexandra Cleland. 1999. Syntactic priming in written production: Evidence for rapid decay. *Psychological Bulletin and Review* 6. 635-640.
- Breiman, Leo. 2001. Random forests. *Machine Learning* 45. 5-32.
- Breiman, Leo, Jerome Friedman, Charles J. Stone and R.A. Olshen. 1984. *Classification and Regression Trees*. New York: Chapman and Hall.
- Bybee, Joan L. 2003. Mechanisms of change in grammaticization: the role of frequency. In Janda, Richard 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 Rena Torres Cacoullos. 2009. The role of prefabs in grammaticization: how the particular and the general interact in language change. In Corrigan, Roberta, Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley (eds.), *Formulaic Language, Volume I: Distribution and Historical Change*, 187-217. Amsterdam, Philadelphia: Benjamins.
- Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford Univ. Press.
- Davies, Mark. 2002. *Corpus del español (100 million words, 1200s-1900s)*. Available online at <http://www.corpusdelespanol.org>. Last access: 22 December 2015.
- Day, Meagan. 2011. *Variation in the use of the -ra and -se forms of the imperfect subjunctive in Modern Spoken Peninsular Spanish*. NWAV 40, Georgetown University.
- DeMello, George. 1993. *-ra vs. -se subjunctive: a new look at an old topic*. *Hispania* 76(2). 235-243.
- Diessel, Holger. 2011. Review article of 'Language, usage and cognition' by Joan Bybee. *Language* 87. 830-844.
- Eckardt, Regine. 2008. Concept priming in language change. *Theoretical Linguistics* 34(2). 123-133.
- Erker, Daniel and Gregory R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3). 526-557.
- Ferreira, Victor S. 2003. The persistence of optional complementizer mention: Why saying a "that" is not saying "that" at all. *Journal of Memory and Language* 48. 379-398.
- Fillmore, Charles J., Paul Kay and Mary C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions. *Language* 64 64. 501-538.

- Forster, Kenneth I. and Chris Davis. 1984. Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10. 680-698.
- Goldberg, Adele E. 1995. *Constructions: a Construction Grammar Approach to Argument Structure*. Chicago, London: The University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E. 2013. Constructionist approaches. In Trousdale, Graeme and Thomas Hoffman (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
- Gries, Stefan Th. 2005. Syntactic priming: a corpus-based approach. *Journal of Psycholinguistic Research* 34(4). 365-399.
- Gries, Stefan Th. and Stefanie Wulff. 2009. Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics* 7. 163-186.
- Guzmán Naranjo, Matías. to appear. The se-ra Alternation in Spanish Subjunctive. *Corpus Linguistics and Linguistic Theory*.
- Hartsuiker, Robert J. and Herman J. Kolk. 1998. Syntactic persistence in Dutch. *Language and Speech* 41(2). 143-184.
- Haverkate, Henk. 2002. *The Syntax, Semantics and Pragmatics of Spanish Mood*. Amsterdam, Philadelphia: John Benjamins.
- Hilpert, Martin. 2014. *Relating language change to language processing: a second look at asymmetric priming*. Paper presented at ICEHL18. Leuven.
- Hothorn, Torsten, Kurt Hornik and Achim Zeileis. 2006. Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651-674.
- Jaeger, T. Florian and Neal Snider. 2008. Implicit learning and syntactic persistence: surprisal and cumulativity. In Love, B. C., K. McRae and V. M. Sloutsky (eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 1061-1066. Austin, TX: Cognitive Science Society.
- Jäger, Gerhard and Anette Rosenbach. 2008. Priming and unidirectional language change. *Theoretical Linguistics* 34(2). 85-113.
- Kempas. 2011. Sobre la variación en el marco de la libre elección entre *cantara* y *cantase* en el español peninsular. *Moenia* 17. 243-264.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith. 2009. *Change in Contemporary English. A Grammatical Study*. Cambridge: Cambridge University Press.
- Lemon, Francis. 1925. The relative frequency of the subjunctive forms in *-se* and *-ra*. *Hispania* 8. 300-302.
- Lope Blanch, Juan M. (ed.). 1977. *Estudios sobre el español hablado en las principales ciudades de América*. México: Universidad Nacional Autónoma de México.
- Lunn, Patricia V. 1995. The evaluative function of the Spanish subjunctive. In Bybee, Joan and Suzanne Fleischman (eds.), *Modality in Grammar and Discourse*, 429-449. Amsterdam, Philadelphia: John Benjamins.
- Pickering, Martin J. and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39(4). 633-651.

- R Development Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at <http://www.R-project.org>. Last access 26 December 2015.
- Real Academia Española. 2010. *Nueva gramática de la lengua española. Manual*. Madrid: Escasa Libros.
- AUTHOR. 2014.
- AUTHOR. 2015.
- Schwarz, Christian. 2016. Recency as a factor of phonological variation. In Behrens, Heike and Stefan Pfänder (eds.), *Experience Counts: Frequency Effects in Language*, 91-109. Berlin, New York: De Gruyter.
- AUTHOR. 2013.
- Strobl, Carolin, James Malley and Gerhard Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348.
- Sussman Goldberg, Barbara. 1995. The *-ra* and *-se* opposition in Spanish. In Contini-Morava, Ellen and Barbara Sussman Goldberg (eds.), *Meaning as Explanation. Advances in Linguistic Sign Theory*, 381-404. Berlin, New York: De Gruyter.
- Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: a corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1). 113-150.
- Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin, New York: De Gruyter.
- Tagliamonte, Sali and Harald Baayen. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135-178.
- Tamminga, Meredith. 2014. *Persistence in the production of linguistic variation*. Ph.D. thesis, University of Pennsylvania.
- Torres Cacoullos, Rena. 2015. Gradual loss of analyzability: Diachronic priming effects. In Adli, Aria, Marco García García and Göz Kaufmann (eds.), *Variation in Language: System- and Usage-based Approaches*, 265-287. Berlin, New York: De Gruyter.
- Valeš, Miroslav. 2006. El prestigio desigual de las formas del imperfecto de subjuntivo *cantara/cantase*. In Moya Corral, Juan A. and Marcin Sosinski (eds.), *Lexicografía y enseñanza de la lengua española. Actas de las XI Jornadas sobre la enseñanza de la lengua española*, 303-311. Granada: Universidad de Granada.
- Wright, Leavitt O. 1926. The indicative forms in *-ra* in Spanish America. *Hispania* 9. 288-293.