

Article

Anna Mastrantuono and Brendan Regan*

Present perfect and preterit variation in the Spanish of Lima and Mexico city: findings from a corpus analysis

<https://doi.org/10.1515/cllt-2022-0060>

Received July 11, 2022; accepted July 13, 2023; published online August 3, 2023

Abstract: In many languages, the present perfect has grammaticalized, gradually displacing the preterit. Within Spanish, this has been documented with the grammaticalization of the present perfect in Peninsular Spanish. To examine this possibility in two Latin American varieties, this study examined present perfect/preterit variation of 36 speakers from Lima and Mexico City from the PRESEEA corpus. While Lima Spanish presented overall more present perfect than Mexico City Spanish, a similar internal constraint hierarchy is predictive of present perfect use in both speech communities. However, Lima Spanish demonstrated a change in progress toward an expansion of the preterit among younger speakers with the indeterminate temporal reference as locus of change. The findings suggest that present perfect grammaticalization may not always be the most common cross-linguistic pathway but rather is subject to source constraints, which may lead to another pathway in which the preterit expands at the expense of the present perfect.

Keywords: grammaticalization; present perfect; Peruvian Spanish; preterit; variationist sociolinguistics

1 Introduction

Many languages present the alternation between present perfect (e.g., I have gone) and preterit (e.g., I went) (Bybee and Dahl 1989). A cross-linguistic trait of the present perfect (henceforth, PP) is its semantic instability: it tends to assume new grammatical functions, expanding to new linguistic contexts through processes of grammaticalization (Bybee et al. 1994: 81; Squartini and Bertinetto 2000). In Romance

***Corresponding author: Brendan Regan**, Classical & Modern Languages & Literatures, Texas Tech University, Lubbock, TX, USA, E-mail: brendan.regan@ttu.edu, <https://orcid.org/0000-0002-4766-6921>

Anna Mastrantuono, Classical & Modern Languages & Literatures, Texas Tech University, Lubbock, TX, USA; and Pietro Verri High School, Milan, Italy

languages, the PP was originally employed to express present states resulting from past events, but in many varieties, it gradually displaced the preterit in the expression of past events and states (*aoristic drift*) (Squartini and Bertinetto 2000). The PP variation across varieties of Spanish has drawn the attention of several scholars. The PP is at an advanced stage of grammaticalization in many Peninsular¹ varieties, where it is the preferred form to express hodiernal (same-day) past (Burgo 2012; Howe 2006; Schwenter 1994; Schwenter and Torres Cacoullós 2008; Serrano 1994), demonstrating a change in *apparent time* (Labov 1994: 45) toward a further expansion of the PP in increasingly distant past contexts (Schwenter 1994). On the contrary, Latin American varieties of Spanish use the PP infrequently (Fløgstad 2016; Schwenter and Torres Cacoullós 2008; Westmoreland 1988). An exception is Peruvian Spanish, in which, although the preterit is the most frequent form, the PP is used in everyday speech (Escobar 1997; Escobar and Roy 2022; Howe 2006; Jara Yupanqui 2011a). Previous research suggests that the PP is undergoing different paths of grammaticalization in Peru and Spain, as different linguistic constraints condition its distribution (Howe 2013; Howe and Scott 2008). However, there is a lack of quantitative variationist studies analyzing PP/preterit variation in Latin American Spanish, thus it is unclear if there is a change in progress in Peruvian Spanish.

The present study aims to fill this gap by examining the factors that condition PP/preterit variation in two Latin American capital cities, Lima (Peru) and Mexico City (Mexico). Mexico City is included as a point of comparison as it has been documented as having one of the highest rates of preterit of all Spanish dialects (Lope Blanch 1972; Moreno de Alba 1978). The aims were twofold: (1) to determine the social and linguistic factors that govern PP/preterit variation in Lima Spanish in comparison to Mexico City Spanish; and (2) to quantitatively examine previous claims of the grammaticalization of the PP or the preterit in Lima Spanish.

2 Background

2.1 Grammaticalization

The semantic instability of the PP is considered revealing of a more general process of semantic change in modern languages, i.e., grammaticalization. When a lexical

¹ Varieties of Peninsular Spanish that demonstrate this advanced stage of grammaticalization include those in Madrid (Howe 2006; Schwenter and Torres Cacoullós 2008; Serrano 1994), Alicante (Schwenter 1994), Bilbao (Burgo 2012), and Valencia (Howe 2006). However, some Peninsular varieties do not demonstrate this pattern such as in La Laguna, Tenerife (Serrano 1995), or parts of Asturias and León (Azpiazu-Torres 2017; Veiga Rodríguez 2014). Thus, “varieties of Peninsular Spanish” here only refers to varieties that demonstrate PP grammaticalization.

form or grammatical item enters into grammaticalization, it gradually develops (new) grammatical functions, expanding to new domains of use (Bybee et al. 1994: 4; Heine et al. 1991: 2; Hopper and Traugott 2003: 1; Lehmann 2015: 13). An example is the English periphrasis “be going,” which was initially used with a literal meaning of “moving through space toward a goal,” but gradually acquired the additional meaning of future temporality, a process known as *generalization of meaning* (Bybee et al. 1994: 5; Hopper and Traugott 2003: 103). The two meanings, the spatial and the temporal, coexisted for some time (*layering*), but eventually, the former meaning demised (*semantic bleaching* or *erosion*). At this point, “be going” became a fully grammaticalized form expressing future. Such a development from spatial to temporal semantic value has been attested in several languages (Bybee et al. 1994: 268). The evidence that similar source material (e.g., verbs of movement) tends to undergo similar processes of semantic change across languages led to the hypothesis of source determination, according to which grammaticalization processes are not random but rather determined by the meaning of the construction that enters into grammaticalization.

One major finding in grammaticalization is *unidirectionality*, according to which once a grammaticalization process starts, it cannot be reversed. For instance, verbs of “movement toward a goal” tend to grammaticalize to futures cross-linguistically, but do not go back to their previous meaning (Bybee et al. 1994: 12–13). Similarly, it has been observed that, cross-linguistically, the PP first evolves from resultative constructions and then tends to acquire other semantic functions, being the encroachment into past perfective functions (typically carried out by means of the preterit) the most common outcome (Bybee and Dahl 1989; Bybee et al. 1994: 81; Squartini and Bertinetto 2000). However, the opposite direction of evolution is not attested (Bybee et al. 1994: 12). The hypotheses of source determination and unidirectionality predict that there are cross-linguistic clines, or universal paths, of semantic change (Bybee et al. 1994: 14). In other words, there are common patterns of semantic evolution of the linguistic material that enters into grammaticalization even in languages that are not geographically or genetically related, which may be due to “common cognitive and communicative patterns underlying the use of language” (Bybee et al. 1994: 14).

The validity of such hypotheses has been object of debate (Fløgstad 2017; Giacalone Ramat 1998; González et al. 2019; Janda 2001). On the one hand, Hopper and Traugott (2003: 139) state that the evidence for unidirectionality is systematic cross-linguistically, whereas counterexamples are sporadic and language-specific. On the other hand, Drinka (2017: 65–72) claims that there is weak evidence of the perfect-to-perfective shift as a universal process. She suggests that this linguistic innovation arose in the 12th and 13th centuries in the Parisian French vernacular in which the “perfect first came to be used to connote the past” (Drinka 2003: 20, 2017: 257).

Drinka claims that this innovation later expanded across Europe in the areas that lay under the Parisian cultural influence (2017: 257–261). Therefore, Drinka indicates that the spread of the PP with preterit value can be better explained by areal diffusion and linguistic contact rather than by universal paths of development.

Most recently, Haspelmath (2019: 10) has critiqued Bybee's (2006) "common pathways of change in the tense-aspect domain." Haspelmath specifically criticized Bybee's (2006: 179–180) claim that "the true universals of language are the mechanisms of change that propel the constant creation and recreation of grammar." Haspelmath (2019: 10) states that Bybee's argument does not differentiate enough between "recurrent pathways of change" and the constraints that lead to such pathways of change. Haspelmath (2019: 3) claims that "universal tendencies cannot be explained by common pathways of change, only by change constraints." He refers to change constraints as "mutational constraints," which he proposes are two general types: source constraints and directionality constraints (2019: 6). The irreversibility of grammaticalization is the most well-known directionality constraint (Haspelmath 1999: 1043, 2004: 17, 2019: 6). As Haspelmath indicates, source constraints, such as internal and external (i.e., language contact), impact the pathway of any grammatical change. Reinhöhl and Himmelmann (2017: 408) further echo the importance of source constraints in grammaticalization as "one needs to analyze change in context" in which "grammaticalization constructions are shaped by forms and structures they are built up from as well as by constructions they interact with." Thus, it is important to consider source and directionality constraints in understanding pathways of change.

2.2 Present perfect and preterit variation

According to prescriptive grammars, the preterit expresses perfectivity; it denotes a situation temporally bounded in the past and reported for its own sake, while the PP indicates the present relevance of a past situation (Bybee et al. 1994: 54; Comrie 1976; Harris 1982).² Comrie (1976) reports four cross-linguistic prototypical uses of the PP: *resultative*, which allows to infer a current state resulting from a past action; *continuative*, which marks past events that continue up to the present; *experiential*,

2 Bybee et al. (1994: 61) refer to perfects as "anterior," and define them as "a past action with current relevance," although conceding that they can also cover other semantic functions (1994: 62). As we chose a form-based methodological approach (see Section 3), avoiding the notion of current relevance (see Section 3.4), we follow the common terminology "perfect" used in grammar textbooks to refer to the periphrastic past (in Spanish, auxiliary verb *haber* "to have" + past participle), adding the adjective "present" to distinguish it from the past perfect forms; and "preterit" to refer to the synthetic past which typically carries perfective functions.

which describes an event that occurred at least once in a certain period of time; and, *recent past*, which reports an event that occurred in a temporal interval that includes the moment of speech. As for Romance languages, the PP originated from a resultative construction attested in Pre-Classical Latin, as in the example by Harris (1982: 47), “*Multa bona bene parta habemus*” “Many goods well obtained we have” (Plautus, Trin. 347). Over time, these constructions lost their original resultative function to become a marker of a past event considered relevant to the present moment (Bybee and Dahl 1989). However, the indeterminacy of the notion of current relevance brought variable uses of the PP across Romance varieties (Cartagena 1999: 2945–6; Harris 1982).

Harris (1982) offered an account of the distribution of the PP in Romance varieties as being organized in four stages of grammaticalization. At stage I, the PP expresses a present state resulting from past actions (resultative function), as seen in Pre-Classical Latin. At stage II, it presents an aspectual function of durativity or iteration, referring to states and events that started in the past and continue up to the speech time. This stage is representative of Latin American Spanish varieties, Galician, and Portuguese. At stage III, the PP expands to an array of contexts that denote perfectivity, namely, an action or situation presented as completed in the past. This is what Squartini and Bertinetto (2000: 404) call the aoristic drift of the PP. Typically, the PP starts expanding to recent past contexts because they are more likely to be perceived as currently relevant. Several Peninsular Spanish varieties are at this stage. At stage IV, the PP expresses past actions regardless of the distance with the moment of speech, whereas the preterit is restricted to formal and written registers. This is the situation of French, Romanian, and Northern Italian. Of note is that Harris considered the varieties at stages III and IV as being “more developed” (i.e., some varieties of Peninsular Spanish), whereas those at earlier stages (i.e., varieties of Latin American Spanish) will eventually reach the more developed stages. However, Squartini and Bertinetto (2000) interpret Harris’ stages as independent developmental paths of a grammaticalizing PP. Therefore, a variety that is at stage II will not necessarily move toward stages III and IV. Similarly, Schwenter (1994: 78) claims that PP forms can be “frozen in their development” and that synchronic investigation is needed “to determine their ongoing development, if any, and to advance prognoses of future grammaticalization.” In the following sections, we provide an overview of PP/preterit variation in four varieties of Spanish: Peninsular Spanish that exhibits the most PP realizations with a grammaticalization in progress; Mexican Spanish that exhibits the most preterit realizations without signs of grammaticalization; and Peruvian and Argentinian Spanish that exhibit a high degree of PP variation different from other Latin American Spanish varieties.

2.3 Present perfect and preterit variation in Spanish

2.3.1 Peninsular Spanish

Several varieties of Peninsular Spanish, such as in Madrid (Howe 2006; Schwenter and Torres Cacoullós 2008; Serrano 1994), Alicante (Schwenter 1994), Bilbao (Burgo 2012), and Valencia (Howe 2006), show an advanced stage of grammaticalization of the PP as a perfective past, given that it is now the preferred form to express hodiernal past (i.e., past events that occurred in the same day of the utterance), whereas the preterit expresses prehodiernal past (i.e., past events that occurred before the day of the utterance).³ Moreover, there is evidence of a change⁴ in *apparent time* as younger generations⁵ use more PP than older generations (Burgo 2012; Schwenter 1994; Serrano 1994) and middle-class speakers use more PP than those from working and upper social classes (Burgo 2012; Serrano 1994). Schwenter and Torres Cacoullós (2008) found that the PP in Madrid is favored not only in hodiernal past contexts but also in indeterminate reference contexts, when the temporal anchoring of the recounted event is not specified. They hypothesize that the indeterminate reference may be an important locus of change for the PP, as it represents an intermediate step leading to the erosion of the current relevance requirement associated with the PP, which can consequently reach the status of prehodiernal perfective past (see Copple 2011).

2.3.2 Mexican Spanish

The PP in Mexican Spanish is semantically associated with durative and repeated actions leading into the present (Harris' [1982] stage II), whereas the preterit is used for completed actions regardless of the temporal distance with the moment of speech (Lope Blanch 1972; Moreno de Alba 1978; Westmoreland 1988). Schwenter and Torres Cacoullós (2008) found that the PP in Mexico City is favored by linguistic factors such as the presence of a frequency adverbial (e.g., *siempre* "always"), a plural direct object, and durative verbs, whereas the preterit is used categorically in the

3 Not all Peninsular Spanish varieties follow this tendency. In La Laguna, Tenerife, the preterit is the preferred form to express hodiernal past (Serrano 1995). In Galicia, and parts of Asturias and León, the preterit is strongly favored to the PP (Azpiazu-Torres 2017; Veiga Rodríguez 2014).

4 From a diachronic perspective, Copple (2011) examined dramatic texts written in Spain between the XV and XIX centuries and found the expansion of the PP at expenses of the preterit. Contrastingly, Moreno de Alba (1998) examined written correspondence from New Spain between the XVI and XIX centuries and found that PP frequency decreased overtime.

5 Schwenter (1994) found that younger generations used PP nearly categorically in the hodiernal past contexts and even relatively frequently in pre-hodiernal contexts.

expression of hodiernal past. Furthermore, they found that, despite the different overall frequencies of the PP in Mexico City and Madrid, irrelevant temporal reference (when the recounted event is not temporally bounded) and indeterminate temporal reference favored the PP in both speech communities. They also found that the adverb *ya* “already” significantly favored the PP in Madrid Spanish, whereas in the Mexican data, it was slightly below the level of significance, and nonetheless, it appeared more frequently with the preterit than the PP (also Howe 2013).

2.3.3 Peruvian Spanish

Whereas PP usage in Mexican Spanish is representative of a more generalized Latin American Spanish tendency (Harris 1982; Howe 2013; Westmoreland 1988), Peruvian Spanish differs from this tendency. In this variety, although the preterit is the more frequent form, the PP is used in everyday speech with nonprototypical functions (Escobar 1997; Escobar and Crespo del Río 2020; Escobar and Roy 2022; García Tesoro and Jang 2018; González et al. 2019; Howe 2006; Howe 2013; Howe and Schwenter 2003; Howe and Scott 2008; Jara Yupanqui 2011a; Rodríguez Louro and Jara Yupanqui 2011).⁶ It can be used to highlight the emotional involvement of the speaker toward the recounted events (González et al. 2019; Jara Yupanqui 2011a) and to interrupt narrations (which are typically recounted in preterit) (Bybee et al. 1994) to perform different pragmatic-discursive functions, such as citing, commenting, or summarizing (Howe and Schwenter 2003; Jara Yupanqui 2011a). Jara Yupanqui (2011a) proposes that Peruvian speakers take advantage of the status of PP as marked form (less frequent) to assign pragmatic functions of psychological relevance to it. In a sentence judgment task, Howe (2013) found that participants from Cuzco had the lowest rate of acceptance of the PP in co-occurrence with “today” adverbials, and the highest rate of acceptance of the PP in co-occurrence with “pre-today” adverbials, as compared to informants from Madrid and Valencia. He concluded that the PP is following a different path of grammaticalization in Peruvian and Peninsular Spanish. In Spain, it is following the same perfect-to-perfective path attested in other Romance languages, being used in contexts that are perceived as not temporally distant from the present (i.e., hodiernal and recent past). In Peru, however, a process of pragmatic extension of the concept of current relevance has occurred; therefore, the PP is admitted in temporally distant contexts, provided that the speaker considers the event as relevant to their current situation (Howe 2013; Howe and Scott 2008).

⁶ Higher frequencies of PP use as compared to the Latin American Spanish tendency have been found in all the Andean region, such as in Ecuador (Bustamante 1991; Pfänder and Palacios 2013), Northeastern Argentina (Kempas 2013; Terán and Kanwitt 2018), and Bolivia (Howe and Schwenter 2003).

Moreover, several studies found that the PP in Peruvian Spanish carries evidential functions, namely, it is used to mark events directly witnessed by the speaker (Escobar 1997; Escobar and Crespo del Río 2020; Escobar and Roy 2022; Klee and Ocampo 1995). Evidential functions were also found in Ecuador Spanish (Bustamante 1991; Pfänder and Palacios 2013). They are likely due to language contact with Quechua, akin to the spatial relevance marker function⁷ found by Escobar (1997). The development of evidential features is one of the possible outcomes of a grammaticalizing PP, although it typically acquires functions of indirect witness marker (Bybee and Dahl 1989). Howe (2013: 115–116) suggests that linguistic contact affected the development of evidentiality. The innovative PP functions in Peruvian Spanish have been found in the speech of Spanish-Quechua bilinguals (Escobar 1997; García Tesoro and Jang 2018; Howe 2013) and Spanish monolinguals (Howe and Scott 2008; Jara Yupanqui 2011a; Jara Yupanqui 2011b).

It remains unclear whether the PP is expanding in Peruvian Spanish. Only a handful of studies have included social variables when examining preterit/PP variation in Latin American Spanish. Howe (2013: 164) hypothesized that Peruvian Spanish presents stable variation of PP and preterit, but he does not quantitatively explore this question and proposes that others examine this possibility in future research. Jara Yupanqui (2006) interviewed forty participants from Lima and found that younger speakers and upper-middle class speakers used the preterit more frequently than older speakers and working-class speakers. She concluded that the use of preterit is expanding at the expense of the PP in Lima. However, these results were not replicated when Jara Yupanqui administered a questionnaire to sixty-four participants, which she believes is due to the nonauthentic contexts elicited by the questionnaire (2006: 216). Bateman (2022), using a questionnaire and sociolinguistic interviews with monolingual Spanish and bilingual Spanish-Quechua speakers in Cusco, found an innovative “contact-induced subjectivization” PP behavior among Quechua-Spanish bilinguals in Cusco. However, Bateman found that Spanish-dominant speakers overall disfavored PP usage as well as those with higher education levels.

2.3.4 Argentine Spanish

Other studies of Latin American Spanish that have examined the role of social factors in PP/preterit variation have examined Argentine Spanish, specifically *porteño* Spanish in Buenos Aires. Kubarth (1992) and Fløgstad (2016) suggest that there is less PP production among younger generations, and that the PP will eventually disappear

⁷ Escobar (1997) found that with Spanish in contact with Quechua the PP is used to recount events that occurred in the same location where the utterance occurs, whereas the preterit marks events that occurred in a different location.

in Buenos Aires Spanish. Rodríguez Louro (2009) did not find any significant age effect in her sociolinguistic interviews, although older informants used more PP than younger participants in a questionnaire. Rodríguez Louro (2009) also examined headlines from Argentinian newspapers between the XVIII and XX centuries and found a significant decrease in the use of the PP across centuries.

Thus, while European Romance languages are moving along the typical PP aoristic drift (Squartini and Bertinetto 2000), some Latin American varieties, such as *porteño* Spanish and possibly Peruvian Spanish, are moving in the opposite direction, with the preterit expanding at the expenses of the PP. In light of this hypothesis, to further explore the issue of the PP/preterit variation in American Spanish varieties, it is necessary to include social variables in the analysis to explore a possible *change in progress* in the linguistic system.

2.4 Research questions

Building on previous studies, the current endeavor aims to further explore the social and linguistic factors that affect the PP/preterit distribution in Lima, Peru Spanish (henceforth, LPS), including Mexico City Spanish (henceforth, MCS) as a point of comparison. Mexico Spanish is considered prototypical of the Latin American norm of use of the PP, whereas Peruvian Spanish shows an increased use of PP as compared to other Latin American varieties, although not as frequent as Peninsular Spanish (Howe and Scott 2008). To further explore this issue, we analyzed the PRESEEA⁸ Corpus (2014–) sociolinguistic interviews conducted in Lima and Mexico City. The current study sought to answer the following research questions:

RQ1: What is the overall frequency of PP and preterit in LPS and MCS? While we expect the preterit to be the most common form in both varieties, we hypothesize that the PP will be significantly more frequent in LPS than in MCS based on previous studies (Howe and Scott 2008; Rodríguez Louro and Jara Yupanqui 2011; Schwenter and Torres Cacoullos 2008).

RQ2: What linguistic factors govern the use of the PP in LPS and MCS? We hypothesize that the temporal reference will be the most significant factor in governing the PP/preterit distribution, in which the irrelevant reference will favor PP and the prehodiernal will favor the preterit as found previously (Howe and Scott 2008; Schwenter and Torres Cacoullos 2008). Moreover, we expect plural direct object to favor the PP as they favor iterative meanings (Howe and Scott 2008; Schwenter and

⁸ The PRESEEA Corpus (*Proyecto para el estudio sociolingüístico del español de España y de América*, 2014–) is constituted of sociolinguistic interviews from cities of in Latin America and Spain (Moreno Fernández 1996).

Torres Cacoulllos 2008). The presence of *ya* “already” is expected to favor the preterit in both varieties (Howe 2013; Howe and Scott 2008; Schwenter and Torres Cacoulllos 2008).

RQ3: What social (extra-linguistic) factors govern the use of the PP in LPS and MCS? Is there any evidence of a change in progress in either speech community? As only a few studies (Bateman 2022; Howe 2013; Jara Yupanqui 2006) examined how social factors constrain PP/preterit variation in Peruvian Spanish and found contrasting results, there was no formal hypothesis as there was no reason to assume a priori that social factors would govern PP variation.

3 Methods

Similar to Schwenter and Torres Cacoulllos (2008), here the variationist comparative method (Poplack and Tagliamonte 2001: Chap. 5) was implemented to compare the linguistic factors that govern PP and preterit use in MCS and LPS. This method compares and contrasts the results of multivariate analyses on different data sets (Torres Cacoulllos 2011: 161), with the aim of examining whether a linguistic factor patterns in the same way across dialects (Tagliamonte 2012: 162). In addition, we included social variables that have been shown to predict the use of PP/preterit in previous studies (Burgo 2012; Fløgstad 2017; Kubarth 1992; Schwenter 1994; Serrano 1994). Following previous studies (Schwenter and Torres Cacoulllos 2008), we used a form-based approach (Torres Cacoulllos 2011: 160), including all instances of verbs conjugated in the preterit or PP. This approach is motivated by the form-function asymmetry that characterizes morphosyntax in oral discourse: a single form can cover a range of meanings and different forms can serve the same grammatical function.⁹ Therefore, it is only by considering all the contexts in which the grammaticalized structure may appear (categorically or in alternation with the other variant) that we can identify its functions and directions of semantical development, if any (Poplack and Tagliamonte 2001: 7; Schwenter and Torres Cacoulllos 2008; Torres Cacoulllos 2011).

⁹ From a grammaticalization theory perspective, this form-function asymmetry is the manifestation of two processes characteristic of grammaticalization: retention of meaning (when a grammaticalized form maintains the ancient meaning along with the new, grammaticalized one) and layering (when different forms, an older one and a new one, fulfill the same function) (Bybee et al. 1994: 15–22).

3.1 Participants

A total of 2,534 tokens of verbs in PP and preterit were coded from the transcriptions of 36 sociolinguistic semistructured interviews in the PRESEEA Corpus. 18 participants were from Lima and 18 from México City (online Appendices A and B for participant demographics). The PRESEEA corpus has 18 speakers for each speech community that are socially stratified by gender (male, female), age (20–34, 35–54, 55+), and educational levels (basic, medium, high). Table 1 summarizes the characteristics of participants in each speech community.

The semi-structured sociolinguistic interviews were collected between 2005 and 2007 in Mexico City and between 2008 and 2009 in Lima. All interviews were recorded in quiet environment such as a classroom, the interviewee’s house, or an office, with the exception of one interview in Lima (conducted in a street) and another in Mexico City (in a *tianguis* “street market”). The interviews are structured in eight thematic modules (online Appendix C) about each speaker’s life to encourage naturalistic speech following PRESEEA methods (Moreno Fernández 1996). While the Lima interviews closely follow this structure, the Mexico City interviews are a bit more variable in content. The interviews conducted in Lima present a similar length (40–50 min), whereas the Mexican ones are longer overall (60 min) but also present more variability in length. Despite these methodological discrepancies, the interviews were transcribed following the same coding procedure, and the same amount of participant demographic information is provided; therefore, we consider the two data sets comparable.

3.2 Dependent variable, data collection, and data coding

The alternation between preterit and PP was coded as a binary dependent variable. Coding began at the start of each interview. Although we included all instances of verbs conjugated in the preterit or PP, we considered a number of contexts outside

Table 1: Participants demographics for each speech community based on the PRESEEA norms.

Education level	Generation 1 (20–34)		Generation 2 (35–54)		Generation 3 (55–83)	
	Male	Female	Male	Female	Male	Female
Basic (0–5 years)	1	1	1	1	1	1
Medium (10–12 years)	1	1	1	1	1	1
High (15+ years)	1	1	1	1	1	1

of the “envelope of variation” (Milroy and Gordon 2003), including progressive structure *estar* + gerund (1), morphologically ambiguous (preterit or present) -AR and -IR verbs in first person plural (2), false starts (3), following previous studies (Poplack and Tagliamonte 2001; Schwenter and Torres Cacoullos 2008), and direct repetitions of the interviewer’s question (4). These contexts were, therefore, not included in the analysis.

- (1) *Fui aprendiendo a que no me importara esas cosas* “I was learning that I didn’t care about those things” (MEXI_H22_054).
- (2) *Pues ya nos casamos* “Well, we already married” (LIMA_H23_050).
- (3) *Ya pasó esa ... ya pasó esa etapa* “That already passed, that stage already passed” (LIMA_H12_028).
- (4) Interviewer: *¿Cómo entraste a trabajar?* “How did you enter to work?” – Interviewee: *¿Cómo entré a trabajar?* “How did I enter to work?” (MEXI_M11_084).

After excluding these tokens, there were 2,534 tokens. However, due to too few tokens per level, we eliminated the hodiernal temporal reference level for the temporal reference variable (16 Mexico City, 19 for Lima) as well as the adverbial of recency level of the adverb type variable (7 Mexico City, 35 Lima). With so few tokens in these levels, we could not reliability claim a main effect or lack thereof when they were placed into the regression models, as it could skew the data with a Type I or Type II error. Thus, the final analysis was conducted with 2,457 total tokens (1,230 México City; 1,227 Lima; *M*: 68/speaker).

3.3 Independent variables

The study included three social variables, four linguistic variables, and speaker as a random factor. Following PRESEEA norms (Moreno Fernández 1996; PRESEEA 2014-), participants were divided by AGE GROUP¹⁰ according to three generations: younger (20–34), middle-aged (35–54), and older (55+); GENDER: male and female; and EDUCATIONAL ATTAINMENT, which included basic (0–5 years of schooling), medium (10–12 years of schooling), and high (15+ years of schooling). This study treats education only as educational attainment without reference to social class as operationalizing social class is quite complex, as it includes different measures of economic power, social prestige, and status (Ash 2013), and varies per the context of each community.

¹⁰ Age was also treated as a continuous variable.

As for linguistic variables, we selected four factors that have been shown to be significant predictors of PP/preterit variation (Howe and Scott 2008; Schwenter and Torres Cacoulllos 2008): temporal reference, temporal adverbial, presence of *ya*, and direct object. The choice of these linguistic variables is motivated by the methodological need of avoiding the notion of current relevance (Torres Cacoulllos 2011), as it is impossible to establish objectively which tokens are currently relevant above intuitive characterizations (Schwenter and Torres Cacoulllos 2008). Similarly, coding the tokens according to the four prototypical functions of the PP (resultative, continuative, experiential, and recent past; see Comrie 1976) is problematic when dealing with spontaneous speech (Howe 2006, 2013; Schwenter and Torres Cacoulllos 2008).¹¹ Linguistic variables such as the temporal reference and the presence of certain temporal adverbials are more objective measures of the constraints that govern PP/preterit variation. Each independent variable is explained below.

First, TEMPORAL REFERENCE included four levels. (i) *Hodiernal*: the event occurred in the same day in which the speech is taking place¹² (5). (ii) *Prehodiernal*: the event occurred in the day before the day of the speech, or in a more remote past (6). (iii) *Indeterminate*: the analyst is unable to determine when an action occurred (7). It is possible to ask “when?” to the speaker to resolve the ambiguity; however, the speaker does not need to specify the temporal reference for their discursive purposes (Torres Cacoulllos 2011: 160). (iv) *Irrelevant*: it cannot be queried when something occurred, as the event lacks a specific temporal relevance, as in the case of permanent states that continue up to the present, repeated actions, yes/no questions, and negative polarity (when an event has not occurred) (8).

- (5) *Hoy les dije que no iba a llegar temprano* “Today I told them that I was not going to arrive early” (MEXI_M11_084).
- (6) *Nos casamos en diciembre de ese año* “We married in December of that year” (LIMA_H33_053).
- (7) *He llamado a I. [nombre] para que viera la casa* “I have called I. [name] so that s/he sees the house” (LIMA_M32_025).
- (8) *Nunca me ha gustado el calor* “I have never liked the heat” (LIMA_M32_025).

¹¹ For this reason, a number of previous studies analyzed PP functions using grammaticality judgment tasks or other experimentally controlled tasks that force specific elicitations of the PP in controlled linguistic contexts.

¹² The hodiernal temporal reference was removed from the analysis because the scarce number of tokens would skew the regression model.

Second, TEMPORAL ADVERBIAL included five levels. (i) *Specific time adverbials*: they answer to the question *when?* (9). (ii) *Adverbials of recency*: specific time adverbials that express that the event is perceived as recent (i.e., *este verano* “this summer,” *ahora* “now”) (10).¹³ (iii) *Atelic adverbials*: frequency and continuous time adverbials, which answer to questions: *for how long?* or *how many times?* (i.e., *siempre* “always,” *a veces* “sometimes,” *hasta* “until”) (11). The atelic context for temporal adverbials was almost always coded as irrelevant for the independent variable of temporal reference. Thus, these two independent variables are not fully independent, but the current study follows previous studies (Howe and Scott 2008; Schwenter and Torres Cacoullos 2008) in incorporating both independent variables. (iv) *Connective adverbials*: typically used to join the parts of a narration (i.e., *después* “after,” *entonces* “then,” *luego* “then”) (12). (v) *Absence of adverbial*: there is no temporal adverbial next to the verb (7). Atelic adverbs, which express duration and frequency, are expected to favor the PP, whereas specific time and connective adverbials are expected to favor the preterit (Howe and Scott 2008).

(9) *Ingresé a la facultad en el 79* I began at the university in 79 (MEXI_H23_018).

(10) *Mi hija recién ha dado a luz* “My daughter recently has given birth” (LIMA_M21_005).

(11) *Ahí duramos como dos, tres años* “There we lasted like two, three years” (MEXI_H32_066).

(12) *Le habló por teléfono y ya después llegó ella* “S/he spoke to her/him on the phone and then afterwards she arrived” (MEXI_H11_078).

Third, PRESENCE OF THE ADVERB *YA* “already,” included two levels. (i) *Present*: *ya* was present in the sentence (13); (ii) *Absent*: there was no *ya* (14).

(13) *Ya me ... me ... me fui del trabajo porque ya salí embarazada* “I already left work because I was already pregnant” (LIMA_M21_005).

(14) *Dejé la carrera y me dediqué a trabajar* “I left my studies at the university and I dedicated myself to work” (MEXI_H32_066).

¹³ The level of adverbials of recency was removed from the analysis because the scarce number of tokens would skew the regression model.

Fourth, DIRECT OBJECT, included three levels. (i) *Singular*: the direct object was singular (15). (ii) *Plural*: the direct object was plural (16). (iii) *Absent*: there was no direct object (17).

- (15) *Nunca se levantó de su silla mientras que escribió la tesis* “He never got up from his chair while he wrote their thesis” (LIMA_M22_023).
- (16) *Me gané unos cuantos pleitos con algunos militares* “I won a number of disputes with some military personnel” (LIMA_H12_028).
- (17) *Esos diez meses fueron espectaculares, ¿no?* “Those ten months were spectacular, no?” (LIMA_M22_023).

3.4 Statistical analysis

Following the variationist comparative method (Poplack and Tagliamonte 2001: Chap. 5; Tagliamonte 2012) and previous PP/preterit variation studies (Schwenter and Torres Cacoullos 2008), the current study conducted separate multivariate analyses for each speech community. This allows for one to understand the underlying linguistic and social factors governing the use of the PP versus the preterit in each community, comparing similarities and differences. For each data set, following Tagliamonte and Bayen (2012), prior to any regression models, a random forest was calculated using the *cforest* function from the *party* package (Hothorn et al. 2020) to determine variable importance. This allows the researcher to order in importance each independent variable in the regression modeling, as order impacts the output of the regression. Following the random forest, a mixed-effects logistic regression model was fitted for each speech community using the *lmer* function (Bates et al. 2015) and *lmerTest* (Kuznetsova et al. 2017) in R (R Core Team 2022). The marginal *R*-squared (R^2m) and conditional *R*-squared (R^2c) values are listed to provide a goodness-of-fit of the variation for each final model (Nakagawa and Schielzeth 2013). All independent variables were tested in the model construction. Nonsignificant interactions were discarded from each subsequent model. For categorical independent variables with more than two levels, post-hoc analyses were conducted with estimated marginal means (Lenth et al. 2018). Different models were compared using ANOVA testing, relying on the lowest AIC being the best explanation of the data. While several interactions were included in the model, in general these models did not converge. Thus, to explore the relationship between significant main effects, a conditional inference tree was conducted using the *cforest* function from the *party* package (Hothorn et al. 2020).

4 Results

4.1 Overall descriptive statistics

Of the 1,230 tokens in MCS, 111 (9.0 %) were realized in the PP and 1,119 (91.0 %) in the preterit. Of the 1,227 tokens in LPS, 303 (24.7 %) were realized in the PP and 924 (75.3 %) in the preterit. Figure 1 visualizes the proportional differences in realizations between communities. A Chi-square test revealed a significant difference between the two speech communities in which speakers from Mexico City produce significantly less PP than speakers from Lima, $\chi^2(1) = 106.5, p < 0.0001$.

4.2 Mexico City

The random forest results indicate that for PP/preterit realizations, the most important predictor is temporal reference, followed by direct object, gender, speaker, education, age (continuous), adverb type, polarity, age group, and the presence of *ya* (Figure 2).

The best-fit mixed-effects logistic regression is presented in Table 2, displaying the estimate, standard error (SE), *t*-value, percent of PP per level, number (*n*) of tokens per level, and *p*-value. As the model is in reference to PP, positive estimates indicate that the listed factor level favors PP more than the reference level and

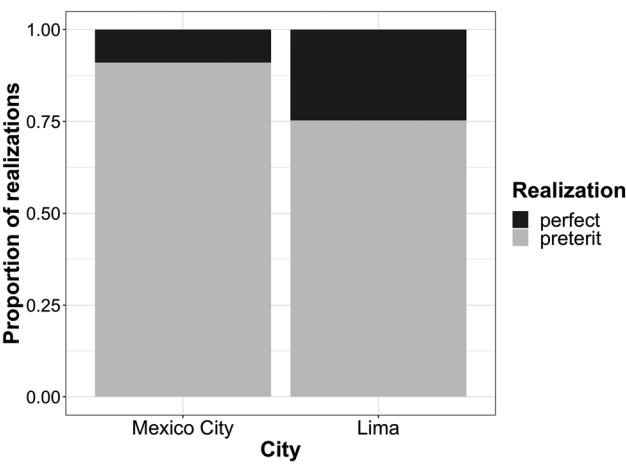


Figure 1: Proportion of realizations per speech community.

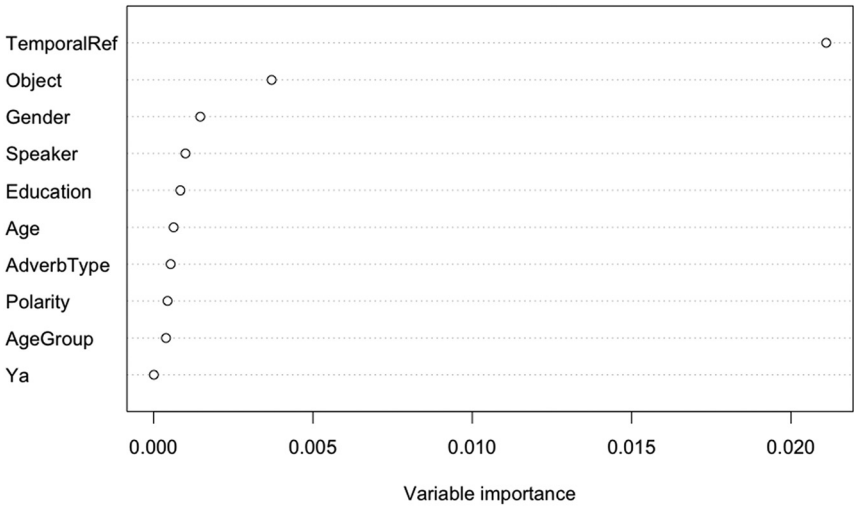


Figure 2: Random forest of PP and preterit variation for Mexico City.

negative estimates indicate that the listed factor level favors preterit realizations more than the reference level. Reference levels are displayed in the table.

The model demonstrated a significant main effect for temporal reference in which the irrelevant temporal reference most favored PP, followed by the indeterminate temporal reference, and then finally the prehodiernal temporal reference (Figure 3A). A post-hoc analysis of the estimated marginal means revealed that the irrelevant temporal reference significantly favored PP realizations more than both indeterminate ($p < 0.001$) and prehodiernal temporal references ($p < 0.001$). Additionally, the indeterminate temporal reference also favored more PP realizations than the prehodiernal temporal reference ($p < 0.01$).

Table 2: Summary of mixed-effects logistic regression for PP and preterit realizations for Mexico City in reference to PP realizations, speaker as a random factor, $n = 1,230$ (R^2m : 0.52, R^2c : 0.58).

Predictors	Estimate	SE	z-value	% PP	n	p-value
(Intercept)	−6.44	1.01	−6.37	–	–	***
TemporalRef (Ref = prehodiernal)	–	–	–	0.24	420	–
Indeterminate	3.60	1.00	3.59	8.77	673	***
Irrelevant	5.57	1.01	5.51	37.23	137	***
Object (Ref = none)	–	–	–	7.19	862	–
Plural	1.32	0.38	3.48	17.57	74	***
Singular	0.55	0.25	2.19	13.95	294	*

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. n = total tokens (PP & preterit) per level.

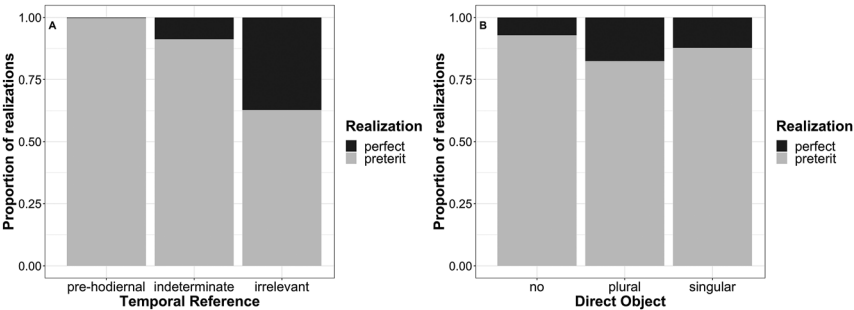


Figure 3: Main effects of temporal reference (A) and direct object (B) for PP and preterit realizations for Mexico City.

The model also demonstrated a significant main effect for direct object in which plural objects most favored PP realizations, followed by singular objects, and then lastly, no objects following the verb (Figure 3B). A post-hoc analysis of the estimated marginal means revealed that the only statistically significant difference between the three levels was that the plural object significantly favored PP realizations compared to no object ($p < 0.01$), but no significant differences compared to singular objects ($p = 0.13$). Furthermore, there were no significant differences between singular objects and no objects ($p = 0.07$).

To further explore the relationship between the two significant main effects, a conditional inference tree was conducted (Figure 4). The conditional inference

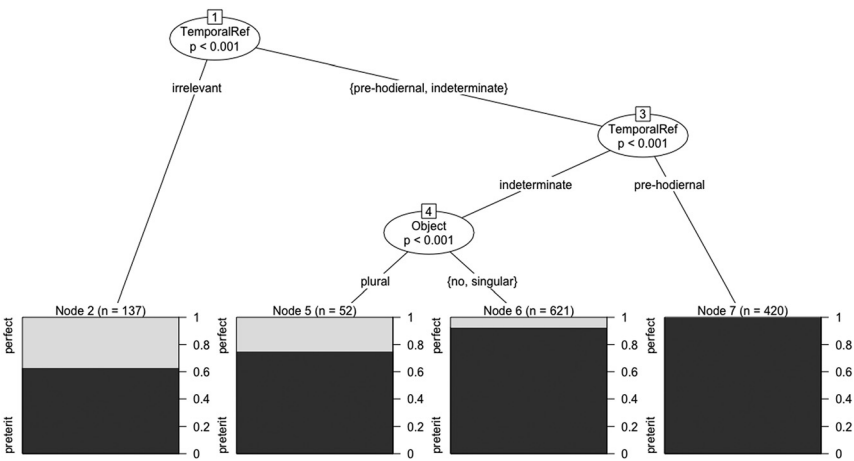


Figure 4: Conditional inference tree of the factors that predict PP/preterit variation for Mexico City (dark gray = preterit, light gray = PP).

tree indicates that temporal reference is the most significant predictor in which the irrelevant temporal reference demonstrates the most PP. When considering the prehodiernal and indeterminate temporal references, only the indeterminate context shows variation as the prehodiernal temporal reference demonstrates nearly categorical preterit. For the indeterminate context, it interacts with the presence of a direct object. Specifically, the indeterminate temporal reference with a plural direct object favors more PP than those of a singular or no direct object.

4.3 Lima

The random forest results indicate that for PP/preterit realizations, the most important predictor is temporal reference, followed by speaker, age group, age (continuous), education, direct object, adverb type, gender, *ya*, and polarity (Figure 5).

Age group and continuous age were placed into different regression models and compared to determine which version on this independent variable best explained the variation. The best model for both age group and continuous age included main effects for temporal reference, adverb type, *ya*, and either age group or continuous age. These model were compared using ANOVA testing, but there was no significant difference between the models and virtually identical AIC values. The R^2c values were identical as well (0.31), with only the R^2m values being slightly higher for the

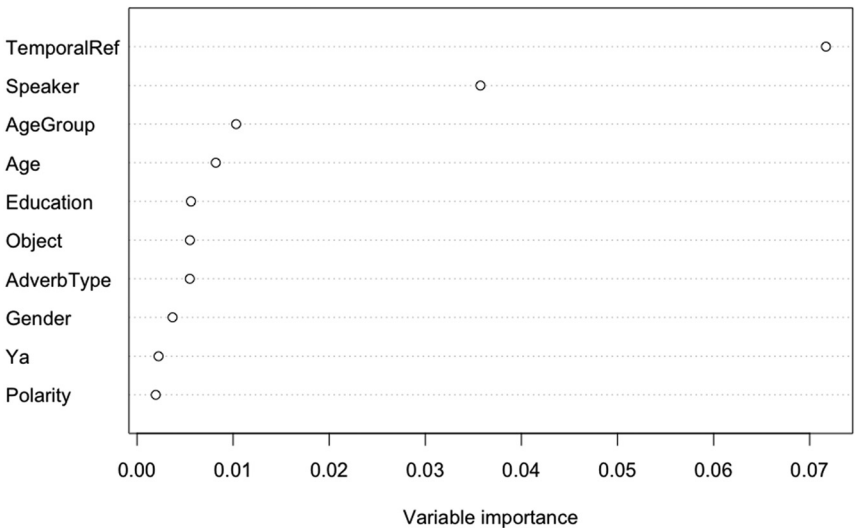


Figure 5: Random forest of PP and preterit variation for Lima.

model with age group (0.23) compared to that of continuous age (0.22). Given this difference, as well as the higher variable importance value given to age group in the random forest, it was decided to use age group in the final regression model. However, follow-up analyses below explore continuous age.

The best-fit mixed-effects logistic regression is presented in Table 3, displaying the estimate, standard error (SE), *t*-value, percent of PP per level, number (*n*) of tokens per level, and *p*-value. As the model is in reference to PP, positive estimates indicate that the listed factor level favors PP more than the reference level and negative estimates indicate that the listed factor level favors preterit realizations more than the reference level. Reference levels are displayed in the table.

The model demonstrated three significant main effects for the linguistic factors. First, the model exhibited a main effect for temporal reference in which the irrelevant temporal reference most favored PP, followed by the indeterminate temporal reference, and then finally the prehodiernal temporal reference (Figure 6A). A post-hoc analysis of the estimated marginal means revealed that the irrelevant temporal reference significantly favored PP more than both the indeterminate ($p < 0.001$) and the prehodiernal temporal reference ($p < 0.001$). Additionally, the indeterminate temporal reference also favored PP more than the prehodiernal temporal reference ($p < 0.001$). Second, the model demonstrated a main effect for adverb type in which the presence of an atelic adverb most favored the PP (Figure 6B). However, a post-hoc analysis of the estimated marginal means only revealed one significant difference between levels in which specific adverbs favored the PP more than the connective adverbs ($p < 0.05$). Third, the model demonstrated a

Table 3: Summary of mixed-effects logistic regression for PP and preterit realizations for Lima in reference to PP realizations, speaker as a random factor, $n = 1,227$ (R^2m : 0.23, R^2c : 0.31).

Predictors	Estimate	SE	z-value	% PP	<i>n</i>	<i>p</i> -value
(Intercept)	−3.11	0.38	−8.25	–	–	***
TemporalRef (Ref = prehodiernal)	–	–	–	8.8	375	–
Indeterminate	1.66	0.26	6.41	27.3	737	***
Irrelevant	2.97	0.34	8.76	60.0	115	***
Age group (Ref = 18–34)	–	–	–	16.9	415	–
35–54	0.32	0.40	0.78	30.0	403	0.43
>55	1.10	0.40	2.76	34.2	409	**
AdverbType (Ref = none)	–	–	–	26.1	811	–
Atelic	−0.05	0.25	−0.20	38.0	142	0.84
Connective	−0.82	0.35	−2.38	11.1	99	*
Specific	0.46	0.29	1.56	14.9	175	0.12
Ya (Ref = no)	–	–	–	25.9	1,125	–
Yes	−0.87	0.34	−2.59	11.8	102	**

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *n* = total tokens (PP & preterit) per level.

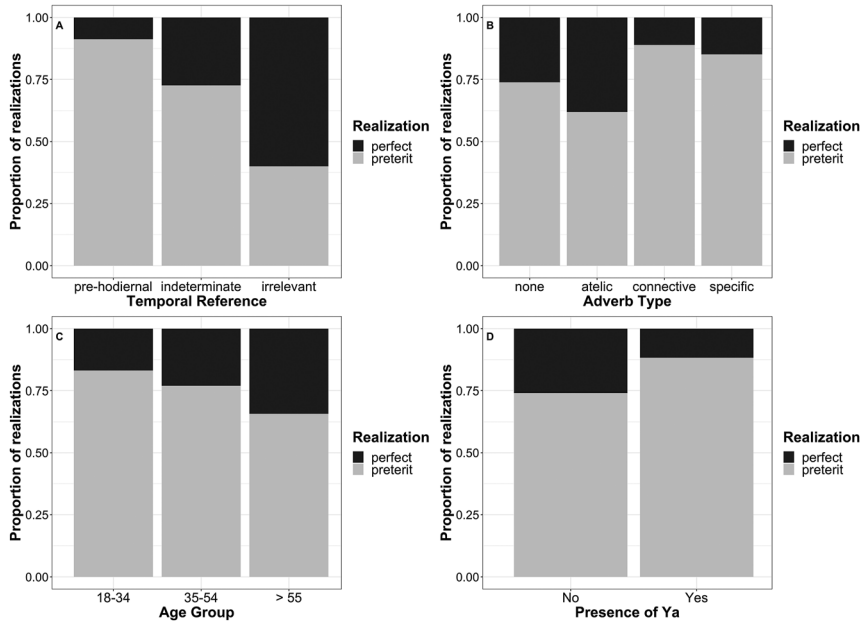


Figure 6: Main effects of temporal reference (A), adverb type (B), age group (C), and direct object (D) for PP and preterit realizations for Lima.

main effect for *ya* in which the absence of *ya* favored the PP more than the presence of *ya* ($p < 0.01$) (Figure 6D).

Different from the MCS, the model for LPS data demonstrated one significant main effect for a social factor: age group. Specifically, those in the youngest age group (18–34) produced less PP than those in the middle age group (35–54) and the oldest age group (55 and older) (Figure 6C). Those aged 18–34 produced statistically significant less PP than those aged 55 and older ($p < 0.05$) but not statistically less than those aged 35–54 ($p = 0.71$). There were no significant differences in PP realizations between those aged 35–54 and those aged 55 and older ($p = 0.11$).

To further explore the relationship between the four significant main effects, a conditional inference tree was conducted (Figure 7). The conditional inference tree indicates that temporal reference is the most significant predictor in which the irrelevant context demonstrates the most PP production. Then when considering only the prehodiernal and indeterminate temporal references, only the indeterminate context shows variation as the prehodiernal context nearly categorically exhibits the preterit. Thus, the variation is most seen in the indeterminate temporal reference as it interacts with age group, *ya*, and adverb type. The first branch here is

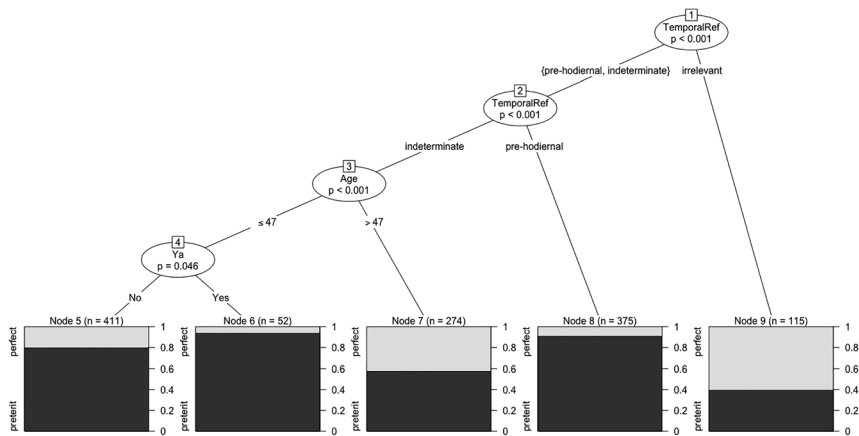


Figure 7: Conditional inference tree of the factors (with categorical age) that predict PP/preterit variation for Lima (dark gray = preterit, light gray = PP).

with age group, in which the oldest age group overall shows higher rates of PP, but even more so with all adverb types that are not connective. Then among the younger age groups, it depends upon the presence of *ya* in which the presence of *ya* favors more preterit, while the absence of *ya* shows less preterit, but this interacts with adverb type in which non-*ya* contexts with atelic, connective, and specific adverbs favors more preterit than non-*ya* contexts with no adverbs. However, as one can observe, the interactions with adverb type appear to be weak significant effects, with values just under $p < 0.05$, and thus these should be taken with caution.

As mentioned previously, the model above opted for age group based on its higher R^2m value in the regression model and higher variable importance in the random forest than continuous age. However, as regression models showed no statistical difference between categorical and continuous age, here we conduct follow-up analyses with continuous age. Below we first visualize all the data points with continuous age (Figure 8A). A Pearson correlation was conducted and found a significant weak positive correlation between age and PP realizations ($n = 1,227$, $df = 1,225$, $r = 0.15$, $R^2 = 0.02$, $p < 0.001$). Then we visualize the data as the percent of PP usage per individual speaker; thus, one data point per person (Figure 8B). A Pearson correlation was conducted and found a nonsignificant moderate positive association between age and PP frequency ($n = 18$, $df = 16$, $r = 0.46$, $R^2 = 0.21$, $p = 0.054$). As this result was nearly significant, a follow-up Pearson correlation was conducted without the outlier LIMA_M11 (26-year-old, female, basic education), who although was part of the youngest group demonstrated the highest PP usage of all speakers. The follow-

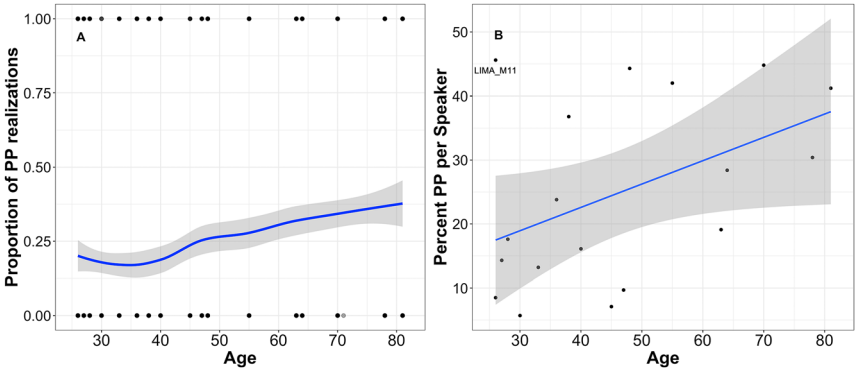


Figure 8: Scatterplot of all data points by continuous age (A) and scatterplot of percent PP per individual speaker by continuous age (B).

up Pearson correlation indicated a significant moderate positive association between age and PP frequency ($n = 17$, $df = 15$, $r = 0.625$, $R^2 = 0.39$, $p < 0.01$).

A conditional inference tree was conducted to explore the relationship between continuous age and three significant linguistic factors (temporal reference, adverb type, *ya*) for PP/preterit variation (Figure 9). Similar to the previous conditional inference tree, the most important factor is temporal reference in which the irrelevant context shows the most PP. Then the prehodiernal temporal reference shows nearly categorical preterit. Similar to the other model, within the indeterminate

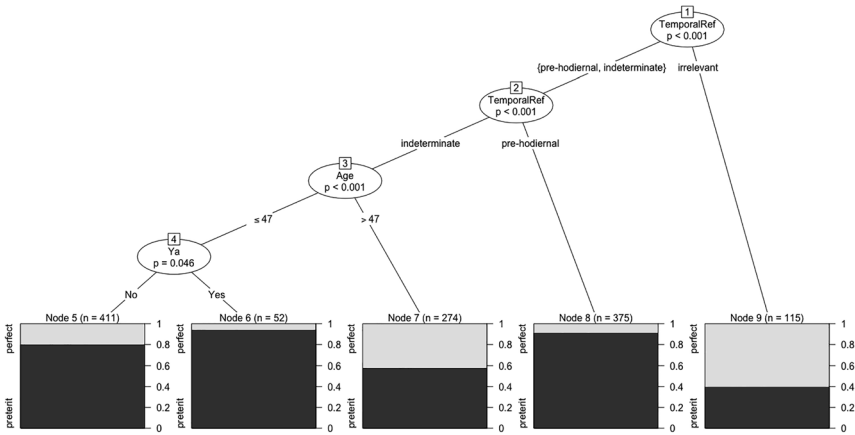


Figure 9: Conditional inference tree of the factors (with continuous age) that predict PP and preterit variation for Lima (dark gray = preterit, light gray = PP).

temporal reference, variation here is predicted by age and *ya*. Specifically, the model branches based under and over 47 years of age, showing that those over 47 years of age produce the most PP even in the indeterminate temporal reference. Then those 47 years of age and younger use more preterit in the presence of *ya* as compared to the absence of *ya* within the indeterminate temporal reference. Adverb type does not appear to interact with continuous age as it did age group. Of note is the model's branching at 47, as opposed to the categorical value of 55 or older, indicating why the regression models yielded similar results.

5 Discussion

Regarding RQ1 (*what is the overall frequency of PP/preterit in LPS and MCS?*), the results revealed that the preterit is the most common form in both speech communities. However, the PP is significantly more frequent in LPS (24.7%) than in MCS (9%). This supports previous research that Peruvian Spanish presents a higher frequency of the PP compared to other Latin American varieties (Howe 2006; Howe and Scott 2008; Jara Yupanqui 2011a; Rodríguez Louro and Jara Yupanqui 2011).

As for RQ2 (*which linguistic factors influence the use of PP/preterit in LPS and MCS?*), the temporal reference is the only significant factor shared by the two speech communities, showing the greatest magnitude of effect in both. These results confirm that the analysis of the temporal reference is crucial to understand preterit/PP variation (Coppole 2011; Howe and Scott 2008; Schwenter and Torres Cacoullos 2008). Despite the different overall frequencies, the two speech communities also share the same constraint hierarchy, with irrelevant temporal reference most strongly favoring the PP, followed by indeterminate temporal reference, and lastly, pre-hodiernal reference, where the preterit is chosen categorically in MCS and nearly categorically in LPS. This is indicative of shared principles of “past-time expression” (Poplack and Tagliamonte 2001: 129) in the two speech communities, and it is consistent with the claim that Latin American varieties, despite different overall frequencies, share semantic-pragmatic features of the PP, as opposed to several Peninsular varieties which present a different functional distribution of PP and preterit (Howe 2013; Howe and Scott 2008).

The plural direct object is the other factor that significantly favors the PP in MCS. Given that plural direct objects are compatible with continuative and iterative meanings (Schwenter and Torres Cacoullos 2008), this result confirms that the PP has a semantic value of durativity and iteration in Mexican Spanish (Harris 1982; Lope Blanch 1972; Squartini and Bertinetto 2000; Westmoreland 1988). In LPS, type of temporal adverbial and adverb *ya* were also significant predictors. Connective adverbials (e.g., *luego*, *después*) disfavor the PP, as they are used to sequence events in a

narration, a function that is typically fulfilled by means of preterit (Howe 2013). As for the adverb *ya*, its absence favors the PP more than its presence. The adverb *ya* indicates that an event occurred at an indefinite point before the moment of speech (Schwenter and Torres Cacoullos 2008: 26). Previous studies found that the presence of *ya* favors the PP in Peninsular Spanish, unlike Latin American Spanish (Howe 2006, 2013: 129–30; Schwenter and Torres Cacoullos 2008). According to Schwenter and Torres Cacoullos (2008: 26), the favoring of PP with *ya* is evidence of the perfect-to-perfective semantic shift of the PP in Peninsular varieties in which PP has become the “default past perfective form.” The current results demonstrate the opposite trend for LPS in which the *ya* disfavored PP realizations indicating that preterit may be the default past perfective form in this variety.

The indeterminate temporal reference stands out for interacting with other factors in both communities. In MCS, it favors more PP when it is accompanied by a plural direct object. In LPS, it interacts with age group. The two youngest age groups produced the preterit almost categorically in presence of the adverb *ya*. They used more PP in the absence of *ya*, and especially when there was no temporal adverbial. The oldest age group did not show an effect of the adverb *ya* within the indeterminate temporal reference, but it showed a disfavoring effect of connective adverbials on the selection of PP. Follow-up analyses with continuous age substantially mirrored these results, confirming that, within the indeterminate temporal reference, those over 47 years of age produced the most PP regardless of the presence of *ya*, whereas those under 47 years of age used more preterit in the presence of *ya* as compared to the absence of *ya*.

In reference to RQ3 (*which social factors influence the use of PP/preterit in LPS and MCS?*), akin to the results of the oral interviews by Jara Yupanqui (2006) and by Bateman (2022), our analysis revealed that the youngest group of LPS speakers produced less instances of PP compared to the oldest group. This age effect in *apparent time* is suggestive of a change in progress in LPS. Importantly, this change is in the opposite direction found in several Peninsular Spanish varieties, and nonetheless, it mirrors the results of other studies regarding Latin American Spanish varieties (Fløgstad 2017; González et al. 2019; Kubarth 1992).

These results have two theoretical implications. Firstly, if we accept that the PP expands through indeterminate temporal reference contexts (Coppole 2011; Howe and Scott 2008; Schwenter and Torres Cacoullos 2008), the data suggest that the PP is retreating from LPS, given its lower frequency in the youngest age group as compared to the oldest age group within the indeterminate reference level. Secondly, this provides interesting insights into the tenet of the unidirectionality of grammaticalization. Grammaticalization theory predicts that when a form grammaticalizes, the process is unidirectional, with the innovative more recent form displacing the older form such as the PP displacing the preterit in several Romance languages

(Bybee et al. 1994: 12; Hopper and Traugott 2003: 100). Given we observe the opposite trend here, the expansion of the preterit at the expense of the PP in LPS, one might claim that this provides a counter example to the unidirectionality of grammaticalization. A true case of “antigrammaticalization” (Haspelmath 2004: 34), however, would be if a grammatical feature after grammaticalizing returns to its former state at a later stage. While there is a tendency among several Peninsular varieties (as well as among other Romance languages) in which the PP is grammaticalizing to overtake the default perfective function, we cannot claim a true counterargument to the irreversibility of grammaticalization as these are separate varieties and, as suggested by previous scholars, may have different pathways to change (Escobar and Crespo del Río 2020; Escobar and Roy 2022; Howe 2013). Rather than reversing the directionality of grammatical change, it appears that there are divergent paths in which some Peninsular varieties are expanding their use of the PP while several Latin American varieties are increasing their use of the preterit as the default perfective. Even though there is quite a bit of evidence of the perfect-to-perfective shift as a universal tendency, rather than focus on one common pathway of change in the tense-aspect domain (Bybee 2006), it is important here to focus on change constraints such as source and directionality constraints (Haspelmath 2019: 3). Following this argument, the grammaticalization of the PP in some Peninsular varieties may be due to (external) source constraints, perhaps as Drinka (2003, 2017) suggests, the contact with the Parisian French vernacular and its cultural influence. The rationale for the lack of a similar grammaticalization pathway, and perhaps opposite direction of grammaticalization in LPS (and other varieties of Latin American Spanish), may be due to different constraints. Our results, in line with previous studies (Bateman 2022; Giacalone Ramat 1998; González et al. 2019; Janda 2001), suggest that grammaticalization is occurring in the opposite direction. Future work is needed to better understand the constraints that are leading to the expansion of the preterit over the PP across several varieties of Latin American Spanish.

While the current results found that overall PP appears to be losing ground to the preterit among younger speakers (similar to Bateman 2022), it is important to recognize a limitation of the study in that we are unable to know whether these speakers from Lima are Quechumaran¹⁴-Spanish bilinguals or monolingual speakers. This is important information as the case of PP in Peruvian Spanish is quite complex as previous studies have demonstrated that PP has an evidentiality function due to its contact with Quechumaran languages (Escobar 1997, 2000; Escobar and Crespo del Río 2020; Escobar and Roy 2022). Additionally, Bateman (2022) has found that there is an innovative “contact-induced subjectivization” PP behavior among Quechumaran-Spanish bilinguals in Cusco, but that Spanish-dominant speakers

¹⁴ Quechumaran refers to the language-family that includes Quechua and Aymara.

(as opposed to Quechua-dominant speakers) overall disfavor PP usage. Thus, what we might be observing in Peruvian Spanish is a tendency toward less PP among Spanish monolinguals or Spanish-dominant speakers, not unlike other parts of Latin America. However, as suggested by several scholars (Escobar 1997, 2000; Escobar and Crespo del Río 2020; Escobar and Roy 2022), PP among Quechumaran-Spanish speakers is taking on evidentiality functions acquired through an evidentiality pathway as opposed to a tempo-aspectual pathway. Future work is needed to further tease apart what may be a larger trend among Latin American Spanish in the expansion of the preterit and the innovative PP evidential and subjectivation functions due to contact with Quechumaran languages in Peru as well as in other Andean region varieties (Escobar and Roy 2022).

The present study is not without a few other limitations. First, the inclusion of more tokens at the hodiernal reference level for the temporal reference factor group, and at the adverbials of recency level for the temporal adverbial factor group, would have allowed for a more complete picture of the PP/preterit variation in each speech community. Also, the limited number of participants per community should call for caution in the interpretation of the results. Future work should continue to examine the possible reduction of PP across varieties of Latin American Spanish using both sociolinguistic interviews as well as experimentally elicited tasks while also taking into consideration possible differences in source constraints between monolingual and multilingual speakers. As PP/preterit variation across Latin American Spanish varieties has been investigated only partially, future research may provide new insights into the processes involved in grammaticalization and how language contact and other source constraints affect it.

In conclusion, the results indicate a change in progress toward an expansion of the preterit in LPS among younger speakers with the indeterminate temporal reference as locus of change for PP usage (Copple 2011; Schwenter and Torres Cacoullous 2008). The findings suggest that perfect-to-perfective grammaticalization may not always be the only cross-linguistic pathway but rather is subject to source constraints, which may lead to another pathway in which the preterit expands at the expense of the PP.

Acknowledgments: The authors would like to thank the PRESEEA research teams in Mexico City and Lima, Peru for providing open-access to the sociolinguistic interview transcripts. We are also indebted to two reviewers for their invaluable comments. We would also like to thank Danny Erker as well as audience members at NWAV 50 for their insightful input. All errors remain are our own.

References

- Ash, Sharon. 2013. Social class. In Jack K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*, 2nd edn. 350–367. Oxford, UK: Wiley.
- Azpiazu-Torres, Susana. 2017. Actitudes lingüísticas de los hablantes gallegos de español frente a la oposición Pretérito Perfecto Compuesto (PPC)/Pretérito Perfecto Simple (PPS). *Lingüística en la Red* 15. 1–34.
- Bateman, Bethany D. 2022. *Contact-induced subjectivization in the compound past: An examination of PP/PRET variation in Cusco Spanish*. Athens, GA: University of Georgia dissertation.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Burgo, Clara. 2012. Social influence in the use of the present perfect in Bilbao. *Spanish in Context* 9(1). 88–108.
- Bustamante, Isabel. 1991. El presente perfecto o pretérito perfecto compuesto en el español quiteño. *Lexis* 15. 195–231.
- Bybee, Joan L. 2006. Language change and universals. In Ricardo Mairal & Juana Gil (eds.), *Linguistic universals*, 179–194. Cambridge: Cambridge University Press.
- Bybee, Joan L. & Östen Dahl. 1989. The creation of tense and aspect in the languages of the world. *Studies in Language* 13(1). 51–103.
- Bybee, Joan L., Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Cartagena, Nelson. 1999. Los tiempos compuestos. In Ignacio Bosque & Barreto Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, II, 2935–2976. Madrid: Espasa Calpe.
- Copple, Mary T. 2011. Tracking the constraints on a grammaticalizing perfect(ive). *Language Variation and Change* 23(2). 163–191.
- Comrie, Bernard. 1976. *Aspect: An introduction to verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Drinka, Bridget. 2003. Areal factors in the development of the European periphrastic perfect. *Word* 54. 1–38.
- Drinka, Bridget. 2017. *Language contact in Europe: The periphrastic perfect through history*. Cambridge: Cambridge University Press.
- Escobar, Anna María. 1997. Contrastive and innovative uses of the present perfect and the preterit in Spanish in contact with Quechua. *Hispania* 80(4). 859–870.
- Escobar, Anna María. 2000. *Contacto social y lingüístico: El español en contacto con el quechua en el Perú*. Lima: Pontificia Universidad Católica del Perú: Fondo Editorial.
- Escobar, Anna María & Claudia Crespo del Río. 2020. La gramaticalización de la “subjektividad” en el español andino. El pretérito perfecto compuesto con valor evidencial. In Luis Andrade Ciudad & Sandro Sessarego (eds.), *Los castellanos del Perú: Historia, variación y contacto lingüístico*, 156–205. Oxford/ New York: Routledge.
- Escobar, Anna María & Joseph Roy. 2022. Grammaticalization in the evidential pathway of the Andean Spanish present perfect: Language contact as a trigger for language change. *Isogloss. Open Journal of Romance Linguistics* 8(4). 1–38.
- Fløgstad, Guro N. 2016. *Preterit expansion and perfect demise in Porteo Spanish and beyond*. Leiden/ Boston: BRILL.
- Fløgstad, Guro N. 2017. Revisiting perfect/preterit instability across Romance. On functional motivations for diverging paths. *Acta Linguistica Hafniensia* 49(2). 195–211.

- García Tesoro, Ana I. & Ji S. Jang. 2018. El pretérito perfecto compuesto en el español andino peruano: usos innovadores y extensión a contextos de aoristo. *Forma y Función* 31(1). 93–123.
- Giacalone Ramat, Anna. 1998. Testing the boundaries of grammaticalization. In Anna Giacalone Ramat & Paul J. Hopper (eds.), *Limits of Grammaticalization*, 107–127. Amsterdam/ Philadelphia: John Benjamins.
- González, Paz, Margarita Jara Yupanqui & Carmen Kleinherenbrink. 2019. The microvariation of the Spanish perfect in three varieties. *Isogloss. Open Journal of Romance Linguistics* 4(1). 115–133.
- Harris, Martin. 1982. The ‘past simple’ and ‘present perfect’ in Romance. In Nigel Vincent & Martin Harris (eds.), *Studies in the Romance verb*, 42–70. London: Croom Helm.
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068.
- Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In Olga Fischer, Muriel Norde & Harry Perridon (eds.), *Up and down the cline: The nature of grammaticalization*, 17–44. Amsterdam/ Philadelphia: John Benjamins.
- Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by constraints on change? In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 1–23. Berlin: Language Science Press.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: A conceptual framework*. Cambridge: University of Chicago Press.
- Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press.
- Hothorn, Torsten, Kurt Hornik, Carolin Strobl & Achim Zeileis. 2020. *Party: A laboratory for recursive partytioning, R package version 1.3–4*. <https://cran.r-project.org/web/packages/party/index.html>.
- Howe, Chad. 2006. *Cross-dialectal features of the Spanish present perfect: A typological analysis of form and function*. Columbus, OH: Ohio State University dissertation.
- Howe, Chad. 2013. *The Spanish perfects: Pathways of emergent meaning*. London/ New York: Palgrave Macmillan.
- Howe, Chad & Scott Schwenter. 2003. Present perfect for preterit across Spanish dialects. *University of Pennsylvania Working Papers in Linguistics* 9(2). 61–75.
- Howe, Chadd & Scott Schwenter. 2008. Variable constraints on past reference in dialects of Spanish. In Maurice Westmoreland & Juan Antonio Thomas (eds.), *Selected proceedings of the 4th workshop on Spanish sociolinguistics*, 100–108. Somerville, MA: Cascadia Proceedings.
- Janda, Richard. 2001. Beyond “pathways” and “unidirectionality”: On the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences* 23(2–3). 265–340.
- Jara Yupanqui, Margarita. 2006. *The use of the preterit and the present perfect in the Spanish of Lima*. Pittsburgh, PA: University of Pittsburgh dissertation.
- Jara Yupanqui, Margarita. 2011a. Funciones discursivas y gramaticalización del pretérito perfecto compuesto en el español de Lima. *Spanish in Context* 8(1). 95–118.
- Jara Yupanqui, Margarita. 2011b. Present perfect usage in Peruvian Spanish and perfective readings in narratives. *Revista Internacional de Lingüística Iberoamericana* IX(18). 213–235.
- Kempas, Ilpo. 2013. Sobre la elección entre ‘canté’ y ‘he cantado’ en el español hablado de Santiago de Estero. In Laura Colantoni & Celeste Rodríguez Louro (eds.), *Perspectivas teóricas y experimentales sobre el español de la Argentina*, 269–282. Frankfurt/ Madrid: Iberoamericana/Vervuert.
- Klee, Carol & Alicia M. Ocampo. 1995. The expression of past reference in Spanish narratives of Spanish-Quechua bilingual speakers. In Carmen Silva-Corvalán (ed.), *Spanish in four continents: Studies in language contact and bilingualism*, 52–70. Washington D.C.: Georgetown University Press.

- Kubarth, Hugo. 1992. El uso del pretérito simple y compuesto en el español hablado de Buenos Aires. In H.L. Traill (ed.), *Scripta philologica*, 553–566. Mexico City: Universidad Autónoma de México.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26.
- Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Malden, MA: Blackwell.
- Lehmann, Christian. 2015. *Thoughts on grammaticalization*, 3rd edn Berlin: Language Science Press.
- Lenth, Russell, Henrik Singmann, Jonathan Love, Paul Buerkner & Maxime Herve. 2018. *Emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.5*. <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>.
- Lope Blanch, Juan M. 1972. Sobre el uso del pretérito en el español hablado en México. In J.M. Lope Blanch (ed.), *Estudios sobre el español de México*. Mexico City: Universidad Nacional Autónoma de México.
- Milroy, Lesley & Matthew Gordon. 2003. *Sociolinguistics: Method and interpretation*. Oxford: Blackwell.
- Moreno de Alba, José G. 1998. La oposición pretérito indefinido/pretérito perfecto compuesto en documentos novohispanos de los siglos XVI-XIX. *Actas del IV congreso internacional de historia de la lengua española*, 619–629. Logroño: Universidad de La Rioja.
- Moreno de Alba, José G. 1978. *Valores de las formas verbales en el español de México*. Mexico City: Universidad Nacional Autónoma de México.
- Moreno Fernández, Francisco. 1996. Metodología del “Proyecto para el Estudio Sociolingüístico del Español de España y de América (PRESEEA)”. *Lingüística* 8. 257–287.
- Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution/British Ecological Society* 4. 133–142.
- Pfänder, Stefan & Azucena Palacios. 2013. Evidencialidad y validación en los pretéritos del español andino ecuatoriano. *Círculo de Lingüística Aplicada a la Comunicación* 54. 65–98.
- Poplack, Shana & Sali A. Tagliamonte. 2001. *African American English in the diaspora: Tense and aspect*. Oxford: Blackwell.
- PRESEEA. 2014. *Corpus del Proyecto para el estudio sociolingüístico del español de España y América*. Alcalá de Henares: Universidad de Alcalá. [<http://preseea.uah.es>].
- R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reinöhl, Uta & Nikolaus P. Himmelmann. 2017. Renewal: A figure of speech or a process sui generis? *Language* 93(2). 381–413.
- Rodríguez Louro, Celeste. 2009. *Perfect evolution and change: A sociolinguistic study of preterit and present perfect usage in contemporary and earlier Argentina*. Melbourne, Australia: University of Melbourne dissertation.
- Rodríguez Louro, Celeste & Margarita Jara Yupanqui. 2011. Otra mirada a los procesos de gramaticalización del Presente Perfecto en español: Perú y Argentina. *Studies in Hispanic and Lusophone Linguistics* 4(1). 55–80.
- Schwenter, Scott. 1994. The grammaticalization of an anterior in progress. *Studies in Language* 18(1). 71–111.
- Schwenter, Scott & Rena Torres Cacoullos. 2008. Defaults and indeterminacy in temporal grammaticalization: The ‘perfect’ road to perfective. *Language Variation and Change* 20(1). 1–39.
- Serrano, María José. 1994. Del pretérito indefinido al pretérito perfecto: un caso de cambio y gramaticalización en el español de Canarias y Madrid. *Lingüística Española Actual* 16(1). 37–57.
- Serrano, María José. 1995. Sobre el uso del pretérito perfecto y pretérito indefinido en el español de Canarias: Pragmática y variación. *Boletín de Filología* 35(1). 534–566.

- Squartini, Mario & Pier Marco Bertinetto. 2000. The simple and compound past in Romance languages. In Ö. Dahl (ed.), *Tense and aspect in the languages of Europe*, 403–440. Berlin: Mouton de Gruyter.
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA: Wiley.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practices. *Language Variation and Change* 24. 135–178.
- Terán, Virginia & Matthew Kanwitt. 2018. Variable past-time expression across multiple tasks in Tucumán, Argentina. *Revista Española de Lingüística Aplicada* 31(2). 605–637.
- Torres Cacoullos, Rena. 2011. Grammaticalization and neutralization-in-discourse hypothesis. In Manuel Díaz-Campos (ed.), *The Handbook of Hispanic Sociolinguistics*, 148–167. Oxford/ Boston: Wiley-Blackwell.
- Veiga Rodríguez, Alexandre. 2014. Diacronía de “he cantado”/“canté” en el sistema verbal español. Subsistema y variantes. In José Luis Ramírez Luengo & Eva Patricia Velásquez Upegui (eds.), *La historia del español hoy*, 151–179. Lugo: Axac.
- Westmoreland, Maurice. 1988. The distribution and the use of the present perfect and the past perfect forms in American Spanish. *Hispania* 71(2). 379–384.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/cllt-2022-0060>).