Studies in Language Companion Series
234

CHAPTER 9

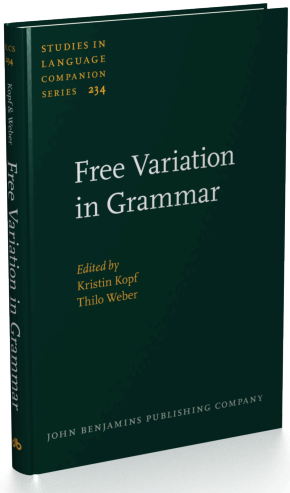
# Syntactic priming and individual preferences

## A corpus-based analysis

Malte Rosemeyer | Freie Universität Berlin

https://doi.org/10.1075/slcs.234.09ros






Pages 260–283 of
**Free Variation in Grammar: Empirical and theoretical approaches**
**Edited by Kristin Kopf and Thilo Weber**
[**Studies in Language Companion Series**, 234]
2023. vi, 352 pp.

CHAPTER 9

# Syntactic priming and individual preferences

## A corpus-based analysis

Malte Rosemeyer
Freie Universität Berlin

This paper analyses the relation between syntactic priming/persistence and individual preferences in the variation between the two forms of the Spanish past subjunctive (*-ra* vs *-se*). The analysis finds that the probability of repetition of one of these variants is governed by an interaction between individual preferences and the difference between self- and other-priming. Previous use of a variant by another speaker is more likely to lead to persistence than previous use by the current speaker in the case that the current speaker does not usually prefer usage of the variant in question. However, if the current speaker prefers usage of the variant in question, this situation is inverted. These findings suggest that for speakers who prefer usage of the primed variant, self-priming in corpus data might actually be more adequately explained as an effect of individual preferences.

**Keywords:** variation, persistence, priming, subjunctive, Spanish, sociolinguistics

### 1. Introduction

Variationist linguistics argues that what is commonly perceived as noise in linguistic data is not 'free' variation in the sense that it cannot be accounted for; rather, "speakers are rarely entirely free to make a choice between the alternatives, since each grammatical option is usually subject to a variety of constraints, some of which may very subtly guide speakers to make the choices that they do" (Cappelle 2009: 183). Free variation consequently equals unexplained variation, and raising the question of free variation is tantamount to a methodological challenge: how can we refine statistical models in such a way as to explain unexplained variation? This paper focuses on one aspect of this problem, namely, the relevance of individual variation. It argues that in order to enhance our models of variation in lin-

https://doi.org/10.1075/slcs.234.09ros

guistic production, it is not sufficient to control for inter-speaker variation on the one hand and syntactic priming/persistence – the repetition of a previously used variant in discourse due to activation effect – on the other hand. Rather, individual preferences are predicted to heavily influence syntactic priming effects. Selection of one variant over another is not only dependent on individual preferences and syntactic persistence *per se* but also the interaction between these two effects.

In order to demonstrate this point, the paper analyses the variation between the two forms of the Spanish past subjunctive (*cant-ara* 'sing-PST.SBJ.3SG' vs *cant-ase* 'sing-PST.SBJ.3SG') in a corpus of spoken Peninsular Spanish. Variation between the two forms is thought to be mostly free, although substantial dialectal preferences are documented. In a previous study (Rosemeyer & Schwenter 2019), it was shown that persistence heavily influences the variation between *-ra* and *-se*. This study advances previous analyses by taking into account the role of individual speaker preferences (i.e. the degree to which a speaker prefers using *-ra* or *-se*) for the variation between *-ra* and *-se*. It demonstrates the relevance of distinguishing between self- and other-priming (i.e. whether or not the prime is uttered by the speaker of the target construction or a different speaker) in such analyses. Results show that speakers who disprefer the primed variant are more likely to repeat the primed variant when that variant has been used by the previous speaker (other-priming) than when they have used it themselves (self-priming). In contrast, speakers who prefer the primed variant are shown to be less likely to repeat the primed variant when that variant has been used by the previous speaker (other-priming) than when they have used it themselves (self-priming). This finding is explained in terms of an activation-based account of persistence. Preference of a speaker for a variant over another can be described in terms of a higher baseline activation of the variant in question, which is why, for speakers who prefer the primed variant, the recent use of that variant impacts the probability of that speaker repeating it less than for speakers who disprefer the primed variant. In contrast, previous use of a variant is more likely to lead to repetition of that variant by speakers who prefer it than for speakers who disprefer it. I propose that such self-repetition processes should not be described as syntactic priming but, rather, are likely to reflect individual preferences. My results suggest that individual preferences reinforce themselves in usage, constituting the speaker's linguistic style via entrenchment processes.

This paper is structured as follows. In Section 2, I describe the relevance of modelling activation effects such as persistence in variationist analyses and develop a model of the interaction between individual preferences and persistence. Section 3 presents the case study, an analysis of the variation between the two Spanish past subjunctive forms in terms of the predictions developed in Section 2. Section 4 discusses the relevance of these findings for current variationist approaches to the description of language, as well as language change.

## 2. Persistence and individual variation

Drawing on the psycholinguistic notion of priming (Dell 1986; Dell et al. 1997; Bargh & Chartrand 2000), variationist studies have shown that, in a situation in which the speaker can select between two or more variants that express a similar function in discourse, she or he will display a tendency towards selecting the variant that she or he has last heard or read. This probabilistic constraint is termed persistence (Szmrecsanyi 2005, 2006), syntactic priming (Bock 1986; Mahowald et al. 2016) or perseveration (Cameron & Flores-Ferrán 2004), and has first been identified in sociolinguistic variationist studies (Poplack 1980; Weiner & Labov 1983; Cameron 1995; Pereira Scherre 2001; Cameron & Flores-Ferrán 2004; Tamminga 2016; Callaghan & Travis 2020; among many others). Persistence escapes any notion of functional or formal differences between linguistic constructions and consequently needs to be dealt as an additional predictor in variationist analyses. In the words of Szmrecsanyi (2005: 140), "persistence […] poses a problem to some varieties of quantitative linguistic research in that text frequencies of some linguistic pattern may be misleading unless, for instance, textual distances between the individual hits are factored in". Indeed, Szmrecsanyi shows that inclusion of predictor variables operationalising persistence increases the quality of statistical models of linguistic variation substantially. To give an example from Spanish, which will be studied in more detail later, Spanish has two past subjunctive forms, which express similar meanings. Thus, Example (1a), taken from the Corpus del español (Davies 2019) can easily be paraphrased using (1b). A corpus-based study by Rosemeyer & Schwenter (2019) demonstrated that use of either of these variants leads to an increased probability of use of the same variant in the subsequent context, as compared to the other variant.

(1) a. *le había-n ped-i-do que* ***cant-a-ra***
him have.PST.IPFV-3PL ask-TH-PTCP that 'sing-TH-PST.SBJ.3SG'
b. *le había-n ped-i-do que* ***cant-a-se***
him have.PST.IPFV-3PL ask-TH-PTCP that 'sing-TH-PST.SBJ.3SG'
'They had asked him to sing.'

Crucially for our purposes, persistence is moderated by additional factors, many of which have been studied in the relevant nature. First, persistence becomes less likely with increasing distance between the prime and the token of the construction under study (Gries 2005; Szmrecsanyi 2005, 2006; Gries & Kootstra 2017; Rosemeyer & Schwenter 2019). This means that the increase in the probability of use of the primed construction is exponentially higher the lower the distance between the prime and the target. Second, formal identity between the prime and the target likewise increases the odds for persistence (Gries 2005; Jaeger & Snider

2008). In other words, repetition at the token level is more likely than repetition at the more abstract type level. Third, various studies have demonstrated that persistence is governed by the individual factors such as age of the speaker (see Szmrecsanyi 2006: 196–198 and references therein). Szmrecsanyi's results suggest an interaction between age and distance in words between prime and target, in that the decay of the activation effect seems stronger for older than younger speakers. These effects are usually explained in terms of an activation-based account of persistence, which assumes that "grammatical structures vary in baseline activation, and speakers prefer structures that are more activated over structures that are less activated" (Cho et al. 2020: 477). The influence of the moderator variables described in the last paragraph can, consequently, either be explained in terms of assumptions with respect to memory constraints (distance between prime and target, age) or the degree to which a prime activates the cognitive representation of the respective construction in the listener's mind (formal identity).

A fourth important factor, which will be the focus of the present paper, is the difference between self-priming and other-priming. Self-priming refers to a situation in which the prime, i.e. preceding instance of the token, is uttered by the same speaker as the target token. In contrast, other-priming describes the situation in which the prime is uttered by a different speaker from the one uttering the target token. Corpus-based studies have found that self-priming is more likely than other-priming (Gries 2005; Fricke & Kootstra 2016). Put simply, it appears that speakers are more likely to repeat their own words than other people's. Within the activation-based account of persistence, one might assume that self-use of a linguistic element activates the cognitive representation of that element in the speaker's mind to a greater degree than other-use of the element. However, this explanation is challenged by experimental evidence that comprehension-to-production priming (referring to a situation in which production of a linguistic element is facilitated by previous exposure to that element) has similar effect sizes as production-to-production priming (see Bock et al. 2007; Pickering & Ferreira 2008: 440–441).

In this paper, I will develop and discuss the hypothesis that in some cases, the finding that speakers are more likely to repeat themselves in corpus data is an artifact of the nature of production data. When analysing corpus data, a situation in which a speaker repeats her or his selection of one variant can mean one of two things. First, the repetition could be described as persistence, in that activation of the representation of the variant in question by the previous utterance has led to easier processing and thus easier production of the same variant in later discourse. Second, it might mean that a speaker simply uses that variant frequently. In the latter case, repetition of the variant would need to be explained

as an effect of individual preferences, i.e. entrenchment of the usage of that variant in the speaker's idiolect.

As an illustration, consider the made-up examples in (2), where speakers have the choice between the prepositional dative (*give the watch to Susan*) and ditransitive dative (*give Susan the watch*) (see e.g. Szmrecsanyi et al. 2017). Whereas the use of *gave the laptop to Lisa* in (2a) would usually be described as an instance of self-priming, (2b) would be described as other-priming. However, the fact that Mary has already used a prepositional dative in (2a) might lead us to suspect that Mary generally prefers using the prepositional dative. In contrast, John's use of the prepositional dative in (2b) cannot be interpreted along the same lines.

(2) a. **Mary:** *First I gave the watch to Susan and then I gave the laptop to Lisa.*
b. **Mary:** *First I gave the watch to Susan.*
**John:** *And then you gave the laptop to Lisa.*

These considerations are important because they suggest an asymmetry between self-priming and other-priming in terms of the relevance of individual preferences. This might account for the fact that previous studies have found self-priming to be more likely than other-priming. In a recent study, Gradoville (2019) analyses the relation between persistence and individual variation in the reduction of *para* 'to, for, in order to' to *p(r)a* in spoken Brazilian Portuguese. He finds that failure to account for the individual speaker leads to an overestimation of persistence. In Gradoville's (2019: 120) words,

> In the case of the present study, although priming is suggested to play a role in the variation surrounding the reduction of *para*, its effect is nowhere near as strong as is suggested when individual variation is not accounted for in the statistical model.

This can be easily explained using our example in (2a). Imagine a situation in which the speaker Mary demonstrates a categorical selection rate of the prepositional dative. In this case, the selection of *gave the laptop to Lisa* over *gave Lisa the laptop* in (2a) cannot be attributed to persistence (or indeed, any other constraint). In contrast, consider a situation in which Mary's idiolect displays variation between the prepositional and the ditransitive dative. Here, the use of the prepositional dative in (2a) might well be a persistence phenomenon.

In this paper, I will argue that Gradoville's (2019) findings indicate more than just a methodological problem. In particular, I propose that the influence of individual variation on the probability for repetition to occur can be explained in terms of entrenchment processes at the speaker level. Central to this idea is the notion, expounded in Jaeger & Snider (2008, 2013), that persistence is not only contingent on activation in the hearer's mind, but also on expectedness. Put sim-

ply, linguistic stimuli that the hearer is less accustomed to will be more unexpected, i.e. surprising, to the hearer. This is due to the fact that language users "implicitly maintain and update probabilistic distributions over linguistic structures" (Jaeger & Snider 2008: 1061). Within a given context, a speaker will have expectations based on her idiolect as to which linguistic elements are likely and unlikely to be used, a fact that might be described as differences in the baseline activation levels of these elements in the given context. Persistence can thus be described as a symptom of implicit learning, i.e. "longer term adaptation within the cognitive mechanisms for creating sentences" (Bock & Griffin 2000: 177). As a result, "less probable syntactic structures, if observed, lead to a bigger change in the probability distribution, which in turn leads to an increased probability of reusing the same structure" (Jaeger & Snider 2008: 1062). In line with these assumptions, Jaeger & Snider (2008) found that the use of unusual passives such as *was ferried* leads to stronger persistence than more frequent passives such as *was made*.

Rosemeyer & Schwenter (2019) interpreted these results as proof that the notion of expectedness is intimately related to the degree of entrenchment of linguistic elements in the language user's mind. Entrenchment may be defined as "the storage of concepts and constructions as (variably) routinized items in long-term memory" (Schmid 2010: 121). Syntagms such as *I don't know, I don't think, do you want* or and *I said* are likely to receive a holistic representation in the speakers' minds due to their high usage frequency. Entrenchment can be framed in terms of predictability (see Diessel 2011: 833): in a string of words such as *I don't know*, the separate words are highly predictable from each other, a fact that corpus-based approaches typically describe using measures of collocation strength such as transitional probabilities or pointwise mutual information.

Since entrenchment is essentially a function of the usage frequency of a linguistic element, linguistic elements that have a low degree of entrenchment in the language user's mind are expected to be less predictable and therefore lead to stronger persistence than elements that have a weak degree of entrenchment. Indeed, in their corpus-based analysis, Rosemeyer & Schwenter (2019) were able to show that while Spanish speakers can select between two variants of the past subjunctive (cf. Example (1) above), the *ra*-form is much more frequent, leading to stronger entrenchment and weaker persistence compared to the *se*-form. The authors also showed that this effect is lexically specific: highly entrenched *se*-forms (especially *fuese* 'be.PST.SBJ.3SG') lead to weaker persistence than less entrenched *se*-forms (such as *ca-yese* 'fall-PST.SUBJ.3SG'). A recent study by Jacobs et al. (2019) suggests that this finding can indeed be characterised as an inverse frequency effect. In particular, Jacobs et al. (2019) present evidence from psycholinguistic experiments that find that in comprehension-to-production priming, rarer syntactic structures are more likely to persist than more frequent

structures. As a result of these findings, they propose a hybrid model that integrates both activation and error-driven/implicit learning accounts of persistence.

The fact that usage frequency moderates the probability for a linguistic element to persist has important repercussions for the analysis of persistence in corpus-based studies. This is due to the fact that speakers may exhibit difference in individual preferences with respect to the use of linguistic elements. For instance, in Gradoville's (2019) study about reduction of Portuguese *para*, the author found that more than half of the speakers in the sample displayed near-categorical reduction rates. In contrast, a significant number of speakers showed rather low regular reduction rates (Gradoville 2019: 107). As in the case of the Spanish past subjunctive, such high rates of inter-individual variation are indicative of the fact that the variation between these forms is governed less by structural constraints than by individual preferences, i.e. it can be characterised as 'free' variation. In line with the implicit learning account of persistence, however, this 'free' variation should be constrained by an interaction between persistence and inter-individual differences in terms of baseline activation levels of such constructions. If the degree of unexpectedness of a linguistic element influences the probability for persistence to occur, the implicit learning account to persistence would predict that the probability to repeat a reduced *para* token would be higher for speakers who regularly display low reduction rates (for which reduced *para* is unexpected) than for speakers who regularly display high reduction rates (for which reduced *para* is expected).

However, it is also to be expected that the influence of individual preferences on persistence is moderated by the difference between self- and other-priming. As noted by Jacobs et al. (2019: Section 6.1), implicit learning accounts to persistence are assumed to be unable to explain self-priming because trivially, speakers are generally less likely to be surprised by their own choice of words than by the choice of words of other speakers (cf. also Cho et al. 2020: 477). As a result, the prediction formulated in the previous paragraph is likely to only be relevant for other-priming, where the prime can actually be unexpected to the hearer. For self-priming, we should expect individual preferences to have the opposite effect. In terms of Gradoville's data, the probability to self-repeat a reduced *para* token would be higher for speakers who regularly display high reduction rates than for speakers who regularly display low reduction rates. Indeed, individual preferences can best be described in terms of this self-reinforcing repetition that are indicative of the speaker's language style. Consistent and selection of one variant over another is an expression of the speaker's linguistic *habitus* (see, e.g, Adli 2013). Note that this assumption is in line with historical studies to persistence effects, which have hypothesised that persistence can have a conserving effect in language change (Rosemeyer 2015; Rosemeyer & Schwenter 2019), by which linguistic elements that have become uncommon are maintained. In the same way, a conserva-

tive speaker might preserve a variant otherwise uncommon in her or his linguistic environment by self-reinforcing repetition.

In this section I have developed the hypothesis of a complex interaction between persistence and individual preferences/entrenchment at the speaker level. If strength of persistence is partially governed by the degree of expectedness of the stimulus, individual preferences should massively influence persistence. This hypothesis leads to the prediction that the difference between self- and other-priming is moderated by individual preferences, such that other-priming is expected to be stronger than self-priming when the speaker generally prefers usage of the competing variant. In the next section of this paper, I will test this prediction on the basis of an analysis of natural data from spoken Spanish.

## 3. The case study: Variation in the Spanish past subjunctive

I carried out a corpus-based analysis of the variation between the Spanish past subjunctive forms *-ra* and *-se* (see 3).

(3) *cant-a-ra* 'sing-TH-PST.SBJ.3SG' vs
*cant-a-se* 'sing-TH-PST.SBJ.3SG'

The *-ra*/*-se* alternation has received much attention in previous research (Lemon 1925; Wright 1926; Bolinger 1956; DeMello 1993; Sussman Goldberg 1995; Asratián 2007; Anderson 2017; Guzmán Naranjo 2017; Bermejo 2019; Rosemeyer & Schwenter 2019; Christ & Feldhausen 2021; Guajardo 2021). There is a consensus in the literature that the *-ra* subjunctive generally has a higher usage frequency than the *-se* subjunctive. This is due to the fact that *-ra* is an innovative variant that underwent semantic change from an original pluperfect reading towards expression of the subjunctive (Veiga Rodríguez 1996; Becker 2008; Rojo & Vázquez Rozas 2014; Rosemeyer 2021), starting to replace *-se* in Early Modern Spanish. Likewise, there is solid evidence for dialectal differences, such that Latin American Spanish dialects use *-se* less than Spanish dialects. There is much less consensus in the literature as to the contextual predictors that govern this variation. Indeed, many authors claim that there is little semantic difference between the two variants (Alarcos Llorach 1999: 158–160; Kempas 2011: 158; Guzmán Naranjo 2017; Rosemeyer & Schwenter 2019), reducing the variation to style and individual preferences (but see Bolinger 1956). Consequently, the *-ra*/*-se* alternation might represent one of the cases that can best be described as free variation, and a perfect test case for studying the interaction between persistence and individual preferences.

### 3.1 Data

I extracted all occurrences of the *-ra* and *-se* forms from the PRESEEA corpus of spoken Spanish (PRESEEA 2014). The PRESEEA corpus contains semi-structured sociolinguistic interviews from a variety of Spanish dialects. Given the strong impact of dialectal variation on the *-ra*/*-se* alternation, I decided to restrict the search query to interviews from Spain. I used available recordings from Alcalá de Henares ($n = 18$), Granada ($n = 17$), Madrid ($n = 18$), Málaga ($n = 18$), Palma de Mallorca ($n = 54$), Santiago de Compostela ($n = 53$) and Valencia ($n = 17$). Together, these transcripts have a length of more than 2.2 million words. Extraction procedures from these data lead to a total number of $n = 4066$ *-ra* and *-se* tokens.

### 3.2 Persistence as a predictor of the variation between -ra and -se

Table 1 gives the distribution of *-ra* and *-se* in the seven sub-corpora. The results from the table suggest substantial dialectal variation even within the Peninsular conversations, ranging from 6.2% *-se*-selection in the conversations from Granada to 23.7% in the conversations from Madrid. It has been a consistent finding in the literature that Southern European Spanish dialects frequently pattern with Latin American dialects, and the distribution in Table 2 seems to suggest that the same is the case here.

**Table 1.** Variation between *-ra* and *-se* forms, by sub-corpus

| | Granada | Alcalá | Málaga | Santiago | Valencia | Palma | Madrid |
|---|---|---|---|---|---|---|---|
| *n -ra* | 212 | 266 | 277 | 1077 | 278 | 994 | 270 |
| *n -se* | 14 | 36 | 41 | 234 | 61 | 222 | 84 |
| % *-se* | 6.2 | 11.9 | 12.9 | 17.8 | 18.0 | 18.2 | 23.7 |

It is important to note that within these sub-corpora representing different dialectal regions, individual variation abounds. Figure 1 visualises the percentage of *se*-usage (compared to *ra*-usage) for each of the $n = 361$ speakers in the corpus. In a boxplot such as the one in Figure 1, the boxes show the middle 50% of scores (i.e. the range between the 25th and 75th percentile). With the exception of the data from Granada, where almost all speakers tend towards using *-ra* over *-se*, all sub-corpora show substantial inter-speaker variation. Likewise, all sub-corpora contain speakers who only use *se*-forms in the respective interview. At the same time, the amount of inter-speaker variation displayed by the speakers in each corpus seems to correlate with the overall mean of *se*-usage described in Table 1. In other words, those dialects that display most *-se*-usage also display most variation

between *-se* and *-ra*. Consequently, while it appears that to individual preferences are to some degree predicted by dialectal variation, this correlation is not absolute.

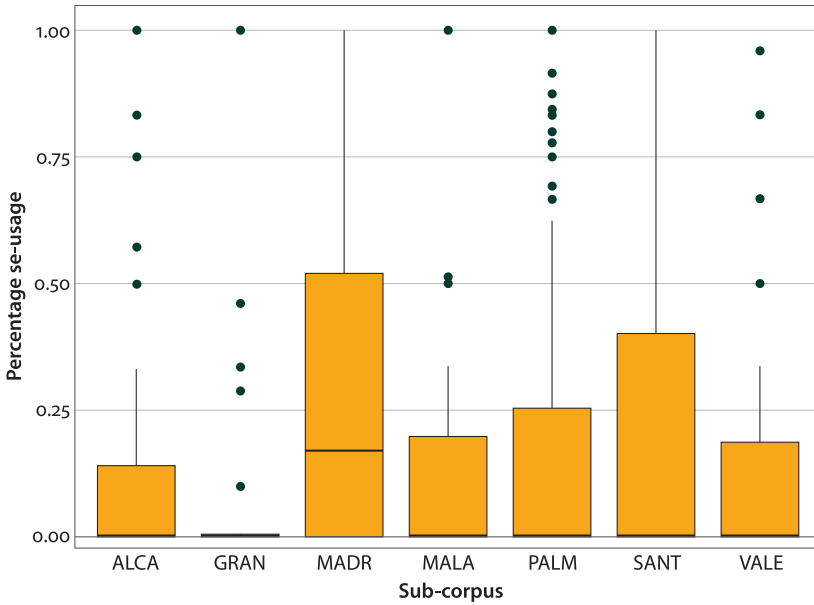



**Figure 1.** Individual variation in *-ra*/*-se* usage, by sub-corpus

Rosemeyer & Schwenter (2019) found that due to its status as a minority variant, previous use of *-se* is more likely to lead to persistence than the use of *-ra* in their corpus of written Spanish texts. Table 2 describes the distribution of *-ra* and *-se* in the PRESEEA corpus by previous use of a past subjunctive form. Tokens from the beginning of the conversation, in which no previous past subjunctive form is attested, were annotated as 'none', and can be used as a reference level. In comparison to contexts with unknown previous use of a past subjunctive form, previous use of *-ra* only increases the mean usage frequency of *-ra* from 87.2% to 90.4%. In contrast, previous use of *-se* increases the mean usage frequency of *-se* from 12.8% to 54.8%. The effect reaches statistical significance according to a $\chi^2$ test ($\chi^2(2) = 790.9$, $p < .001^{***}$). The analysis thus reproduces Rosemeyer & Schwenter's (2019) findings on the basis of spoken data; previous use of *-se* is more likely to lead to persistence than previous use of *-ra*.

Table 3 again examines the distribution of *-ra* and *-se* by previous use of a past subjunctive form, distinguishing this time between self- and other-priming. For this analysis, the $n = 195$ tokens without a prime, i.e. from the beginning of the conversation, were excluded (see Table 1). The analysis replicates the results by Gries (2005) and Fricke & Kootstra (2016), where self-priming was found to

**Table 2.** Variation between *-ra* and *-se* forms, by previous past subjunctive form

| Previous past subjunctive form | *n* | % *-ra* | % *-se* |
|---|---|---|---|
| none | 195 | 87.2 | 12.8 |
| *-ra* | 3216 | 90.4 | 9.6 |
| *-se* | 655 | 45.2 | 54.8 |

be more likely than other-priming. Indeed, according to $\chi^2$ tests, only the distribution for self-priming reaches statistical significance ($\chi^2(1) = 1030.6$, $p < .001^{***}$), whereas the distribution for other-priming fails to do so ($\chi^2(1) = 2.5$, $p > .05$).

**Table 3.** Variation between *-ra* and *-se* forms, by previous past subjunctive form and the difference between self- and other-priming

| Previous past subjunctive form | *n* | Self-priming | | Other-priming | |
|---|---|---|---|---|---|
| | | % *-ra* | % *-se* | % *-ra* | % *-se* |
| *-ra* | 3216 | 93.6 | 6.4 | 82.3 | 17.7 |
| *-se* | 655 | 34.2 | 65.8 | 77 | 23 |

## 3.3 Modelling the influence of individual preferences

Recall that discussion in Section 2 led to the prediction that the difference between self- and other-priming is moderated by individual preferences, such that other-priming is expected to be stronger than self-priming if the speaker generally prefers usage of the competing variant. As regards to methodology, this implies that in order to evaluate these predictions the dependent variable should not be *-ra*/*-se* selection *per se*, but rather whether or not the target is the same form as the prime. In establishing such a dependent variable – termed PERSIST – it is thus possible to predict the probability of persistence as a function of various predictor variables and their interactions by using a logistic regression model. The variable PERSIST was assigned the value 'True' in the case that a *-ra* token was preceded by a *-ra* token or a *-se* token was preceded by a *-se* token. In all other cases, the variable was assigned the value 'False'.

The nature of the dependent variable PERSIST dictates that the first *-ra*/*-se* token at the beginning of each conversation could not be taken into account. I consequently deleted these $n = 195$ tokens (one per conversation) from the dataset, leaving us with a final dataset of $n = 3871$ tokens of *-ra* or *-se* that are preceded by at least one other token of *-ra* or *-se*.

Table 4 summarises the predictor variables used in the regression model. Inclusion of the first two predictor variables was motivated in terms of the predictions established in Section 3. First, the variable PreviousSpeaker measures whether or not the speaker of the *-ra*/*-se* target is identical to the speaker of the previous *-ra*/*-se* token. Second, the variable SpeakerPreference measures the percentage with which the current speaker used the target form (*-ra* or *-se*) in the conversation (including the first tokens of *-ra* or *-se* in the conversation). Speaker preferences are operationalised the same way in Gradoville's (2019) study. I am well aware of the fact that this operationalisation procedure is inherently circular, in that the speaker's choice of one variant over the other is motivated by the overall usage frequency of that variant in the speaker's production data. Put simply, it is unsurprising that a speaker who generally prefers usage of *-ra* in a conversation should select also select *-ra* at one point in this conversation. However, recall that inclusion of SpeakerPreference in the model was not motivated by an interest about the simple correlation between SpeakerPreference and Persist, but rather the assumption that the effect of PreviousSpeaker (i.e. self- vs other-priming) on Persist is moderated by SpeakerPreference (see prediction (b) above). In other words, the fact that my results show SpeakerPreference to have a significant main effect on the choice between *-ra* and *-se* is trivial. However, the effect of the interaction effect between PreviousSpeaker and SpeakerPreference is not: it is not circular to assume that the strength of the effect of PreviousSpeaker is moderated by SpeakerPreference.[1]

The third predictor variable, Identity, was included on the basis of previous studies on persistence. As outlined in Section 2, studies such as Gries (2005) and Jaeger & Snider (2008) showed that persistence is more likely to occur in situations in which the prime and the target are formally identical. A past subjunctive form such as *fuese* 'be.pst.sbj.3sg' is more likely to prime subsequent use of *fuese* than of *cantase* 'sing-th-pst.sbj.3sg'. Gries (2005) moreover found an interaction effect between this variable and the difference between self- and other-priming, in that the identity effect was stronger for self- than other-priming. The variable Identity received the value 'True' when the prime and the target had the same verb lemma and 'False' in all other cases. Morphological information was neglected because this

1. One of the anonymous reviewers proposed to include a predictor operationalising the dialect or subcorpus in question. While this proposal seems reasonable in the light of the substantial dialectal variation (see discussion of Table 1), including such a variable in the regression model described in (4) is problematic due to collinearity between that variable and individual preferences visualised in Figure 1. In other words, individual preferences to some degree reflect dialectal variation. I tested including a Dialect variable as a random effect in the otherwise unchanged model equation in (4). As expected, inclusion of Dialect resulted in a singular model fit indicative of collinearity.

stricter operationalisation resulted in few cases coded as 'True'. In line with the results from previous studies, the probability of the value 'True' for Persist was expected to increase for cases with the value 'True' for Identity.

**Table 4.** Summary of predictor variables in the logistic regression model

| Name | Type | Description |
|---|---|---|
| PreviousSpeaker | Factorial | Identity of the speaker of the previous *-ra*/*-se* token ('Other' = speaker different than current speaker, 'Self' = Same speaker as current speaker) |
| SpeakerPreference | Numerical | Percentage of use (0–100) of the target variant for current speaker within the conversation. |
| Identity | Factorial | Whether or not the target has the same verb lemma as the prime ('True') or not ('False') |

At least one interaction effect was expected: an interaction between SpeakerPreference and PreviousSpeaker. The results from Gries (2005), described in the last paragraph, additionally led to the expectation of an interaction effect between PreviousSpeaker and Identity. In a first step, I consequently calculated a maximal logistic regression model in R (R Development Core Team 2021) predicting the dependent variable Persist from the three variables described in Table 4 and all possible interactions between these variables. This model is described in (4).

(4) Persist ~ SpeakerPreference + PreviousSpeaker + Identity + SpeakerPreference : PreviousSpeaker + PreviousSpeaker : Identity + SpeakerPreference : Identity + SpeakerPreference : PreviousSpeaker : Identity

In order to select the most parsimonious model, I then employed an automatic backward selection process using the function *pdredge* from the R package MuMIn (Bartón 2020). This model selection process lead to the exclusion of the interactions SpeakerPreference : Identity and SpeakerPreference : PreviousSpeaker : Identity and, thus, the final model formula in (5).

(5) Persist ~ SpeakerPreference + PreviousSpeaker + Identity + SpeakerPreference : PreviousSpeaker + PreviousSpeaker : Identity

Table 5 summarises the results from the logistic regression model. Three significant main effects were found. First, likelihood of the value 'True' for the dependent variable Persist increases with higher values for SpeakerPreference. As mentioned above, this is a result of the circularity inherent in predicting Persist

from SpeakerPreference. Second, the main effect of PreviousSpeaker indicates that Persist is more likely to receive the value 'True' when the current speaker is identical to the speaker of the previous *-ra* or *-se* token, reproducing the results from the descriptive analysis (Section 3.2). Third, the expected significant correlation between the variable Identity – measuring formal identity of the prime and the target token – and Persist was found. Thus, Persist is more likely to receive the value 'True' when the verb lemma of the prime was identical to the verb lemma of the target.

**Table 5.** Summary of results of the logistic regression model (dependent variable Persist). Note: COEF = coefficient, SE = standard error, Z = *z* value, P = *p* value. All *p* values below .05 were assumed to be significant

| Predictor | | Level | COEF | SE | Z | P |
|---|---|---|---|---|---|---|
| Main effects | (Intercept) | | −1.28 | 0.3 | −4.6 | <.001 |
| | Speaker-Preference (SP) | | 0.03 | 0.0 | 7.8 | <.001 |
| | PreviousSpeaker | Other | *Reference level* | | | |
| | | Self | 2.23 | 0.3 | −6.5 | <.001 |
| | Identity | False | *Reference level* | | | |
| | | True | 2.0 | 0.4 | 5.4 | <.001 |
| Interaction effects | Speaker-Preference (SP) : PreviousSpeaker (PS) | SP : PS=Self | 0.05 | 0.0 | 11.0 | <.001 |
| | Speaker-Preference (SP) : Identity (ID) | SP : ID=TRUE | −0.02 | 0.00 | −4.5 | <.001 |

The model found two significant interaction effects, which I interpret in terms of a moderator approach to interactions (see Jaccard 2001:30–41).[2] First, the effect of SpeakerPreference on Persist is moderated by PreviousSpeaker such that the positive correlation between Persist and SpeakerPreference is predicted to be even stronger if the current speaker is identical to the previous speaker. Second, the effect of SpeakerPreference on Persist is also moderated

**2.** In this approach, an interaction effect B : C on the dependent variable A is taken to indicate that the strength of the effect of the predictor B on the dependent variable A is moderated by the predictor C. To give a simple example, A might represent quality of taste, B the difference between cookie and broccoli, and C variation in whether or not a cup of coffee is consumed at the same time. A subject might like cookies better than broccoli (main effect of B on A) but might not generally like eating better together with a cup of coffee (main effect of C on A). However, under a moderator approach, if we find a statistically significant interaction effect of B : C on A, it might be that cookies taste even better with a cup of coffee than without one.

by Identity, such that the positive correlation between Persist and SpeakerPreference is predicted to be weaker in cases in which the prime and the target token are formally identical.

Figure 2 visualises the influence of the interaction SpeakerPreference : PreviousSpeaker on the distribution of the dependent variable Persist. In terms of the main effects of the variables, the effect of SpeakerPreference suggests that a speaker is more likely to select the same past subjunctive form (target) as the recent past subjunctive form (prime) when the data from the corpus suggest that this speaker generally prefers using this past subjunctive form. As noted above, this effect is circular, since the choice of the variants is motivated by the tendency for the speaker to select the variant in question over the entire course of the conversation. Interestingly, however, the regression analysis suggests that the effect is governed by the variable PreviousSpeaker in two ways. First, the analysis shows that the increase of Persist for each percentage of SpeakerPreference is very much linear for other-priming (black line). In contrast, for self-priming (orange line), we find a nonlinear trend: the correlation between SpeakerPreference and Persist is much stronger for speakers who either strongly disprefer (SpeakerPreference between 0 and 25) or prefer (SpeakerPreference between 75 and 100) use of the past subjunctive form than for speakers who show more variable behaviour (SpeakerPreference between 25 and 75).

Second, for speakers who do not prefer usage of the target variant (SpeakerPreference between 0 and 50), we can see that previous production of that variant by another speaker is more likely to lead to a repetition of that variant by the current speaker than previous production of that variant by the same speaker. This effect is particularly strong for speakers who strongly disprefer using the target variant (SpeakerPreference between 0 and 25). As we will see in the discussion of the findings below, this finding can be motivated in terms of an implicit learning account to persistence.

The model predicts a very different situation for speakers who prefer usage of the target variant (SpeakerPreference between 50 and 100). In this speaker group, self-priming is consistently stronger than other-priming. It appears that speakers in this group select the variant they prefer independently of the other speakers' linguistic behaviour. Put simply, speakers who show strongly entrenched usage patterns (e.g. speakers who almost always use *-ra*) are less affected by previous usage of that variant because they would have selected that variant anyways. Arguably, the variable Persist does not measure persistence in this speaker group anymore but is directly correlated to individual preferences.

Figure 3 visualises the influence of the interaction SpeakerPreference : Identity on the distribution of the dependent variable Persist. Again, the figure demonstrates the main effect of SpeakerPreference, in that speakers who prefer

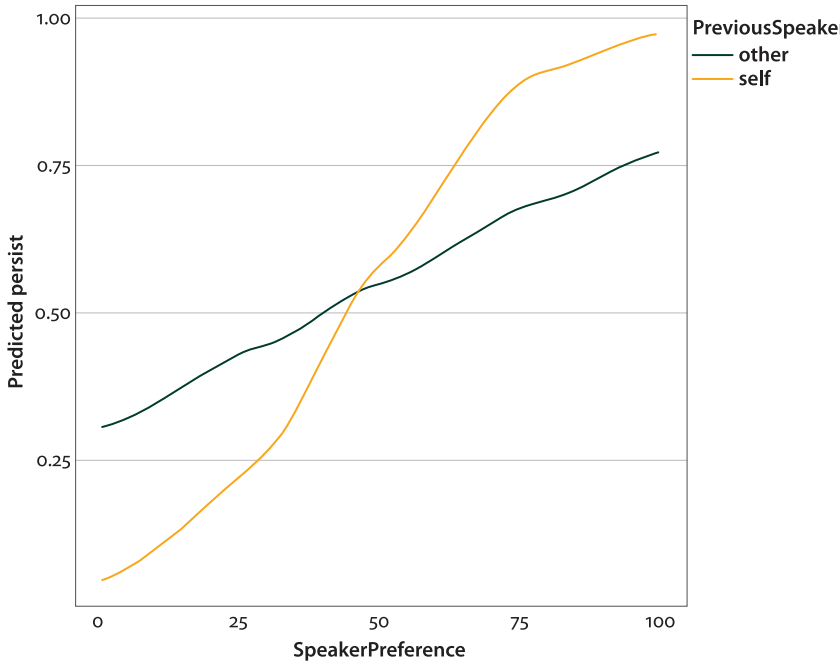



**Figure 2.** Predicted PERSIST (use of *-ra* or *-se* following the same variant in the previous discourse), by SPEAKERPREFERENCE and PREVIOUSSPEAKER

using the target variant will typically do so as well when that variant is used in the prime. It also illustrates the main effect of the variable IDENTITY, in formal similarity (same verb lemma) between the prime and the target generally increases the likelihood of persistence (i.e. repetition of the primed form in the target).

As to the interaction between SPEAKERPREFERENCE and IDENTITY, Figure 3 demonstrates that the correlation between IDENTITY and the dependent variable PERSIST is moderated by SPEAKERPREFERENCE. In particular, the effect of IDENTITY on PERSIST is much stronger for speakers who do not prefer usage of the target variant (SPEAKERPREFERENCE between 0 and 75) than for speakers who do (SPEAKERPREFERENCE between 75 and 100). As in the case of the interaction effect between SPEAKERPREFERENCE and PREVIOUSSPEAKER, I will argue in the discussion that this effect can be motivated in terms of entrenchment, on the basis of the assumption that the productivity of a variant that is weakly entrenched in a speaker's mind relies more on exact repetition than the productivity of a variant that is strongly entrenched in a speaker's mind. In terms of the statistical description, this leads to a ceiling effect; speakers who almost always use the primed variant are necessarily insensitive to factors such as lexical identity.

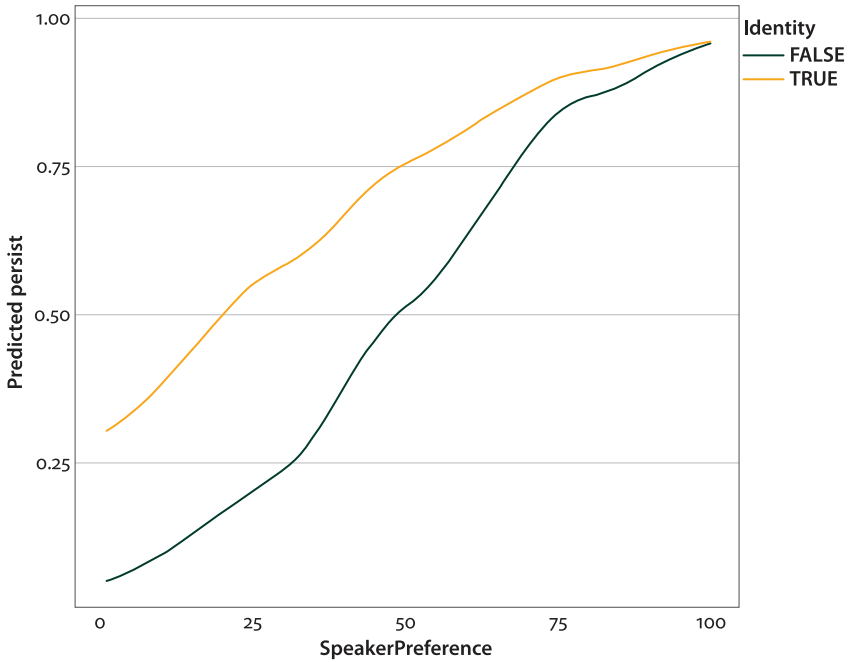



**Figure 3.** Predicted PERSIST (use of *-ra* or *-se* following the same variant in the previous discourse), by SPEAKERPREFERENCE and IDENTITY

## 3.4 Discussion of results

The analysis presented in this section of the paper tested the prediction that the influence of individual preferences on persistence is moderated by the difference between self- and other-priming, such that other-priming is expected to be stronger than self-priming when the speaker generally prefers usage of the competing variant. As predicted, the analysis has found that other-persistence is stronger than self-persistence for speakers who do not prefer usage of the primed variant (in terms of my operationalisation of SPEAKERPREFERENCE, speakers who use the primed variant in less than 50% of the cases). This finding is in line with the assumption that strength of persistence is correlated to expectedness and implicit learning: when another speaker uses a variant that is only weakly entrenched in the hearer's grammar, this usage event will lead to more surprisal and concomitantly, stronger persistence than previous use of a variant that the hearer prefers.

50% of SPEAKERPREFERENCE was found to be the tipping point after which self-priming becomes stronger than other-priming. When a hearer listens another speaker use a variant that she or he prefers using herself, this previous use of the

variant leads to weaker persistence. This means that for speakers who strongly prefer using one variant, repeated use of this variant is less likely to reflect persistence in the strict sense than these individual preferences. From the perspective developed in this paper, this result warrants the conclusion that for speakers who show a strong preference for the primed element, the phenomenon of self-priming in corpus data is more adequately explained as individual-level entrenchment or style. In line with these assumptions, studies such as Kaschak (2007) and Kaschak, Kutta & Schatschneider (2011) report that persistence operates over time spans as long as one week.

This assumption explains why, on average, self-priming (measured in terms of the probability of a variant being used after previous use of that variant by the same speaker) has a stronger effect than other-priming in corpus data. As hypothesised by Gradoville (2019), failure to distinguish persistence from individual preferences leads to an overestimation of persistence in the data. Crucially however, this problem affects self-priming to a much greater degree than other-priming, which is less easily attributed to individual preferences.

Finally, the analysis also found a significant interaction between speaker preferences and the variable IDENTITY measuring whether or not the previous past subjunctive token is formally similar (same verb lemma) to the target past subjunctive token. As expected, formal similarity was found to increase the probability of an element to persist. However, this main effect is moderated by speaker preferences: formal similarity has a stronger impact of persistence for speakers who disprefer using the primed variant. This effect can again be explained in terms of an implicit learning account to persistence. Thus, speakers who disprefer using the primed variant have a weaker representation of this variant in their mind (entrenchment), leading to comparably lower cognitive ease of accessibility and production of the entire morphological paradigm of that variant. For such speakers, previous use of a form such as *fueseis* 'be.PST.SBJ.2PL') is more likely to lead be unexpected and consequently cause persistence than for speakers who display strongly entrenched usage of *-se*.

## 4. Conclusions

Variationist linguistics assumes that "a lot of what appears to be free variation can be accounted for if linguists take social factors into account as well as linguistic factors" (Meyerhoff 2006: 10). Free variation is indeed unexplained variation, and it is up to the analyst to develop methods and concepts that allow identifying constraints that explain away 'free' variation. The results from this paper suggest that

one way of attaining this goal is to explore the constraints that shape individual preferences in language usage. In particular, it was proposed that individual preferences and syntactic priming, two variables that are usually modelled as separate predictors and whose effects are typically attributed to 'free' variation are, in fact, collinear. Consequently, 'free variation' is not just an effect of individual preferences but results from a complex interaction between the speaker's previous experience with language and social constraints.

Focusing on one specific case of variation, namely, the alternation between two Spanish past subjunctive forms, an analysis was developed that allowed formulating predictions about whether or not a speaker will use the same subjunctive form as a previously subjunctive form. The analysis thus effectively measured the probability for persistence to occur depending on three predictor variables: individual speaker preferences as regards selection of these variants, the difference between self- and other-priming, and formal similarity between the prime and the target. The analysis demonstrated a strong influence of these parameters on strength of persistence, which was explained in terms of the notions of persistence and cognitive entrenchment. In particular, it was found that previous use of a variant by another speaker is more likely to lead to persistence than previous use by the same speaker in the case that this speaker does not usually prefer usage of the variant in question. In contrast, previous use of a variant by the same speaker is more likely to lead to persistence if the speaker generally displays high rates of selection of this variant. Similarly, the effect of formal similarity between the prime and the target was found to be moderated by individual preferences, such that formal similarity leads to stronger persistence for speakers who disprefer usage of the primed variant than for speakers who prefer usage of the respective variant.

These findings suggest that, in order to explain more 'free' variation in such cases, it is necessary to assume more complex models of the influence of individual preferences on linguistic variation. It has been proposed in this paper that for speakers who prefer the primed variant, self-priming can be described as linguistic style, in that consistent and self-reinforcing selection of one variant over another is an expression of the speaker's linguistic *habitus* (Adli 2013). Consequently, the results from this paper call for a finer-grained perspective on operationalising individual preferences in variationist sociolinguistics. In particular, the desideratum can be formulated that individual preferences be operationalised not only of the basis of the production data of the respective speaker within one conversation (as in the analysis reported in this paper), but ideally in more production data or even pre-tests. Such a procedure would allow escaping the inherent circularity of trying to predict usage of one variant over another by a speaker in terms of her or his production of these variants in the same stretch of speech, which arguably weakens the analysis presented in this article.

Finally, the results from this paper have interesting implications for studies on language change that discuss the role of asymmetric priming (Jäger & Rosenbach 2008; Hilpert & Correia Saavedra 2016). Hilpert & Correia Saavedra (2016) tested the prediction, established in Jäger & Rosenbach (2008), that the unidirectionality of grammaticalisation is due to the fact that in grammaticalisation processes, the original lexical item (e.g. *go*) should prime the constructions that have resulted from grammaticalisation processes affecting *go* (e.g. *be going to*), but not vice versa (Hilpert & Correia Saavedra 2016: 358). As experimental paradigm, the authors employed a maze task in which participants' response times to primed and unprimed elements were measured, but failed to find evidence for the assumed asymmetric priming effect. In the lights of the results from this paper, one possibility for this null finding is that the maze task only involved other-priming (participants were given stimuli by the experimenters that they had to respond to). However, the results from this paper suggest that it is self-priming, not other-priming, that can lead to the conventionalisation of usage of variants or grammatical constructions. It thus stands to reason that, in order to test the hypothesis of asymmetric priming, it might be necessary to consider the difference between self- and other-priming.

## Acknowledgements

I am grateful to Hendrik De Smet, Scott Schwenter, Kristin Kopf, Thilo Weber, and the three anonymous reviewers for helpful comments and discussion, which have improved the paper both in terms of the analysis and the content. All remaining errors are my own.

## References

Adli, Aria. 2013. Syntactic variation in French *Wh*-questions: A quantitative study from the angle of Bourdieu's sociocultural theory. *Linguistics* 51(3): 473–515.

Alarcos Llorach, Emilio. 1999. *Gramática de la lengua española*. Madrid: Espasa Calpe.

Anderson, Ana M. 2017. Receding or Resurgent? On the Use of Cantase (and Cantara) in Galician Spanish. Doctoral dissertation. Minnesota: University of Minnesota.

Asratián, Arucia. 2007. Variación *-ra/-se* en el español hablado en Caracas. *Boletín de lingüística* 19(5): 5–41.

Bargh, John A. & Chartrand, Tanya L. 2000. The mind in the middle: A practical guide to priming and automaticity research. In *Handbook of Research Methods in Social and Personality Psychology Psychology*, Harry T. Reis & Charles M. Judd (eds), 253–285. Cambridge: Cambridge University Press.

Bartón, Kamil. 2020. *MuMIn: Multi-Model Inference. R package version 1.43.17*. https://cran.r-project.org/package=MuMIn (3 March 2023).

Becker, Martin. 2008. From temporal to modal: Divergent fates of the Latin synthetic pluperfect in Spanish and Portuguese. In *The Paradox of Grammatical Change. Perspectives from Romance* [Current Issues in Linguistic Theory 293], Ulrich Detges & Richard Waltereit (eds), 147–179. Amsterdam, Philadelphia: John Benjamins.

Bermejo, Víctor L. 2019. El pretérito imperfecto de subjuntivo en la Península Ibérica del siglo XX. *Verba* 46: 313–338.

Bock, Kathryn J. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3): 355–387.

Bock, Kathrin J., Dell, Gary S., Chang, Franklin & Onishi, Kristine H. 2007. Persistent structural priming from language comprehension to language production. *Cognition* 104(3): 437–458.

Bock, Kathrin J. & Griffin, Zenzi M. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129(2): 177–192.

Bolinger, Dwight L. 1956. Subjunctive *-ra* and *-se*: "Free Variation"? *Hispania* 39(3): 345–349.

Callaghan, Matthew & Travis, Catherine E. 2020. Priming as a diagnostic of grammatical constructions: Second-person singular in Chilean Spanish. *Languages* 6(1): 1–22.

Cameron, Richard. 1995. The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6(7): 1–27.

Cameron, Richard & Flores-Ferrán, Nydia. 2004. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1): 41–65.

Cappelle, Bert. 2009. Can we factor out free choice? In *Describing and Modeling Variation in Grammar* [Trends in Linguistics. Studies and Monographs [TiLSM] 204], Andreas Dufter, Jürg Fleischer & Guido Seiler (eds), 183–201. Berlin, New York: De Gruyter.

Cho, Sun-Joo, Watson, Duane, Jacobs, Cassandra & Naveiras, Matthew. 2020. A Markov Mixed-Effect Multinomial Logistic Regression Model for nominal repeated measures with an application to syntactic self-priming effects. *Multivariate Behavioral Research* 56(3): 476–495.

Christ, Christina & Feldhausen, Ingo. 2021. Die *-ra/-se*-Alternation beim subjuntivo imperfecto im europäischen Spanisch: Eine soziolinguistische Korpusstudie. *Zeitschrift für romanische Philologie* 137(3): 827–859.

Davies, Mark. 2019. Corpus del español, News on the Web. http://www.corpusdelespanol.org/ (3 March 2023).

Dell, Gary. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93(3): 283–321.

Dell, Gary, Burger, Lisa & Svec, William. 1997. Language production and serial order: A functional analysis and a model. *Psychological Review* 104(1): 123–147.

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.

Fricke, Melinda & Kootstra, Gerrit J. 2016. Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language* 91: 181–201.

Gradoville, Michael. 2019. The role of individual variation in variationist corpus-based studies of priming. *Italian Journal of Linguistics* 31(1): 93–124.

Gries, Stefan T. 2005. Syntactic priming: a corpus-based approach. *Journal of Psycholinguistic Research* 34(4): 365–399.

Gries, Stefan T. & Kootstra, Gerrit J. 2017. Structural priming within and across languages. A corpus-based perspective. *Bilingualism: Language and Cognition* 20(2): 235–250.

Guajardo, Gustavo. 2021. Two subjunctives or three? A multimodel analysis of subjunctive tense variation in complement clauses in Spanish. *International Journal of Corpus Linguistics* 26(2): 248–283.

Guzmán Naranjo, Matías. 2017. The *se-ra* Alternation in Spanish Subjunctive. *Corpus Linguistics and Linguistic Theory* 13(1): 97–134.

Hilpert, Martin & Correia Saavedra, David. 2016. The unidirectionality of semantic changes in grammaticalization: An experimental approach to the asymmetric priming hypothesis. *English Language and Linguistics* 22(3): 357–380.

Jaccard, James. 2001. *Interaction Effects in Logistic Regression* [Quantitative Applications in the Social Sciences 135, Sage university papers series 07–135]. Thousand Oaks, London, New Delhi: Sage University Papers.

Jacobs, Cassandra L., Cho, Sun-Joo & Watson, Duane G. 2019. Self-priming in production: Evidence for a hybrid model of syntactic priming. *Cognitive Science* 43(7): e12749.

Jaeger, Florian & Snider, Neal. 2013. Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127(1): 57–83.

Jaeger, Florian & Snider, Neal. 2008. Implicit learning and syntactic persistence: surprisal and cumulativity. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society, July 23 – 26, 2008 Washington, D.C., USA*, Bradley C. Love, Ken McRae & Vladimir M. Sloutsky (eds), 1028–1033. Austin, TX: Cognitive Science Society.

Jäger, Gerhard & Rosenbach, Anette. 2008. Priming and unidirectional language change. *Theoretical Linguistics* 34(2): 85–113.

Kaschak, Michael P. 2007. Long-term structural priming affects subsequent patterns of language production. *Memory & Cognition* 35(5): 925–937.

Kaschak, Michael P., Kutta, Timothy J. & Schatschneider, Christopher. 2011. Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition* 39(3): 381–388.

Kempas, Ilpo. 2011. Sobre la variación en el marco de la libre elección entre *cantara* y *cantase* en el español peninsular. *Moenia* 17: 243–264.

Lemon, Francis. 1925. The relative frequency of the subjunctive forms in *–se* and *–ra*. *Hispania* 8(5): 300–302.

Mahowald, Kyle, James, Ariel, Futrell, Richard & Gibson, Edward. 2016. A meta-analysis of syntactic priming in language production. *Journal of Memory and Language* 91: 5–27.

Meyerhoff, Miriam. 2006. *Introducing Sociolinguistics*. London, New York, Routledge.

Pereira Scherre, María. 2001. Phrase level parallelism effect on noun phrase number agreement. *Language Variation and Change* 13(1): 91–107.

Pickering, Martin J. and Ferreira, Victor S. 2008. Structural priming: a critical review. *Psychological Bulletin* 134(3): 427–459.

Poplack, Shana. 1980. The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. In *Locating Language in Time and Space*, William Labov (ed), 55–67. New York: Academic Press.

PRESEEA. 2014. *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. https://preseea.uah.es (17 May 2023).

R Development Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (3 March 2023).

Rojo, Guillermo & Vázquez Rozas, Victoria. 2014. Sobre las formas en «-ra» en el español de Galicia. In *Perspectives in the study of Spanish language variation. Papers in the honor of Carmen Silva-Corvalán* [Verba: Anuario Galego de Filoloxía 72], Andrés Enrique-Arias, Manuel Gutiérrez, Alazne Landa & Francisco Ocampo (eds), 237–270. Santiago de Compostela: Universidade de Santiago de Compostela.

doi Rosemeyer, Malte. 2015. How usage rescues the system: persistence as conservation. In *Variation in Language: System- and Usage-based Approaches* [linguae & litterae 50], Aria Adli, Marco García García & Göz Kaufmann (eds), 289–311. Berlin, Boston: De Gruyter.

Rosemeyer, Malte. 2021. Variación diatópica en la evolución del pluscuamperfecto sintético. In *Variación diatópica y morfosintaxis en la historia del español*, Blanca Garrido Martín, María del Carmen Moral del Hoyo & Matthias Raab (eds), 197–222. Santiago de Compostela: Universidade de Santiago de Compostela.

doi Rosemeyer, Malte & Schwenter, Scott A. 2019. Entrenchment and persistence in language change: the Spanish past subjunctive. *Corpus Linguistics and Linguistic Theory* 15(1): 167–204.

Schmid, Hans-Jörg. 2010. Entrenchment, salience, and basic levels. In *The Oxford Handbook of Cognitive Linguistics* [Oxford Handbooks], Dirk Geeraerts & Hubert Cuyckens (eds), 117–138. Oxford: Oxford University Press.

doi Sussman Goldberg, Barbara. 1995. The *-ra* and *-se* opposition in Spanish. In *Meaning as Explanation. Advances in Linguistic Sign Theory* [Trends in Linguistics. Studies and Monographs [TiLSM] 84], Ellen Contini-Morava & Barbara Sussman Goldberg (eds), 381–404. Berlin, New York: De Gruyter.

doi 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.

doi Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis* [Trends in Linguistics. Studies and Monographs [TiLSM] 177]. Berlin, New York: De Gruyter.

Szmrecsanyi, Benedikt, Grafmiller, Jason, Bresnan, Joan, Rosenbach, Anette, Tagliamonte, Sali A. & Todd, Simon. 2017. Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa* 86(1): 1–27.

doi Tamminga, Meredith. 2016. Persistence in phonological and morphological variation. *Language Variation and Change* 28(3): 335–356.

Veiga Rodríguez, Alexandre. 1996. *La forma verbal española cantara en su diacronía*. Santiago de Compostela: Universidade de Santiago de Compostela.

Weiner, E. Judith & Labov, William. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(1): 29–58.

Wright, Leavitt O. 1926. The indicative forms in *–ra* in Spanish America. *Hispania* 9(5): 288–293.