

Present Perfect and Preterit Variation in the Spanish of Lima and Mexico City

by

Anna Mastrantuono, M.A.

A Thesis

In

Hispanic Linguistics

Submitted to the Graduate Faculty  
of Texas Tech University in  
Partial Fulfillment of  
the Requirements for  
the Degree of

Master of Arts

Approved

Dr. Brendan Regan  
Chair of the Committee

Dr. Idoia Elola

Dr. James Lee

Mark Sheridan  
Dean of the Graduate School

August, 2021

Copyright 2021, Anna Mastrantuono

## ACKNOWLEDGMENTS

These have been two very intense years, in which many things occurred, including some that no one was expecting. I would like to thank people who accompanied me in my way toward graduating at TTU.

Firstly, I am grateful to my advisor Dr. Brendan Regan for his extremely valuable advice, continuous encouragement, and psychological support. His commitment to students and ability to transmit his passion for sociolinguistic research definitely impacted my personal and professional growth.

Moreover, I would like to thank my professors in the Department of Classical & Modern Languages & Literatures (CMLL): Dr. Idoia Elola, Dr. Sara Guengerich, Dr. James Lee, Dr. Alicia Luque, and Dr. Raychel Vasseur. Their passion, creativity, and willingness to share their expertise allowed me to conclude this two-year program having gained important amounts of new knowledge and skills. I may say that they opened a completely new world to me.

Furthermore, I am grateful to my family (mother, father and brother) for their support in all these years.

And lastly, a special thanks goes to my life partner and best friend Davide, without whom I would not be here, writing these lines. Thank you for teaching me to always push beyond my limits and embrace new challenges.

## TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>v</b>
<b>LIST OF TABLES .....</b>	<b>vii</b>
<b>LIST OF FIGURES .....</b>	<b>viii</b>
<b>I. INTRODUCTION .....</b>	<b>1</b>
<b>II. BACKGROUND LITERATURE .....</b>	<b>3</b>
2.1 Cross-linguistic instability of the PP .....	3
2.2 A historical sketch of grammaticalization theory .....	4
2.3 Processes and hypotheses related to grammaticalization .....	6
2.4 Grammaticalization of the PP .....	8
2.5 Grammaticalization of the PP in Romance .....	10
2.6 Grammaticalization of the PP across Spanish varieties .....	13
2.6.1 Prototypical uses across Spanish dialects .....	13
2.6.2 PP in Peninsular Spanish .....	16
2.6.3 PP in American Spanish .....	19
2.6.3.1 <i>Mexican Spanish</i> .....	19
2.6.3.2 <i>Peruvian Spanish</i> .....	22
2.6.3.3 <i>Is the PP undergoing change in American Spanish?</i> .....	27
2.7 Research questions .....	29
<b>III. METHODOLOGY .....</b>	<b>33</b>
3.1 Materials and participants .....	33
3.2 Procedure .....	34
3.3 Variables .....	36
3.3.1 Dependent variable .....	36
3.3.2 Independent variables .....	38
3.3.2.1 <i>Social variables</i> .....	38
3.3.2.2 <i>Linguistic variables</i> .....	43
3.4 Statistical analysis .....	47
<b>IV. RESULTS .....</b>	<b>49</b>
4.1 Mexico City .....	49
4.2 Lima .....	53

4.3 A closer look at the age effect in Lima .....	58
4.4 PP in Lima and Mexico City: similarities and differences .....	64
4.5 Excluded tokens .....	67
<b>V. DISCUSSION .....</b>	<b>69</b>
5.1 Research questions .....	69
5.1.1 RQ 1. What is the overall frequency of PP and preterit in Lima and Mexico City? .....	69
5.1.2 RQ 2. Which linguistic and extra-linguistic factors govern the usage of the PP in Lima and Mexico City? .....	70
5.1.3 RQ 3. Is there any evidence of change in progress in the usage of the PP in Lima and Mexico City? .....	74
5.2 Theoretical implications .....	76
5.3 Pedagogical implications .....	78
5.4 Limitations .....	78
5.5 Future directions .....	79
<b>VI. CONCLUSION .....</b>	<b>81</b>
<b>REFERENCES.....</b>	<b>83</b>
<b>APPENDICES: PARTICIPANT DEMOGRAPHIC DATA .....</b>	<b>89</b>
<b>APPENDIX A .....</b>	<b>90</b>
<b>APPENDIX B .....</b>	<b>92</b>

## ABSTRACT

In many Romance languages the present perfect (e.g., *I have gone*) tends to undergo a process of grammaticalization, gradually acquiring perfective functions and displacing the preterit (e.g., *I went*). In Spanish, the grammaticalization of the present perfect (henceforth, PP) as a perfective past has been observed in many Peninsular varieties. By contrast, American Spanish varieties favor the preterit, although some dialects, as Andean Spanish, present a higher frequency of use of the PP, possibly as a result of the contact with Quechua. Overall, there is a scarcity of quantitative variationist studies that investigate the distribution of the PP and the preterit in Spanish American dialects. This study aims to fill this gap by examining PP and preterit variation in Lima (Peru) and Mexico City (Mexico).

First, I examined the overall frequency of the PP in each variety. Afterwards, I examined a set of linguistic and social variables in order to determine whether they influence the PP and preterit distribution. I extracted the data from 36 semi-structured sociolinguistic interviews (18 Lima + 18 Mexico City) in the PRESEEA corpus (Project for the Sociolinguistic Study of Spanish in Spain and America, 2014-). I collected 2,457 tokens (1,230 México City; 1,227 Lima) of verbs in PP and preterit, which represent the binary dependent variable. I included four linguistic independent variables (temporal reference, temporal adverbial, presence of the adverb *ya*, and presence of direct object) and three social independent variables (gender, age group, and education level of the participants). I analyzed the data with a mixed effects logistic regression in Rbrul (Johnson, 2009) in the statistical package R (R Core Team, 2020) with linguistic and social variables as fixed factors and participant as random factor. Separate regression models were conducted for each speech community and then results were compared in order to best examine the social and linguistic factors that were most important in each speech community.

Results show that the PP occurs in the 24.7% of the tokens in Lima, and in the 9% in México City. In both dialects, the linguistic contexts that more significantly favor the PP are the irrelevant temporal reference (59.7% in Lima, 37.2% in Mexico), and the presence of an atelic adverb (38% in Lima, 22.3% in Mexico City). As for

social factors, the age group is significant in Lima, where the oldest group favor the use of the PP (34.2%), whereas the youngest group disfavor it (16.9%). There were no significant effects of social variables in Mexico. Such results suggest that, despite the difference in overall frequency, similar linguistic constraints rule the distribution of the PP and the preterit in these two varieties of Spanish, although Lima Spanish also presents innovative uses of the PP. However, whereas there is stable variation in Mexico, in Lima there are signs of ongoing semantic change through the expansion of the preterit observed through a change in *apparent time* (Labov, 1994, 2001). The possibility that the preterit is expanding in Peruvian Spanish is discussed in the final part of this work.

## LIST OF TABLES

1. Participants information in each speech community based on the PRESEEA norms.....	34
2. Summary of Rbrul step-up/step-down mixed effects logistic regression of PP/preterit realizations in Mexico City, with speaker as a random factor, in reference to PP realizations; $N= 1,230$ ; $R^2$ fixed= 0.978, $R^2$ random= 0.002, $R^2$ total = 0.98. ....	50
3. Summary of Rbrul step-up/step-down mixed effects logistic regression of PP/preterit realizations in Lima, with speaker as a random factor, in reference to PP realizations; $N = 1227$ ; $R^2$ fixed = 0.23, $R^2$ random = 0.076, $R^2$ total = 0.306. ....	54
4. Present perfect usage by participant in the oldest generational group in Lima .....	58
5. Present perfect usage by participant in the intermediate generational group in Lima. ....	59
6. Present perfect usage by participant in the youngest generational group in Lima...	60
7. Present perfect usage by age group and linguistic context in Lima.....	61
8. Factor weights in reference to PP realizations in Mexico City and Lima.....	65
9. Participant demographic data for Lima (PRESEEA).....	90
10. Participant demographic data for Mexico City (PRESEEA).....	92



## LIST OF FIGURES

1. Overall frequency of PP and preterit in Mexico City. ....	49
2. Main effect of temporal reference on PP and preterit use in Mexico City. ....	51
3. Main effect of temporal adverbial on PP and preterit use in Mexico City. ....	52
4. Main effect of direct object on PP and preterit use in Mexico City.....	52
5. Overall frequency of PP and preterit in Lima. ....	53
6. Main effect of temporal reference on PP and preterit use in Lima. ....	55
7. Main effect of the adverb “ya” on PP and preterit use in Lima. ....	55
8. Main effect of temporal adverbial on PP and preterit use in Lima. ....	56
9. Main effect of age group on PP and preterit use in Lima (bar chart).....	57
10. Main effect of age group on PP and preterit use in Lima (line chart).....	57
11. Main effect of age on PP use in Lima (note the outlier). ....	60
12. Overall frequency of PP and preterit per speech community. ....	64
13. Main effect of temporal reference on PP usage in each speech community.....	66
14. Main effect of temporal adverbial on PP use in each speech community. ....	67

## CHAPTER 1

### INTRODUCTION

Many languages in the world present the alternation between present perfect (e.g., *I have gone*) and preterit (e.g., *I went*) (Bybee & Dahl, 1989, p. 67). 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 (Bybee et al., 1994, p. 84; Squartini & Bertinetto, 2000). This process is known as grammaticalization (Bybee et al., 1994; Hopper & Traugott, 2003; Schwenter, 1994). In the context of Romance languages, the PP was originally employed to express resultative states of past actions, but in many varieties, it developed new semantic functions, gradually displacing the preterit (*aoristic drift*) in the expression of past actions and states (Squartini & Bertinetto, 2000).

The instability in the use of PP across Spanish dialects has drawn the attention of several scholars. PP is at an advanced stage of grammaticalization in some peninsular varieties, for example, Madrid (Serrano, 1994; Howe, 2006), Alicante (Schwenter, 1994a), Valencia (Howe, 2006) and Bilbao (Burgo, 2012). However, most American varieties of Spanish do not show the same tendency and use the PP infrequently (Fløgstad, 2016; Schwenter & Torres Cacoullós, 2008; Westmoreland, 1988; among others). An exception is Andean Spanish, in which, although preterit is the most frequent form, the PP is used in everyday speech (Escobar, 1997; García Tesoro & Jang, 2018; Jara Yupanqui, 2006, 2011). Previous research showed that the PP is undergoing different paths of grammaticalization in Peru and Spain, as different linguistic constraints condition its distribution (Howe & Schwenter, 2008, p. 107), and that the linguistic contact with Quechua may account for non-prototypical functions of Peruvian PP (Escobar, 1997; Klee & Ocampo, 1995).

Overall, there is a lack of studies that analyze the distribution of the PP and the preterit in American Spanish from a quantitative variationist perspective. The present study aims to fill this gap by examining the factors that condition PP and preterit variability in two Hispanic 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, 1998).

This work is organized as follows: in chapter 2, I will discuss the background literature. I will start by introducing the fundamental concepts related to the theory of grammaticalization, and how they apply to the grammaticalization of the PP in Romance languages in general, and in Peninsular and American Spanish in particular. In chapter 3, I will describe the methodology of the study. Chapter 4 presents the results, which are further discussed in chapter 5 along with the limitations of this study and the venues for future research. Finally, the conclusion is presented in chapter 6.

## CHAPTER II

### BACKGROUND LITERATURE

#### 2.1 CROSS-LINGUISTIC INSTABILITY OF THE PP

The present perfect is an unstable grammatical category, as its functions and domains of use tend to vary not only across languages, but also within a same language throughout its diachronic evolution (Harris, 1982, p. 44). Cross-linguistically, it is often the case that the functions of the PP overlap, at least partially, with those of the preterit. The overlap may be restricted to some contexts, but the PP may also go as far as substituting the preterit in the expression of past events in all contexts (Dahl, 2000; Schwenter, 1994a; Squartini & Bertinetto, 2000; among others). This phenomenon is considered revealing of more general processes of semantic change in modern languages. Specifically, the evolution of the PP from a resultative construction to a more widely used marker of past events is considered a manifestation of grammaticalization.

Grammaticalization is defined as the process through which lexical forms gradually develop grammatical functions; similarly, grammatical items can also develop new grammatical functions in a set of linguistic contexts over time (Bybee et al., 1994, p. 4; Hopper & Traugott, 2003, p. 1; Lehmann, 2015 [1982], p. 13). A classic example is the English periphrasis ‘be going’. Originally, the verb ‘to go’ was used literally, meaning that the subject is moving through space toward a goal (Bybee et al. 1994, p. 5), but the construction gradually acquired the additional meaning of future temporality. The two meanings, the spatial and the temporal, coexisted for some time (a process known as *layering*), but eventually, the former meaning demised (*semantic bleaching*). At this point, ‘be going’ is a fully grammaticalized form which expresses future. A similar path of semantic development of verbs expressing ‘an agent moving on a path toward a goal’ from spatial to temporal has been attested in several languages (Bybee et al. 1994, p. 268).

The evidence concerning cross-linguistic patterns of evolution of the grammatical systems of unrelated languages is of interest for linguistic research as it

may help shed light onto general diachronic processes of evolution of languages. After all, the final goal of grammaticalization theory is to explain how grammar systems emerged out of lexical words in human languages, and how such systems evolved over time. The attention toward grammaticalization in linguistics arose especially in the last forty years, however, the theory of grammaticalization is rooted much earlier in the context of Indo-European studies. Before turning our attention to the grammaticalization of the PP, let me provide a brief overview of the history of grammaticalization theory, and the main processes and hypotheses related to it.

## **2.2 A HISTORICAL SKETCH OF GRAMMATICALIZATION THEORY**

The idea of grammar being created from lexical, concrete items that gradually acquire formal, grammatical meaning can be drawn back to the German philosopher Humboldt (1767-1835; Hopper & Traugott, 2003, pp. 19-20). However, the first to name this phenomenon as ‘grammaticalization’ was the French Indo-Europeanist Antoine Meillet in his article *L'Évolution des Formes Grammaticales* (1912).

Here, Meillet asserted that new grammatical forms arise through two main processes: analogical innovation, and the attribution of grammatical character to a previously autonomous word. Analogy, that is, the extension of a morphological paradigm to a linguistic element in conformity with another element perceived as a model, was a well-known mechanism at the time, from which Neogrammarians drew heavily to explain morphological change in modern languages. An example is the plural form *shoes*, which replaced the ancient form *shoen* for analogy with other plurals as *stone-stones* (Hopper & Traugott, 2003, p. 22). As for the transformation of lexical words into grammatical morphemes, Meillet highlighted that, unlike analogy, this process had been largely ignored by linguists, but there was a need to reconsider it, given its potential for explaining the creation of new grammatical categories which did not have linguistic expression in previous stages of language, leading to a transformation of the whole language system (Meillet, 1958 [1912], p. 133).

In subsequent years, the study of grammaticalization was limited to Indo-European studies. This reflects the spread of structuralism, in which it was assumed

that the mechanisms behind linguistic change are unpredictable and driven by system external factors (Lehmann, 2015 [1982], p. 6). Among the main exponents of Indo-European studies dealing with grammaticalization there are Benveniste (a student of Meillet), Kuryłowicz, and particularly Givón, who is considered a precursor of modern approaches to grammaticalization. In his work<sup>1</sup>, the author investigates grammaticalization processes (which he calls *syntacticization* and *morphologization*) in several African languages verbal systems and in the evolution from pidgins to creoles.

The resurgence of interest toward grammaticalization takes place in the last two decades of the twentieth century. This reflects, on the one hand, the abandonment of the structuralist framework, and a new approach to synchronic and diachronic dimensions of language change, conceived as inseparable. On the other hand, the establishment of new technologies allowed for the collection of linguistic data in large computerized corpora and the implementation of extended cross-linguistic projects. Linguistic research started to draw on quantitative analysis of data from historical and synchronic corpora, in order to statistically identify cross-linguistic patterns in grammatical systems. Two seminal works in this field are those of Bybee (1985) and Dahl (1985).

Bybee (1985) examined verbal morphology in fifty languages, controlling for genetic and areal bias, and tested the cross-linguistic comparability of grammatical morphemes (which she calls “grams”) meaning. Her data are drawn from a corpus of reference grammars of multiple languages created by Perkins (1980). Similarly, Dahl (1985) explored tense-aspect-mood relationships by administering a grammatical judgement questionnaire to native speakers of 64 languages, mostly European. Despite the different methodologies, they found similar results, which were later summarized in a joint article (Bybee & Dahl, 1989). They found that there is a set of “gram” classes (or, in other words, grammatical morphemes marking tense and aspect) which recur more often cross-linguistically. These are perfective, imperfective, progressive,

---

<sup>1</sup> For example: *Historical syntax and synchronic morphology: an archaeologist's field trip* (1971); *On Understanding Grammar* (1979).

future, past, and perfect morphemes. The main implication is that it is possible to make cross-linguistic generalizations about the semantic content of grammatical morphemes of tense and aspect, which in turn is interpreted as evidence that the languages of the world present form/ meaning correlations (Bybee & Dahl, 1989, p. 56). These regularities of development of grammatical morphemes across otherwise not comparable languages were further explored in the book *The Evolution of Grammar* by Bybee, Perkins, and Pagliuca (1994). Here, the authors identify grammaticalization (which they call “grammaticization”) as a basic mechanism by which semantic substance is molded into grammatical meaning cross-linguistically (Bybee et al., 1994, p. 3). In the same years, the European Science Foundation sponsored the EUROTYP (Typology of Languages in Europe) Project (1990-1994), which aimed to investigate the typological variation in European languages with respect to nine domains of language, including tense and aspect, focusing on “the study of grammaticalization processes as manifested in European languages, and on the identification, description and explanation of tendencies peculiar to the tense-aspect systems in European languages” (Dahl, 2000, p. 3). These lines of research provide a theoretical framework to explore the functions of the PP as a semantic category which commonly shifts in its functions not only across languages, but also within the varieties of a same language.

### **2.3 PROCESSES AND HYPOTHESES RELATED TO GRAMMATICALIZATION**

Grammaticalization is recognized for being a slow process that can reach different generations (Bybee et al., 1994; Hopper & Traugott, 2003). A grammaticalizing form overgoes several stages of development, as we have seen with the periphrasis ‘be going’. First, this structure gradually acquired the new temporal meaning. This phenomenon is known as “semantic generalization”, as there is an extension of the contexts in which it is possible to use the structure. This process is understood as being essentially metaphoric, as the two meanings present a metaphorical relation from more concrete (the original meaning) to more abstract (the new, grammaticalized meaning; Bybee et al., 1994, p. 283). A corollary of semantic generalization is “layering”, or “semantic retention”. For some time, ‘be going’

maintained its earlier meaning, so that the temporal and the spatial meanings coexisted. Eventually, the spatial meaning demised (a process called “semantic bleaching”, “semantic reduction”, or “erosion”), and ‘be going’ became a fully grammaticalized form which expresses future (Bybee et al., 1994, p. 6).

The observation that there is a group of verbs which tend to undergo similar processes of semantic change cross-linguistically 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. For example, it has been observed that verbs expressing general meanings as ‘go’, ‘come’, ‘do’, ‘be’, and ‘have’, represent the main lexical sources for grammar morphemes cross-linguistically (Bybee et al., 1994, p. 9). Moreover, verbs expressing movement present similar patterns of meaning shift from spatial to future temporality in a set of unrelated languages. In other words, the source verb (or, more precisely, its semantic class) determines “the path that grammaticalization follows and, consequently, the resulting grammatical meanings” (Bybee et al., 1994, p. 9).

Another hypothesis based on empirical data is that of unidirectionality, according to which once a grammaticalization process starts, it cannot be reversed. In other words, elements that grammaticalize do not go back to their lexical independent meaning (Bybee et al., 1994, p. 14). This hypothesis is based on empirical observations, but it has been object of debate (see for example Fløgstad, 2017; Giacalone Ramat, 1998; González et al., 2019; Janda, 2001). However, the evidence for unidirectionality is systematic cross-linguistically, whereas evidence of counterexamples has been considered by many scholars too sporadic and language-specific to be significant (Hopper & Traugott, 2003, p. 138). Nevertheless, Hopper and Traugott (2003) call for caution in making assumptions about unidirectionality in the reconstruction of diachronic changes, especially when we lack historical data and therefore must reconstruct the process of linguistic change inferentially (Hopper & Traugott, 2003, p. 139).

Taken together, the hypotheses of source determination and unidirectionality predict that there are cross-linguistic clines, or universal paths, of semantic change



(Bybee et al., 1994, p. 15). Although there exist language-specific instances of grammaticalization, “given that the source material that enters into grammaticalization is similar cross-linguistically, it predicts cross-linguistic similarity in paths of development” (Bybee et al., 1994, p. 15). In the following section, I will provide an overview of how the theory of grammaticalization has been applied to the PP, first across world languages, and then in Romance languages.

## **2.4 GRAMMATICALIZATION OF THE PP**

According to Bybee and Dahl (1989), the PP is present in 25 to 35 per cent of the languages in their samples (p. 67), including most European languages and other unrelated languages. Moreover, the PP is often expressed periphrastically across languages, with an auxiliary verb (generally ‘be’ or ‘have’) and a participle, and in most languages, it evolved from resultative constructions (Bybee & Dahl, 1989, pp. 71-72). Resultative constructions describe a present state resulting from a past event, and were present in both Latin and Old English (Bybee et al., 1994, p. 68). In resultative constructions, the participle carries an adjectival function, whereas the verb (generally ‘to have’ or ‘to be’) has a literal meaning. In Old English, ‘to be’ was used when the adjective was referred to the subject of the sentence, and ‘to have’ when it referred to the object, as in example 1:

- (1) Ic haefde hine gebundenne (*I had him in a state of being tied*; Bybee et al., 1994, p. 68).

Resultative constructions were initially limited to a specific semantic class of verb, namely, change-of-state verbs. As we have seen, an important mechanism of semantic change is the generalization of meaning, which implies the expansion to new classes of verbs. Perfects developed out of resultative constructions when the adjective lost its original value to become part of the verb, whereas the verb lost its semantic meaning to become an auxiliary (Bybee et al., 1994, p. 68). According to Bybee and Dahl (1989), there are three cross-linguistic tendencies of grammaticalization of the perfect: the development of evidential functions, the development as a past perfective marker function, and the evolution of perfect categories to express remoteness

distinctions (Bybee & Dahl, 1989, p. 73). The evidential function marks the distinction between an event directly witnessed by the speaker, and an event reported to, or inferred by, the speaker. This evolution is motivated by the semantical closeness of the resultative meaning and the evidential meaning of inference, as the former indicates a state that occurs due to past actions, while the latter indicates that a past action is known or inferred from results (Bybee et al., 1994, p. 96). The development of perfects into markers of indirect evidence has been reported in a number of languages, as Turkish, Bulgarian, Macedonian, Udmurt (west-central Russia), Inuit (Greenland), and Tucano (Colombia), among others (Bybee et al., 1994, p. 95). In the present study I will not focus specifically on this function of the PP, although I will mention it again when treating Peruvian Spanish, as some studies ascribe non-prototypical uses of the PP in Peruvian Spanish to the transfer of evidential functions which arose from the linguistic contact with Quechua (Escobar, 1994; Klee & Ocampo, 1995).

Likewise, the analysis of remoteness distinctions through the evolution of perfect categories is above the scope of this work. I will focus my analysis on the development of past perfective functions, which leads the PP and the preterit to overlap in some linguistic contexts. This tendency has been attested cross-linguistically in very different language families, such as Mandarin Chinese, German languages (English, German and Dutch), Romance languages (French, standard Italian and standard Romanian), and African languages of the Kru and Bantu group (Bybee et al., 1994, p. 81). However, the amount and pattern of the overlap between PP and preterit may vary greatly across languages. As for Romance, Squartini and Bertinetto (2000) claim that the evolution from resultative to perfect constructions, and the subsequent creation of a series of compound tenses for analogy with the PP, represent “the point where a comprehensive story of the Romance Compound Past (and, concomitantly, of the simple past) becomes impossible, for its evolution is different in each language” (p. 405). As I shall show in the next section, the PP differs in its functions and contexts of use across Romance languages, and even within regional varieties of the same language. Spanish does not represent an exception to this tendency.

## 2.5 GRAMMATICALIZATION OF THE PP IN ROMANCE

In Classical Latin, the past form (e.g., *FECI*, ‘I did/ have done’) encoded both the perfective and the “past with current relevance” functions. However, a resultative structure which gave rise to the PP paradigm is already attested in the Pre-Classical period, as in example 2:

- (2) *Multa bona bene parta habemus* (Plautus, *Trin.* 347) (‘*Many goods well obtained we have*’: we possess many well obtained goods)<sup>2</sup>.

In the resultative construction, the verb ‘to have’ (*habemus*) maintains its lexical meaning of possession, and the past participle (*parta*) is a complement of the direct object (*multa bona*), with whom it agrees in gender and number. In modern Spanish, the resultative construction is expressed by means of *tener* + past participle with adjectival function, as in example 3:

- (3) *Tengo la tarea hecha* (‘*I have the homework done*’).

Moreover, Classical Latin presented a periphrastic structure with cognition verbs, as in *cognitum habeo* ‘I (have) learned/ I know well’, which may also have served as a basis for the development of the present perfect (Squartini & Bertinetto, 2000, p. 405). Gradually, the resultative structure underwent a process of reanalysis and was extended to new contexts. Initially it admitted only telic verbs, but in Vulgar Latin it came to be used with any kind of verb. The past participle gradually lost its adjectival function and became part of the verb, which in turn lost its lexical meaning and became an auxiliary. At this point, the foreground meaning of the resultative construction (a present state resulting from a past event) shifted to that of a marker of past action which is considered relevant to the present moment (Bybee et al., 1994, p. 69). Scholars agree in that the indeterminacy of the concept of “current relevance”, which is not objective or absolute in any sense, brought to variable uses of the PP in Romance (Bosque & Demonte, 1999, p. 2945; Harris, 1982, p. 45). A famous account

---

<sup>2</sup> The example is reported in Squartini, M., & Bertinetto, P. M. (2000).

of the distribution of the PP in Romance languages is provided by Harris (1982), who described the PP as being organized in four stages of grammaticalization. Given the importance of this classification in studies about PP grammaticalization in Romance languages, I will report them here.

At stage I, the PP has a purely resultative function: it expresses a present state resulting from past actions. In modern Spanish, as we have seen, the PP has lost this semantic function, which is expressed by means of *tener* + past participle. Harris presents some vernaculars of Southern Italy (Sicilian and Calabrian) as being at stage I, and cites examples in Calabrian vernacular taken from Rohlfs (1966-1969), which I report in examples 4 and 5:

(4) Non m'a scrivutu (*'he hasn't written to me'*, so I don't have news of him).

(5) L'aju fattu (*'I've done it'*, sometime in my life, it is part of my experience; Harris, 1982, p. 50).

However, Squartini and Bertinetto (2000) claim that no contemporary Romance language exhibits a PP with a purely resultative function, and provide examples of uses of perfect in Sicilian and Calabrian that are compatible with the second stage. Moreover, they highlight the similarity of PP functions in Sicilian dialect and Mexican Spanish (p.407).

At stage II, the PP presents an aspectual function of durativity or iteration. It refers to states and events that started in the past and continue up to the speech time. It is often accompanied by adverbs such as *siempre* 'always', *nunca* 'never', *muchas veces* 'many times', *con frecuencia* 'often', as in example 6:

(6) Siempre me ha gustado vivir aquí. (*'I always liked [:PP] living here'*).

This stage is representative of the PP uses in Spanish American varieties, Galician and Portuguese (Harris, 1982, p. 50).

At stage III, the PP expands to an array of contexts that denote perfective aspect, namely, an action or situation presented as finished and completed in the past. This is

what Squartini and Bertinetto (2000) call the *aoristic drift* of the PP (p. 404). Typically, the PP gradually expands to very recent past contexts because these are more likely to be perceived as “currently relevant”. This leads to the bleaching of the meaning component related to the current relevance of the action, and conversely the semantic connection of the PP with concluded events is reinforced (Bybee et al., 1994, p. 87; Schwenter, 1994b). Many Peninsular Spanish dialects are at this stage, where the PP is the preferred form for expressing past actions that occurred in the same day of the speech as seen in example 7, or in a moment that is considered connected to the moment of speech. The connection with the moment of speech is often expressed by means of temporal adverbials that mark an interval of time which includes the present moment, as in example 8:

(7) Esta mañana he madrugado. (*I woke up [:PP] early this morning*’).

(8) Este año no hemos viajado. (*We haven’t travelled [:PP] this year*’).

At stage IV, the PP expresses past actions regardless of the distance with the moment of speech, and the preterit is restricted to formal and written registers. This is the situation of Standard French, Standard Romanian, and Standard Italian. Several scholars noted that European Spanish speakers show increasing acceptance of this use of the PP, although its frequency in pre-hodiernal temporal contexts is still very low (Howe, 2013; Serrano, 1994). Pre-hodiernal temporal contexts refer to events and situations which occurred at any point in time before the day of the speech, whereas hodiernal temporal contexts refer to events and situations which occurred the same day in which they are being recounted. Schwenter (1994a, pp. 46-47) claims that once the PP has established as an hodiernal perfective (namely, when it marks hodiernal past reference), the tendency will be to further extend to more remote (pre-hodiernal) temporal contexts, and predicts that the PP in Alicante Spanish is moving from the third stage of grammaticalization toward the fourth stage.

It is worth mentioning that a common interpretation of the classification by Harris puts the varieties at stages III and IV (as Peninsular Spanish) at a “more developed” stage of evolution, whereas those at earlier stages (as American Spanish) will eventually reach the more developed ones. As pointed out by some scholars, this

interpretation is misleading. Overall, it is difficult to predict the evolution of linguistic systems as many different factors (including educational policies and mass media) interact. Squartini & Bertinetto (2000) proposed to interpret Harris' stages as independent developmental paths of PP grammaticalization. Therefore, a variety which is at stage II will not necessarily move toward stages III and IV. Similarly, Schwenter (1994a) claims that PP forms can be “frozen in their development” and that synchronic investigation is “necessary in order to determine their ongoing development, if any, and to advance prognoses of future grammaticalization” (p. 78). In fact, the studies by Schwenter (1994a, 1994b) offer an insight into ongoing processes of PP grammaticalization in Spanish, and in particular Peninsular varieties. In the next sub-section, I will provide an overview of the studies concerning PP grammaticalization in different varieties of Spanish.

## **2.6 GRAMMATICALIZATION OF THE PP ACROSS SPANISH VARIETIES**

### **2.6.1 Prototypical uses across Spanish dialects**

In order to examine the prototypical uses of the PP, and the patterns of its semantic overlap with the preterit across Spanish varieties, I made use of two reference grammars: *Manual de la Nueva Gramática Española* issued by the Royal Spanish Academy (2010) and *Gramática Descriptiva de la Lengua Española* by Bosque and Demonte (1999).

According to *Manual de la Nueva Gramática Española*, the opposition between the preterit and the PP reflects the opposition of an absolute and a relative tense (2010, p. 438). An *absolute* tense marks an action or state that started and concluded at some point in the past, for its own sake and with no reference to the present moment (namely, it has a perfective aspectual function). Conversely, a *relative* tense marks an action or state, either punctual or durative, that took place in a moment in the past which extends to, and includes, the present moment (notion of *extended now*, see Howe, 2013, p. 138). This value of the PP gives rise to two different interpretations. The first is that of “open perfect” (also called *antepresente*), which marks a situation continuing up to the present, as in example 9, and also in example 6 presented above.

Conversely, example 10 represents an instance of the second interpretation, where the PP marks past events and, therefore, carries a perfective (or aoristic) value:

- (9) María no ha llegado (*'Maria has not arrived'*).  
(10) El matrimonio ha sido el diez de agosto (*'The wedding was [:PP] the 10th of August'*).<sup>3</sup>

According to the authors, the perfective interpretation shown in example 10 is present in Central and Southern Spain, in the Peruvian coastal area, in the Andean area of Bolivia and Colombia, in North-Eastern and Central Argentina, and, with more restrictions, in Cuba and the Antilles. In the remaining Spanish speaking areas, the PP is limited to the “open perfect” interpretation (p.438) as presented in example 9.

Moreover, *Nueva Gramática Española* presents four main functions of the Spanish PP (pp. 440-441), summarized here:

*Experiential perfect*: it describes an event or situation that occurred one or more times in a certain period of time, which may be expressed through adverbials as *últimamente* ‘lately’, *en estos tiempos* ‘in these times’, *en estos días* ‘in these days’, or remain undefined. In the latter case, the PP may be accompanied by adverbials as *alguna vez* ‘ever, sometime’, *en alguna ocasión* ‘sometime’, *X veces* ‘X times’ (example 11):

- (11) He hablado con él tres veces (*'I spoke with him three times'*).<sup>4</sup>

*Continuative perfect*: it refers to a past situation which extends to the present. It is expressed generally by means of atelic verbs, as in example 12. Telic verbs appear in negative contexts (example 13):

- (12) Siempre he vivido aquí (*'I have always lived here'*).  
(13) María no ha llegado (*'Maria has not arrived'*).

---

<sup>3</sup> Example 9 is taken from *Manual de la Nueva Gramática Española*; example 10 is taken from the interview LIMA-M31-009 in PRESEEA. Each PRESEEA interview is identified as follows: first, there is a code indicating the city where the interview was conducted (“LIMA” or “MEXI”), followed by “M” (male) or “F”(female), and a numerical code.

<sup>4</sup> Examples 11 through 16 are taken from *Manual de la Nueva Gramática Española*.

*Perfect of recent past (and recent news perfect)*: it allows to report an event that occurred in a temporal interval that includes the moment of speech. It is used for recent events, often without temporal specification (example 14):

(14) He comido con Luis (*'I ate [:PP] with Luis'*).

The temporal frame may be specified by adverbials expressing recentness, as: *hoy* 'today', *este verano* 'this summer', *este año* 'this year', *hace un momento* 'a moment ago'. This use is present only in some varieties of Spanish (European, Peruvian, Bolivian and the others mentioned above). A sub-type of the perfect of recent past is the "recent news" perfect, which is used in newspapers to mention a recent event for the first time (in contrast, the preterit is used to recount the details of the event development, after the first mention to it has been made).

*Resultative perfect*: it allows to infer a current state resulting from a past action, as in examples 15 and 16:

(15) El jarrón se ha roto (*'the vase broke [:PP] down'*, implying that: *está roto ahora*, 'it is broken now').

(16) Ustedes me han decepcionado (*'You disappointed [:PP] me'*, implying that: *estoy decepcionado ahora*, 'Now I am disappointed').

This classification closely resembles the one proposed by Comrie (1976) for cross-linguistic prototypical functions of the PP, which are: perfect of result, experiential perfect, perfect of persistent situation, and perfect of recent past (p.56).

Furthermore, *Nueva Gramática Española* makes the point that in many Hispanic American countries, the preterit is admitted in all the above contexts instead of the PP (p. 443), and that in fact most Spanish American varieties limit their use of the PP to continuative contexts. This is consistent with the claim of Harris (1982) about Spanish American varieties being at the second stage of development of the PP. In *Nueva Gramática Española* it is also mentioned that River Plate Spanish differentiates from the American norm, as it presents alternation between the preterit and the PP in continuative contexts (see also Rodríguez Louro, 2009, who claims that continuative contexts are expressed by means of preterit in Buenos Aires).



According to Bosque and Demonte (1999), although in literary Spanish the functions of the PP and the preterit are clearly differentiated, the blurry nature of the concept of “anteriority within the present frame”<sup>5</sup> brought to variability in its oral uses (p. 2945), which reflects “parallel and distinct evolutions of the same heritage” (p. 2951). In Spain the situation is relatively homogeneous, and the use of the PP resemble the norms of literary language. However, the contact varieties of Spanish in Galicia (under the influence of Galician-Portuguese), Asturias and León (under the influence of Asturian) diverge from the literary norm and favor the preterit. On the contrary, in Madrid vernacular language, the compound form tends to overlap with prototypical preterit functions. As for American varieties, here the PP maintains the functions characteristic of pre-classic Spanish: it is used in durative and iterative contexts, describing events or states which hold at the moment of speech. Events that occurred and concluded prior to the moment of speech are expressed by means of preterit, regardless of the temporal distance with the moment of speech. The same occurs in Canary Islands variety. However, research suggests that the distribution of the preterit and the PP is unstable in Canary Islands, where it seems to be converging toward the main Peninsular norm, and in the American Southern Cone area, where the preterit is expanding (p. 2948).

### **2.6.2 PP in Peninsular Spanish**

The functions and frequency of use of the PP have been examined in Peninsular dialects both from a diachronic and a synchronic perspective. Copple (2011) compared the distribution of PP and preterit diachronically by analyzing Spanish dramatic texts from XV, XVII and XIX centuries. She found that, across centuries, the PP expanded its contexts of use, appearing in 26% of the XV century data, and in 49% of the XIX century data. In the XV century the PP appeared mainly in temporally irrelevant contexts, which express durative and iterative situations, and in very recent past contexts. In this latter case, the PP appeared primarily with a restricted semantic class of verbs associated with the original resultative function of the PP: verbs of

---

<sup>5</sup> In the original text, ‘Pasado dentro del ámbito de la actualidad del hablante’ (p. 2950).

perception, mental states, change of physical and emotional state, communication, motion and physical violence (Copple, 2011, p. 178). During the XVII century the PP extended to new semantic classes of verbs and increased its frequency, especially in very recent past contexts. This expansion favored the association of the PP with past perfective functions, which in turn allowed for an expansion in hodiernal contexts (that is to say, events that happened in the same day of the moment of speech) in the XIX century, although at this point in time, the preterit is still the most frequent form. This analysis resembles that of Schwenter (1994b), according to whom the “hot news” function (which is comparable to what Copple calls “very recent past”) plays a crucial role in the PP aoristic drift, as it favors the semantic bleaching of the “current relevance” function, which typically marks the PP, and favors the association of the PP with a purely perfective past. In other words, the fact that the PP is used to express very recent events leads to a gradual erosion of the relevance implications associated with the PP, promoting instead a perfective reading of its semantic functions (p.1000).

From a synchronic perspective, several studies examining the distribution of the PP in contemporary Peninsular Spanish found evidence of an advanced stage of grammaticalization. Schwenter (1994a) conducted sociolinguistic interviews in Alicante and found that the PP is the favored form in hodiernal past contexts, where youngest speakers (from 18 to 25 years old) used the PP nearly categorically (94%), whereas the 27% of the oldest group (40 years old and older) used the preterit in this context. Moreover, in pre-hodiernal contexts, although the preterit is the preferred form, the youngest age group uses the PP at a higher frequency than the oldest group (39% versus 17%). The same tendency was found by Serrano (1994) with oral interviews in Madrid. These results are interpreted as evidence of a change in progress toward the grammaticalization of the PP as a perfective past. Schwenter (1994a) claims that the PP in Alicante is following the same aoristic drift path as French, Romanian and Northern Italian, where the perfect has taken on all preterit functions in colloquial language, and the preterit is restricted to written language and formal speech (p. 100). Similar results were found by Schwenter and Torres Cacoullós (2008) in Madrid when comparing the frequency and distribution of PP and preterit in oral

data from the Spanish capital and Mexico City. The authors claim that in Madrid, the PP has become the default (namely, the most common and unmarked) form to express perfective past, as it appears in 54% of the data. In this variety, the PP has established as the only form expressing hodiernal past (p. 23), and also appears frequently (73%) in indeterminate temporal contexts, in which the temporal anchoring of the recounted event is not specified. The PP is also favored by the temporal adverb *ya*, a tendency that put this variety in opposition with the Mexican one, where the same adverb favors the preterit (although not significantly).

Serrano (1996) examined the alternation of PP and preterit in La Laguna, Tenerife (Canary Islands) in hodiernal and very recent contexts. In these contexts, in opposition to the Madrid norm, Canary Islands dialects typically use the preterit, as happens in American Spanish. Serrano found that the second generational group (35-54 years old) and the middle class follow the Madrid norm more closely and use the PP at a higher frequency as compared to other age groups. Moreover, individuals with medium and high level of education use the PP more often than those with a low level of education. The author interprets the results as a sign of the positive social evaluation toward the Madrid norm. Also, she explains the higher frequency of the PP in the second generational group with the fact that many of these individuals have jobs which include close contact with national tourism, so they have an interest in conforming to the norm of the majority. Regarding gender, although women from the upper-middle class showed the highest frequency of use of the PP, this difference was not significant in the statistical model. Gender was neither significant in the sample from Bilbao (Basque Country) explored by Burgo (2012). Here, the leaders in PP usage are the middle class (followed by the low social class) and the youngest generation (20-34 years old).

Taken together, the results of these studies suggest that PP functions are undergoing a change toward increased grammaticalization in the Peninsular varieties examined. Overall, it is predicted that, given the higher frequency of PP use among young generations and (when considered) the middle class, two social variables that are often revealing of linguistic change in progress (Labov, 1994), the PP will further

expand to past perfective functions in the Peninsular dialects examined, as happened in French and Standard Italian. However, the way through which this expansion would take place is object of debate. As we have seen, Schwenter (1994b) underlies the importance of the “hot news” function of the PP, as it allows for the gradual generalization of use of the PP in increasingly distant past reference contexts. Schwenter and Torres Cacoullos (2008) hypothesize that indeterminate reference contexts may be the locus of change, as they constitute an intermediate step before the PP reaches the status of both hodiernal and pre-hodiernal perfective past. On the other hand, Howe (2013) claims that the overall acceptance of the PP in pre-hodiernal contexts is so low in Peninsular data, that it is not possible to make predictions about a further expansion. What is clear is that the use of the PP in Peninsular dialects is much more extended than in American dialects, where it does not carry the hodiernal perfective function. I will now discuss research about the PP and preterit variation in Hispanic America.

### **2.6.3 PP in American Spanish**

#### *2.6.3.1 Mexican Spanish*

According to the classification of PP stages of development proposed by Harris (1982), most Spanish American dialects lay in the second stage, where the PP is “aspectually marked as durative or repetitive” (p. 50). Mexican Spanish is considered prototypical in this respect. Whether the PP in Mexican Spanish carries an imperfective aspectual value, which opposes it to the perfective aspectual value of the preterit, has been object of debate. Some influential accounts present the Mexican PP as carrying imperfective functions (Lope Blanch, 1972; Moreno de Alba, 1978; Westmoreland, 1988), an observation that arises from the fact that the continuative and iterative contexts in which the PP appears may well be interpreted as imperfective. Perfective aspect presents situations as temporally bounded and concluded, whereas imperfective aspect presents situations as not concluded, or open until the moment of speech. In this sense, example 17 (with the verb in preterit) means that Pedro is no

longer living, but example 18 (with the verb in PP) implies that Pedro is still studying at the moment of speech, and will continue to do so<sup>6</sup>:

(17) Pedro estudió toda su vida (*'Pedro studied [:Preterit] all his life'*).

(18) Pedro ha estudiado toda su vida (*'Pedro has been studying [:PP] all his life'*).

The idea that the PP carries an imperfective aspect function has been criticized by other scholars, e.g. Serrano (1996) and Kubarth (1992), according to whom the iterative and durative functions of American Spanish have a temporal, rather than aspectual, value, as showed by the fact that PP functions do not overlap with the imperfect tense, which indeed carries a real imperfective aspectual value. In any case, it is evident that the PP in most American varieties does not mark concluded actions, regardless of the temporal distance with the moment of speech. This is the main difference with respect to Peninsular varieties, in which the PP functions as an hodiernal perfective.

Schwenter and Torres Cacoulios (2008) collected tokens of verbs in PP and preterit from a corpus of oral interviews in Mexico City (2234 tokens), and Madrid (1783 tokens). In Mexico City, they found a frequency of 15% for the PP and 85% for the preterit. The authors examined a set of linguistic factors (temporal reference, temporal adverbial, presence of the adverb *ya*, Aktionsart verb class, clause type, quality of the direct object, polarity) and their impact on the selection of PP or preterit in the two cities. In Mexico City, the factors that resulted significant are the temporal reference, the temporal adverbial, the noun number (plurals favor the PP), the clause type (yes-no questions and relative clauses favor the PP), and the Aktionsart class of the verb (durative verbs slightly favor the PP, while punctual verbs disfavor it). The irrelevant temporal reference is the most significant factor favoring the PP. Irrelevant reference marks those situations in which it is not pertinent to ask when an event occurred for not being temporally bounded, and includes continuative and iterative contexts. Continuative contexts marked by atelic adverbials as *siempre* 'always',

---

<sup>6</sup> The examples are taken from Moreno de Alba (1978), p. 57.

*nunca* ‘never’, *desde* ‘since’ show the highest rate of PP use. Frequency adverbials also favor the PP, although the preterit is the most frequent form. Indeterminate temporal reference contexts, in which temporal anchoring is left unspecified by the speaker, also show a favoring effect for the PP. In Madrid only four factors are significant (temporal reference, temporal adverbial, noun number and presence of adverb *ya*). This is considered an evidence of a more advanced stage of grammaticalization in Madrid, given that the semantic generalization typical of grammaticalization processes involves loosen linguistic restriction in the use of the form, and consequently, the extension of the linguistic contexts in which the PP is acceptable (Bybee et al. 1994, p. 6).

Moreno de Alba (1998) explored the issue diachronically, examining the PP and preterit distribution in written correspondence from New Spain between the XVI and the XIX centuries. Results show that the PP frequency steadily decreased as compared to the preterit over time. In fact, the PP has a frequency of 39% in the XVI century, 26% in the XVII, 20% in the XVIII and 15% in the XIX. Recall that Copple (2011) found the opposed tendency in dramatic texts written in Spain between the XV and the XIX century. In the study by Moreno de Alba (1998), the PP initially appeared in three main contexts: first, in continuative contexts (events and states that continue to the moment of speech); second, with an open interpretation, so that a past event or state is perceived as connected to the present (the *extended now* interpretation), and it may happen again in the future; and third, in continuative- resultative contexts, when the effects of events that occurred in the past still hold true in the present, because the past events are recent or especially significative. It is especially in this last context, where the connection of the PP with the present is less straightforward, that the preterit gradually expanded at expenses of the PP. Therefore, it seems reasonable to affirm that, in American Spanish, a semantic development opposed to that registered in Peninsular Spanish has occurred, and may still be in process; such a phenomenon would implicate that not only the grammaticalization of the PP as a perfective past is possible, but also the opposite process (the preterit expanding to PP functions) may occur. This evidence may provide an argument to the opponents of the hypothesis of

unidirectionality of semantic change, as it shows that the PP initially held perfective functions that gradually came to be expressed by means of preterit. However, there is the need for more studies which examine the evolution in the functions of the PP and the preterit in Mexican dialect.

#### 2.6.3.2 *Peruvian Spanish*

Peruvian Spanish is considered an exception to the Spanish American tendency. In this variety, although the preterit is still the most frequent form, the PP is used in everyday speech with a perfective value (Escobar, 1997; García Tesoro & Jang, 2018; Howe, 2006, 2013; Howe & Schwenter, 2008; Jara Yupanqui, 2011a, 2011b). For example, Rodríguez Louro and Howe (2010) found that in narrative contexts, which typically require the use of the preterit, the PP appears in the 54% of the data in Peninsular Spanish, 26% in Peru, and 10% in Argentina. Similarly to Peru, higher frequencies of use of the PP as compared to the Spanish American tendency have been found in all the Andean region, for example in Ecuador (Bustamante, 1991; Pfänder & Palacios, 2013), North-Eastern Argentina (Kempas, 2013; Terán & Kanwitt, 2018), and Bolivia (Howe & Schwenter, 2003).

Some accounts of PP usage in the Andean area highlight that it resembles the Peninsular norm (Alonso & Henriquez Ureña, 1964, p. 155; López Morales, 1996, p., 25<sup>7</sup>, among others). For example, *Nueva Gramática Española* claims that these varieties share with Peninsular Spanish the use of PP for recent past events. In fact, at a first sight, it appears that the Peruvian use of the PP falls somewhere between the Peninsular norm and the Mexican norm. Howe and Schwenter (2008) found a frequency of 26.4% for the PP in the Lima corpus by Caravedo (1989). In the sentence judgement task carried out by Howe (2013), the PP represented the 23% of the tokens in Cuzco, and the 27.1% in Lima.

However, Howe (2006, 2013) underlies the different nature of the PP use in Peru as compared to Peninsular Spanish. In fact, in Peruvian Spanish the use of the PP in hodiernal temporal references is not widespread (Howe & Schwenter, 2008;

---

<sup>7</sup> Cited in: Howe & Schwenter, 2003, pp. 55-56.

Rodríguez Louro & Howe, 2010), and conversely, it is accepted in pre-hodiernal contexts, as in example 19:

- (19) El matrimonio ha sido el diez de agosto. (*The wedding was [:PP] the 10<sup>th</sup> of August*; LIMA\_M31\_009).

In Howe (2013), the PP in co-occurrence with pre-hodiernal adverbials (e.g. *ayer* ‘yesterday’, *el otro día* ‘the other day’, *el año pasado* ‘the past year’) appeared in the 11% of the data from Cuzco (Peru), 2.2% of the Madrid data, and 0% in the Valencia data set. Cuzco also presented the highest number of speakers who accepted both the PP and the preterit in such contexts, as compared to Peninsular speakers. On the contrary, only the 25.9% of Cuzco participants used the PP in co-occurrence with “today” adverbials, a much more restricted number compared to Madrid (75.6%) and Valencia (87.5%). Howe concludes that Peruvian PP is following a different path of grammaticalization as compared to its Peninsular counterpart, as the grammaticalization is motivated by two different pragmatic processes. In Spain, the gradual erosion of the current relevance requirement typically associated with perfects brought to the establishment of perfective uses of the PP in hodiernal and recent past contexts, namely, those that are perceived as not temporally distant from the present. Conversely, in Peru, a process of pragmatic extension of the concept of current relevance took place. This led to the “subjectification” of the interpretation of what is currently relevant; therefore, the perfect is admitted in temporally distant contexts, provided that the speaker considers the event as relevant to his current situation.

In line with Howe’s findings, several authors claimed that Peruvian PP, unlike the Peninsular PP, carries evidential functions (Escobar, 1997; García Tesoro & Jang, 2018; González et al., 2019; Jara Yupanqui, 2011a; Rodríguez Louro & Jara Yupanqui, 2011). For example, it is used to mark that the speaker witnessed the events that is narrating (Escobar, 1997; García Tesoro & Jang, 2018; Klee & Ocampo, 1995). Escobar (1997) also found that the PP marks spatial relevance, so it is preferred when describing an event that happened in the same place where it is being recounted. Another function of Peruvian Spanish is to highlight the emotional involvement of the speaker toward the events she/he is recounting, regardless of the temporal distance



from the moment of speech (González et al., 2019; Jara Yupanqui, 2011a). Also, the PP is used to interrupt narrations (which are typically recounted in preterit) to perform different pragmatic-discursive functions, such as citing, commenting or summarizing (Howe & Schwenter, 2003; Jara Yupanqui, 2011a). Jara Yupanqui (2011a) proposes that Peruvian speakers take advantage of the status of PP as marked (namely, less frequent) form to assign pragmatic functions of psychological relevance to it, as in example 20:

- (20) Bueno, yo una vez presencié el asalto de un amigo [...]. Creo que mi pata se estaba olvidando su casaca y fui y se la dejé [...]. Y en ese momento, los han asaltado con pistola, los han tirado al piso, han robado, han manoseado a la chica. (*Well, once I witnessed how a friend was robbed [...]. I think that my buddy was forgetting his jacket and I went and left it for him [...]. And at that moment, they were robbed [:PP] at gunpoint, they were thrown [:PP] to the ground, and robbed [:PP], they got fresh [:PP] with the girl*). (UMO15, Jara Yupanqui, 2011a, p. 228).

These non-prototypical uses of the perfect may be due to the contact with Quechua. This claim is based on the observation that Quechua presents a reduced tense morphology, but has grammatical morphemes to mark the spatial relevance and the reliability of the information (Escobar, 1997; González et al., 2019; Klee & Ocampo, 1995). This latter use is evidential, namely, it provides information about the source of information, which may have been collected through direct experience, experience transmitted by others, or inference (Klee & Ocampo, 1995, p.64). In Quechua, the suffix {-ra} or {-rqa} marks a concluded action that was carried out with the direct participation and conscious control of the speaker. On the other hand, the suffix {-sqa} marks a past action that the speaker did not witness (it was reported to her/him), and that happened without her/his direct participation or out of his control. It is used also to recount events which were dreamt, imagined, and historical and mythological accounts (Bustamante, 1991, p. 223). These features of Quechua verbal morphology are transferred to the Spanish verbal system by bilingual speakers.

Recall also that the development of evidential features is one of the possible evolutions of a grammaticalizing PP (see section 2.4). According to Bybee and Dahl (1989), the PP typically acquires functions of indirect witness marker. This is the

function that Bustamante (1991) finds in the Spanish of Quito (Ecuador): the verbal system has been reorganized so that the preterit marks direct witnessing (it holds the functions of the Quechua morpheme  $\{-rqa\}$ ), whereas the present perfect marks indirect witness, as the Quechua morpheme  $\{-sqa\}$ . The same finding was found in Pfänder and Palacios (2013) with Ecuador Spanish, where the present perfect marks indirect witness with a certain degree of reliability, and the past perfect marks events that were recounted to the speaker by a source of information with low reliability. However, the results of Escobar (1997) are different: she claims that in Spanish in contact with Quechua, the PP marks direct evidentiality, namely, events directly witnessed by the speaker. Also, the PP has the function of marking spatial relevance. Quechua counts with a set of spatial markers of verbs, which provide information about whether the action occurred in the same place where it is being recounted, or in another place (p.867). According to the author, the PP marks events that occurred in the same location where the speaker is talking, whereas the preterit marks events occurred in another location, regardless of the distance with the moment of speech. In a similar fashion, Klee and Ocampo (1995) claim that the PP is used to mark events directly witnessed by the speaker, whereas the past perfect marks reported events, and the preterit does not have any evidential function (p. 64). Overall, the findings by Escobar (1997) and Klee & Ocampo (1995) go against the typical evidential development found for the PP, namely, that of an indirect witness marker. It may well be the case that the situation of linguistic contact affected the development of evidentiality, giving rise to an unusual outcome, as Howe (2013) suggests. However, Howe (2013) puts into question the notion that Quechua past tense morphemes  $\{-sqa\}$  and  $\{-rqa\}$  carry any evidential functions (pp. 115-116). In any case, the innovative functions of the PP in Peruvian Spanish were found both in Spanish-Quechua bilingual communities (Escobar, 1997; García Tesoro & Jang, 2018), and in Spanish monolingual communities, mainly in Lima (Howe 2013; Howe & Schwenter, 2008; Jara Yupanqui, 2011a, 2011b).

It remains an open question whether the PP is expanding in Peruvian Spanish as it is happening in Peninsular dialects. Howe (2013), who did not examine social

variables in his research, hypothesized that Peruvian Spanish presents stable variation of PP and preterit, but as he does not quantitatively explore this question, he proposes that others examine this possibility in future research (p.164). Conversely, Jara Yupanqui (2006) included social variables in her doctoral dissertation examining the frequency of the preterit and the PP in Lima, presenting the results of both an exploratory study and a main study. In the exploratory study, she conducted unstructured interviews with forty participants and collected 6,979 tokens of verbs in PP and preterit. The participants were divided according to two socioeconomic groups (upper-middle class and working class), two genders (male or female), and two age groups (18-27, 28-37). Results showed that young speakers significantly used the preterit more frequently as compared to the oldest group of speakers (86% of verbs in preterit for the former group, and 78% for the latter,  $p < 0.05$ ). The socioeconomic stratum also resulted significant ( $p < 0.05$ ), with the upper-middle class using the preterit more often (86%) than the working class (79%). Jara Yupanqui (2006) concludes that “the use of the preterit over the present perfect seems to be a phenomenon that is increasing in the Spanish of Lima; however [...] a deeper study of the phenomenon is needed” (p. 92). She also explains the higher frequency of the PP among the oldest generation and the working class by the fact that many of these speakers are children of immigrants from the Andean area, which is characterized by Quechua-Spanish bilingualism and a relatively high frequency of use of the PP. Therefore, these speakers were exposed to the Andean norm during their childhood. Nevertheless, the results of the main study did not confirm the findings of the exploratory study. In the main study, sixty-four participants completed a questionnaire eliciting verbs in either preterit or PP, for a total of 1,598 tokens. As in the exploratory study, participants were divided according to two age groups (18-25, 32-42), two socio-economic groups (middle and working class) and two genders (male, female). The effects of social variables did not turn out to be significant, therefore, the hypothesis of linguistic change in progress in the use of the PP/ preterit in Lima Spanish is not supported. Jara Yupanqui (2006) explains these results with the non-authentic, controlled contexts elicited by the questionnaire, which does not promote

the use of expressive and stylistic devices, as happens in natural speech: “Only the analysis of real conversation will give us the opportunity to observe social meanings and ongoing change more accurately [...]. If there is a growing preference for either the preterit or the present perfect, it is at the discoursal-pragmatic level that it will be observed. If there is any change in progress in favor of one or the other, it is at this level that it can be identified” (pp. 216-217).

### 2.6.3.3 *Is the PP undergoing change in American Spanish?*

To the best of my knowledge, only a handful of studies explored the issue of stable variation versus change in progress in American Spanish. One is the doctoral dissertation of Jara Yupanqui (2006) about *Limeño* Spanish, which I summarized in the previous section (2.6.3.2). All others are related to River Plate Spanish. Kubarth (1992) conducted sociolinguistic interviews with 54 speakers from Buenos Aires, and found that middle-aged participants (30-49) and upper-middle class used the PP more frequently than the young users (13-30; p. 565). The most striking result was that, in contrast to what is happening in Spain, Buenos Aires Spanish showed the opposite trend, with the preterit being favored by young speakers. A similar result was found by Fløgstad (2016), who examined the alternation between the PP and the preterit in semantic contexts typically associated to the PP (experiential, continuative, recent past, and resultative contexts). Fløgstad used data from 19 sociolinguistic interviews (14 were conducted in Buenos Aires, and 5 in Uruguay, either in Montevideo or Dolores), and informants were divided into two age groups. The youngest group included participants from 14 to 34 years old, and the oldest group included participants who were 35 to 86 years old at the time of the interview. Moreover, Fløgstad used data from the transcriptions of 6 interviews in the corpus *El Habla Culta de la Ciudad de Buenos Aires* (1987), along with the introductory texts included at the beginning of each interview. She collected 538 tokens in total and she found that the PP is virtually absent in her oral interviews from the youngest group age. Fløgstad agrees with Kubarth (1992) in that the preterit is expanding in Buenos Aires Spanish, and predicts that the PP will eventually disappear in this variety.

In her doctoral dissertation, Rodríguez Louro (2009) collected 3625 tokens of verbs in preterit and PP. She extracted the data from recordings of 30 speakers casual conversations, sociolinguistic interviews with 38 speakers, and a questionnaire in which 100 speakers conjugated verbs (either in PP or preterit) in sentences. All the speakers were from Buenos Aires, and were between 17 and 80 years old. In order to test age effects on PP use in Buenos Aires Spanish, Rodríguez Louro divided the participants in two age groups (17-35 in the youngest group, and 36-77 in the oldest group), but did not find a significant effect of age on the use of the PP in her oral data. Nevertheless, she found that youngest speakers favor the use of the PP in experiential contexts as compared to the oldest, but oldest speakers use the PP at a higher overall rate across different contexts as compared to the young, thus partially confirming the results of Kubarth (1992) and Fløgstad (2016). Moreover, Rodríguez Louro also examined headlines in Argentinian newspapers from the XVIII to the XX century. In this case, she found a significant decrease in the use of the PP across centuries. In conclusion, Rodríguez Louro (2009) makes the claim that the PP in Buenos Aires Spanish is qualitatively distinct from other American varieties of Spanish, as it carries more restrictions on its use. In Buenos Aires, the PP is favored only in experiential contexts, where speakers refer to a situation that has been experienced at least once in the past, and may repeat in the future. As opposed to the general claim that PP carries continuative functions, Rodríguez Louro shows that in the Buenos Aires variety, those function are expressed by means of preterit. Therefore, she defies the idea of American Spanish as being at a “frozen developmental stage” which stands at a lower level of development as compared to Peninsular Spanish, which is the most common interpretation of the developmental stages formulated by Harris (1982).

The finding that the preterit may come to overlap with PP functions, and gradually displace it, is in line with the diachronic research of Moreno de Alba (1998) in New Spain written documents. Some recent studies suggest that there is a general tendency in Romance languages toward the simplification of the verbal system, and whereas European Romances are moving along the perfect-to-perfective path typical of the PP grammaticalization, some American varieties are moving toward the opposite process,

namely, the expansion of the preterit at the expenses of the PP (Fløgstad, 2016, 2017; González et al., 2019). In light of these hypotheses, I propose that, in order to further explore the issue of the instability of the PP/preterit system, it is necessary to include social variables in research on PP grammaticalization, given that social variables have proven to correlate systematically with phenomena of change in progress in the linguistic system.

## 2.7 RESEARCH QUESTIONS

The reviewed literature demonstrates considerable variation in the distribution of the PP and the preterit across Spanish dialects. In the present study, I analyzed data from Lima and Mexico City from the PRESEEA<sup>8</sup> Corpus (2014-). To summarize the background literature, the variety of Mexico City, which has received more attention in the literature than other Spanish American dialects (see for example Lope Blanch, 1972; Moreno de Alba, 1998; Schwenter & Torres Cacoullós, 2008), is considered prototypical of the Spanish American norm in the use of the PP, where the PP is used in durative and iterative contexts (examples 21 and 22), but not in hodiernal contexts (example 23):

(21) Siempre me ha gustado el fútbol (MEXI\_H21\_090) (*I always liked* [:PP] *soccer*) → durative.

(22) Eso lo hemos discutido muchas veces (Lope Blanch, 1972, p. 132) (*We discussed* [:PP] *this many times*) → iterative.

(23) Ahorita me dejaste con la duda (MEXI\_H21\_090) (*Now you left* [:Preterit] *me with the doubt*) → hodiernal.

As for Lima, research showed that in the Peruvian variety (and in the Andean area in general), the PP is used at a higher frequency as compared to other Spanish American varieties, although is not as frequent as in Peninsular Spanish, and that it

---

<sup>8</sup> The Corpus PRESEEA (*Proyecto para el estudio sociolingüístico de español de España y de América*, 2014-) collects sociolinguistic interviews from several Spanish speaking countries and cities. The project is coordinated by Dr. Francisco Moreno and Dr. Ana María Cestero of the University of Alcalá (Alcalá de Henares, Spain). The transcriptions of the interviews, along with methodological recommendations, are available at: <https://preseca.linguas.net>.

shows the tendency of appearing in pre-hodiernal contexts, although the preterit is still the favored form (Howe & Schwenter, 2008, among others).

Importantly, there lack studies that investigate the effects of social variables on PP and preterit use in Lima and Mexico City dialects. Building on the previous findings of Jara Yupanqui (2006), this study aims to further explore the social and linguistic factors that affect the PP/preterit distribution in *Limeño* Spanish, including Mexico City as a point of comparison. In order to better examine the effects of generation on the PP/ preterit distribution, the age range of the participants includes speakers from 20 to 81 years old. Recall that the sample of participants of Jara Yupanqui (2006) ranged from 18 to 37 years old in the preliminary study (based on unstructured interviews), and from 18 to 42 in the main study (based on a questionnaire).

The current study sought to answer the following research questions:

1. What is the overall frequency of PP and preterit in Lima and Mexico City?

Based on the results of previous research, I expect the preterit to be the most common option in both varieties, and nonetheless, the PP will be significantly more common in the Lima data (Howe & Schwenter, 2008; Rodríguez Louro & Jara Yupanqui, 2011; Schwenter & Torres Cacoullós, 2008).

2. Which linguistic and extra-linguistic factors govern the use of the PP in Lima and Mexico City?

The aim of the second research question is to go beyond the mere description of the frequency of the PP and offer an account of:

- a. How the linguistic context, expressed in terms of temporal reference, presence and type of temporal adverbials, and quality of the direct object, may influence the linguistic choices of the speakers, favoring either the PP or the preterit;
- b. Whether social variables, which have shown systematic correlations with linguistic variation in previous sociolinguistic research (including research about the PP/ preterit variation), play a role. If this is the case, we must contemplate the possibility that the PP and preterit are not in stable variation in

the speech communities under exam, and consequently, there could be a linguistic change in progress.

This approach follows Schwenter & Torres Cacoullós (2008), according to whom such an analysis allows to go beyond intuitive characterizations (p. 10) and to shed light on the “pathways” of evolution of the PP in different dialects, as “quantitative similarity in perfect usage is not necessarily reflective of functional similarity” (Howe & Schwenter 2003, p. 67).

My hypothesis is that the temporal reference will be the most significant factor in governing the PP/preterit distribution, as found in previous research (Howe & Schwenter, 2008; Schwenter & Torres Cacoullós 2008). Howe and Schwenter (2008) examined eight factor groups for Lima (temporal reference, type of adverbial, presence of *ya*, polarity, clause type, transitivity, plurality of direct object, and lexical aspect), and found that only temporal reference and plural direct object were significant for PP choice. For Mexico City, Schwenter and Torres Cacoullós (2008) found that type of temporal adverbial, plural direct object, clause type, and verb lexical class were also significant (these last two factors are not included in my analysis).

3. Is there any evidence of change in progress in the usage of the PP in Lima and Mexico City?

I examined age group, gender, and level of education of the participants in order to answer the third research question, as different rates of use of PP and preterit by different social groups have proven to be reliable indicators of linguistic *change in progress* (Labov, 1994). Several studies conducted in Spain showed consistently that young speakers favor the PP, but only a handful of studies examined social factors in the distribution of PP and preterit in American Spanish, with mixed results (see section 2.6). Some studies hypothesize that the preterit may be expanding into PP functions in Latin American Spanish (Fløgstad, 2017; González et al., 2019). Given the inconsistency in the results of studies examining social variables in American Spanish, I adopt a null hypothesis for the third research question, hypothesizing that the PP and the preterit are in stable variation in the dialects of Lima and Mexico City,



and therefore, social variables do not have any effect on the usage of PP and preterit forms.

## **CHAPTER III**

### **METHODOLOGY**

The research of Schwenter and Torres Cacoullós (2008) guided the methodology of the study. Schwenter and Torres Cacoullós (2008) used the variationist comparative method (Poplack & Tagliamonte, 2001), which involves the comparison of multivariate analyses that include the same factor groups on different data sets (Torres Cacoullós, 2011, p. 161), to compare the PP and preterit patterns of variation in Madrid and Mexico City. I replicated their methodology on data from Mexico City and Lima. In addition, I included social variables (age, gender, and level of education) that were not included in Schwenter and Torres Cacoullós (2008), but that have proven to be significant in sociolinguistic studies on the PP/ preterit variation in Peninsular Spanish.

This chapter is organized as follows: first, I will describe the materials I used for the study, and the characteristics of the participants. Then, I will describe the procedure I followed. Next, I will present the dependent and independent variables, and describe the levels within each factor group of the independent variables. Lastly, the statistical analysis will be introduced.

#### **3.1 MATERIALS AND PARTICIPANTS**

I collected tokens of verbs in PP and preterit from the transcriptions of 36 sociolinguistic semi-structured interviews in the PRESEEA Corpus. 18 participants were from Lima and 18 from México City. For each speech community, the 18 participants are socially stratified by gender (male, female), age (young, adult or elderly), and educational level (basic, medium, high) based on the norms of the PRESEEA project. Regarding age, the youngest generation group includes participants from 20 to 34 years of age. The second generational group includes participants from 35 to 54 years of age. The oldest generational group includes participants who were 55 years old or older at the time of the interview. As for the level of education, the informants in the basic education level group ranged from being illiterate to having completed primary school (from 0 to 5 years of school education). Those who

attended at least five years of school after primary school were placed in the medium education level group (10-12 years of education). The high education level group included individuals with a university degree (15 years of education or more). It should be noted that while many studies using the PRESEEA Corpus data use education as a proxy for social class, given the complicated nature of social class per speech community (Ash, 2013), in this study this will only be treated as educational level. Table 1 summarizes the characteristics of participants in each speech community. Appendices A and B provide more detailed biographical information about the participants from both speech communities, which I extracted from the introductory text to the interview transcriptions.

Table 1. Participants information in each speech community based on the PRESEEA norms.

	<b>Generation 1 (20-34)</b>		<b>Generation 2 (35-54)</b>		<b>Generation 3 (55+)</b>	
	<b>M</b>	<b>F</b>	<b>M</b>	<b>F</b>	<b>M</b>	<b>F</b>
Education level 1	1	1	1	1	1	1
Education level 2	1	1	1	1	1	1
Education level 3	1	1	1	1	1	1

### 3.2 PROCEDURE

The semi-structured sociolinguistic interviews were collected between 2005 and 2007 in the Mexican corpus, and between 2008 and 2009 in the Lima corpus, thus we can reasonably compare speaker age groups across speech communities. The recordings were transcribed and each transcription was revised by two different researchers.

In general, the interviews follow the methodological directions issued by PRESEEA coordinators<sup>9</sup>, according to whom the interviews should be structured in eight thematic modules:

---

<sup>9</sup>Moreno Fernández, F. (2021). *Metodología del PRESEEA*, <https://preseea.linguas.net/Metodolog%C3%ADa.aspx/>. See also Moreno-Fernández, F. (2005).

1. Greetings, questions about the preferred treatment form (*tú* or *usted*) and weather preferences.
2. Place of residence: questions about the house, the neighbors and the neighborhood where the interviewee lives. Elicitation of memories about the neighborhood where the interviewee grew up.
3. Family and friends: questions about interviewee's social and everyday life, and important events related with family and friends.
4. Habits: questions about holidays, typical local food, typical meals in celebrations.
5. Danger of death: the interviewer asks the interviewee to recount an extreme situation in which they feared for their life, or a very stressful situation they have been passing through.
6. Important anecdotes: questions about important events in the life of the interviewee.
7. Desire of economic improvement: questions about gambling, lotteries and what one would do if winning much money.
8. Conclusion: concluding question about where to buy a soda nearby the location of the interview, and farewells.

However, it is worth to mention that this structure presents some degree of variability in the two data sets. In Lima, the interviews present a more regular pattern, and closely follow the recommended thematic modules. In Mexico City, they are more varied in content and the interviewer does not strictly follow the PRESEEA thematic modules. In some interviews, participants talk freely about their professional life or events of their life, with minimal intervention by interviewer, who does not try to change the stream of the conversation and only occasionally asks follow-up questions. Also, in the Mexican data there are frequent interventions of secondary participants, such as family members in the household where the interview is taking place. Conversely, in the Lima corpus the intervention of external people happens only occasionally. Furthermore, the interviews conducted in Lima present a similar length (between 40 and 50 minutes approximately), whereas the Mexican ones are longer

overall (around 60 minutes), but also present more variability in length (one interview lasts 110 minutes, and another one more than three hours). Most interviews were recorded in a quiet environment as the house of the participant, a classroom, or the interviewee's office. However, one interview with a Mexican street seller takes place in the *tianguis* 'street market' where he works, with interventions from different secondary participants (customers, other sellers, friends of the interviewee), and one interview in the Lima corpus is held in the street. Despite some methodological discrepancies, the interviews were transcribed following the same coding procedure, and the same amount of information about the social characteristics of each participant is provided (see Appendices A and B); therefore, I consider that the two data sets present an acceptable level of comparability.

I extracted a total of 2534 tokens of verbs in PP and preterit from the 36 interviews. In the next section, I will provide details about the procedure of collection of tokens. Afterward, I will describe how I coded the tokens according to the independent variables.

### **3.3 VARIABLES**

#### **3.3.1 Dependent variable**

The alternation between preterit and PP represents the binary dependent variable. Following Schwenter and Torres Cacoullos (2008), I defined the variable context broadly, including all the instances of verbs conjugated in preterit or in PP, with the aim of including in the analysis all the array of meanings fulfilled by these verb forms. In other words, the "envelope of variation" (Milroy & Gordon, 2003, p. 180) is form-based and includes all the occurrences of the grammaticalizing structure (PP), as well as its morphosyntactic alternative (the preterit) (Torres Cacoullos, 2011, p. 160). This approach is called "grammaticalization-path approach to the envelope of variation" (Torres Cacoullos, 2011, p. 160) and is motivated by the form-function asymmetry that characterizes morphosyntax in oral discourse: as well as a single form can cover a range of meanings, different forms can serve the same grammatical function (Schwenter & Torres Cacoullos, 2008; Torres Cacoullos, 2011). From a

grammaticalization theory perspective, this form-function asymmetry is the manifestation of two processes characteristic of grammaticalization, namely, 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, fulfil the same function; Bybee et al., 1994). Therefore, it is only by considering all the contexts in which the grammaticalized structure may appear (categorically or in alternation with the alternative variant, the preterit) that we can identify its functional domains and directions of semantical development, if any.

However, there were a few contexts that were considered outside of the envelope of variation. These were a total of 93 tokens, which I excluded from the analysis:

- Progressive structure *estar* + gerund (example 24):

(24) De ahí yo.. este como.. fui aprendiendo a que no me importara esas cosas (MEXI\_H22\_054).

- Morphologically ambiguous -AR and -IR verbs in first person plural (preterit or present) (example 25):

(25) Un día nos casaremos.. pues ya nos casamos (LIMA\_H23\_050).

- False starts (example 26):

(26) Ya lo superó, ya pasó esa.. ya pasó esa etapa (LIMA\_H12\_028).

- Repetitions of the interviewer's question (example 27):

(27) *A (Interviewer): ¿Cómo entraste a trabajar? – B (interviewee): ¿Cómo entré a trabajar?* (MEXI\_M11\_084).

I coded a total of 2534 tokens of verbs in PP and preterit. Coding began at the start of each interview. However, 77 tokens were removed prior to the multivariate analyses due to eliminating two levels from the independent variables, given the few tokens at those levels, which was skewing the statistical analysis. In other words, with so few tokens in a level, one cannot reliably claim a main effect or lack thereof when placed into the regression model, given there are not enough tokens of particular

levels; in addition to producing a false main effect, it could also negate a true main effect of another level from the same independent variable. This led me to eliminate the two levels of the independent variables which presented the lowest number of tokens. Specifically, I eliminated the hodiernal temporal reference level, which included 35 tokens (16 in Mexico, 19 in Lima), and 42 tokens (7 in Mexico, 35 in Lima) which occurred with a temporal adverbial of recency (e.g., *ahora, recién, últimamente, en estos días*), for a total of 77 tokens. Thus, the final analysis was conducted on 2,457 tokens (1,230 México City; 1,227 Lima), an average of 68 tokens per participant. The individual speaker token number presents small fluctuations, depending on how many tokens with hodiernal temporal reference and adverbials of recency I eliminated for each participant.

### **3.3.2 Independent variables**

The independent variables are divided in two categories: social (extra- linguistic) and linguistic (language- internal). Following the variationist comparative method (Poplack & Tagliamonte, 2001), I conducted separate analyses, with the same variables, on each speech community (Lima and Mexico City), in order to disclose differences not only in the overall frequency of the PP, but also in the social and linguistic constraints that govern its use in each dialect. The study included three social variables and four linguistic variables, which are listed below.

#### *3.3.2.1 Social variables*

Participants were divided according to three generation groups: youngest generation (20-34 years old), middle-aged (35-54 years old), and oldest generation (55+ years old). Additionally, they were divided by gender (male, female) and level of education (basic, medium, and high).

I included the age group as an independent variable as it is considered one of the main indicators of linguistic change in progress in sociolinguistic studies in *apparent time* (Labov, 2001). According to the *apparent time* construct, it is possible to observe linguistic change at a unique point in time by capturing differences in the frequency of use of a linguistic variant across different generations. In other words, change in

progress allows to make inferences about linguistic change that may have taken place in the recent past, and to predict the evolution of a linguistic variable in the language system (Tagliamonte, 2012, p. 43). When a linguistic variant is used at a similar frequency by different age groups, we are in front of stable variation. The variants of a variable may have some stylistic or social values associated, but no variant is spreading over the other, as variability is inherent to the linguistic system. Conversely, if a variant increases or decreases its frequency of use according to the age of the speaker, we may be in presence of a change in progress. The variant most frequently used by young generations may become prevalent in the future, whereas the variant most frequently used by the oldest generations will disappear, or will be maintained only in a restricted number of contexts. The increased rates of use of the PP among the young compared to the old found in variationist studies about Peninsular Spanish (Schwenter, 1994; Serrano, 1996; Howe, 2006) has been interpreted as a manifestation of change in progress toward the grammaticalization of the PP as a past perfective. Conversely, some studies about Buenos Aires Spanish found the opposite tendency, with increased rates of use of the preterit between the young generation (Fløgstad, 2016; Kubarth, 1992). This evidence led the authors to claim that the preterit is taking over the PP functions in this dialect. Therefore, the finding of age effects in the data, in either direction, may be indicative of a change in progress in the speech communities under exam.

Nonetheless, it is worth mentioning that the apparent time construct presents some limits. The different frequency of a variant across age groups may be due to age grading, that is, changes in the ways people speak over their lifespan (Labov, 1994, p. 84). Adults may be encouraged to adopt standard variants instead of local, vernacular variants, because of the pressure of the job market, in order to find (and maintain) good job positions. Conversely, adolescents, but also retired people, tend to use more vernacular, non-standard varieties, including the stigmatized ones. However, age grading typically affects variants with a high degree of social recognition, where a specific variant of a linguistic variable is considered prestigious or, on the contrary, stigmatized (Tagliamonte, 2012, pp. 47-48). This does not seem to be the case of our



morphosyntactic variable (preterit and PP), which alternation in discourse lacks an overt social recognition, similarly to other morphosyntactic variables, as overt subject pronouns in Spanish (Erker, 2017). Erker (2017) claims that linguistic variables may be more or less socially salient to speakers within a speech community, and that salient features undergo dialect levelling, that is to say, a speech community tends to converge toward the norm of another speech community. For example, the opposition between the pronouns *vos* and *tú* is highly salient for Spanish speakers in the US, and therefore carries a social value, which led to the stigmatization of *vos* and a change in progress toward the expansion of the variant *tú* at expenses of *vos*. As Erker points out, the bulk of research about this variable shows that, in US cities where Mexicans are the largest Hispanic group, Mexican Spanish holds the prestige and Salvadoran and Honduran speakers demonstrate linguistic accommodation toward the majority norm, which leads to dialect leveling of the feature. On the other hand, the frequency of use of overt subject pronouns is not salient for speakers of Spanish, and although there is regional variation (with Caribbean dialects using overt subject pronouns more frequently as compared to mainland Spanish American dialects), there is no social value associated to its use. Therefore, regional differences are maintained across bilingual Spanish-English communities in the US, with no group of speakers showing signs of converging to the norm of another group of speakers (although overall long-term residents in the US tend to increase their rate of use of subject pronoun, as a consequence of the linguistic contact with English) (Erker, 2017). I claim that similarly, in Mexico City and Lima there is no social value associated with the use of PP and preterit, as their use does not mark a specific social group (e.g. teenagers, or working class), and therefore, this variable does not carry overt social recognition<sup>10</sup>.

Gender is another key social variable for understanding language change (Labov, 2001, p. 262), as most sociolinguistic variationist studies found systematic patterns of relationship between gender and linguistic choices. In particular, research is consistent

---

<sup>10</sup> Although this may not be the case in other speech communities, where the same feature may be salient (for example, in Galicia and Asturias in contact with speakers from central Spain, where the frequency of PP and preterit could be salient due to differences between the speech communities in the same country).

in that, on one hand, in situations of stable variation women tend to favor prestigious variants, and avoid stigmatized variants. In this sense, they show a conservative behavior (which Labov calls the “conformity of women”, 2001, p. 266). On the other hand, it has been observed that in situations of linguistic change, women favor innovative variants, and therefore, they are leaders in linguistic change (Labov, 2001, pp. 279-280). Although not all sociolinguistic studies are consistent in their results, the regularity of these patterns led Labov to call this phenomenon “the gender paradox”. The reasons of the differential behavior of genders in relation to linguistic choices are above the scope of this work<sup>11</sup>, but I included gender as an independent variable to test whether gender differences are present in my data. However, it is worth mentioning that the studies that took gender into account in the analysis of PP grammaticalization in Peninsular Spanish did not find a significant effect of this variable. For example, Burgo (2012) did not find any effect of gender in Bilbao (Basque Country). Serrano (1996) also did not find a significant main effect of gender, although she did find that the leaders of change (e.g., those who used the PP more frequently) were men from lower-middle class and women from both lower-middle and upper-middle class (which confirms that gender is a social construct that commonly interacts with other social variables; Eckert, 1989; Labov, 1990). Overall, social variables such as age and social class appear to be more significant than gender in these studies. As for Hispanic America, Rodríguez Louro (2009) found that men significantly favor the PP in oral interviews as compared to women, but in questionnaire data, women used the PP slightly more often than men. Jara Yupanqui (2006) did not find a significant effect of gender in her data.

Finally, participants were divided into three groups according to their level of education, which is used as a proxy for socio-economic level. As we have seen in the literature review chapter, the socio-economic status is, together with age and gender, one of the most important variables in variationist research, as different socio-

---

<sup>11</sup> Eckert (1989) offers an exhaustive analysis of the issue; more recently, several third-wave sociophonetic perceptual studies illuminated the gender paradox from the point of view of the differential social judgement that men and women receive when using the same variant of a variable. See for example Barnes (2015) for Asturian Spanish, and Chappell (2016) for Costa Rican Spanish.

economic groups do not speak exactly in the same way (Labov, 2001). Individuals from the middle class (especially women) generally show increased use of innovative variants compared to lower and upper class. Different reasons can help explain the innovative linguistic behavior of the middle-class members, for example, their increased social and geographical mobility. From a social network perspective (Milroy & Milroy, 1992), individuals from the middle-class usually have complex but loose interpersonal ties, a condition that favors the adoption of linguistic innovations. Conversely, the low and the upper class tend to interact in tighter social networks, and consequently disfavor the introduction of new phonological and morphosyntactic features in their language system (Milroy & Milroy, 1992).

However, from a methodological point of view, operationalizing social class is a difficult task, as it includes different measures of economic power and property ownership, as well as social prestige, reputation, and status (Ash, 2013, p. 351). The factors concurring to define social class are also dependent from the socio-economic context of the speech community under investigation. PRESEEA interviews do not provide information about the social class of the participants, although they include the profession of the interviewee. In the interviews I used, most speakers in the higher education group have highly professional jobs (professor, psychologist, historian, biologist, etc.), although there are also an actor, a dancer and a student. Most individuals in the medium education group work as office employees (e.g.: a secretary, a post office director), two work in security forces (a police officer and a security guard), and there are also a dancer and a shoemaker. Two women with medium level of education are housekeepers. As for the low education group, the informants' occupation range from occasional jobs in the case of housekeeping women (cleaning service worker, babysitter), to low-income jobs as waiter and market seller. Overall, the level of education of the participants seems a reliable approximation to their socio-economic level.

In conclusion to this section, it is worth mentioning that, according to some scholars, in the Spanish American context more educated individuals may use the PP more often than individuals with low education, as the PP is perceived as a prestigious

literary form, whereas the preterit is more common in popular speech (Westmoreland, 1988). However, Schwenter and Torres Cacoullós (2008), who extracted their Mexican data from an *Habla culta* ‘Educated Speech’ corpus (Lope Blanch, 1971) and a *Habla popular* ‘Popular Speech’ corpus (Lope Blanch, 1976), did not find different rates in the use of the PP across the two corpora.

### 3.3.2.2 *Linguistic variables*

The selection of the linguistic independent variables was based on Schwenter and Torres Cacoullós (2008). Specifically, I selected four factors that were the most significant in their study, namely, temporal reference, temporal adverbial, presence of *ya*, and number (singular, plural or absent) of the direct object. The choice of these linguistic variables is motivated by the methodological need of avoiding the notion of “current relevance”, which in most research about the PP is considered the main feature distinguishing PP functions from the preterit ones. However, this notion is “empirically intractable” (Schwenter & Torres Cacoullós, 2008; Torres Cacoullós 2011, p. 160), as it is impossible to establish objectively which tokens are currently relevant above intuitive characterizations; similarly, codifying the tokens according to the four prototypical functions of the PP (experiential, continuative, hot news, and resultative) is problematic when dealing with spontaneous speech and above ideal examples, given the frequent ambiguity between continuative and experiential context, and between hot news and resultative contexts (Howe, 2013; Schwenter & Torres Cacoullós, 2008)<sup>12</sup>. Therefore, the linguistic variables I selected, such as the temporal reference and the presence of certain temporal adverbials, are motivated by the need of using more objective measures of the constraints that govern PP and preterit alternation than the notion of “current relevance”.

Below I listed each independent variable with their levels in bullet points below them.

---

<sup>12</sup> For this reason, a number of previous studies analyzed PP functions using grammaticality judgement tasks, or other tasks that were experimentally controlled, in order to force specific elicitations of the PP in controlled linguistic contexts (e.g., González et al., 2019; Howe 2006, 2013; Jara Yupanqui, 2006).

1. Temporal reference:

○ *Hodiernal*: the event occurred in the same day in which the speech is taking place (example 28): +<sup>13</sup>

(28) Hoy les dije que no iba a llegar temprano (MEXI\_M11\_084).

○ *Pre-hodiernal*: the event occurred at a point in time before the day of the speech (example 29):

(29) Nos casamos en diciembre de ese año (LIMA\_H33\_053).

○ *Indeterminate*: the analyst (and maybe even the speaker) is unable to determine when an action occurred. 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 her/his discursive purposes (Torres Cacoullos, 2011: 160), as in example 30:

(30) He llamado a I. para que viera la casa (LIMA\_M32\_025).

○ *Irrelevant*: it cannot be queried when something occurred, as the event lacks a specific temporal relevance, for instance in the case of permanent states that continue up to the present, repeated actions (for example with the adverb *muchas veces* ‘many times’), yes/no questions, and negative polarity (namely, when an event has not occurred), as in example 31:

(31) Nunca me ha gustado el calor (LIMA\_M32\_025).

2. Temporal adverbial:

○ *Specific time adverbials*: they answer to the question *when?*, e.g.: *ayer*, *en 1992*, *hace una semana*, *cuando era pequeño* (example 32):

(32) Ingresé a la facultad en el ’79 más o menos (MEXI\_H23\_018).

---

<sup>13</sup> The symbol “+” indicates that I coded the data, but subsequently I removed this level from the statistical analysis because of the scarce number of tokens.

- *Adverbials of recency*: adverbials that express that the event is perceived as recent (e.g.: *este verano, esta mañana, ahora, recién, últimamente*; example 33): +

(33) Mi hija recién ha dado a luz (LIMA\_M21\_005).

- *Atelic adverbials*: frequency and continuous time adverbials, which answer to questions: *for how long?*, or: *how many times?* (E.g. : *siempre, nunca, a veces, durante una hora, desde, hasta..*; example 34). The atelic context for temporal adverbials was almost always coded as irrelevant for the independent variable of temporal reference. Thus, it is worth noting that these two independent variables are not fully independent, but the current study follows previous studies (Schwenter & Torres Cacoullós, 2008; Howe & Schwenter, 2008) in incorporating both independent variables.

(34) Ahí duramos como dos, tres años (MEXI\_H32\_066).

- *Connective adverbials*: typically used to join the parts of a narration (e.g.: *antes, después, entonces, luego*; example 35):

(35) Le habló por teléfono y ya después llegó ella (MEXI\_H11\_078).

- *Absence of adverbial*: there is no temporal adverbial next to the verb, as in example 36:

(36) He llamado a I. para que viera la casa (LIMA\_M32\_025).

### 3. Presence of the adverb *ya*:

- Present, as in example 37:

(37) Ya me.. me.. me fui del trabajo porque ya salí embarazada (LIMA\_M21\_005).

- Absent, as in example 38:

(38) Dejé la carrera y me dediqué a trabajar (MEXI\_H32\_066).

I coded *ya* as a separate factor group as *ya* can appear together with other temporal adverbials, as seen in in example 33.

4. Presence of direct object:

- Singular (example 39):

(39) Nunca se levantó de su silla mientras que escribió la tesis (LIMA\_M22\_023).

- Plural (example 40):

(40) Me gané unos cuantos pleitos con algunos militares (LIMA\_H12\_028).

- Absence of direct object (example 41):

(41) Esos diez meses fueron espectaculares, ¿no? (LIMA\_M22\_023).

According to Howe and Schwenter (2008), the analysis of temporal reference is crucial to understand the processes of semantic change in the past reference in Spanish (p. 107). In previous research, the temporal reference turned out to be highly significant in Lima (Howe & Schwenter, 2008), as well as in Madrid and Mexico City (Schwenter & Torres Cacoullos, 2008). According to these studies, the irrelevant and indeterminate temporal references favor the PP in all dialects, whereas the hodiernal temporal reference favors the PP only in Peninsular Spanish; some authors claim that the indeterminate temporal reference is a key context in the grammaticalization of the PP, as these contexts straighten the association of the PP with a perfective past meaning (Schwenter & Torres Cacoullos, 2008; Copple, 2009). I initially coded 4 levels in this factor group. Indeterminate temporal contexts turned out to be by far the most common contexts ( $1410/2457 = 57.3\%$  of total tokens, 54.7% in Mexico City, 60% in Lima), followed by pre-hodiernal ( $795 / 2457 = 32.3\%$ , 34.1 in Mexico City, 30.5% in Lima), and finally irrelevant ( $252 / 2557 = 10.2\%$ , 11.1% in Mexico City, 9.3% in Lima). As for hodiernal temporal reference, I found only 35 tokens (16 in Mexico City, 19 in Lima), which I excluded from the statistical analysis because of their scarce number, which was skewing the regression model.

As for the temporal adverbials factor group, atelic adverbs, which express duration and frequency, are expected to favor the PP (Howe & Schwenter, 2008), whereas specific time adverbials are expected to favor the preterit. Again, I excluded

one level from the analysis, that is to say, adverbials of recency (e.g., *ahora, recién, últimamente, en estos días*). The reason was again the scarce number of tokens (42 in total, 7 in Mexico City and 35 in Lima), a condition that may hinder the reliability of the statistical analysis. The majority of tokens in this factor group occurred without any temporal adverbial (1663/2457 = 67.6% of total tokens, 69.2% in Mexico City, 66% in Lima), followed by atelic (299/ 2457 = 12.1% of total tokens, 12.7% in Mexico City, 11.5% in Lima), specific (295/2457 = 12% of total tokens, 9.7% in Mexico City, 14.2% in Lima) and connective adverbials (200/ 2457= 8.1% of total tokens, 8.2% in Mexico City, 8% in Lima).

The adverb *ya* was coded separately as it may co-occur with other temporal adverbials. Schwenter and Torres Cacoullos (2008) found that *ya* significantly favors the PP in Madrid Spanish. However, they found that in the Mexican data, *ya* did not turned out to be significant, and nonetheless, it appeared more frequently with the preterit. Therefore, we are in front of a variable that constrains the use of PP and preterit in opposed directions in Madrid as compared to Mexico City.

Another factor that significantly favored the PP in both Schwenter and Torres Cacoullos (2008) and Howe and Schwenter (2008) was the plural direct object, as plural object favor iterative meanings, as in example 42. Recall that that the PP in American varieties fulfills functions of continuity and repetition:

(42) Yo te he oído canciones tuyas [...] muy bonitas (Schwenter & Torres Cacoullos, 2008, p. 22).

Finally, the participants were included as a random factor in the analysis.

### **3.4 STATISTICAL ANALYSIS**

Following the variationist comparative method (Poplack & Tagliamonte, 2001, Tagliamonte, 2002), separate multivariate analyses including the same factor groups were conducted on each speech community, with the objective of understanding the underlying grammatical features that constrain the use of the PP, and which are the similarities and differences between the two speech communities in terms of PP and preterit usage.



I ran a mixed-effects logistic regression model using the statistical package Rbrul (Johnson 2009) in the program R (R Development Core Team 2020) to analyze the most predictive factors that contribute to the distribution of preterit and PP in Lima and Mexico City. Many previous studies on PP/preterit variation used Goldvarb X (Sankoff et al., 2005). The advantage of Rbrul lays in the possibility of including the speaker as a random effect in addition to the fixed effects (linguistic and social factors), which allows Rbrul to be more conservative compared to GoldVarb in avoiding Type 1 errors (that is, the finding of significant effects where there are none; Johnson 2009: p. 365). I ran the regression in reference to realizations of the PP. First, I compared the overall frequency of use of preterit and PP in each speech community. Then, I compared the strength and direction of constraints for each factor group in Lima and Mexico City. The results are presented in the next chapter.

## CHAPTER IV

### RESULTS

In this chapter, I will show the results of the two independent multivariate regression analyses carried out for each speech community. First, I will present the contextual factors contributing to the choice of the PP or preterit in Mexico City. Then, I will illustrate the results in the Lima data set. Finally, I will contrast the two sets of results, comparing the overall rates of use of PP and preterit in each dialect, as well as the direction and strength of the effects of the factor groups which turned out to be significant in constraining the use of PP and preterit.

#### 4.1 MEXICO CITY

Overall, in the Mexico City data set, 9% (111/1230) of the tokens are in PP and 91% (1119/1230) are in preterit, as seen in Figure 1.

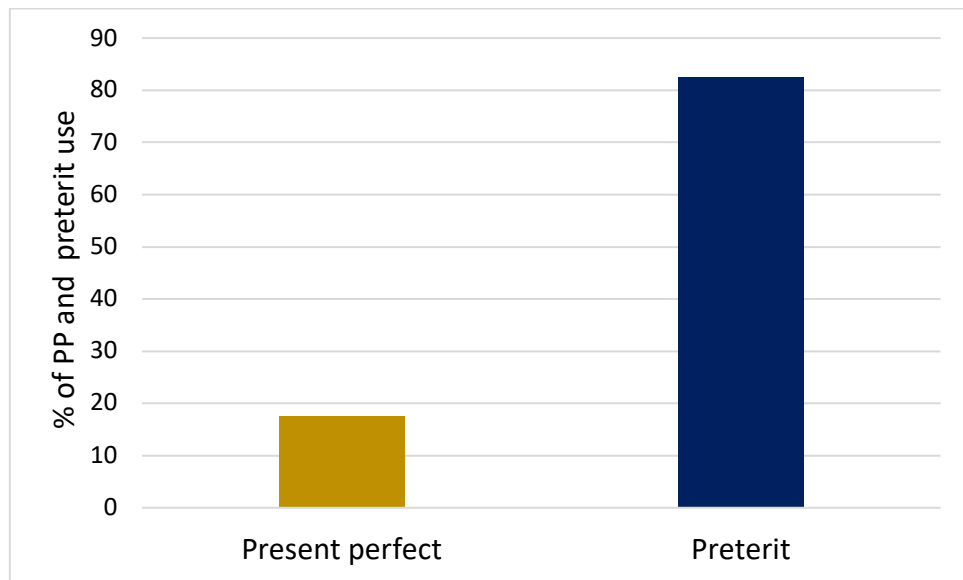


Figure 1. Overall frequency of PP and preterit in Mexico City.

Table 2 shows the effects of the factor groups selected as significant by Rbrul step up/step down logistic regression, which included temporal reference ( $p < 0.0001$ ), temporal adverbial ( $p < 0.001$ ), and direct object ( $p < 0.05$ ). Table 2 includes both log-odds and factor weights for convenience. Positive log-odd values must be interpreted as favoring the PP, while negative log-odds disfavor it (and, consequently, favor the

preterit). As for factor weights, values above 0.5 favor the PP, and those below 0.5 disfavor it. This model has an overall  $R^2$  of 0.98 (fixed factor  $R^2$  of 0.978, random factor  $R^2$  of 0.002).

Table 2. Summary of Rbrul step-up/step-down mixed effects logistic regression model of PP/preterit realizations in Mexico City, with speaker as a random factor, in reference to PP realizations;  $N = 1,230$ ;  $R^2$  fixed = 0.978,  $R^2$  random = 0.002,  $R^2$  total = 0.98.

<b>Predictors</b>	<b>Log-odds</b>	<b>Factor weights</b>	<b>% PP</b>	<b>Tokens (N)</b>
<b>Temporal Reference</b> <i>p</i> < 0.0001				
Irrelevant	7.550	0.999	37.2%	137
Indeterminate	5.482	0.996	8.9%	674
Pre-hodiernal	-13.032	<0.001	0%	419
<b>Temporal adverbial</b> <i>p</i> < 0.001				
No	8.975	> 0.999	8.9%	853
Atelic	8.601	> 0.999	22.3%	157
Specific	-8.330	< 0.001	0%	119
Connective	-9.246	<0.001	0%	101
<b>Direct object</b> <i>p</i> < 0.005				
Plural	0.716	0.672	17.6%	74
Singular	-0.094	0.477	12.2%	294
No	-0.622	0.349	7.2%	862

No social factor (age, gender, and level of education) was selected as significant.

As for linguistic variables, three of the four factor groups included in the analysis were selected as significant: temporal reference, temporal adverbial, and direct object. The effect of the adverb *ya* was not significant. As expected, the temporal reference was the factor group with the strongest effect ( $p < 0.0001$ ) in constraining the use of the PP and the preterit as seen in Figure 2. Also consistent with predictions is the strong favoring effect for PP use in irrelevant temporal contexts

(51/137 = 37.2%). Nevertheless, the preterit is still the most common form in irrelevant contexts, a pattern that, as we shall see in the next section, differentiates this variety of Spanish from the Lima one (at least in the current data). Indeterminate temporal contexts (in which the speaker does not specify the temporal anchoring of the event) also favor the PP, which occurs in 8.9% (60/ 674) of the data. Conversely, the PP does not appear in pre-hodiernal contexts (0/ 419 = 0%), which are expressed exclusively by means of preterit.

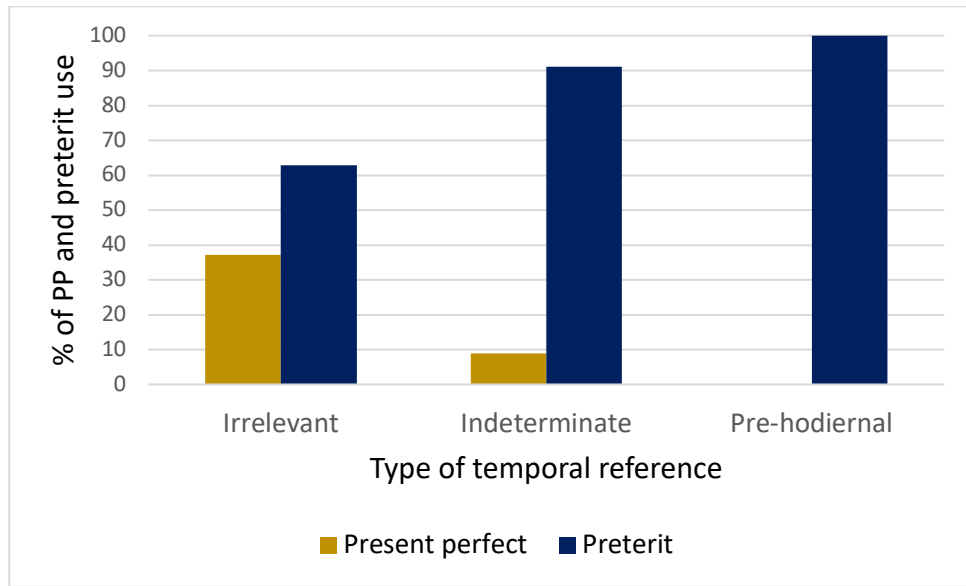


Figure 2. Main effect of temporal reference on PP and preterit use in Mexico City.

The factor group with the second magnitude effect is the temporal adverbial ( $p < 0.001$ ). As shown in Figure 3, atelic adverbials favor the PP (35/ 157 = 22.3%). Recall that atelic adverbials include adverbials of duration (as *desde* ‘since’, *hasta* ‘until’, *durante x tiempo* ‘during x time’) and frequency (like *siempre* ‘always’, *nunca* ‘never’, *alguna vez* ‘ever/sometime’). The absence of a temporal adverbial also favors the PP, which is used at a rate of 8.9% (76/ 853). Conversely, the presence of a specific time adverbial (as *el año pasado* ‘last year’, *en 1992* ‘in 1992’, *cuando era pequeño* ‘when I was a child’) or a connective adverbial (as *antes* ‘before’, *después* ‘after’, *luego* ‘then’, *entonces* ‘then’) strongly disfavors the PP. The preterit is used categorically in presence of these adverbials, in a similar fashion to the pre-hodiernal temporal reference.

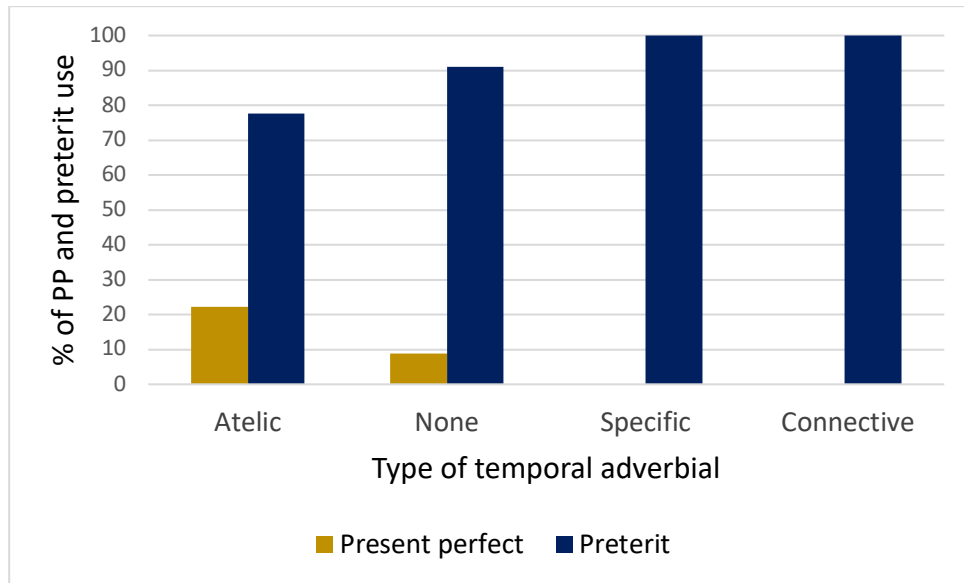


Figure 3. Main effect of temporal adverbial on PP and preterit use in Mexico City.

Lastly, there was a significant effect of the presence and quality (singular or plural) of the direct object ( $p < 0.005$ ), which is depicted in Figure 4. Plural direct objects favor the PP (13/74 = 17.6%), whereas singular direct objects (36/294 = 12.2%) and absence of direct object (62/862 = 7.2%) disfavor it.

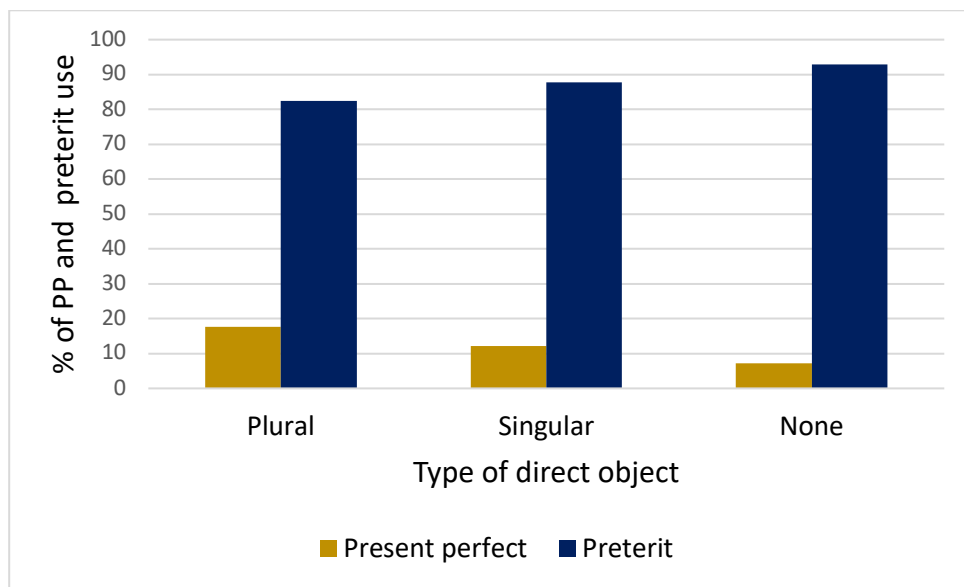


Figure 4. Main effect of direct object on PP and preterit use in Mexico City.

## 4.2 LIMA

The overall frequency of PP and preterit in the data from Lima is shown in Figure 5. The PP occurred in 24.7% of the tokens (303/1227), and the preterit in the remaining 75.3% (924/ 1227). This means that in Lima the PP is almost three times more frequent than in Mexico City.

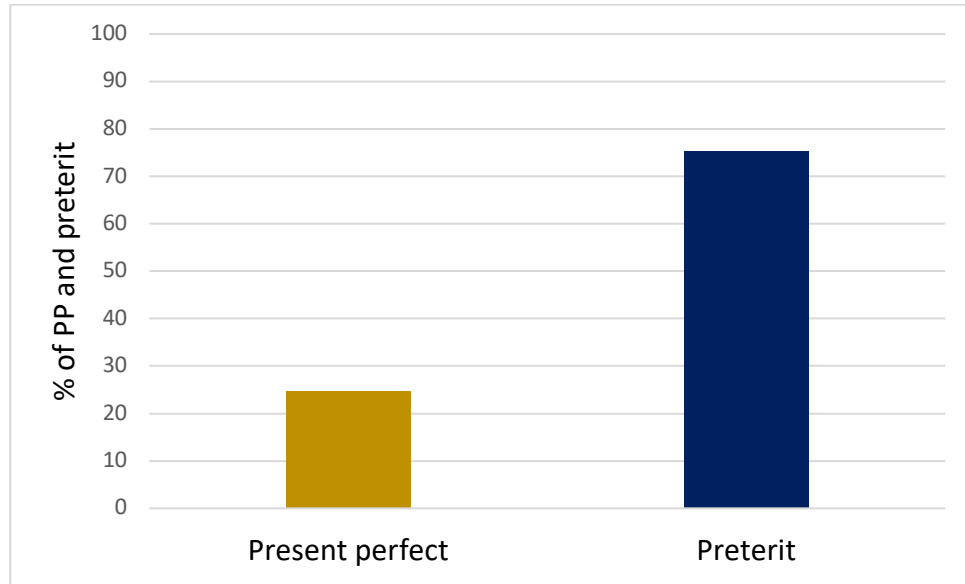


Figure 5. Overall frequency of PP and preterit in Lima.

A separated mixed effects logistic regression was run for the Lima data including the three social factors, four linguistic factors, and speaker as a random factor. Out of the four linguistic factors, temporal reference, temporal adverbial, and presence of *ya* turned out as significant, whereas the presence of direct object did not. As for the three social variables, age group was significant, but gender and level of education were not. Table 3 shows the effects of the factor groups selected as significant by Rbrul step up/step down logistic regression. The Lima model has an overall  $R^2$  of 0.306 (fixed factor  $R^2$  of 0.23, random factor  $R^2$  of 0.076). What is striking here is that, although the same type of analysis, with the same factor groups, was conducted on two comparable sets of data, the  $R^2$  value is considerably higher in the data from Mexico City than in those from Lima.

Table 3. Summary of Rbrul step-up/step-down mixed effects logistic regression model of PP/preterit realizations in Lima, with speaker as a random factor, in reference to PP realizations;  $N = 1227$ ;  $R^2$  fixed = 0.23,  $R^2$  random = 0.076,  $R^2$  total = 0.306.

<b>Predictors</b>	<b>Log-odds</b>	<b>Factor weights</b>	<b>% PP</b>	<b>Tokens (N)</b>
<b>Temporal Reference</b> <i>p</i> < 0.0001				
Irrelevant	1.455	0.811	59.7%	119
Indeterminate	0.216	0.554	27.5%	726
Pre-hodiernal	-1.671	0.158	8.9%	383
<b>Ya</b> <i>p</i> < 0.01				
Absent	0.435	0.607	25.9%	1125
Present	-0.435	0.393	11.8%	102
<b>Temporal adverbial</b> <i>p</i> < 0.05				
Specific	0.562	0.637	14.9%	175
No	0.102	0.525	26.1%	811
Atelic	0.052	0.513	38%	142
Connective	-0.716	0.328	11.1%	99
<b>Age group</b> <i>p</i> < 0.05				
Oldest	0.626	0.652	34.2%	409
Adult	-0.155	0.461	23.1%	403
Youngest	-0.471	0.384	16.9%	415

The factor group which showed the strongest main effect was the temporal reference ( $p < 0.0001$ ) as seen in Figure 6, in a similar fashion to Mexico City. The levels in this factor group (irrelevant, indeterminate, and pre-hodiernal temporal reference) showed the same order and direction of effect as in Mexico City, with irrelevant and indeterminate temporal reference favoring the PP, and the pre-hodiernal reference disfavoring it. In Lima, the PP is the most common variant in irrelevant temporal references ( $71/119 = 59.7\%$ ). In indeterminate temporal contexts, it occurs

in 27.5% of the tokens (200/ 726). As for pre-hodiernal contexts, a small but not a negligible amount of PP tokens is present (34/383 = 8.9%), a pattern that distinguishes this variety from the Mexico City one.

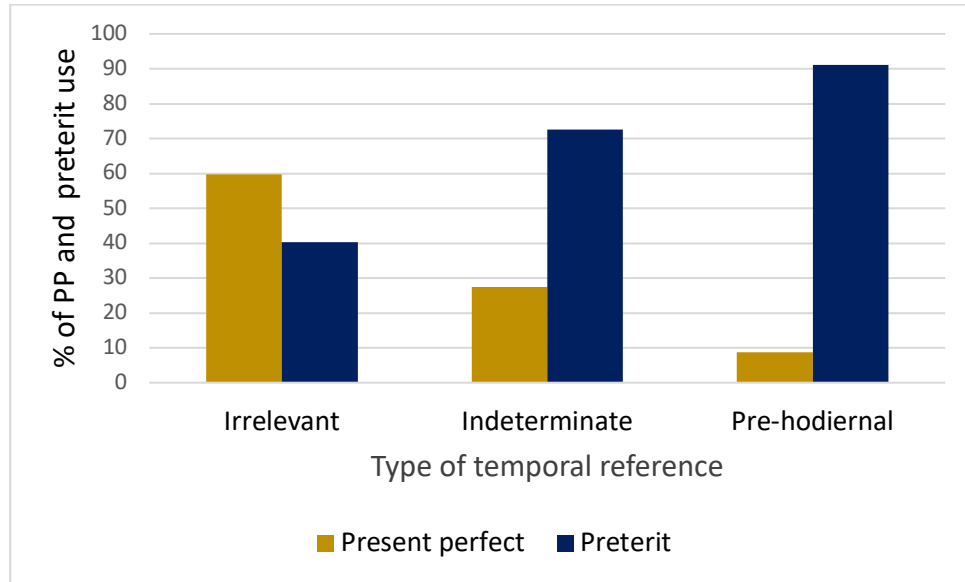


Figure 6. Main effect of temporal reference on PP and preterit use in Lima.

Figure 7 shows the effects of the second strongest main effect on PP/preterit realizations, the presence of the adverb *ya* ( $p > 0.005$ ). The absence of *ya* favors the PP (291/1125 = 25.9%), while the presence of *ya* disfavors it (12/102 = 11.8%).

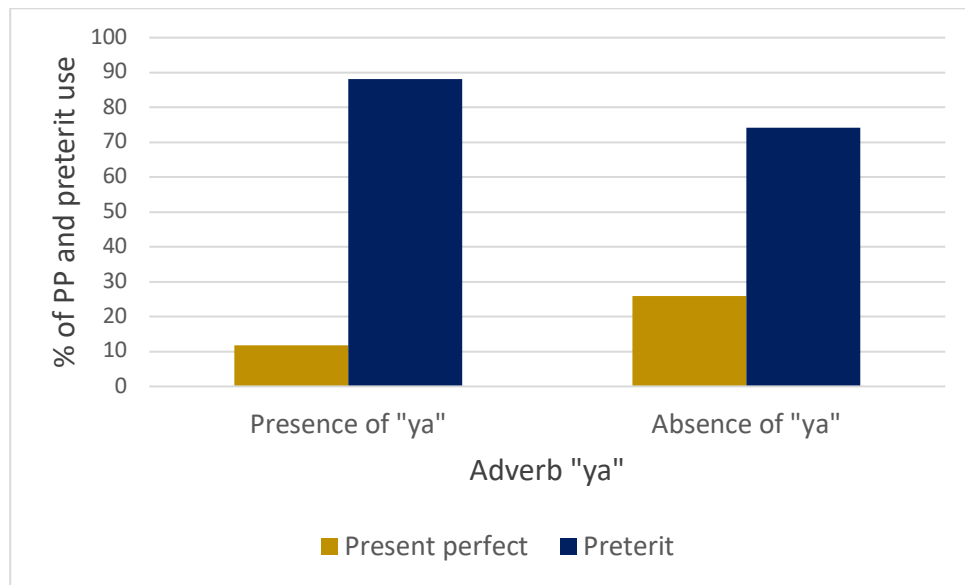


Figure 7. Main effect of the adverb “ya” on PP and preterit use in Lima.



The third most significant main effect was the type of temporal adverbial ( $p = 0.02$ ) as seen in Figure 8. The PP is favored with atelic adverbials (54/142 = 38%), and in absence of temporal adverbial (212/ 811 = 26.1%). As opposed to the Mexico City data, the PP also occurs, although in few tokens, with adverbials of specific past time (26/ 175 = 14.9%), and connective adverbials (11/ 99 = 11.1%).

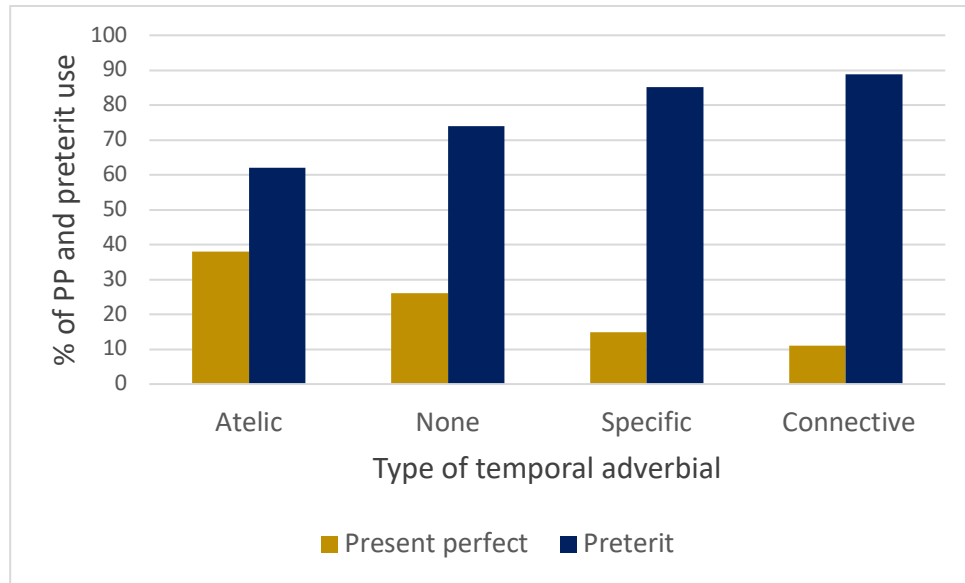


Figure 8. Main effect of temporal adverbial on PP and preterit use in Lima.

The last significant factor group selected by the logistic regression is the social variable of age group ( $p = 0.03$ ), as seen in Figure 9. Interestingly, the direction of the effect is the opposite as compared to the one found in several studies about Peninsular Spanish (see for example Burgo 2012; Schwenter, 1994; Serrano 1996). In the oldest generation group, which includes participants from 55 to 81 years old, the PP occurs in 34.2% of the tokens (140/409), around twice more often as compared to the youngest age group (70/ 415 = 16.9%), which disfavors the PP. The middle age group slightly disfavors the PP, which occurs in 23.1% (93/ 310) of the tokens, therefore showing an intermediate frequency of use of the PP. Figure 10 presents the same data, but with the goal of showing a clear depiction of the frequencies in PP and preterit across generations.

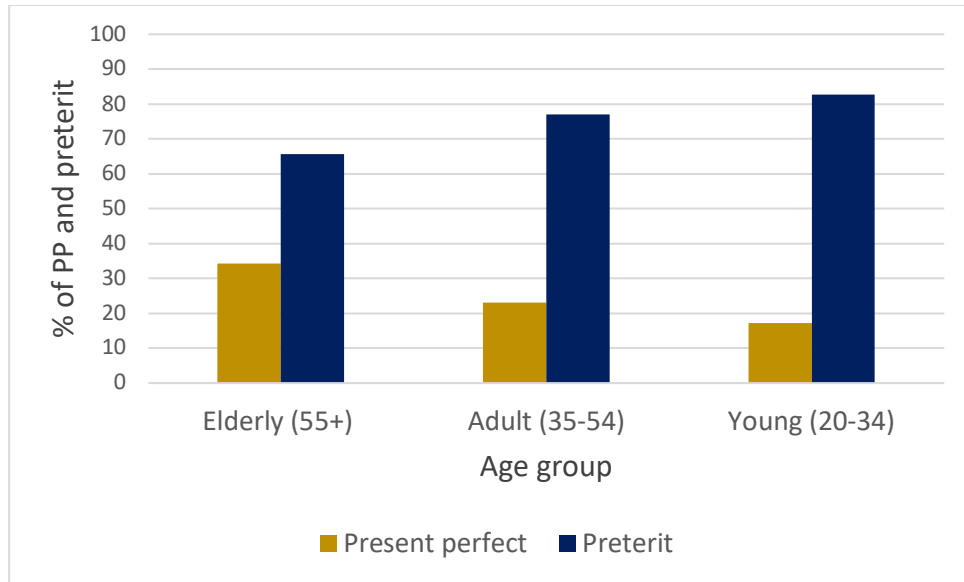


Figure 9. Main effect of age group on PP and preterit use in Lima (bar chart).

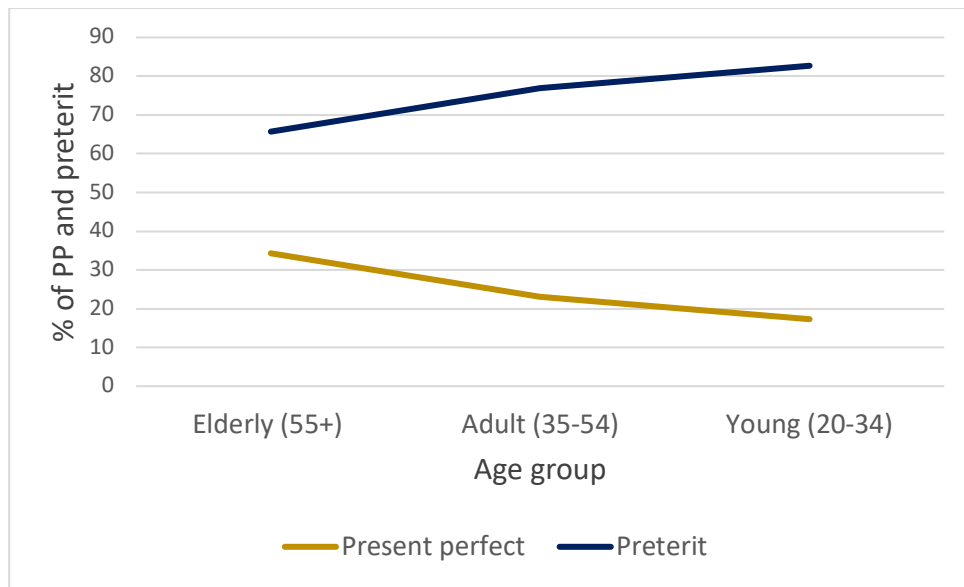


Figure 10. Main effect of age group on PP and preterit use in Lima (line chart).

The overall effect of age group ( $p = 0.03$ ) may be revealing of subtle changes in the semantic functions associated with the PP (and consequently, the preterit) in this variety of Spanish. Moreover, from the perspective of the grammaticalization theory, the extension of the semantic functions of the preterit at expenses of the PP is a controversial phenomenon. This possibility is further discussed in chapter 5. In the

next section, I will examine more in detail the effects of age group in Lima. Finally, I will compare the results of the analyses conducted on the two speech communities.

### 4.3 A CLOSER LOOK AT THE AGE EFFECT IN LIMA

In order to have a better understanding of the PP/preterit distribution in Lima, I will now present the frequency of use of the PP by participant, starting from the oldest generational group. Table 4 reports the average usage of the PP by participant. Recall that the overall frequency of use of the PP across all the age groups in Lima is 24.7%. In the oldest generational group, the average is 34.2%, therefore, it is higher than the overall average. Within this age group, two women with basic and medium level of education show the highest rates of PP (42% and 44.8% respectively). The man with low level of education also shows a high frequency of use of the PP (41.2%), whereas the male participant with high level of education has the lowest rate of use of the PP (18.8%). At an intermediate frequency of use, there are the woman with high level of education (28.4%) and the man with medium level of education (30.4%). The standard deviation in this age group is 10.07.

Table 4. Present perfect usage by participant in the oldest generational group in Lima.

<b>Participant</b>	<b>Age</b>	<b>Gender</b>	<b>Education level</b>	<b>% PP</b>
LIMA_H31	81	M	Basic	41.2
LIMA_H32	78	M	Medium	30.4
LIMA_H33	63	M	High	18.8
LIMA_M31	55	F	Basic	42
LIMA_M32	70	F	Medium	44.8
LIMA_M33	64	F	High	28.4
Average = 34.2%				
St. dev. = 10.07				

Table 5 shows the values for the middle-aged generational group. Here, the average frequency of the PP is 23% and there is high variability across participants. A woman with medium level of education has the lowest rate of use (7.1%) and a man with medium level of education shows the highest rate (44.3%). For the other participants, it is difficult to find a pattern of usage: the man with high level of

education also shows a high frequency of use of the PP (37.3%), the woman with low level of education has a frequency that reflects the average of the group (23.4%), and the woman with the highest level of education along with the man with the lowest level of education use the PP at a lower frequency than the average (16.1% and 10% respectively). The standard deviation in this group is 15, the highest of all groups.

Table 5. Present perfect usage by participant in the intermediate generational group in Lima.

<b>Participant</b>	<b>Age</b>	<b>Gender</b>	<b>Education level</b>	<b>% PP</b>
LIMA_H21	47	M	Basic	10
LIMA_H22	48	M	Medium	44.3
LIMA_H23	38	M	High	37.3
LIMA_M21	36	F	Basic	23.4
LIMA_M22	45	F	Medium	7.1
LIMA_M23	40	F	High	16.1
Average = 23%				
St. dev. = 15.01				

As for the youngest generational group (table 6), the first thing that stands out is the presence of an outlier. The woman with low educational level (speaker LIMA\_M11) shows a frequency of use of the PP (44.6%), which is considerably higher compared to all other participants, who have a frequency of use between 4.3% (represented by the woman with medium education) and 18.6% (represented by the man with high level of education). In this group, the average frequency of use of the PP is 17.2%, with a standard deviation of 14.28. If we take out the outlier and consider only the remaining five participants, then the average use of the PP lowers to 11.7%, with a standard deviation of 5.48.

Table 6. Present perfect usage by participant in the youngest generational group in Lima.

Participant	Age	Gender	Education level	% PP
LIMA_H11	26	M	Basic	8.6
LIMA_H12	33	M	Medium	12.9
LIMA_H13	28	M	High	18.6
LIMA_M11	26	F	Basic	44.6
LIMA_M12	30	F	Medium	4.3
LIMA_M13	27	F	High	14.3
Average = 17.2%				
St. dev. = 14.28				

To assess the relationship between age as a continuous factor and the percent of PP usage, a Pearson correlation was conducted that indicated a non-significant 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, elementary education) indicating a significant moderate positive association between age and PP frequency ( $n = 17$ ,  $df = 15$ ,  $r = 0.625$ ,  $R^2 = 0.39$ ,  $p < 0.01$ ). Figure 11 shows the scatterplot of PP usage by participant, where the outlier is clearly visible.

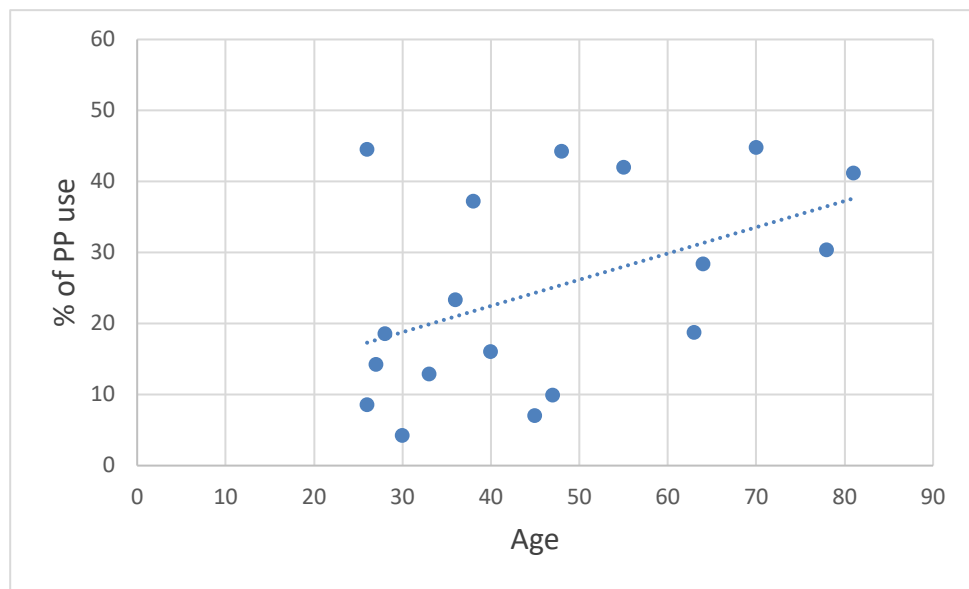


Figure 11. Main effect of age on PP use in Lima (note the outlier).

Table 7 shows the frequency of use of the PP, expressed both in number of tokens and percentage, by age group and linguistic variables (those selected as significant in the regression analysis). Overall, the frequency of use of the PP tends to increase in each level of the factor groups as the age group increases. In other words, the middle-aged group shows higher rates of PP use as compared to the youngest age group, and the same is true for the oldest group as compared with the middle-aged group.

Table 7. Present perfect usage by age group and linguistic context in Lima.

	Youngest (PP/total = %)	Middle-aged (PP/total = %)	Oldest (PP/total = %)
<b>Temporal Reference</b>			
Irrelevant	14/ 29 = 48.2%	24/ 39 = 61.5%	31/ 47 = 65.9%
Indeterminate	44/ 258 = 17%	65/ 259 = 25%	92/ 220 = 41.8%
Pre-hodiernal	12/ 128 = 9.3%	4/ 105 = 3.8%	17/ 142 = 11.9%
<b>Temporal adverbial</b>			
Atelic	8/ 40 = 20%	21/ 51 = 41.1%	25/ 51 = 49%
None	51/ 280 = 18.2%	67/ 272 = 24.6%	94/ 259 = 36.2%
Specific	8/ 58 = 13.7%	1/ 48 = 2%	17/ 69 = 24.6%
Connective	3/ 37 = 8.1%	4/ 32 = 12.5%	4/ 30 = 13.3%
<b>Ya</b>			
Present	2/ 44 = 4.5%	2/ 32 = 6.2%	8/ 26 = 30.7%
Absent	68/371= 18.3%	91/ 371 = 24.5%	132/ 383 = 34.4%

The only exception to this trend was found in the two levels where the PP is used in a non-conventional way: with a pre-hodiernal temporal reference, and with specific time adverbials. In both cases, the youngest age group shows a higher rate of PP compared to the middle-aged group. With pre-hodiernal temporal reference, the youngest generation had a PP rate of 9.3% (twelve tokens), very closed to the oldest group rate (11.9%, seventeen tokens), but the middle-aged group used the PP only in

3.8% (four tokens) of the tokens in this context. As for specific time adverbials, the youngest group used the PP in eight tokens (13.7%), the middle-aged group in only one token (2%), and the oldest group in seventeen tokens (24.7%). Therefore, for these non-conventional (or innovative) uses of the PP, the oldest age group members are the leaders, followed by the youngest group, and the middle-aged group shows the most conservative behavior. Given the low number of tokens in these categories, I controlled whether the occurrences of PP came primarily from a specific participant. For instance, is it the case that the twelve PP tokens with pre-hodiernal temporal reference in the youngest age group come mainly from the young woman with low education, namely, the outlier who considerably overused the PP as compared to the other participants in the same age group? Apparently, this is not the case, and the tokens in PP are distributed across participants. Recall that each age group has six speakers. For the pre-hodiernal temporal reference, the twelve tokens in the youngest group come from three speakers, that is, the man with medium education and the two women with low and high education (four tokens each). Therefore, the woman with low education (the outlier) does not show an increased rate of PP use in pre-hodiernal contexts as compared to other participants. The four tokens in the middle-aged group also come from three participants, again, the man with medium education (two tokens) and the two women with low and high education (one token each). As for the oldest age group, all participants used the PP in this context. Of the seventeen tokens, four come from male participants and thirteen from female participants.

I found similar results for the PP in presence of specific time adverbials. In the youngest age group, the eight tokens come from three participants, again, the man with medium education (four tokens) and the two women with low and high education (two tokens each). Again, the outlier (the woman with low education) did not use the PP at a considerably higher rate as compared to other participants in the same group. In the middle-aged group, there is only one occurrence of PP with a specific adverbial, which comes from the man with medium level of education. As for the oldest group, the tokens in PP come from all participants but one, the man with a low educational level. Six tokens come from the other two men and the eleven remaining tokens come

from the three women. In conclusion, the tokens in PP in these contexts are distributed across participants and the results are not particularly skewed because of the different linguistic behavior of one of the participants.

Overall, the oldest age group shows the higher rate of use of the PP in all contexts. The sharpest differences between generational groups can be found in presence of the adverb *ya*: the youngest generation almost categorically selects the preterit in this linguistic context (with only the 4.5% of PP), the middle-aged group shows a similar behavior (with the PP appearing in 6.2% of the tokens), but the oldest group used the PP in 30.7% of the tokens. Recall that this adverb is considered revealing of dialect differences, with the PP being selected significantly more often in presence of *ya* in Peninsular Spanish (Schwenter & Torres Cacoullós, 2008: 21), and the preterit being preferred in American Spanish (Howe, 2013: 129-130). In my data from Lima, the preterit is the most common form for all groups, and nonetheless, the two youngest groups are consistent in strongly favoring the preterit, whereas the oldest age group shows much more variability, selecting the PP in approximately one third of the cases.

The increased rates of PP usage among the oldest generational group is also evident in indeterminate temporal contexts, where the youngest group used the preterit in 17% of the tokens, but the oldest group used it in 41.8% of the tokens. The middle-aged group shows an intermediate behavior, using the PP in 25% of the tokens. A similar pattern appears with atelic adverbs. The youngest generation has a PP rate of 20%, the middle generation 41.1%, and the oldest generation used the PP in 49% of the tokens.

The theoretical implications of this gradual decline in the use of the PP in Lima are discussed in the next chapter. I will now focus on comparing the two speech communities, contrasting the effects and direction of the variables which affect the selection of PP and preterit.



#### 4.4 PP IN LIMA AND MEXICO CITY: SIMILARITIES AND DIFFERENCES

As demonstrated by the analyses, the PP is used almost three times more often in Lima than in Mexico City as seen in Figure 12.

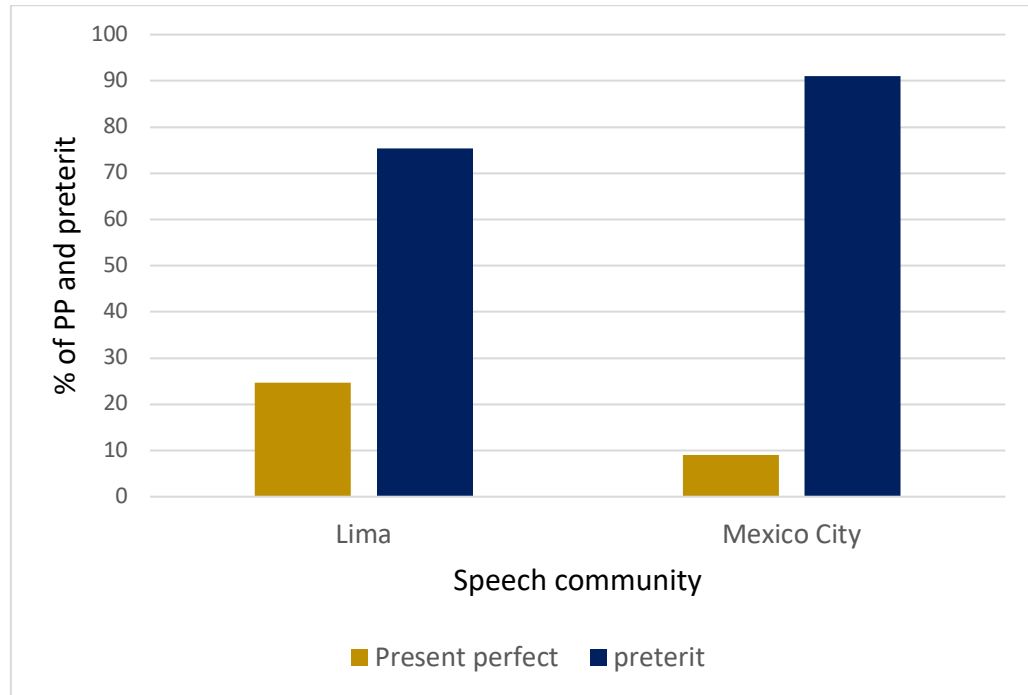


Figure 12. Overall frequency of PP and preterit per speech community.

Following Schwenter and Torres-Cacoullós (2008), Table 8 presents the factor weights values in reference to PP usage for each level of the factor groups in Mexico City and Lima. Recall that a factor weights of 0.5 or above favors the dependent variable (the PP), whereas a factor weight below 0.5 disfavors it. Non-significant factor groups are marked with “n.s.”.

Table 8. Factor weights in reference to PP realizations in Mexico City and Lima.

	<b>MEXICO CITY</b>	<b>LIMA</b>
<b>Temporal reference</b>		
Irrelevant	0.99	0.81
Indeterminate	0.99	0.55
Pre-hodiernal	0.00	0.15
<b>Temporal adverbial</b>		
Atelic	0.99	0.51
None	0.99	0.52
Specific	0.00	0.63
Connective	0.00	0.32
<b>Ya</b>		
Present	n.s.	0.39
Absent		0.60
<b>Direct object</b>		
Singular	0.47	
Plural	0.67	n.s.
None	0.34	
<b>Age</b>		
Youngest	n.s.	0.38
Middle		0.46
Oldest		0.65
<b>Gender</b>		
Male	n.s.	n.s.
Female		
<b>Level of education</b>		
Basic		
Medium	n.s.	n.s.
High		

Common to the two dialects are the effects of temporal reference and temporal adverbial. The temporal reference has the strongest effect in both speech communities, a result in line with previous research (Schwenter & Torres Cacoullós, 2008; Howe & Schwenter, 2008). The direction of the effects is also the same, with irrelevant and indeterminate temporal reference favoring the PP, and the pre-hodiernal temporal reference favoring the preterit. However, in the Lima speech community there are

instances of PP in pre-hodiernal contexts, whereas Mexico City follows the standard norm more closely, and no tokens in PP are present in pre-hodiernal contexts, as seen in Figure 13.

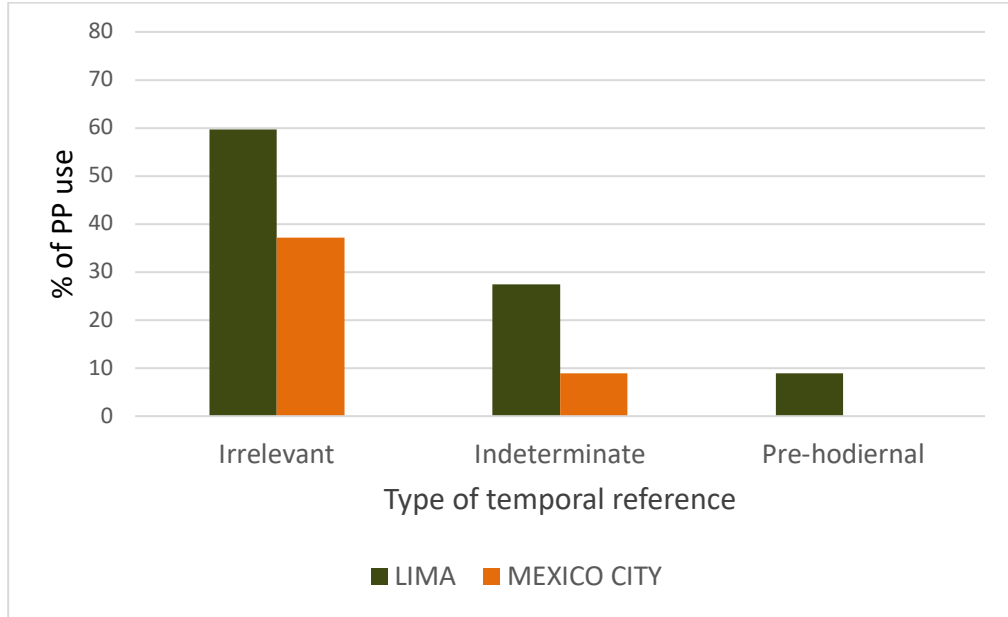


Figure 13. Main effect of temporal reference on PP usage in each speech community.

Temporal adverbial is the second strongest predictor of PP/preterit variation in Mexico City, and the third in Lima (after the adverb *ya*). Here, the direction of the effects is very similar, with only one difference. In both speech communities, the conditions that favor the PP are the presence of an atelic adverbial, and the absence of any temporal adverbial. Conversely, connective adverbs favor the preterit. This was expected as connective adverbs are used in narrations, which are typically carried out by means of preterit, as the PP lacks of the sequencing effects required for narration (Bybee et al., 1994: 62). As for specific time adverbials, the expectation is that they disfavor the PP, as they locate the situation or event in a pre-hodiernal past. This is what I found in Mexico City, where specific adverbs strongly disfavor the PP. However, in Lima a not negligible rate of PP (14.9%, 34 tokens) is present in co-occurrence with specific time adverbials as seen in Figure 14.

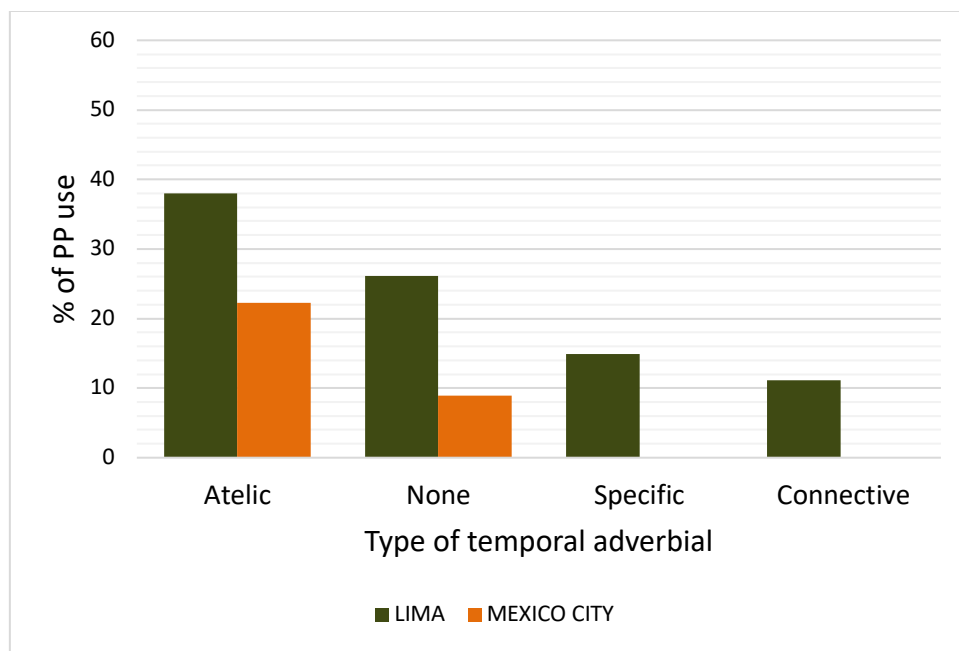


Figure 14. Main effect of temporal adverbial on PP use in each speech community.

As for the other factor groups, *ya* was significant in Lima but not in Mexico City. Schwenter and Torres Cacoullos (2008) also did not find a significant effect of *ya* in Mexico, although the *p* value was only slightly above the level of significance, and they noted that the PP occurred more frequently in absence of *ya* than in its presence. We have seen that in Lima *ya* favors the preterit, and nonetheless, different generations have different linguistic behaviors, with the oldest generation using the PP in co-occurrence with *ya* at a considerably higher rate than the participants in the youngest group, who almost categorically use the preterit in this linguistic context.

Social variables, as we have seen, are not significant in constraining the variation between present perfect and preterit, except for the age group in Lima.

#### 4.5 EXCLUDED TOKENS

Before concluding this chapter, I would like to turn the attention to the linguistic contexts I excluded from the statistical analysis, namely, hodiernal temporal reference and co-occurring adverbials of recency (e.g. *ahora* ‘now’, *recién* ‘recently’, *últimamente* ‘lately’). These were a total of 77 tokens, of which 35 appeared in hodiernal context (16 in Mexico City, 19 in Lima), and 42 which co-occurred with an adverbial of recency (7 in Mexico, 35 in Lima).

In Mexico City, I excluded 23 tokens. 16 are hodiernal and 7 co-occur with an adverbial of recency. 20 tokens are in preterit and only 3 in PP (15%). Of these three tokens in PP, two are found in the hodiernal temporal reference ( $2/16 = 12.5\%$ ) and one with an adverbial of recency ( $1/7 = 14.2\%$ ). Overall, the frequency of the PP is very low, and nonetheless, it is higher than the overall frequency of 9% found in the statistical model. In this sense, these results are in line with the claim that adverbials of recency favor the PP in Mexican Spanish (Schwenter & Torres-Cacoullas, 2008).

In Lima there is a different pattern of use: the PP occurs in the two thirds of tokens, 36 out of 54 (66.6%). In the hodiernal reference context, 10 out of 19 tokens are in PP (52.6%). The co-occurrence of an adverbial of recency seems to strongly favor the PP, with 26 tokens out of 35 (74.2%). Overall, it seems that in Lima dialect there is a high level of intra-speaker variation, with participants using both the PP and the preterit in the same linguistic contexts, whereas Mexico City appears more conservative in the choice of the preterit.

Nonetheless, the number of tokens in these linguistic contexts is so low that these results must be interpreted with caution.

## CHAPTER V

### DISCUSSION

This chapter is organized as follows. In section 1, I will go back to the research questions which guided this study and summarize the results. I will use previous research which guided the methodology used here (Howe & Schwenter 2008; Schwenter & Torres Cacoullos, 2008) as points of comparison. In section 2, I will discuss the theoretical implications of my findings, in particular, regarding the age effect found in Lima. On the one hand, the fact that the preterit seems to be expanding in this variety is in contrast with a fundamental tenet of grammaticalization theory, namely, the unidirectionality of the process. On the other hand, there is previous research (although scarce) which found a similar expansion of the preterit in the semantic territory of the PP. In section 3, I will dedicate some words to the pedagogical implications of my findings for the teaching of the PP/ preterit opposition in the L2 (and Heritage) Spanish Language classroom, in particular in relation to the idea that Latin American Spanish constitutes a uniform variety of Spanish, and to the teaching of present perfect as expressing “recent past”, a definition that does not adequately describe the uses of the PP in the speech communities I examined. Finally, the limitations of this study and the venues for future research are presented.

#### 5.1 RESEARCH QUESTIONS

I report here the research questions and the main findings related to each of them.

##### 5.1.1 RQ 1. What is the overall frequency of PP and preterit in Lima and Mexico City?

As predicted, the preterit was the most common form in both speech communities, and the PP was significantly more common in the Lima data. In Mexico City, 9% (111/1230) of the tokens (111/1230) were PP, and 91% (1119/1230) were preterit. This is a lower frequency compared to the Mexican data analyzed by Schwenter & Torres Cacoullos (2008), where the PP appeared in 15% of the tokens. They used a sample of the *Habla Culta* and *Habla Popular* corpora published by Lope Blanch in 1971 and 1976, respectively. Mexico City PRESEEA corpus was collected

around thirty years later, between 2005 and 2007. It is therefore possible that this dialect has moved toward a linguistic change with the expansion of preterit at the expenses of the PP. However, the lack of any social variable significantly affecting the variation between preterit and PP calls to caution in this interpretation. Nor did Schwenter and Torres-Cacoullós (2008) find significantly different frequencies across the *Habla Culta* and *Habla Popular* corpora. Therefore, the different frequencies found in Schwenter and Torres-Cacoullós (2008) as compared to the current study are more likely due to a different interview structure.

As for Lima, 24.7% of the tokens (303/1227) are in PP, and the preterit occurs in the remaining 75.3% (924/ 1227). This is also coherent with predictions, as Peruvian Spanish is known to have an increased frequency of the PP. Howe (2006) reports a frequency of 29.5% for the perfect in Lima, based on the corpus of Caravedo (1989), and of 23% in Cusco, based on the interview data collected by Howe. However, Howe (2006, p. 121) warns that there is a discrepancy in the number of tokens in the two data set (443 in Cusco, 3,667 in Lima). In Howe and Schwenter (2008), which again used the corpus by Caravedo (1989), the distribution is 26.4% for the PP and 73.6% for the preterit. Jara Yupanqui (2006) found a frequency of 82% for the preterit and 18% for the PP in her unstructured interviews. Overall, the current findings are consistent with these results, confirming that Peruvian Spanish presents a consistently increased use of the PP as compared to other varieties as Mexican and River Plate Spanish (see Rodríguez Louro & Jara Yupanqui, 2011, among others).

### **5.1.2 RQ 2. Which linguistic and extra-linguistic factors govern the usage of the PP in Lima and Mexico City?**

The temporal reference was the factor group with the strongest main effect in both Lima and Mexico City ( $p < 0.0001$  in each community). In both speech communities, the irrelevant temporal reference most strongly favors the PP, followed by the indeterminate temporal reference. However, in Mexico City the preterit is preferred in all contexts. In Lima, the PP is preferred in irrelevant contexts, and the preterit in all other contexts. As for pre-hodiernal contexts, the PP is strongly disfavored in Mexico City (no PP tokens were present in the data), however it occurs

in the 8.9% of the tokens in Lima. Overall, Lima presents increased frequencies of the PP as compared to Mexico in all contexts.

The temporal adverbial also affects significantly the choice of the PP or preterit in both speech communities, although the effect was stronger in Mexico City ( $p < 0.001$ ) than in Lima ( $p < 0.05$ ). In both cities, the PP is favored with atelic adverbials, and in absence of a temporal adverbial, and is disfavored with connective adverbials, however, in Mexico the effect is categorical (100% of preterit), whereas in Lima there are tokens in PP in these linguistic contexts. As for specific time adverbials, they disfavor the PP in Mexico City (0%), but they slightly favor the PP in Lima (14.9%).

The other significant linguistic factors are the presence of *ya* in Lima and the direct object in Mexico City. The results in Mexico City overall resemble those of Schwenter and Torres Cacoullos (2008), although they found higher frequencies of the PP at each level of the factor groups. They also found that temporal reference, temporal adverbial, and direct object significantly constrain the use of the PP, whereas the adverb *ya* does not. However, their analysis included more linguistic factor groups that resulted significant, that is, clause type and Aktionsart class of the verb. I left these factors out of my analysis for two reasons: first, I was interested in testing the effects of social factors, which had been not tested in previous research. The inclusion of extra-linguistic factor groups forced me to reduce the number of linguistic factor groups to avoid overtaxing the statistical model, also in consideration of the number of tokens, with the aim of reaching an acceptable statistical power (for the same reason I eliminated the two levels with fewer tokens, that is, hodiernal temporal reference and co-occurring adverbials of recency). Additionally, clause type and Aktionsart verb class showed the weakest effect in the study of Schwenter and Torres Cacoullos (2008).

As for Lima, the results partially resemble those of Howe and Schwenter (2008). In common with their analysis there is the PP favoring effect of the irrelevant and indeterminate temporal reference, however, they did not find an effect of *ya*, nor of the temporal adverbials. In contrast, they found an effect of the plural direct object. In the present study, temporal adverbials showed the same direction of effect as in Mexico



City, with atelic adverbials favoring the PP. Overall, these results are consistent with the claim that the PP in American Spanish carries continuative and iterative functions (Harris, 1982; Lope Blanch, 1972; Squartini & Bertinetto, 2000; Westmoreland, 1988), as attested by the favorable effect of irrelevant temporal reference and atelic adverbials.

As for the effect of direct object, the fact that I did not find a significant effect may be due to the coding procedures. I coded as singular all direct objects that were grammatically singular, although semantically plural, as in example 43:

(43) He conocido a mucha gente (LIMA\_H12\_028).

Therefore, I coded *mucha gente* as singular. Possibly, the effects of plural direct object may have been hidden because of the coding procedure I followed.

With regard to the adverb *ya*, recall that Schwenter & Torres Cacoulios (2008) found that it favors the PP in Peninsular Spanish; however, it was not significant in their Mexico City data (with a  $p < 0.503$ , p. 35), where nonetheless the PP occurred more frequently without the adverb *ya* than in its presence. According to Howe (2013) the direction of the effect of the adverb *ya* is one of the main indicators of the difference between Peninsular Spanish and American Spanish, given that this adverb favors the PP in the former, but disfavor it (favoring instead the preterit) in the latter. In this work, the most striking result regarding the effect of the presence *ya* in Lima is the sharp difference between generations. The youngest used the preterit almost categorically in presence of the adverb *ya*, however, the oldest generations selected the PP in one third of the tokens; therefore, the disfavoring effect for the PP in co-occurrence with *ya* is stronger among the young.

As anticipated in chapter 2, the fact that an increased frequency of the PP has been found in different zones of the Andean area (Bustamante, 1991; Howe & Schwenter, 2003; Pfander & Palacios, 2013, among others) led a number of scholars to search for a motivation of this phenomenon in the linguistic contact, especially with Quechua. Recall that the results of this research are mixed: whereas Escobar (1997) and Klee and Ocampo (1995) claim that the Peruvian PP is used to mark events

directly witnessed by the speaker, Bustamante (1991) and Pfänder and Palacios (2013) found that in the Spanish of Ecuador, the PP is a marker of indirect witnessing.

When coding the data, I was surprised by the high incidence of the PP with perfective verbs as *nacer* ‘to be born’. This verb typically requires the preterit (*nací*), as it marks a perfective context. However, in my data from Lima, I often found the PP *he nacido*. I considered the possibility that this was due to the spatial value of the PP presented by Escobar (1997). However, in my data, most Lima born interviewees used the preterit when talking about their date and place of birth, and there are some instances of *he nacido* when the place of birth is not Lima, as in examples 44 and 45:

(44) He nacido allá.. allá, este.. en Huacayo, ¿no? (LIMA\_H31\_18).

(45) He nacido en Villa MT.. cumpliendo mis dieciséis, diecisiete.. ya me vine pa’ acá (LIMA\_M21\_005).

These examples seem to contradict the value of spatial marker the PP as formulated by Escobar (1997), and are more compatible with the analysis of Bustamante (1991) and Pfänder and Palacios (2013) for Ecuador Spanish: the PP marks an event on which the speaker does not have conscious control, as being born, regardless of the location. However, in example 46 there is an instance of the verb *nacer* used in PP by a Lima-born speaker (San Cosme is a neighborhood of Lima):

(46) He nacido en San Cosme.. En San Cosme he vivido hasta los cuatro años (LIMA\_M11\_003).

Also, there are other verbs in PP used to talk about pre-hodiernal events that happened in Lima, as in example 47, which may have a spatial value:

(47) Acá ha sido el matrimonio... ya tienen un año, ha sido el diez de agosto (LIMA\_M31\_009).

Overall, the results of investigation regarding PP functions in situations of Quechua-Spanish in contact are mixed, and more research is needed. Although it is reasonable that Quechua has some kind of influence on the use of the PP in Spanish, it is still not clear how this influence operates. Also, the majority of the studies were

conducted in areas of Quechua-Spanish bilingualism, however, Lima is considered a Spanish monolingual area. On one hand, it is likely that national immigration from the Andean area to the capital city has brought linguistic innovations (Jara Yupanqui, 2006). On the other hand, the participants in PRESEEA corpus were all Lima-born, or had been living in Lima since their childhood. Therefore, we lack not only of an account of how Quechua and Spanish verbal systems interact, but also of how contact-related phenomena spread to areas with no linguistic contact.

### **5.1.3 RQ 3. Is there any evidence of change in progress in the usage of the PP in Lima and Mexico City?**

As we have seen, in Mexico City I did not find any effect of social variables, and therefore I can claim that the PP and the preterit are in stable variation in this dialect. This also confirm the observation by Schwenter and Torres Cacoullós (2008) that no difference was found in the *Habla Culta* and *Habla Popular* corpora (Lope Blanch, 1971, 1976), which is reflected in my data in an absence of effect of the level of education.

As for Lima Spanish, the age effect found in *apparent time* (Labov, 1994) may be indicative of a change in progress. I have found that the oldest age group (55+ years old) favors the PP (with an average of 34%), whereas the youngest age group (20-34 years old) favors the preterit (and only used the PP in 17% of the tokens). The middle-aged group shows the highest variability in the use of the PP and the preterit, whereas the oldest group is more homogeneous, and so does the youngest group, especially if we exclude the outlier speaker (the woman with low level education), whose frequency of use of the PP is among the highest of all the speaker sample (together with most participants form the oldest group). The oldest group showed the highest frequencies of use of the PP in all the significant speech contexts, and their increased use in comparison to other groups is especially striking in presence of the adverb *ya*, which in the statistical analysis has an overall disfavoring effect for the PP. Nonetheless, the oldest group selected the PP in one third of the tokens approximately (30.7%) in presence of *ya*, while the youngest age group used the preterit in the 95.5% of the tokens. The findings regarding the age effect are consistent with those of the

preliminary study of Jara Yupanqui (2006), who found that in Lima, youngest speakers use the PP less often than oldest speakers (14% for the former, and 22% for the latter group) in oral interviews. I found the same direction of the age effect, although with an overall higher frequency of PP usage in the oldest group of speakers; this can be accounted by the fact that the pool of participants of Jara Yupanqui (2006) was divided into two age groups and included speakers 37 years of age and younger, whereas in this study there are three generational groups, with the oldest group being represented by speakers who were between 55 and 81 years old at the time of the interview. Additionally, there are two differences in my results as compared to those of Jara Yupanqui (2006): first, I did not find a significant effect of education, whereas Jara Yupanqui (2006) found that the lowest social economic stratum uses the PP at a higher frequency as compared to the upper-middle class. This may be due to the fact that level of education and socio-economic class do not reflect each other completely. In the study of Jara Yupanqui (2006), the socioeconomic level included both the level of education and the neighborhood of residence of participants (which was not taken into consideration in the current study). Nonetheless, the data I analyzed also revealed a higher frequency of the PP in the lowest educational group as compared to the highest, although this difference did not result significant. Second, Jara Yupanqui (2006) did not find an effect of age and social stratum in the questionnaire she administered to a separate group of participants. However, she claims that this discrepancy is likely due to the controlled contexts elicited by a questionnaire, as opposed to spontaneous conversational contexts, and that only the analysis of oral interactions may shed further light on the issue of linguistic change in progress of a morphosyntactic variable as the PP/preterit semantic overlap.

Most studies about PP grammaticalization in Spain rely on the age effect found in *apparent time* (Labov, 2001) to support the claim that the PP is grammaticalizing as a perfective past, given the evidence that the youngest generations use the PP more frequently than the oldest. Moreover, the youngest generations, as opposed to the oldest, occasionally use the PP in pre-hodiernal contexts (Burgo, 2012; Howe, 2006; Schwenter 1994a). This evidence is considered strong enough to claim that there is a

linguistic change in progress in Madrid Spanish and other Peninsular dialects, where the PP is undergoing the “perfect-to-perfective path” (Schwenter & Torres Cacoullos, 2008) also attested in other Romance languages as French, Italian and Romanian. The age factor is therefore considered a key factor for the observation of linguistic change in progress. However, the change I found in Lima has the opposite direction of that attested in Peninsular Spanish. This finding may have different implications, which will be further discussed in the next section.

## **5.2 THEORETICAL IMPLICATIONS**

One of the main tenets of grammaticalization is that it is unidirectional (Bybee et al., 1994, p. 12; Hopper & Traugott, 2003, p. 99). This means that when an innovative (read: more recent) structure, as is the PP in Romance languages, starts taking on the functions of a more ancient one (the preterit), this process cannot be reversed. Therefore, once the PP has expanded to past perfective functions, displacing the preterit, it is not possible that the preterit expands again to its previous functions. This assumption has been criticized by a number of scholars (Giacalone Ramat & Hopper, 1998; Janda 2011; among others), who present counterexamples to the unidirectionality of grammaticalization and sustain the idea that the opposite process may occur.

As for Spanish, some scholars claim that Romance languages are moving toward a simplification of the verbal system whose outcomes may vary, including the possibility that the preterit expands to PP functions (Fløgstad, 2017; González et al., 2019). Fløgstad (2016) examined data from a set of studies on PP and preterit variation in 40 Romance dialects and languages and found that, although the grammaticalization of the PP as a past perfective is the most common type of linguistic change, the extension of the preterit in the functions of the PP is also a recurring phenomenon. González et al. (2019) in their analysis of Peninsular, Peruvian, and Argentinian Spanish also claim that, contrary to the prediction according to which the more recent form (the PP) displaces the old (the preterit), several Latin American dialects are showing signs of the preterit encroaching the PP territory. Moreover, a handful of studies conducted on historical documents written in

Hispanic America across centuries shows that in fact, the PP was more widespread in American Spanish in past centuries than nowadays, and that the preterit has gradually taken over functions typically associated with the PP, as the “hot news” function.

Rodríguez Louro (2009) examined Argentinian Spanish in newspapers from XVIII to XX century, while Moreno de Alba (1998) analyzed personal correspondence in New Mexico from XVI to XIX century. Both found significantly higher frequencies of the PP in the past centuries than in more recent times. Furthermore, Jara Yupanqui (2006) and two studies about Buenos Aires Spanish (Fløgstad, 2017; Kubarth, 1992) found a decrease in the use of PP among young generations. Rodríguez Louro (2009) did not find such a significant effect of age in Buenos Aires, although she found that the PP is favored in experiential contexts by young generations, whereas older generations tend to use it more in pre-hodiernal contexts (a phenomenon that the author interprets as hypercorrection). Taken together, this evidence seems to contradict the tenet of the unidirectionality of grammaticalization. Even if we consider the extension of the preterit to functions typically associated with the PP as a separate phenomenon independent from grammaticalization, we lack of a theoretical framework to explain why the instability of the PP/ preterit system may result in an extension of the preterit at expenses of the PP. This may be a venue for future research.

Another point connected to the age effect in Lima is the claim that the PP extends through indeterminate contexts. According to Schwenter and Torres Cacoullós (2008), the indeterminate temporal reference may be a significant locus of change for the PP. These contexts are not clearly temporally anchored, as the speaker does not need to specify the temporal distance with the moment of speech. An increasing use of the PP in these contexts may gradually reinforce the semantic connection of the PP with a perfective past value, and then allow for the appearance of the PP in both hodiernal and pre-hodiernal past contexts. However, in my data from Lima, the temporal reference is one of the contexts in which the different usage across generations is more patent. The oldest group uses the PP in 42% of the tokens in this context, and the young group in 17%. If we accept the hypothesis by Schwenter and Torres Cacoullós (2008) regarding the semantic extension of the PP through

temporally indeterminate contexts, we must conclude that the PP is actually retreating from the Spanish of Lima, as demonstrated by the low frequency of the PP in the youngest generational group.

### **5.3 PEDAGOGICAL IMPLICATIONS**

The teaching of the variation between PP and preterit in the L2 Spanish classroom has been characterized by a prescriptive orientation. The PP is often presented as a marker of recent past contexts, including the hodiernal contexts. However, we have seen that this is not the case of American Spanish, where the PP principally marks iterative and durative contexts, and, in any case, it is rarely used in varieties as Mexican and River Plate Spanish.

Moreover, many accounts of PP use present the Spanish American norm as essentially uniform, representing a frozen stage of development of Spanish language, whereas Peninsular norm has evolved. As I tried to show, American varieties are not frozen in their development, and do not represent a homogeneous system from the point of view of morphology, as different factors, including language contact, may affect their evolution.

In the field of Heritage Spanish, scholars advocate for a pedagogical approach which fosters sociolinguistic and critical language awareness (see, for example, Leeman & Serafini, 2016). By presenting the PP and preterit as structures with variable uses according to the speech community, instead of giving rules which reflect the prescriptive canon and the literary norm, we may teach Spanish in a critical way, not prescriptive, but descriptive, and inclusive of different varieties. Teaching the variability as a natural correlate of language in the Heritage and L2 classroom may be an effective way of developing critical sociolinguistic awareness in the Heritage and L2 Spanish classroom.

### **5.4 LIMITATIONS**

A first limit in this study is the lack of inter-rater reliability. Data from natural speech present ambiguities which need to be resolved by the researcher. The presence

of a second rater is the most common way to guarantee an objective and reliable analysis.

The scarcity of tokens with hodiernal reference and with adverbials of recency also constitute a limit of this work. In case of conducting a follow-up study with sociolinguistic interviews in Lima, I would add some questions eliciting these types of contexts, as “*cuéntame tu día hoy*” ‘tell me about your day today’ and “*cuéntame algún evento reciente en tu vida*” ‘tell me about some recent event in your life’, in order to have a sufficient number of tokens for both the hodiernal context (for the independent variable temporal reference) and the co-occurring adverbial of recency context (for the independent variable temporal adverbial) to conduct a reliable statistical analysis. The inclusion of tokens with hodiernal reference and adverbials of recency would allow for a more complete picture of the PP/preterit variation in the speech communities under exam.

The differences between the two corpora represent another limitation. Although the two corpora (Lima and Mexico City) follow the same methodological directions of PRESEEA, there are slight differences in the structure of the interviews (see section 3.2).

Finally, an important limit is represented by the number of tokens and participants. In particular, an increased number of participants would allow to better control for the age effect in Lima. Moreover, data in *real time* (Labov, 1994) would allow to confirm my finding in *apparent time* regarding the decreased frequency of the PP among young generations in Lima. This will be further discussed in the next section.

## **5.5 FUTURE DIRECTIONS**

As anticipated in the previous section, data obtained in *apparent time* benefit greatly from the inclusion of data in *real time*. This may be reached by conducting a trend study in Lima, with a pool of participants who are comparable (for age, level of education and gender) with the sample utilized in the present study. If the findings of this study are confirmed in a follow-up trend study, it would allow to support the evidence of a gradual decrease of PP usage in this speech community.



Another potential topic of investigation is the extension of the preterit to functions typically fulfilled by means of PP from a cross-dialectal and cross-linguistic point of view. The theory of grammaticalization allowed to explain for the common phenomenon of the expansion of the PP in several languages, including Romance languages and Peninsular Spanish. However, we lack a model that accounts for the expansion of the preterit at expenses of the PP forms, which seems to be the case of some Spanish American varieties, either in the present (as in our Lima data; see also: (Fløgstad, 2016; Jara Yupanqui, 2006; Kubarth, 1992) or in the past, as Moreno de Alba (1998) found for Mexican Spanish, and Rodríguez Louro (2009) for Argentinian Spanish. Overall, the linguistic variation across Spanish American varieties have been investigated only partially, and further research may shed light and provide new insights onto the processes and hypotheses related to the grammaticalization of the PP.

## CHAPTER VI

### CONCLUSION

In this work, by examining the speech communities of Lima and Mexico City, I have shed further light on the processes that govern the PP and preterit variation in Spanish. I have followed Schwenter and Torres Cacoullos (2008) in analyzing the effects of a set of linguistic variables on the PP or preterit choice, with the aim of overcoming the methodological difficulties of examining the functions of the PP from the subjective notion of “current relevance”. Moreover, I have included in my analysis three social variables (age, gender, and level of education) which have often proven correlate significantly with language variation in processes of linguistic change (including PP/ preterit variation in Peninsular Spanish), in order to find out whether the PP and the preterit are in stable variation, or show signs of a change in progress in the speech communities of Lima and Mexico City.

My results regarding the linguistic constraints in PP/preterit variation mirror the results of previous research. The PP in Mexico City and Lima shows similar effects of the most statistically significant linguistic constraints. It is favored in temporally irrelevant and indeterminate contexts and in presence of atelic adverbs. However, the overall frequency is higher in Lima, which also shows presence of the PP in contexts that do not allow the PP in Mexico City, as in pre-hodiernal contexts and with co-occurring connective adverbials.

The contribution to previous research is represented by the age effect found in Lima. To the best of my knowledge, only Jara Yupanqui (2006) found a similar effect in her exploratory study in Lima; however, her sample of participants only included speakers 37 years old and younger, whereas this study also includes older participants. Of course, this must be considered as a preliminary result, and more research is needed in order to confirm it. In this sense, a trend study to be conducted in the upcoming years, with a sample of speakers comparable to the one from the PRESEEA corpus examined here, would be an ideal way to test the age effect found in the current study.

In conclusion, I hope to have added a small piece to the complex puzzle of the variation between PP and preterit in Spanish, and more in general, in Romance languages.

## REFERENCES

- Alonso, A. & Henríquez Ureña, P. (1951). *Gramática castellana. Segundo curso* (10<sup>th</sup> ed.). Losada.
- A.S.A.L.E. & Real Academia Española. (2010). *Nueva Gramática de la Lengua Española: Manual*. Espasa Calpe.
- Ash, S. (2013). Social class. In J. K. Chamber, & N. Schilling-Estes (Eds.), *The Handbook of language variation and change* (2<sup>nd</sup> ed.) (pp. 350-367). Wiley.
- Barnes, S. (2015). Perceptual salience and social categorization of contact features in Asturian Spanish. *Studies in Hispanic and Lusophone Linguistics*, 8(2), 213–241. <https://doi.org/10.1515/shll-2015-0009>
- Bosque, I., & Demonte, V. (1999). *Gramática descriptiva de la lengua española* (Vol. 2). Espasa Calpe.
- Burgo, C. (2012). Social influence in the use of the present perfect in Bilbao. *Spanish in Context*, 9(1), 88–108. <https://doi.org/10.1075/sic.9.1.04bur>
- Bustamante, I. (1991). El presente perfecto o pretérito perfecto compuesto en el español quiteño. *Lexis*, 15, 195-231.
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins.
- Bybee, J. L., & Dahl, O. (1989). The creation of tense and aspect in the languages of the world. *Studies in Language*, 13(1), 51-103.
- Bybee, J. L., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. The University of Chicago Press.
- Caravedo, R. (1989). *El español de Lima: Materiales para el estudio del habla culta*. Pontificia Universidad Católica del Perú.
- Chappell, W. (2016). On the social perception of intervocalic /s/ voicing in Costa Rican Spanish. *Language Variation and Change*, 28(3), 357–378. <https://doi.org/10.1017/S0954394516000107>
- Copple, M. T. (2011). Tracking the constraints on a grammaticalizing perfect(ive). *Language Variation and Change*, 23(2), 163–191. <https://doi.org/10.1017/S0954394511000044>
- Dahl, Ö. (1985). *Tense and Aspect Systems*. Blackwell.

- Dahl, Ö. (Ed.). (2000). *Tense and Aspect in the Languages of Europe*. Mouton de Gruyter.
- Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, 1, 245–267. <https://doi.org/10.4324/9780429496288>
- El Habla Culta de la Ciudad de Buenos Aires. Materiales para su Estudio*. (1987). Universidad Nacional de Buenos Aires, Facultad de Filosofía y Letras.
- Erker, D. (2017). The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language and Linguistics Compass*, 11(1), 1–20. <https://doi.org/10.1111/lnc3.12232>
- Escobar, A. M. (1997). Contrastive and Innovative Uses of the Present Perfect and the Preterit in Spanish in Contact with Quechua. *Hispania*, 80(4), 859-870. <https://doi.org/10.2307/345107>
- Flogstad, G. N. (2016). *Preterit Expansion and Perfect Demise in Porteño Spanish and Beyond*. BRILL. <https://brill.com/view/title/32623>
- Flogstad, G. (2017). Revisiting Perfect/Preterit instability across Romance. On functional motivations for diverging paths. *Acta Linguistica Hafniensia*, 49(2), 195–211. <https://doi.org/10.1080/03740463.2017.1353200>
- García Tesoro, A. I., & Jang, J. S. (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. <https://doi.org/10.15446/fyf.v31n1.70444>
- Giacalone Ramat, A., & Hopper, P. J. (Eds.). (1998). *Limits of Grammaticalization*, John Benjamins Publishing Company.
- Givón, T. (1971). Historical syntax and synchronic morphology: an archaeologist's field trip. *Papers from the 7<sup>th</sup> meeting of the Chicago linguistic society*, pp. 394-415.
- Givón, T. (1979). *On Understanding Grammar*. Academic Press.
- González, P., Jara Yupanqui, M., & Kleinherenbrink, C. (2019). The microvariation of the Spanish perfect in three varieties. *Isogloss*, 4(1), 115-133. <https://doi.org/10.5565/rev/isogloss.60>
- Harris, M. (1982). The 'past simple' and 'present perfect' in Romance. In N. Vincent & M. Harris (Eds.), *Studies in the Romance verb* (pp. 42–70). Croom Helm.
- Hopper, P. J., & Traugott, E. (2003). *Grammaticalization* (2<sup>nd</sup> ed.). Cambridge University Press.

- Howe, C. (2006). Cross-dialectal features of the Spanish present perfect: A typological analysis of form and function [doctoral dissertation, Ohio State University]. <https://www.semanticscholar.org/paper/Cross-dialectal-features-of-the-Spanish-present-a-Howe/43cbfa8af54545a5867c94c31ce85a791256e0a7>
- Howe, C. (2013). *The Spanish Perfects: Pathways of Emergent Meaning*. Palgrave Macmillan.
- Howe, C., & Schwenter, S. A. (2003). Present Perfect for Preterit across Spanish Dialects. *University of Pennsylvania Working Papers in Linguistics*, 9(2), 61–75.
- Howe, C., & Schwenter, S. A. (2008). Variable constraints on past reference in dialects of Spanish. In: M. Westmoreland, & J. A. Thomas (Eds.), *Selected Proceedings of the 4<sup>th</sup> Workshop on Spanish Sociolinguistics* (pp.100-108). Cascadilla Proceedings Project.
- Janda, R. D. (2001). Beyond “pathways” and “unidirectionality”: On the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences*, 23(2–3), 265–340. [https://doi.org/10.1016/s0388-0001\(00\)00023-1](https://doi.org/10.1016/s0388-0001(00)00023-1)
- Jara Yupanqui, I. M. (2011a). Funciones discursivas y gramaticalización del pretérito perfecto compuesto en el español de Lima. *Spanish in Context*, 8(1), 95–118. <https://doi.org/10.1075/sic.8.1.05jar>
- Jara Yupanqui, I. M. (2011b). Present perfect usage in Peruvian Spanish and perfective readings in narratives. *Revista Internacional de Lingüística Iberoamericana*, Vol.9, 2(18), 213-235. <https://www.jstor.org/stable/41678478?seq=1>
- Jara Yupanqui, I. M. (2006). The use of the preterit and the present perfect in the Spanish of Lima [doctoral dissertation, University of Pittsburgh]. <http://d-scholarship.pitt.edu/8224/1/JaraIleanaMay042006.pdf>
- Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1), 359-383.
- Kempas, I. (2013). Sobre la elección entre ‘canté’ y ‘he cantado’ en el español hablado de Santiago de Estero. In L. Colantoni & C. Rodríguez Louro (Eds.), *Perspectivas teóricas y experimentales sobre el español de la Argentina* (pp. 269–70). Iberoamericana Vervuert.

- Klee, C., & Ocampo, A. (1995). The expression of past reference in Spanish narratives of Spanish-Quechua bilingual speakers. In C. Silva-Corvalán (Ed.), *Spanish in four continents: Studies in language contact and bilingualism* (pp. 52-70). Georgetown University Press.
- Kubarth, H. (1992). El uso del pretérito simple y compuesto en el español hablado de Buenos Aires. In H.L. Traill (Ed.), *Scripta Philologica*(pp. 553-566). Universidad Autónoma de México.
- Labov, W. (1990). The intersection of sex and social class in the course of the linguistic change. *Language Variation and Change*, 2, 205-254.
- Labov, W. (1994). *Principles of linguistic change. Internal factors*. Blackwell.
- Labov, W. (2001). *Principles of linguistic change. Social factors*. Blackwell.
- Leeman J., & Serafini, E. (2016). Sociolinguistics and heritage language education: A model for promoting critical translanguing competence. In M. Fairlough & S. Beaudrie (Eds.), *Innovative strategies for heritage language teaching* (pp.56-79). Georgetown University Press.
- Lehmann, C. (2015). *Thoughts on grammaticalization* (3<sup>rd</sup> ed.). Language Science Press. <https://library.oapen.org/handle/20.500.12657/32841>
- Lope Blanch, J. M. (Ed.). (1971). *El habla de la Ciudad de México: Materiales para su estudio*. Universidad Nacional Autónoma de México.
- Lope Blanch, J. M. (Ed.). (1972). *Estudios sobre el español de México*. Universidad Nacional Autónoma de México.
- Lope Blanch, J. M. (Ed.). (1976). *El habla popular de México: Materiales para su estudio*. Universidad Nacional Autónoma de México.
- López Morales, H. (1996). Rasgos generales. In M. Alvar (Ed.), *Manual de dialectología hispánica: El español de América*. Ariel.
- Meillet, A. (1912). L'évolution des formes grammaticales. *Scientia*, vol. XII, 24(6), reprinted in A. Meillet (Ed.), (1958). *Linguistique historique et linguistique générale* (pp. 130-148). Champion.
- Milroy, L. & Gordon, M. (2003). *Sociolinguistics: Method and Interpretation*. Blackwell.
- Milroy, L., & Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21(1), 1-26.

- Moreno de Alba, J. G. (1978). *Valores de las formas verbales en el español de México*. Universidad Nacional Autónoma de México.
- Moreno de Alba, J. G. (1998). La oposición Pretérito Indefinido / Pretérito Perfecto Compuesto en documentos novohispanos de los siglos XVI-XIX. In *Actas Del IV Congreso Internacional de Historia de La Lengua Española* (Vol. I) (pp. 619–629). Universidad de La Rioja.
- Moreno Fernández, F. (2005). Project for the sociolinguistic study of Spanish from Spain and America (PRESEEA): A corpus with a grammar and discourse bias. In T. Takagaki, S Zaima, Y. Tsuruga, & F. Moreno-Fernández (Eds.), *Corpus-based approaches to sentence structures* (pp. 265-288). John Benjamin Publishing Company.
- Perkins, R. D. (1980). *The evolution of culture and grammar* [Doctoral dissertation, State University of New York at Buffalo].
- Pfänder, S., & Palacios, A. (2013). Evidencialidad y validación en los pretéritos del español andino ecuatoriano. *Círculo de Lingüística Aplicada a la Comunicación*, 54, 65-98.
- Poplack, S., & Tagliamonte, S. A. (2001). *African American English in the Diaspora: Tense and Aspect*. Blackwell.
- PRESEEA (2014-): *Corpus del Proyecto para el estudio sociolingüístico del español de España y América*. Universidad de Alcalá.  
<https://preseea.linguas.net/Corpus.aspx>
- R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org>
- Rodríguez Louro, C. (2009). Perfect evolution and change: A sociolinguistic study of Preterit and Present Perfect usage in contemporary and earlier Argentina [Doctoral dissertation, University of Melbourne]. [https://minerva-access.unimelb.edu.au/bitstream/handle/11343/35140/114322\\_Rodriguez%20Louro%20PhD%20Thesis.pdf?sequence=1](https://minerva-access.unimelb.edu.au/bitstream/handle/11343/35140/114322_Rodriguez%20Louro%20PhD%20Thesis.pdf?sequence=1)
- Rodríguez Louro, C., & Howe, C. (2010). Perfect potential: Semantic change in narrative contexts across Spanish. *Revista Internacional de Lingüística Iberoamericana*, 8(2), 157–174
- Rodríguez Louro, C., & Jara Yupanqui, I. M. (2011). Otra mirada a los procesos de gramaticalización del Presente Perfecto en español: Perú y Argentina. *Studies in Hispanic and Lusophone Linguistics*, 4(1). <https://doi.org/10.1515/shll-2011-1091>



- Rohlfs, G. (1966-1969). *Grammatica storica della lingua italiana e dei suoi dialetti*. Einaudi.
- Sankoff, D., Tagliamonte, S., & Smith, E. (2005). Goldvarb X. Computer program. Department of Linguistics, University of Toronto, Canada.  
<http://individual.utoronto.ca/tagliamonte/goldvarb.html>
- Schwenter, S. A. (1994a). The Grammaticalization of an Anterior in Progress. *Studies in Language* 18(1), 71–111. <https://doi.org/10.1075/sl.18.1.05sch>.
- Schwenter, S. A. (1994b). “Hot news” and the grammaticalization of perfects. *Linguistics*, 32(6), 995-1028.
- Schwenter, S. A., & Torres Cacoullos, R. (2008). Defaults and Indeterminacy in Temporal Grammaticalization: The ‘perfect’ Road to Perfective. *Language Variation and Change* 20(1), 1–39.  
<https://doi.org/10.1017/S0954394508000057>
- Serrano, M. J. (1996). Sobre el uso del pretérito perfecto y pretérito indefinido en el español de Canarias: pragmática y variación. *Boletín de Filología*, 35(1), 534-566. <https://boletinfilologia.uchile.cl/index.php/BDF/article/view/19197/20320>
- Serrano, M. J. (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.
- Squartini, M., & Bertinetto, P. M. (2000). The Simple and Compound Past in Romance Languages. In: Ö. Dahl (Ed.), *Tense and Aspect in the Languages of Europe* (pp. 403–40). Mouton de Gruyter.
- Tagliamonte, S. A. (2012). Comparative sociolinguistics. In S. A. Tagliamonte (Ed.). *Variationist Sociolinguistics : Change, observation, interpretation*, John Wiley & Sons, Inc.
- Terán, V., & Kanwit, M. (2018). Variable past-time expression across multiple tasks in Tucumán, Argentina. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 31(2), 605–637.  
<https://doi.org/10.1075/resla.16013.ter>
- Torres Cacoullos, R. (2011). Grammaticalization and Neutralization-in-Discourse Hypothesis. In: M. Díaz-Campos (Ed.). *The Handbook of Hispanic Sociolinguistics* (pp. 148-67). Wiley-Blackwell.
- Westmoreland, M. (1988). The distribution and the use of the present perfect and the past perfect forms in American Spanish. *Hispania*, 71(2), 379-84.

**APPENDICES: PARTICIPANT DEMOGRAPHIC DATA**

**APPENDIX A****PARTICIPANT DEMOGRAPHIC DATA FOR LIMA (PRESEEA)**

Table 9. Participant demographic data for Lima (PRESEEA)

<b>Participant</b>	<b>Gender</b>	<b>Age</b>	<b>Age group</b>	<b>Education</b>	<b>Education level</b>	<b>Profession</b>	<b>Origin</b>
LIMA_H11_011-1	Male	26	Youngest	Elementary school	Basic	Waiter and laborer	Lima
LIMA_H12_028	Male	33	Youngest	High school	Medium	Actor	Lima
LIMA_H13_047	Male	28	Youngest	University	High	Historian	Lima
LIMA_M11_003	Female	26	Youngest	Elementary school	Basic	Housekeeper and seller	Lima
LIMA_M12_020	Female	30	Youngest	High school	Medium	Secretary	Lima
LIMA_M13_037	Female	27	Youngest	University	High	Dancer	Lima
LIMA_H21_014	Male	47	Middle	Elementary school	Basic	Welder	Lima
LIMA_H22_033	Male	48	Middle	High school	Medium	Police officer	Lima
LIMA_H23_050	Male	38	Middle	University	High	Historian	Lima
LIMA_M21_005	Female	36	Middle	Elementary school	Basic	Housekeeper	Lima
LIMA_M22_023	Female	45	Middle	High school	Medium	Housekeeper	Lima

Table 9. Continued

<b>Participant</b>	<b>Gender</b>	<b>Age</b>	<b>Age group</b>	<b>Education</b>	<b>Education level</b>	<b>Profession</b>	<b>Origin</b>
LIMA_M23_042	Female	40	Middle	University	High	Psychologist	Lima
LIMA_H31_018	Male	81	Oldest	Elementary school	Basic	Retired	Lima
LIMA_H32_036	Male	78	Oldest	High school	Medium	Shoemaker	Lima
LIMA_H33_053	Male	63	Oldest	University	High	Engineer	Lima
LIMA_M31_009	Female	55	Oldest	Elementary school	Basic	Housekeeper	Lima
LIMA_M32_025	Female	70	Oldest	High school	Medium	Housekeeper	Lima
LIMA_M33_043	Female	64	Oldest	University	High	Social worker	Lima

**APPENDIX B****PARTICIPANT DEMOGRAPHIC DATA FOR MEXICO CITY (PRESEEA)**

Table 10. Participant demographic data for Mexico City (PRESEEA)

<b>Participant</b>	<b>Gender</b>	<b>Age</b>	<b>Age group</b>	<b>Education</b>	<b>Education level</b>	<b>Profession</b>	<b>Origin</b>
MEXI_H11_078	Male	27	Youngest	Elementary school	Basic	Waiter	Mexico City
MEXI_H12_042	Male	20	Youngest	High school	Medium	Student	Mexico City
MEXI_H13_006	Male	24	Youngest	University	High	Architect	Mexico City
MEXI_M11_084	Female	21	Youngest	Elementary school (not completed)	Basic	Cleaner	Mexico City
MEXI_M12_048	Female	27	Youngest	High school (not completed)	Medium	Dancer and model	Mexico City
MEXI_M13_012	Female	21	Youngest	Medical school	High	Student	Mexico City
MEXI_H21_090	Male	52	Middle	Elementary school	Basic	Blacksmith	Mexico City
MEXI_H22_054	Male	40	Middle	High school	Medium	Security guard, painter and musician	Mexico City
MEXI_H23_018	Male	45	Middle	University	High	Biologist	Mexico City

Table 10. Continued

<b>Participant</b>	<b>Gender</b>	<b>Age</b>	<b>Age group</b>	<b>Education</b>	<b>Education level</b>	<b>Profession</b>	<b>Origin</b>
MEXI_M21_096	Female	37	Middle	Elementary school	Basic	Cleaner	Mexico City
MEXI_M22_060	Female	51	Middle	High school	Medium	Post office director	Mexico City
MEXI_M23_024	Female	40	Middle	University	High	Historian and dancer	Mexico City
MEXI_H31_102	Male	75	Oldest	No education	Basic	Book promoter	Mexico City
MEXI_H32_066	Male	59	Oldest	High school of business	Medium	Retired	Tepechitlán, Zacatecas. He lives in Mexico City since he was 7 years old.
MEXI_H33_030	Male	72	Oldest	School of theatre	High	Actor and musician	Puebla. He moved to Mexico City before he was 7 years old.
MEXI_M31_108	Female	78	Oldest	No education	Basic	Housekeeper	El Campanario, Hidalgo
MEXI_M32_072	Female	56	Oldest	High school	Medium	Teacher and seller	Mexico City
MEXI_M33_036	Female	58	Oldest	University	High	Teacher and researcher	Estado de Jalisco