Perfect-perfective variation across Spanish dialects: a parallel-corpus study

Fuchs, M.; Gonzalez Gonzalez, P.

languages

MDPI

Article

Perfect-Perfective Variation across Spanish Dialects: A Parallel-Corpus Study

Martín Fuchs [1,*] and Paz González [2]

[1] Utrecht Institute for Linguistics OTS, Utrecht University, 3512 JK Utrecht, The Netherlands
[2] Leiden University Centre for Linguistics, Leiden University, 2311 BX Leiden, The Netherlands; p.gonzalez@hum.leidenuniv.nl
* Correspondence: m.fuchs@uu.nl

Abstract: To analyze crossdialectal variation between the use of a Present Perfect form (*Pretérito Perfecto Compuesto*) and a Perfective Past form (*Pretérito Indefinido*) in Spanish, we make use of two converging methodologies: (i) parallel corpus research, where we compare different translations of the same text (*Harry Potter and the Philosopher's Stone*) into specific standardized written varieties of Spanish (Peninsular, Mexican, Argentinian), and (ii) an elicitation forced-choice task, where native speakers of each of the cities in which these standardized written norms are produced (Madrid, Mexico City, Buenos Aires) have to choose between the *Pretérito Perfecto Compuesto* and the *Pretérito Indefinido* as the most natural filler for a blank in contexts extracted from the novel. Results from these two tasks do not align completely. While the data from our parallel corpus work indicate a wider distribution of Perfect use in the Mexican translation than in the Peninsular and the Argentinian ones, the elicitation task shows that only the choices of the speakers of Madrid (Castilian Spanish) and Buenos Aires (Rioplatense Spanish) converge with their respective translations patterns. Since the distribution observed in the Mexican translation not only goes against the elicitation data, but also contradicts previous findings in the literature, we abandon it in further analyses. In the second part of the paper, through a detailed annotation of the Peninsular and Argentinian corpora, we show that the constraints allowing Perfect use in each of these standardized varieties respond only to some features previously advanced in the literature. While both dialects allow for experiential and resultative readings of the *Pretérito Perfecto Compuesto*, Castilian Spanish also prefers the use of this marker to locate an event in the hodiernal past. On the other hand, Rioplatense Spanish systematically defaults to the *Pretérito Indefinido* in these cases, displaying a more restricted distribution for the Perfect form. Both dialects also seem to exhibit a preference for the *Pretérito Perfecto Compuesto* in continuative contexts. Our work thus provides two crucial take-home messages: (i) understanding crossdialectal variation in written language is crucial for advancing crosslinguistic generalizations about tense-aspect phenomena; and (ii) combining parallel corpus and experimental methodologies can help us understand in a more thorough way the distribution of Perfect and (Perfective) Past forms across dialects.

Keywords: Perfect; Perfective Past; dialect variation; Spanish; parallel corpus; forced-choice task

Citation: Fuchs, Martín, and Paz González. 2022. Perfect-Perfective Variation across Spanish Dialects: A Parallel-Corpus Study. *Languages* 7: 166. https://doi.org/10.3390/languages7030166

Academic Editors: Henriëtte de Swart and Bert Le Bruyn

Received: 27 September 2021
Accepted: 13 June 2022
Published: 1 July 2022

1. Introduction

Most theoretical accounts that address the variation in use between forms that instantiate Perfect and Perfective meanings are based predominantly on English data (e.g., McCoard 1978; McCawley 1981; Michaelis 1994; Portner 2003; Nishiyama and Koening 2010), although some recent work has started to consider other European languages, such as German, French, and Dutch (e.g., Löbner 2002; Vet 1992; Schaden 2009; Le Bruyn et al. 2019), or even a crosslinguistic perspective (e.g., de Swart 2007; van der Klis et al. 2021; Bertrand et al. 2022). The main goal of this body of research has been to arrive at a concrete semantics for the Perfect, in order to account for the distribution of (PRESENT) PERFECT[1]

and (PERFECTIVE) PAST forms in each of these languages. While this partial movement towards testing the crosslinguistic validity of the proposed semantic generalizations is refreshing, these studies usually take a macro-variation perspective and fail to recognize that languages are not uniform entities, and that dialectal micro-variation might also show relevant constraints at play in the distribution of these forms.

In this study, we propose to take a deeper look at Spanish, a language that presents two forms in competition within the Perfect-Perfective domain: the *Pretérito Perfecto Compuesto* (a periphrastic PERFECT form constituted by the auxiliary *haber* 'to have' and the past participle; PPC, henceforth) and the *Pretérito Indefinido* (a simple, (PERFECTIVE) PAST form; PI, henceforth), as in (1) and (2), respectively:[2]

(1)	Ana	ha	desayunado	en	ese	café
	Ana	have.3SG.PRS	have-breakfast.PTCP	at	that	café
	'Ana has had breakfast at that café'					

(2)	Ana	desayunó	en	ese	café
	Ana	have-breakfast.3SG.PST.PRFV	at	that	café
	'Ana had breakfast at that café'				

These two forms are used in Spanish to refer to past events, and there have been several proposals to account for their distribution (e.g., Alarcos Llorach 1947; Harris 1982; Westmoreland 1988; Schwenter 1994; Schwenter and Torres Cacoullos 2008; González and Verkuyl 2017). The main established generalization is that the PPC can make reference to past events as far as they have occurred in the day of utterance, while the PI needs to be used otherwise (e.g., Berschin 1975, 1976; Fleischman 1983). However, these previous descriptions have generally focused on Castilian[3] or other Peninsular varieties, not recognizing that this population only represents about 10% of the +500 million native speakers of Spanish, and that different dialectal varieties of Spanish can partition this semantic space in similar yet different ways.

To address this shortcoming, here we make use of parallel corpus research, a promising methodology to analyze crosslinguistic and crossdialectal variation, since it relies on comparing translations of the same text into different languages or dialectal varieties. Using this method, we perform a piecemeal comparison of three different standardized written varieties of Spanish: Peninsular, Mexican and Argentinian. We consider that these written varieties aim to represent three dialects spoken within those geographical areas, which are generally considered to constitute their corresponding norms: Castilian, Mexican Altiplano[4], and Rioplatense.[5] We argue that the translator of each written standard variety is compelled to use a dialect-appropriate form in a set of identical contexts across translations that are supposed to convey the same meaning, therefore allowing to compare the varieties with respect to the choice of marker in this tense-aspect domain. In this way, crucial distinctions in the distribution across dialectal varieties can be revealed, providing a more fine-grained picture of the distribution of the markers in the language as a whole. That said, assuming that one translator's idiolect, and the corresponding standardized variety that they write in when they produce a translation for a wide geographical area, represents the dialect of the biggest city in the distribution area with respect to the grammatical feature under analysis can be misguided. Therefore, we also present the results of an elicitation forced-choice task performed by native speakers of the dialects we are considering, and compare them with the results from the parallel corpus study, before moving forward with further annotations on our corpora. In this way, making use of two complementary methodologies allows us to provide a variety of data sources that can shed light on the issue of Perfect-Perfective Past crossdialectal variation in Spanish.

This paper is structured as follows. In Section 2, we first introduce the problem of Perfect-Perfective Past variation crosslinguistically (Section 2.1), we then look at it from the perspective of Spanish and present previous accounts of the alternation between the PPC and the PI (Section 2.2), and finally discuss insights from previous research on dialectal variation on this domain (Section 2.3). In Section 3, we present the primary methodology

of our study: the use of parallel corpora and the application of multidimensional scaling to visualize crossdialectal variation patterns. Section 4 presents the results of applying this method to our data, which motivate a second round of manual annotation of the tokens under examination. Before proceeding with that annotation, in Section 5, we present an elicitation forced-choice task performed by native speakers of each dialect. This task provides us with an alternative source of data to compare the claims in the previous literature with the patterns emerging from the parallel corpus approach. The results of this task direct us to leave the Mexican translation out from further analysis. Section 6 describes the second annotation, which considers the effect of different linguistic variables in the choice of PPC or PI in the Rioplatense and Castilian translations. Section 7 presents the results of this process, Section 8 discusses them, and Section 9 advances some general conclusions and reflects on the use of two converging methodologies to analyze crossdialectal semantic variation.

2. Perfect-Perfective Past Variation

2.1. Present Perfect and Perfective Past Markers Crosslinguistically

In terms of Reichenbach (1947), it is possible to define the Perfect and the Past in a pretty simple way. While these meanings share their temporal semantics, such that the time of the event precedes the time of speech (E–S), they differ with respect to their relation to reference time: the Past aligns reference time with the event time (E,R–S), and the Perfect does so with the speech time (E–R,S). Thus, the semantic difference between these two meanings is usually understood as being of aspectual nature: the Perfect conveys completion in relation to the Present, while the Past also marks anteriority in an aorist-like way (e.g., Verkuyl 1999; González 2003).

This relation to the Present allows forms that encode Perfect meaning to convey a set of different readings, which have been thoroughly described in the literature on the English *Present Perfect* (e.g., McCawley 1971; McCoard 1978; Portner 2003):

(3) Resultative: Mary has read *Middlemarch*. [Portner 2003: 459, ex. (2)]

(4) Experiential: Mary has gone to that bar (and she might go again).

(5) Continuative: Mary has lived in London for 5 years (and she still does).

(6) Hot news/hodiernal: Mary has won the contest (just now)

The resultative reading, as in (3), focuses on the result state that holds in the present as a result of a past event (i.e., Mary has read Eliot's novel, so now she has that knowledge, she is in the state of having read it). In (4), we observe the experiential reading, where the focus is on the fact that this event occurred at some past point in time, and is now part of Mary's experience. The continuative reading in (5) only occurs with stative predicates and indicates that the state continues to hold into the present. Finally, the so-called 'hot news' reading in (6) implies that the event has just happened.

Arriving at a unified semantics that can account for all these readings has been the focus of most research related to the Perfect (e.g., Klein 1992; Iatridou et al. 2001; Portner 2003). One crucial problem of this approach is that most of this research has focused on English (though Iatridou et al. 2001 also deal with Bulgarian and Greek), even if, for example, French or Dutch use their PERFECT forms with a broader distribution. For instance, a sentence including a locating past-referring adverbial such as *yesterday*, in a sentence such as *Yesterday, Mary *has eaten/ate breakfast at that café* cannot appear with the *Present Perfect* in English and needs to default to the *Simple Past*. However, the same sentence can be easily translated to French (*Hier, Mary a pris le petit-déjeneur dans ce café)* or Dutch (*Gisteren heeft Mary in dat café ontbeten)*, and have both these languages use their corresponding PERFECT markers (French *Passé Composé: 'a pris'*/Dutch *Voltooid Tegenwoordige Tijd*: *'heeft ontbeten'*) instead of their (PERFECTIVE) PAST ones (French *Passé Simple: 'prit'*/Dutch *Onvoltooid Verleden Tijd*: *'ontbeet'*).

2.2. The Use of the Pretérito Perfecto Compuesto and the Pretérito Indefinido

Spanish has an incredible range of dialectal variation in its verbal paradigm; more specifically, it is in the use of its past tenses that Spanish shows most of its variation. For instance, the Real Academia Española—the institution that intends to ensure the stability of the Spanish language—points out that the PPC is the verbal form that shows most geographical variation in its usage (NGRAE 2009). Variation in Spanish temporo-aspectual representations has been a well-studied phenomenon and is still a current research focus in variation studies (Howe 2006; Schwenter and Torres Cacoullos 2008; Azpiazu 2013, 2014, 2015; González et al. 2019; Fuchs 2020; González and Kleinherenbrink 2021; among others).

The distinction between the use of the PPC and the PI is not clear-cut and the constraints on their distribution are not well-understood; that is, it is relatively unclear what conditions trigger the use of one or the other marker. As Azpiazu (2013, p. 10) explains, the PPC and the PI are said to be opposed under temporal criteria. The Castilian norm in the Grammar of the Real Academia (NGRAE 2009) states that the PI describes an event in the past whose boundary is also in the past, while the PPC designates an event in the past but that is approximated to the 'now'. A few studies have shown that Castilian Spanish uses the PPC not only for typical Perfect readings (e.g., resultative, continuative), expressing a completed event with current relevance (e.g., Alarcos Llorach 1970), but also for Perfective Past meanings, such as the expression of bounded events, as long as they are contained in the hodiernal past; i.e., events that have occurred within the day of utterance (Schwenter 1994). Conversely, the PI is said to express completed events but without current relevance, detached from the present, and anchored to a specific past reference point before the day of utterance. Relatedly, González and Verkuyl (2017) characterize the Spanish tense forms binarily except for the PI, a semantic outsider, with aorist interpretation.

In any case, most of this research has been carried out in Peninsular dialects, with specific attention to Central ones, which are spoken by only one tenth of the Spanish speaking population, while constraints on tense-aspect marker variation have been shown to reflect large dialectal differences (e.g., Howe 2006; González et al. 2019; Fuchs 2020). A few studies have analyzed the distribution of these forms in Latin American dialects in comparison to their (Central) Peninsular use. We summarize them directly below.

2.3. Dialectal Variation in Spanish in the Perfect-Perfective Domain: Synchrony and Diachrony

The general observation that shows up in all dialectal studies on PPC-PI variation is that the different Peninsular varieties in general, and the Spanish spoken in Central Spain in particular, allow for more uses of the PPC than all Latin American varieties when considered as a whole, though there is no consensus on how to explain this crossdialectal variation (e.g., Kany 1945; Westmoreland 1988; De Kock 1989; Howe 2006; Schwenter and Torres Cacoullos 2008; among others).

Schwenter (1994) shows through a series of production tasks that in Alicante, Spain, the PPC is the preferred form to refer to past events in *hodiernal* temporal contexts (i.e., past events that occurred within the day of utterance). This study also shows that this marker can appear in *prehodiernal* contexts too (i.e., temporal contexts before the day of utterance), but it is not the preferred marker in those cases, where the dialect resorts to the PI. This picture contrasts with other dialectal varieties, especially Latin American ones, where variation has also been reported. Squartini and Bertinetto (2000) and Camus Bergareche (2008) point out that the main characteristic of the PPC across all Latin America is that it is used in durative and iterative contexts but never in recent past contexts, such as the ones described by Schwenter (1994). However, De Jonge (2001) finds that in Argentina the PPC can highlight a situation within a narration and be used to refer to even more remote events. This pattern seems to be further extended in Perú, where, according to Howe (2006), this form can be used to refer to both recent and remote events. On the other hand, Rodríguez Louro and Jara Yupanqui (2011) and Jara Yupanqui (2012) show that Andean varieties of Peruvian Spanish have a preference for the PI, even if they show presence of the PPC in almost all contexts of everyday use. In the case of the Mexican Altiplano dialectal variety,

there is a predominance of the use of the PPC only in continuative contexts, while most past events are expressed with the PI (Schwenter and Torres Cacoullos 2008; Soto 2015; among others). More recently, González et al. (2019) show that for native speakers of Castilian, Peruvian and Rioplatense Spanish, the PI is also the preferred form in a variety of contexts. For instance, in Rioplatense Spanish the PI is much more frequent than the PPC, on a 9:1 proportion, but the authors show that the PPC competes with the PI in continuative contexts (1:1 proportion), and only fails to appear in experiential, resultative, and hodiernal contexts. This contrasts with previous studies in this variety, where the PPC was shown to appear in experiential indefinite past contexts, and only minimally in resultative and continuative settings (Rodríguez Louro 2009, 2012). This whole set of studies strengthens the idea that there is no linguistic consistency across the different Spanish dialectal varieties. However, one result definitely stands out: PPC forms are often used in Spain, while in Latin America the PI is the default past tense per excellence (Howe 2006; González et al. 2019).

A potential source of dialectal variation between PERFECT-PERFECTIVE forms can be ascribed to different diachronic trajectories in the development of these markers across the different dialects. Crosslinguistically, PERFECT-PERFECTIVE variation has been shown to be partially constrained by a diachronic cline known as the 'aoristic drift' (see e.g., Harris 1982; Fleischman 1983; Bybee et al. 1994; Condoravdi and Deo 2014), in which PERFECT markers first express resultative readings, then continuative and experiential ones, to finally be able to express events located in the recent past/hodiernal past, before encroaching into the domain of PERFECTIVE markers more generally. The hodiernality feature has been said to be particularly relevant in the case of (Romance) PRESENT PERFECT forms (Harris 1982).

This hodiernal/prehodiernal distinction (Dahl 1985) is actually applied by Schwenter (1994) to pursue the idea that the Spanish PPC has evolved from Perfect to hodiernal Perfective meaning, subsequently extending to convey more remote past time events (Bybee et al. 1994; Fleischman 1983). Some years later, Schwenter and Torres Cacoullos (2008) proposed a different path of grammaticalization, where the shift was found in temporally indeterminate past contexts that lack specific temporal reference. In their study, the Castilian PPC is already favored in resultative, continuative, and recent past contexts with present relevance. The increase of PPC use in contexts without temporal reference is justified by these authors by an overestimation of present relevance by the speaker (Schaden 2012). In contexts with overt temporal reference, Azpiazu (2015) finds that the use of the PPC is also constrained by pragmatic or extralinguistic factors: the PPC is allowed to combine with prehodiernal adverbials, with the result of presenting the narrative as closer to the hearer. In a more recent paper, Azpiazu (2021) concurrently shows that the PPC in co-occurrence with prehodiernal adverbials is mainly found in the narration of personal experiences or personal narratives, as opposed to fictional and historical narration.

As González and Verkuyl (2017) discuss, the difference between the PPC and the PI does not appear to be theoretically unbridgeable. There is a recurring idea, already from grammaticalization studies, that there is a tendency to move towards a system where only two past tenses are needed, a division that Spanish assigns to the Perfective-Imperfective aspectual distinction. Thus, the dialectal differences across Spanish varieties seem to indicate different constraints in the use of PERFECT and PERFECTIVE PAST markers across the dialects, ultimately producing diverse diachronic trajectories that will lead into just one PERFECT(IVE) PAST marker. For Castilian Spanish, the prediction is that the PI will be pushed back to a niche, as already has happened in some areas (Schwenter 1994; Howe 2006), while in Latin American Spanish, the PI might adopt the semantic properties of a disappearing PPC. In both cases, though, a clear reduction of tense forms seems to be taking place.

In summary, the picture on the distribution of the PPC and the PI across dialects of Spanish is not that clear. There are two uses in the Castilian variety that do not show up in the Latin American dialects: the 'hodiernal' one (the event is over but it is located on the day of speech), and the 'recent past' one, which locates the event in the immediate recency. In both cases, the event is close to speech time, and there is the speaker's desire to include

them within that 'extended present' (Azpiazu 2013, p. 22). However, even in these cases, there seems to be overlap and alternation with the PI, and it is unclear how frequencies between the two markers are distributed with respect to the different uses.

To advance the understanding of the distribution between PPC and PI in Spanish, we analyze here variation in the use of these markers across three dialects: Castilian, Mexican Altiplano and Rioplatense. A parallel corpus perspective can be useful in clarifying the distributional picture across dialects, since it allows us to keep the meaning/context constant, and observe whether the PPC or the PI is chosen by the translator to replace a given marker in the original source text. When taking this methodological perspective, however, there is a caveat about dialectal variation that is worth mentioning: we are dealing with written standards of these dialects that reflect the norm—and not the actual speech—of each of these geographical regions. However, to the extent that we are dealing with *grammatical* variation, we expect to see in the standardized written translations the same patterns that are part of native speakers' grammars. In any case, when translating, should translators use a local variety of a language or not? Should they use their own dialect or one that can be understood more widely in other surrounding countries where the translated book might also be sold? While this is a hard decision for a translator to make, it is most of the times the publishing house who decides which is the concrete variety that will be used in a translation, which also goes through an editorial process after the translator has done their job (Sundell 2010, p. 7). In the following sections, we will see how this becomes something that needs to be taken into account when studying crossdialectal variation with parallel corpus methods, and how some translations align better than others with other dialectal data sources, such as native speaker's choices in an elicitation task.

3. Methodology: Parallel Corpus and Multidimensional Scaling

We use a new methodology in which parallel corpora are analyzed to compare crosslinguistic data (e.g., Wälchli and Cysouw 2012; van der Klis et al. 2017, 2021; Bogaards 2022; Mulder et al. 2022). We gather our data from three different Spanish translations of chapters 1, 16 and 17 of J.K. Rowling's *Harry Potter and the Philosopher's Stone.* We use these chapters because they contain a fair amount of dialogue, and the alternation between PERFECT and (SIMPLE/PERFECTIVE) PAST forms is not observed in narrative sequences in English and Spanish, where (SIMPLE/PERFECTIVE) PAST forms are used exclusively. This method takes a form-based approach since it identifies specific constructions based on their morphology (e.g., a combination of the verb *have* and a past participle for the English *Present Perfect*). Later, the source text is aligned with the translations and a human annotator needs to identify the matching grammatical form in the target language (e.g., a combination of the verb *haber* 'to have' and a past participle in Spanish, which would be the *Pretérito Perfecto Compuesto*, or a different verbal form that indicates the *Pretérito Indefinido*). Using parallel corpora thus grants a fine-grained comparison between the translations since the context of use is kept stable, while each specific translation is assumed to display the form that better conveys that constant meaning in the language or dialect under consideration.

The three different translations of *Harry Potter and the Philosopher's Stone* into Spanish follow the publishing house decision to subdivide the Spanish speaking world in three large geographical areas and produce three adapted versions, one per region: Spain, Latin America, and the Southern Cone (that is, the region that includes the southernmost part of Latin America (for bibliographical information on the different editions, please see Appendix A)). "Latin America" becomes, under this publishing house decision, every region in the Spanish-speaking portion of The Americas that is not Argentina, Uruguay, and Chile.[6] The publishing cities in these broad regions were Barcelona[7] (for the edition that represents the standard norm of the Spanish spoken in Spain), Mexico City (for the edition that represents the standard norm spoken in "Latin America"), and Buenos Aires, Argentina (for the edition that represents the standard norm spoken in the Southern Cone). We explore whether these standard written norms in the translations are able to

characterize the grammatical constraints that native speakers' adhere to in the following dialects: Castilian Spanish, Mexican Altiplano Spanish, and Rioplatense Spanish.[8]

According to a Harry Potter fan website (Potterglott 2022), which reviewed more than 32 different Spanish editions of *Harry Potter and the Philosopher's Stone*, and to a book that explains the distribution of the Harry Potter series in the Spanish-speaking world (Tarantino 2018, pp. 157–81), the first Spanish translation was done by Alicia Dellepiane Rawson, an Argentinian poet from Buenos Aires, for the Argentinian publishing house Emecé. This Argentinian publishing house was the first rights-holder for the Spanish translation of *Harry Potter and the Philosopher's Stone*, but they published the book in Spain by a subsidiary of them, Salamandra. In early 2001, Emecé was acquired by a larger media conglomerate, but Salamandra was not part of the acquisition, and became an independent publishing house, retaining the rights to *Harry Potter*, and started to publish the whole series from that point on in Spain. Salamandra became then the publisher and the distributor in Spain, but had an agreement with Emecé and Océano (a Mexican publishing house) as the corresponding distribution houses in the Southern Cone and in the rest of Latin America. However, the first translation from 1999—intended for the whole Spanish-speaking world—was highly criticized for having a lot of Argentinian regionalisms (Sundell 2010), so in 2003 Salamandra decided that the book should be produced in three different versions, one for each of the regions described above, which were the intended commercial markets of the publishing house, also determined by the distribution companies.[9]

Potterglott (2022) also explains that identifying the different versions is relatively easy, given two parameters: first, in chapter one, the English original refers to a *tawny owl*, that becomes a *lechuza rojiza* in the Southern Cone translation, and a *lechuza parda* in the translations for Latin America and Spain; second, in the same chapter, Mr. Dursley refers to his son as a *little tyke* in English, which becomes *Chiquilín* in the Southern Cone, *Diablillo* in the rest of Latin America, and *Tunante* in Spain. These two lexical choices make it easy to check which version one is dealing with. In our case, all the versions that we worked with complied with these features. We also confirmed the different versions on the basis of an additional linguistic factor that allows to separate the translation intended for Spain from the other two. One of the most noticeable differences between Castilian Spanish and the Spanish spoken in all of Latin America can be found in the pronoun system. The Spanish 2nd person distinguishes between singular and plural and between familiar and formal; that is, there are potentially 4 different forms that can be translated into English *you*. In the case of the plural, Spain has a formal/familiar distinction at play reflected in the existence of two different words: *vosotros* (familiar) and *ustedes* (formal). This distinction is lost in Latin America, where there is only one word, *ustedes,* that covers both uses. This crossdialectal pattern was observed in the translations we used.

Once we were confident about the different translations we were using, we proceeded to use a specific parallel corpus methodology, *Translation Mining*, based on the work by van der Klis et al. (2017). The first step in this method is to obtain the set of all grammatical forms of interest. To that end, we extracted all cases where either the original in English uses a *Present Perfect* or a *Simple Past*, or at least one of the Spanish translations shows a PPC or a PI. We left aside cases in which at least one Spanish translation uses a *Pretérito Imperfecto* to translate any of the two English forms (n = 229), since this is not an aspectual distinction marked through grammatical forms in English. The extraction was made through the *Perfect Extractor*, a system developed by van der Klis et al. (2017), which searches the original text for different grammatical configurations (e.g., for the PERFECT, it looks for structures that include any form of the auxiliary HAVE and the past participle in each language/translation). The sentences are later automatically aligned across translations, and a human annotator enters the correspondence between grammatical markers in the source language (in this case, English), and the target language (in this case, each Spanish translation), through the *Time Align* application. An example of this application can be seen in Figure 1.

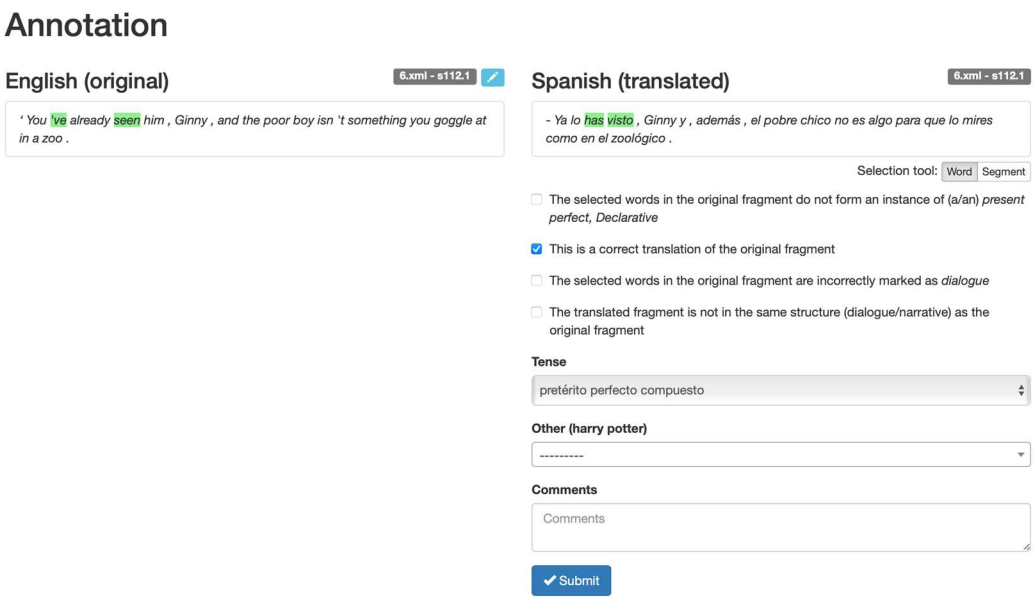

Figure 1. An example of the *Time Align* interface between English and (Castilian) Spanish.

In this case, the *Perfect Extractor* has indicated *'ve [. . .] seen* as the *Present Perfect* form under consideration in English, and the annotator needs to select *has visto* in the (Castilian) Spanish translation. The annotator can also indicate if the translation is not correct (according to their knowledge of the target language), if the extraction has been done incorrectly, or if the alignment is incorrect. Once these parameters are checked, the annotator also needs to indicate the corresponding Spanish marker under the tense subfield (in this case, a *Pretérito Perfecto Compuesto*). Using this method for each of the Spanish translations allows us to combine the annotations in each individual translation and look at the patterns that emerge from these combinations. For instance, are all English *Present Perfect* translated into a PPC in all dialects of Spanish? Or are all of them translated into a PI? Do some dialects use the PPC while others resort to the PI?

Once all translations are annotated, we can create tuples that indicate which grammatical marker appears in that context in each of the languages or dialects we are considering. For instance, for the example in Figure 1, we would have a tuple that for the languages or dialects <English, Castilian Spanish, Mexican Altiplano Spanish, Rioplatense Spanish> would indicate <*Simple Past*, *Pretérito Perfecto Compuesto*, *Pretérito Perfecto Compuesto*, *Pretérito Indefinido*>. We can count the different 4-tuples for all contexts of interest and keep track of how many contexts have a specific selection of tense markers across languages.

To help us in inferring generalizations from four different versions of the book, the second step in the *Translation Mining* method is to make use of a cartographic approach to data visualization known as *Multidimensional Scaling* (MDS, Wälchli and Cysouw 2012; for a summary on linguistic applications and the concrete functioning of MDS, see also van der Klis and Tellings 2022). MDS takes the individual contexts in the dataset as its input and places them on a two-dimensional space based on the similarity of the 4-tuples; that is, if the 4-tuples are more similar to one another, the MDS algorithm will place them closer with respect to each other.[10] Dots that appear together in the maps represent contexts in which the 4-tuples are equivalent, whereas dots that are far from each other represent different 4-tuples. Therefore, applying this method to the crossdialect data generates maps that cluster the different sets of tense markers across different dialects with respect to their contexts of use.

4. Results

Using the methodology described above, we first generate a map that considers the distribution of the *Present Perfect* and the *Simple Past* in English in comparison to the

corresponding PERFECT and (PERFECTIVE) PAST forms in the three Spanish translations (i.e., the *Pretérito Perfecto Compuesto* and the *Pretérito Indefinido*). This will be our baseline map, our initial point of comparison, since it represents the English original. Through the application of MDS, we obtain Figure 2, where blue dots represent the *Present Perfect* and green dots represent the *Simple Past*. In this figure, we can observe that there are 44 cases of the *Present Perfect* (in blue) and 543 cases of the *Simple Past* (in green). To further profit from the use of MDS maps, we follow van der Klis et al. (2021) and generate for each of the translations equivalent maps of Figure 2 adding language-specific markup to each of the new maps. We use the same color coding (PERFECT = blue; PAST = green) to allow for an easier comparison across the different maps.

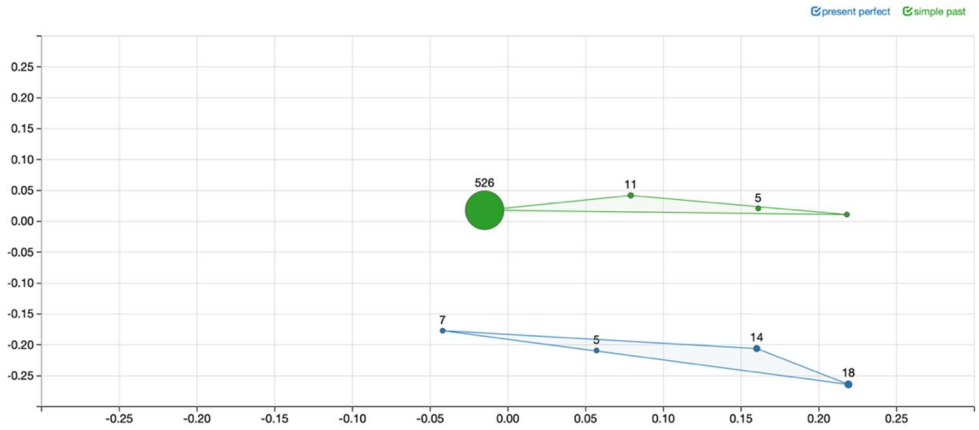

Figure 2. English map of the distribution of PERFECT and PAST forms in *Harry Potter*.

When we produce an analogous map for the Castilian Spanish translation, we obtain Figure 3. The cartographic representation, in combination with Figure 2, allows us to see two major things: 32 *Present Perfect* cases in the English original are translated into the PPC in Castilian Spanish, which are the dots in the low right corner of the map. Above these dots, we see that there are also 6 cases of the English *Simple Past*, previously indicated by green in Figure 2, that become blue, showing that they are translated into the Castilian PPC. On the other hand, there are 12 *Present Perfect* cases in English (in the middle low part of the map) which get translated into the PI in Castilian Spanish, and 537 English *Simple Past* cases that remain a (PERFECTIVE) PAST form in Castilian Spanish.

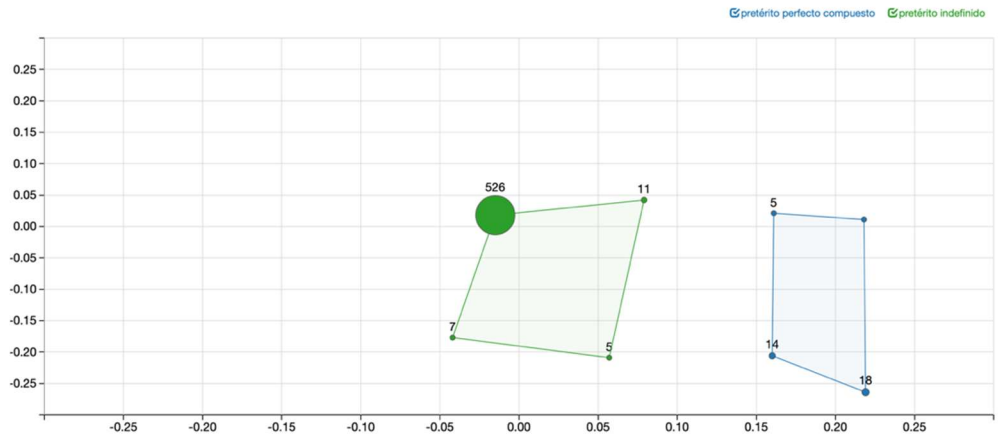

Figure 3. Castilian Spanish map of the distribution of PERFECT and PAST forms in *Harry Potter*.

Applying the same technique, we can observe the patterning in the Mexican Spanish translation in Figure 4. In this case, we see an expansion of PERFECT forms. There are 37 English *Present Perfect* cases that remain a PERFECT form (the PPC) in Mexican Spanish,

but there are additionally 17 cases in which English displays a *Simple Past*, and the translator opts for a PPC in Mexican Spanish. On the other hand, only 7 cases in which English displays a *Present Perfect* get translated into a PI in Mexican Spanish. Finally, 526 English *Simple Past* forms remain a (PERFECTIVE) PAST in this translation.

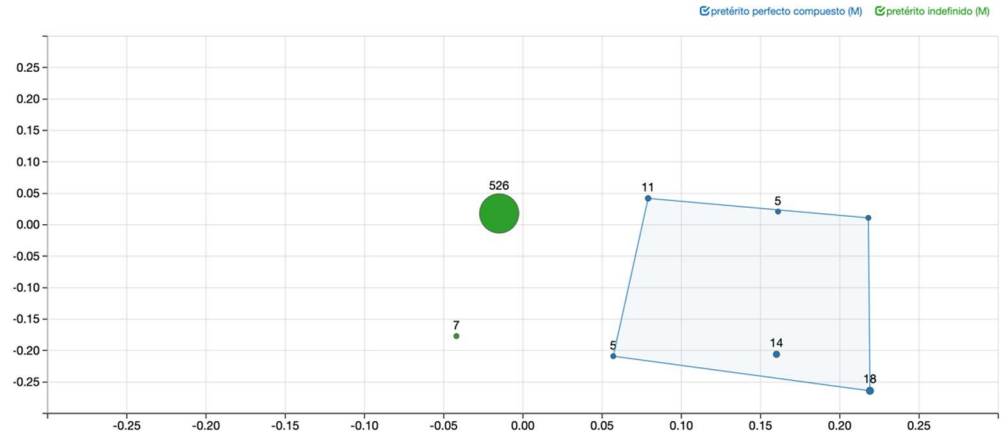

Figure 4. Mexican Spanish map of the distribution of PERFECT and PAST forms in *Harry Potter*.

As for the Rioplatense data, when applying the same procedure, we obtain Figure 5, where we observe less PERFECT forms. While there are still 18 English *Present Perfect* forms that remain a PERFECT in Rioplatense Spanish, the remaining 26 English *Present Perfect* cases get translated into a PI in this translation. Moreover, there is only one case in which English displays a *Simple Past* and the Rioplatense translator opted for a PPC—the other 542 *Simple Past* forms produce an equivalent (PERFECTIVE) PAST form, the PI, in this translation.

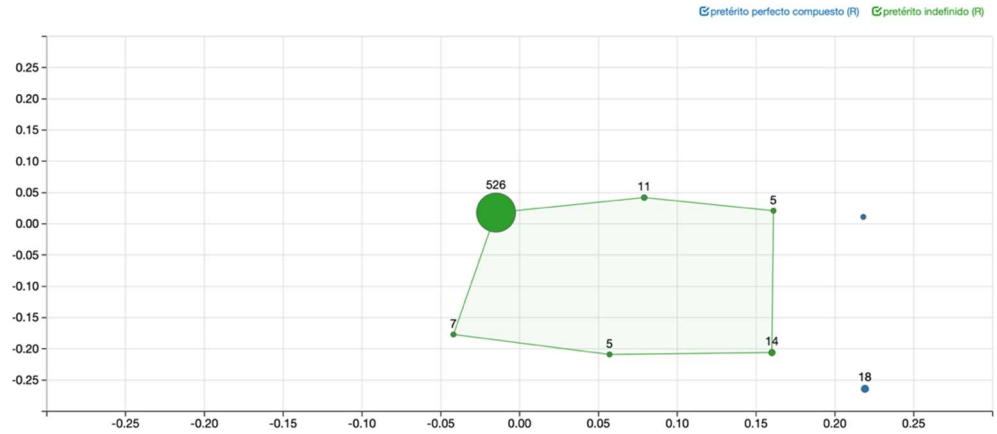

Figure 5. Rioplatense Spanish map of the distribution of PERFECT and PAST forms in *Harry Potter*.

In summary, out of 44 English *Present Perfect* cases in the original text, we find more instances of PPC in the Mexican translation (n = 37), than in the Castilian one (n = 32) or the Rioplatense version (n = 18), which shows a broader use of the PI. In cases in which the English original had a *Simple Past* (n = 543), most tokens are translated to the Spanish PI across Spanish translations, but once again Mexican Spanish shows more PPC tokens (n = 17) than its Castilian (n = 6) and Rioplatense (n = 1) counterparts. Crucially, the maps allow us to uncover a pattern across the different Spanish translations: there is a subset relation in PERFECT use, such that all cases that display the PPC in Rioplatense Spanish are cases that display the same marker in the other two varieties, and all cases that display a PPC marker in Castilian Spanish are cases that also display a PPC marker in Mexican Spanish.

There are also seven cases where English uses a *Present Perfect* and the three Spanish translations use a PI, contra a generalization advanced in van der Klis et al. (2021), which claims that English use of PERFECT forms is a proper subset of Spanish ones. It is remarkable, however, that five of these seven cases have a contracted auxiliary and, moreover, are cases in which the *Simple Past* form is identical to the English *Past Participle*. These graphemic similarities could make the translator opt for a PI in Spanish, since the saliency of the morphological distinction across forms is "lost in translation". We leave a detailed analysis of these counterintuitive cases for future work.

5. Interim Discussion: The Need for a Triangulation of Methods

By applying the *Translation Mining* method to our parallel corpus data, we find a proper subset relation across Spanish translations in terms of PERFECT use: Mexican Spanish has a more extended use of the PPC than Castilian Spanish, and this latter dialect, in turn, displays a broader use of this form than its Rioplatense counterpart. This distributional pattern, however, goes against established descriptions in the literature about the degree of PPC use in Latin American Spanish in general, and in Mexican Spanish in particular (e.g., Moreno de Alba 1978; Schwenter 1994; Schwenter and Torres Cacoullos 2008), which has always been considered a dialect with a more restricted set of usages than those displayed by the varieties of Spanish spoken in Spain.

This result leads us to question the status of the translations with respect to the varieties we had assumed them to represent, specifically in the domain of Spanish PPC-PI distribution. A triangulation of methods can then be useful to allow us to compare a different source of data with the results from our parallel corpus approach. To that end, we ran an elicitation forced-choice task in which native speakers from each of the largest cities within each book distribution area (i.e., Mexico City, Madrid, and Buenos Aires) had to choose between the PPC and the PI forms in a context extracted from the novel. Participants were recruited in each dialect by word-of-mouth (n = 10 per dialect), and the elicitation task was performed on a written page with 15 contexts out of each corresponding translation of the book; that is, the fragments were the same, but they were extracted from each translation in case some lexical item in the surrounding context, specific to a given translation, had an effect on the participants' choices. In each of these contexts, the tense marker under consideration was deleted, and participants had to select one of two choices in each case: either the PPC or the PI, which were written directly below the context. Participants were given written instructions to select the marker that sounded most natural in that given context, and they were told that all the contexts came out of the Harry Potter novel. An example of this task for Rioplatense Spanish is shown below in Figure 6:

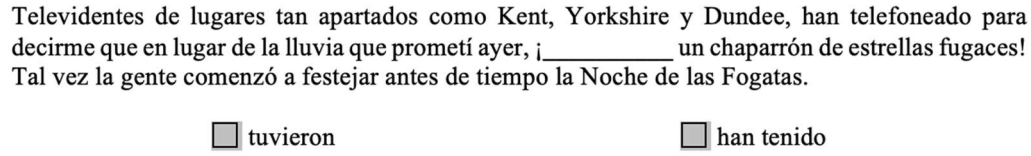

Figure 6. Example item from elicitation forced-choice task (Rioplatense version).

All contexts had been previously extracted from the novel by means of the *Perfect Extractor*. In particular, we included in the forced-choice task all 16 cases where only Mexican Spanish shows a PPC form, 7 cases where all dialects display the PPC, and 7 cases where all dialects display a PI in order to balance the stimuli. This amounted to a total of 30 contexts, which we divided in 2 lists of 15 contexts each.

Results indicating the matching percentages between the form chosen by the participants and the tense form displayed in the specific translation are presented in Table 1. Rows indicate the different translations under scrutiny, while columns indicate the different kind of stimuli (i.e., cases where only the Mexican translation displays a PPC, cases where all translations display a PI, and cases where all translations display a PPC). Finally, the

last column indicates the matching percentage between participants' answers and each translation taking into account all items.

Table 1. Results from forced-choice task (matching percentages between participants' choices and corpus forms).

Dialect	Only Mexican PPC	All Dialects PI	All Dialects PPC	Total
Mexican	26.60%	93.10%	81.50%	55.00%
Rioplatense	81.25%	80.00%	65.70%	77.30%
Castilian	56.25%	57.10%	65.70%	58.60%

We can observe that Rioplatense Spanish shows the most convergence between participants' data and its corresponding translation, while Castilian Spanish shows less of a convergence, but still a constant degree of variability across the three conditions. This we take to be good grounding to further explore the parallel corpus data, and consider them representative of the choices that native speakers of those dialects make when it comes to PPC-PI use in an experimental setting. Conversely, the main finding is that the Mexican data show almost all of its divergence between the translation and the participants' choices in the cases that go against the descriptions established in previous literature (that is, in the cases where only the Mexican Spanish translation displays a PPC while the others translations opt for a PI). In these cases, there is only a 26.6% degree of convergence between the translation and the choices of native speakers of this dialect. We consider that this last result constitutes a firm basis for calling into question our assumption that the translation geared towards the whole of Latin America (except for the Southern Cone) represents the dialect of Mexico City. Consequently, we leave this translation out of our subsequent analysis (since the comparisons both with the previous reports in the literature and with the experimental data reveal a largely divergent pattern), and thus focus on the relation between the Castilian and the Rioplatense Spanish translations.

6. Variables at Play in the Selection of Perfect-Perfective Forms across Spanish Dialects

In order to explain the subset relation that is observed between Rioplatense Spanish and Castilian Spanish with respect to PERFECT use (i.e., the fact that all contexts in which Rioplatense Spanish uses a PPC are a subset of the contexts in which Castilian Spanish uses a PPC), we consider all 46 remaining tokens from the parallel corpus analysis (i.e., cases in which either the English original shows a *Present Perfect* or the Rioplatense Spanish or the Castilian Spanish translations show a PPC regardless of the marker in English). All these cases occurred in dialogue contexts in the novel, where variation between these PERFECT and (PERFECTIVE) PAST forms has been shown to occur in these languages.

We code each of the contexts for the following variables: inherent aspect (Verkuyl 1972, 1993 et seq.), polarity, clause type, grammatical person, grammatical number, sentential force (based on variables analyzed in Schwenter and Torres Cacoullos 2008 to address PPC-PI alternation), and reading conveyed by the sentence under study (based on González et al. 2019). To this end, both authors of the current paper did an independent coding of all 46 tokens in the English original, with a convergence in 44 cases (95.7% of the data).[11] We report results on that subset of data. Each token was analyzed by looking at a full version of the original text to be able to consider contextual features that would indicate the reading that the specific marker was conveying in each particular instance. We describe the coding of the variables, and the levels within them, directly below:

6.1. Inherent Aspect

We code for inherent aspect as it considers both the lexical type of the verb and the nature of the NP at the VP level, which has been shown to affect the availability of tense-aspect markers in different usage patterns (Verkuyl 1972, 1993). With respect to this variable, we differentiate between terminative and durative aspect: the former indicates that an

event is bounded, telic and complete, while the latter presents the event as unbounded, atelic, or incomplete. An English example from the original corpus for the terminative case is given in (7), while an example for the durative context is presented in (8):

(7) Terminative: "And finally, bird watchers everywhere *have reported* that the nation's owls have been behaving very unusually today".

(8) Durative: "I have form only when I can share another's body . . . but there *have always been* those willing to let me into their hearts and minds".

6.2. Polarity

In this case, we consider whether the sentence that contains the event under scrutiny is affirmative or negative. An example of an affirmative context is given in (9), and an example of a negative context is presented in (10):

(9) Affirmative: "Unicorn blood *has strengthened* me, these past weeks . . . "

(10) Negative: "My mistake, my mistake—I *didn't* see you—of course, you're invisible".

6.3. Clause Type

With respect to clause type, we differentiate between main and subordinate clauses, as shown respectively in examples (11) and (12):

(11) Main: "My scar keeps hurting me. It'*s happened* before, but never as often as this".

(12) Subordinate: "Mr. Ronald Weasley and Miss Granger will be most relieved you *have come round*, they have been extremely worried".

6.4. Sentential Force

As for sentential force, we differentiate between declarative and non-declarative (interrogative, imperative) contexts, as respectively illustrated by examples (13) and (14):

(13) Declarative: "Snape'*s* already *got* past Fluffy".

(14) Non-declarative: "'How *did* you *know* it was me?', she asked".

6.5. Grammatical Person

In the case of grammatical person, we code for first, second or third person, as examples (15–17) show:

(15) First: "Snape came out and asked me what I was doing, so I *said* I was waiting for Flitwick".

(16) Second: "*Haven't* you *heard* what it was like when he was trying to take over?".

(17) Third: "A lot of the greatest wizards *haven't got* an ounce of logic; they'd be stuck in here forever".

6.6. Grammatical Number

The sentences under analysis can display either singular or plural grammatical number, as in examples (18) or (19) below:

(18) Singular: "The Potters, that's right, that's what I *heard*".

(19) Plural: "We'*ve had* Sprout, that was the Devil's Snare".

6.7. Reading

With respect to this independent variable, we distinguish four different readings: continuative, experiential, hodiernal, and resultative. These readings have been extensively described in the literature (see Section 2.1 above). Continuative readings are cases in which

the tense-aspect marker is applied to a situation in which the past event keeps holding into the present; it shows up in many cases together with a *since*-adverb clause (e.g., Rodríguez Louro 2012). Experiential readings indicate the occurrence of an event that is seen as part of the "life experience" of the subject at issue; it is compatible with adverbials such as *once*. Hodiernal contexts are related to the day of utterance time (e.g., Schwenter 1994); typical adverbs that show up in these contexts are *today* or *this morning*. Finally, resultative readings include cases in which the tense-aspect marker codifies a result; that is, when a state is the result of a past event. An example of each of these contextual readings from our *Harry Potter* corpus is presented directly below. A continuative reading is presented in (20); an experiential context is given in (21); a hodiernal past context is introduced in (22), indicated by the adverbial clause *since sunrise*, and (23) presents a resultative reading:

(20) Continuative: "Since then, I *have served* him faithfully".

(21) Experiential: "My dear Professor, I'*ve* never *seen* a cat sit so stiffy".

(22) Hodiernal: " . . . there *have been* hundreds of sightings of these birds flying in every direction since sunrise"

(23) Resultative: "As for the Stone, it *has been destroyed*".

7. Results from Variable Coding

We analyze the results of the coding described in Section 6 by performing a linear mixed model effects analysis in R (R Core Team 2021) using the *lme4* package (Bates et al. 2015) with a binominal function, since the dependent variable is binary (i.e., the form that shows up in the translations is either a PPC or a PI). To determine the role of the different independent variables under consideration, we perform Likelihood Ratio Tests of the full model with the effect in question against a model without it (see Winter 2019 for a justification). We also consider potential interactions between the independent variables. Post-hoc tests to analyze the direction of the effect are run with the *multcomp* package (Hothorn et al. 2008) and *p*-values are corrected by Tukey. All models under consideration have random intercepts for item, which in this simple data set is the maximum random effect structure that is justified.

As a summary, remember that we are analyzing 44 cases. These cases are translated into 37 PPC and 7 PI in the Castilian Spanish version, and to 17 PPC and 27 PI in the Rioplatense Spanish one, as Figure 7 shows.

So, first and foremost, we analyze whether the difference in number of PPC tokens across dialects is significant. Comparing a model with dialect as a main effect and a model without the effect in question shows a significant effect of dialect ($\chi^2(1) = 20.146$, $p < 0.001$). A post hoc test corrected by Tukey shows that the difference arises by a greater use of PPC in the Castilian Spanish translation than in the Rioplatense Spanish one ($\beta = 2.128$, se = 0.516, $p < 0.001$), confirming what the raw frequency counts indicate. On the other hand, when studying the different independent variables under consideration, we find no main effect of polarity ($\chi^2(1) = 0.935$, $p = 0.334$), clause type ($\chi^2(1) = 0.163$, $p = 0.686$), sentential force ($\chi^2(1) = 0.149$, $p = 0.928$), grammatical person ($\chi^2(2) = 3.179$, $p = 0.204$), or grammatical number ($\chi^2(1) = 0.001$, $p = 0.98$). While there is a main effect of inherent aspect ($\chi^2(1) = 4.202$, $p < 0.05$) that favors the appearance of PPC in durative contexts in comparison to terminative ones when considering both translations together, a posthoc test indicates that the effect is actually not a significant one ($\beta = 1.2937$, $se = 0.7755$, $p = 0.095$). The same occurs with the variable related to the reading that the context expresses: there appears to be a main effect ($\chi^2(3) = 8.346$, $p < 0.05$), but no individual comparison is significant under a posthoc test with Tukey-corrected *p*-values. In summary, there does not seem to be an independent variable that triggers a greater use of PPC in both dialects at the same time.

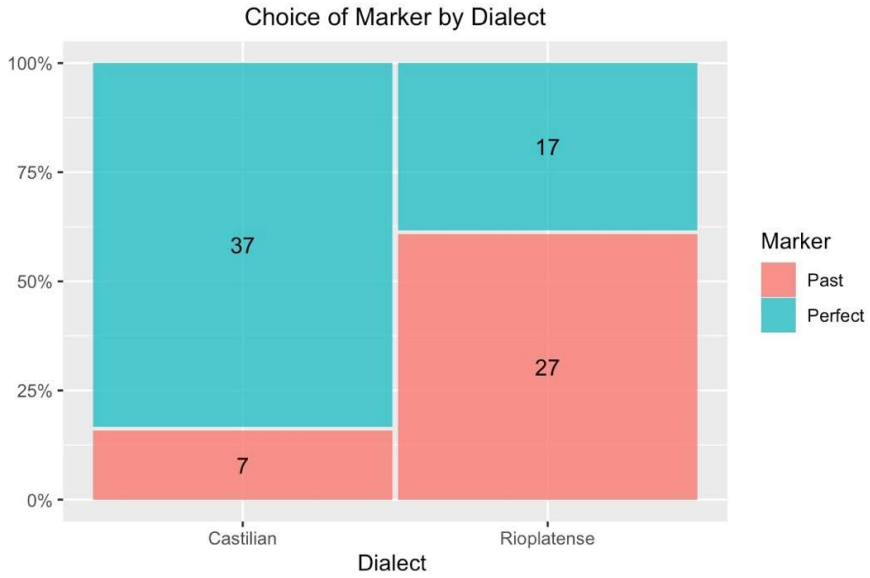

Figure 7. Choice of grammatical marker in each Spanish translation.

So, what variables are driving this big difference across translations, which is the distinction we are interested in? What makes Castilian Spanish have a broader distribution of PPC use than its Rioplatense counterpart? To address that problem, we look at interaction effects, and find the following significant 2-way interactions: between inherent aspect and dialect ($\chi^2(2) = 6.612$, $p < 0.05$), between dialect and grammatical person ($\chi^2(4) = 13.885$, $p < 0.01$), and between reading and dialect ($\chi^2(6) = 10.668$, $p < 0.05$). No other interactions are significant, neither between two independent variables or more.

To assess the source of the difference with respect to inherent aspect, we thus subset the data into terminative and durative aspect, and evaluate the role of dialect as a predictor within each subset. In the case of durative aspect, we find a main effect of dialect ($\chi^2(1) = 8.424$, $p < 0.01$), but that difference turns out to be non-significant when we correct the p-value by Tukey in a post-hoc test ($\beta = 36.06$, se = 993.43, $p = 0.971$). We can see the distribution of markers in durative contexts by dialect in Figure 8.

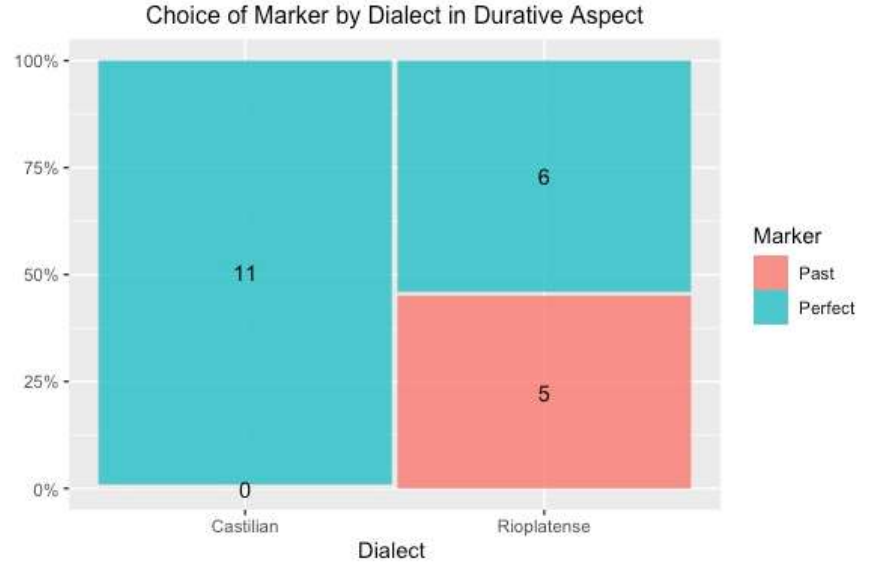

Figure 8. Choice of grammatical marker by dialect in durative aspect contexts.

On the other hand, when it comes to terminative aspect, we find a main effect of dialect ($\chi^2(1) = 14.408$, $p < 0.001$), which is maintained in a post-hoc test corrected by Tukey,

showing that the Castilian Spanish translation favors the appearance of the PPC to a larger extent in these aspectual contexts (β = 2.005, *se* = 0.5636, $p < 0.001$). We can see the counts for each marker in Figure 9.

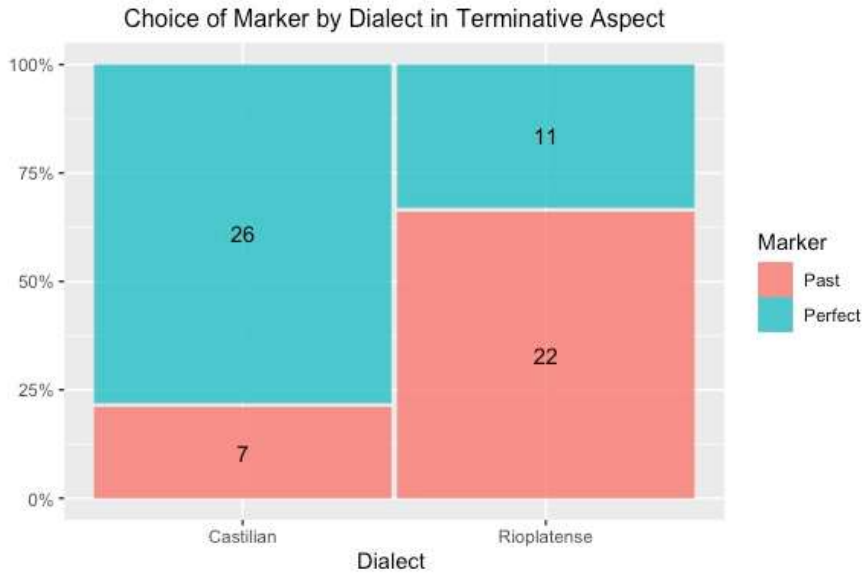

Figure 9. Choice of grammatical marker by dialect in terminative aspect contexts.

This means that the source of the difference between translations in relation to inherent aspect is due to the Castilian Spanish version allowing for more PPC use in terminative contexts than its Rioplatense counterpart. When we are dealing with durative aspect, on the other hand, both dialectal translations behave alike in spite of the frequency counts. In other terms, what this analysis shows is that the Rioplatense Spanish translation, which tends to disfavor PPC use, favors the use of this marker in durative contexts to the point that it is not (significantly) different from the Castilian Spanish one.

To analyze the interaction between dialect and grammatical person, we subset the data by grammatical person. When considering first person cases, we observe a main effect of dialect ($\chi^2(1) = 10.221$, $p < 0.01$), which in a post-hoc test corrected by Tukey shows that the Castilian Spanish translation significantly favors the use of the PPC in these cases in comparison to the Rioplatense Spanish version of the text (β = 2.416, se = 0.836, $p < 0.01$). In the case of second person, a main effect of dialect is also observed ($\chi^2(1) = 19.408$, $p < 0.001$), but no post-hoc test can be performed since the Rioplatense Spanish translation has zero cases of PPC in this condition, and, conversely, the Castilian Spanish translation has zero cases of a PI marker in the second person. Finally, in the case of the third person, there is no main effect of dialect ($\chi^2(1) = 1.933$, $p = 0.1645$), showing that both translations allow for the use of the PPC to the same degree in these cases. We can observe the different counts in each grammatical person by dialect in Figure 10. This figure shows three different patterns: first person cases seem to reproduce the general tendency that the Castilian Spanish translation permits more PPC use than the Rioplatense Spanish one; second person cases show a complete opposite pattern of use in each translation, since the Castilian Spanish one only uses the PPC while the Rioplatense Spanish version only uses the PI; finally, third person cases show no significant difference in PPC use across the two translations.

The last interaction that we need to unpack is the one between reading and dialect. Continuative cases cannot be compared since both translations maximize PPC use in these cases: there are only three cases with this reading, but both versions of the text only use the PPC form in all of them. When analyzing experiential readings, we do observe a main effect of dialect ($\chi^2(1) = 4.557$, $p < 0.05$), but a post-hoc test corrected by Tukey shows that the difference between dialects is only marginally significant (β = 2.457, se = 1.296, $p = 0.0581$), in favor of the Castilian Spanish translation. Hodiernal cases also show a main

effect of dialect with a greater effect size ($\chi^2(1) = 24.234$, $p < 0.001$). In this case, a posthoc test corrected by Tukey shows that in these contexts the Castilian Spanish translation significantly favors the use of the PPC in comparison to the Rioplatense one ($\beta = 25.918$, se = 9.405, $p < 0.01$), or, in other terms, that the Rioplatense Spanish text strongly disfavors the appearance of PPC in these contexts. Finally, in the case of resultative readings, there is again a main effect of dialect ($\chi^2(1) = 4.806$, $p < 0.05$), which holds in a posthoc test corrected by Tukey, showing a greater use of PPC in the Castilian Spanish version with respect to its use in the Rioplatense translation ($\beta = 1.406$, $se = 0.6649$, $p < 0.05$). In the different mosaic graphs in Figure 11, we can see the different frequency counts for each marker by dialect under each of the reading conditions.

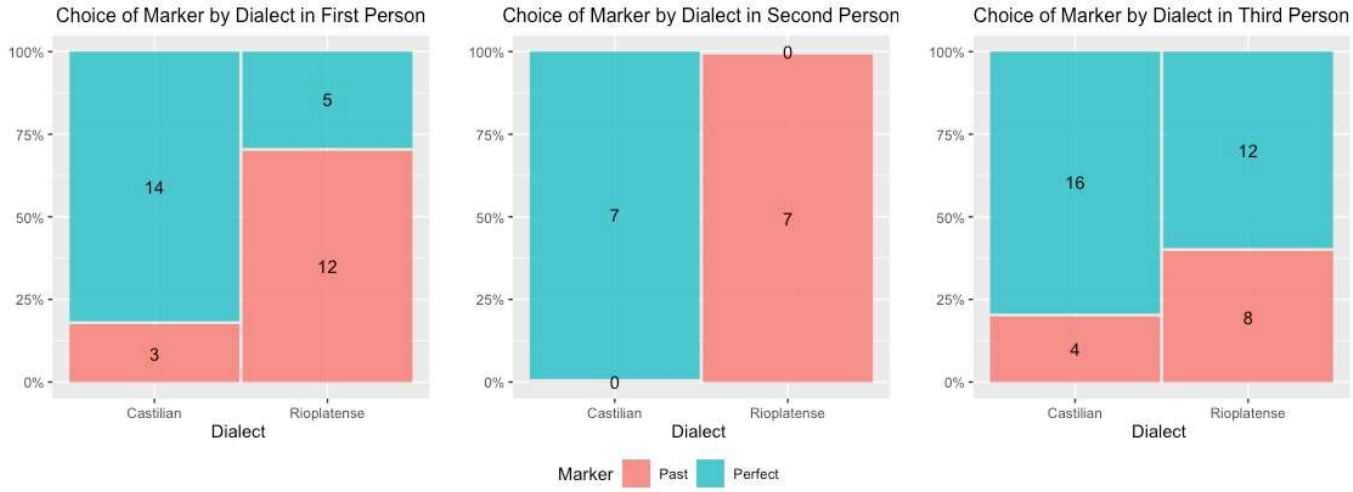

Figure 10. Choice of grammatical marker by dialect by grammatical person.

Figure 11. Choice of grammatical marker by dialect in each contextual reading type.

8. General Discussion

As presented in the previous sections, in order to describe PERFECT use in Rioplatense and Castilian Spanish, we extracted from our parallel corpora all individual tokens of the PPC, and also the cases in which English uses a *Present Perfect* form even if both Spanish dialects use a PI, and coded them for inherent aspect, polarity, clause type, sentential force, grammatical person, grammatical number, and the reading that is conveyed by the sentential meaning in the specific context. We examine below any findings worth discussing.

Results from this coding indicate that in relation to inherent aspect, durative contexts increase the chance that Rioplatense Spanish displays a PPC instead of a PI, given that both dialects are not significantly different from each other in this kind of contexts. So, in a context such as the one presented in (8), repeated below as (24), there is a higher chance that both dialects would display a PPC instead of Rioplatense Spanish displaying a PI. Indeed, when we check the different corpora, this is the case, as shown in the Castilian Spanish translation in (25) and the Rioplatense Spanish translation in (26):

(24) **English**: "I have form only when I can share another's body . . . but there *have always been* those willing to let me into their hearts and minds".

(25) **Castilian Spanish**: "Tengo forma sólo cuando puedo compartir el cuerpo de otro . . . pero siempre *ha habido* seres deseosos de dejarme entrar en sus corazones y en sus mentes."

(26) **Rioplatense Spanish**: "Tengo forma sólo cuando puedo compartir el cuerpo de otro . . . pero siempre *han estado* aquellos deseosos de dejarme entrar en sus corazones y en sus mentes."

However, the Castilian Spanish translation uses the PPC in all durative aspect cases, while the Rioplatense Spanish translation only uses it in a bit more than 50% of them. It seems that for Castilian Spanish, contrary to Rioplatense Spanish, the inherent aspect of the verb-noun compound is more of a decisive factor in favoring the PPC.

Grammatical person shows a pattern worth explaining in relation to the second person: while in the Castilian Spanish translation, it prompts the appearance of the PPC, in the Rioplatense Spanish case, there are zero cases of PPC with a second grammatical person. As for the third person, on the other hand, it seems to favor the appearance of PPC in the Rioplatense Spanish translation. We argue that the absence of second person PPC forms in Rioplatense Spanish might be due to the translator's attempt to follow the publishing house expectation that a valid translation for a wider area of the Southern Cone (Uruguay, Chile, and Argentina more generally) was provided. In Rioplatense Spanish, instead of the widely used 2nd person singular pronoun *tú*, the form *vos* is used. However, the use of *vos* in translations is "considered negative in commercial terms" (Sundell 2010, p. 25). This is indeed reflected in the use of *tú* as the 2nd person singular pronoun across the Rioplatense Spanish translation, even when Buenos Aires is a *voseante* region. Besides the pronoun itself, using *vos* as the 2nd person pronoun would prompt the use of a whole different verbal paradigm (see e.g., Fontanella de Weinberg 1992). We suggest that using both *tú* and the PPC might have felt too much of a 'Castilian' translation for the Rioplatense translator, so that given the constraint to use *tú* to appeal to a broader Southern Cone audience, this was counterbalanced by a constant use of the PI in this set of cases. As for the fact that the third person allows for more uses of PPC in Rioplatense Spanish than what we might expect by looking at the general frequencies of PPC-PI alternation, this is an unexpected finding, but can be related to Azpiazu (2021), who found in Castilian Spanish more PPC first person uses as opposed to third person (fictional) narration ones. In any case, the fact that Rioplatense Spanish allows for more uses of third person PPC remains an open question to us, which we leave for further study.

In line with previous work on dialect variation (e.g., Rodríguez Louro 2009; González et al. 2019), we also show that besides these sentential linguistic features, there are other contextual constraints at play in selecting each form, and that these contextual properties act in diverse ways across different Spanish dialects. The context in which the marker appears,

and the reading that this context is conveying, proves to be a crucial distinction: out of the 11 cases of hodiernal contexts, Castilian Spanish uses a PPC in 10 while Rioplatense Spanish only shows that marker in one. For instance, if we take the context in (22), repeated below in (27) in English, we see that it gets a PPC in the Castilian Spanish translation in (28), but a PI in the Rioplatense one as shown in (29):

(27) **English**: " . . . there *have been* hundreds of sightings of these birds flying in every direction since sunrise".

(28) **Castilian Spanish**: " . . . se *han producido* cientos de avisos sobre el vuelo de estas aves en todas direcciones, desde la salida del sol."

(29) **Rioplatense Spanish**: " . . . *hubo* cientos de avisos sobre el vuelo de esos pájaros en todas direcciones, desde la salida del sol."

Instead, in all continuative contexts, the Rioplatense Spanish translation uses the PPC just as its Castilian counterpart. Therefore, while in Castilian Spanish considerations about hodiernality are relevant, Rioplatense Spanish, in contrast, presents only the prototypical Perfect readings (experiential, continuative, resultative), and defaults to the use of the PI otherwise. It is notable, though, that in the case of experiential readings the difference across dialects is only marginally significant, pointing to a potential advancement of PPC forms in the Rioplatense Spanish dialect through this kind of reading, while continuatives with PPC marking seem to be already fully established, and resultatives lag behind. These results are in line with previous research that has shown that in Rioplatense Spanish the use of the PI is much more frequent than the PPC (e.g., Kubarth 1992; De Jonge 1999; Rodríguez Louro 2009). Since Kubarth (1992), it has been claimed that Rioplatense Spanish uses the PPC for "temporal attainment" contexts; that is, when it is clear that the action continues until the moment of speaking or to its completion before the end of the interval of speech—what we call here continuative and resultative readings. We see in our data that this pattern holds in the case of continuatives, but not so much when it comes to resultatives. In any case, 10 out of 22 resultative cases have a PPC in the Rioplatense Spanish translation, showing that even if different from the Castilian Spanish version, the dialect partially allows this reading to be conveyed with a PPC.

As for the hodiernality constraint, it is also important to look at previous research (e.g., Harris 1982) that indicates that the (Romance) PRESENT PERFECT follows a series of stages in its diachronic evolution in the grammaticalization path knows as the 'aoristic drift'. Once only a resultative marker in Latin, it first evolved into a form that could more generally convey past events that were construed as either durative or repetitive. A third stage represents an extension in the domains of use to all past situations with relevance to the present moment: hodiernality—that is, reference to the day of utterance— clearly invokes by default this kind of current relevance. Rodríguez Louro (2009, p. 249) proposes that in Rioplatense Spanish the PPC does not establish a relation with the present but it is used instead to express indefinite past and generic reference. In González et al. (2019) it is shown that the PPC in this dialectal variety refers to a past situation, although it does not mention a specific moment. Corroborating previous studies, this is what our data ultimately shows with respect to each dialect: in Castilian Spanish, a PPC marker that has extended its use to refer to all past situations as far as they are contained in the day of utterance; in the case of Rioplatense Spanish, a PPC in a previous diachronic stage, which does not allow to convey hodiernal past eventualities.

9. Conclusions

We have shown that different linguistic and contextual factors participate in explaining the crossdialectal Spanish patterns that had been described in the literature in the case of PERFECT-PERFECTIVE PAST variation. The parallel corpus method allowed us to identify new features that underlie these distributional patterns. That said, we decided to check for the reliability of our assumption that the translation intended for the Latin American marker (excluding the Southern Cone) represented the dialect of Mexico City, since it went

against the distributional patterns previously established. Through an experimental forced-choice task, we were able to show that this translation corpus was not convergent with native speakers' choices in the domain of variation between the *Pretérito Perfecto Compuesto* and the *Pretérito Indefinido*, so we decided to leave that translation aside.

Therefore, we conclude that the parallel corpus approach constitutes a valid method that allows to uncover generalizations about tense distribution in a specific language and its varieties. Researchers should, however, keep two things in mind: (i): standardized written varieties might only partially reflect the spoken behavior of native speakers, and (ii) translated corpus might not be representative of native speakers' linguistic behavior, especially when the data goes against previously widely described patterns in the literature. With respect to (i), we have shown that when we are after variation that is *grammatical* in nature, we will find it (even if in less cases) in written standardized varieties. To address (ii), we have proposed a solution: to gather native speakers' choices in an experimental task and compare those results with the translations, expanding the sources of data.

In sum, our study has provided evidence that crossdialectal variation becomes crucial for the proposal of crosslinguistic generalizations that deal with meaning distinctions. Data such as ours are better explained from a bottom-up perspective in which specific tokens are not fitted into predefined categories, but the richness of dialectal variation is wholeheartedly recognized.

Author Contributions: Conceptualization, M.F. and P.G.; methodology, M.F. and P.G.; software, M.F.; validation, M.F. and P.G.; formal analysis, M.F and P.G.; investigation, M.F and P.G.; resources, M.F and P.G.; data curation, M.F and P.G.; writing—original draft preparation, M.F and P.G.; writing—review and editing, M.F and P.G.; visualization, M.F.; supervision, M.F and P.G.; project administration, M.F.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NWO (Dutch Research Council), grant number 360-80-070.

Institutional Review Board Statement: The study was approved by the Faculty Ethics Assessment Committee of the Faculty of Humanities (FEtC-H) of Utrecht University (reference number: 20-249-03, approved on 5 October 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The parallel corpus data reported in this study is privately stored at https://time-in-translation.hum.uu.nl (accessed: 13 June 2022) responding to copyright reasons. The experimental data is privately stored at Utrecht University's secure servers, as per our Ethics Committee policies, and can be made available upon request.

Acknowledgments: We thank three anonymous reviewers for their very helpful feedback and comments, as well as the Time in Translation team at Utrecht University, and audiences at the "Expanding Romance Linguistics—Crossing the boundaries" workshop at the University of Graz, and at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences. We would also like to thank Patricio Tarantino for sharing with us a chapter of his book on the history of the Harry Potter series in Spanish, and for helping us figure out the intricacies of the different Spanish versions and translations of the first book. As per usual, all errors and omissions remain our own.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Argentinian (Rioplatense) edition: *Harry Potter y la Piedra Filosofal* (2003). ISBN: 978-84-7888-861-0 Translated by Alicia Dellepiane. Barcelona: Emecé/Salamandra.

Latin American (Mexican) edition: *Harry Potter y la Piedra Filosofal* (2005). ISBN: 978-84-7888-654-8. Barcelona: Salamandra/Océano.

Peninsular (Castilian) edition: *Harry Potter y la Piedra Filosofal* (2010). ISBN: 978-84-9838-266-2. Barcelona: Salamandra.

Notes

1 We use small caps to indicate a crosslinguistic marker type comprising a set of language-specific forms, such as PERFECT for the *Present Perfect* in English, the *Passé Composé* in French, the *Pretérito Perfecto Compuesto* in Spanish, and so forth. We reserve italics to indicate language-specific forms, and we use plain text to refer to meanings.

2 All interlinear glosses follow Leipizg Glossing Rules.

3 By Castilian Spanish, we refer to the variety of Peninsular Spanish spoken in Central Spain (e.g., Madrid and its surroundings).

4 We use the term Mexican Altiplano Spanish to refer to the Spanish spoken in Mexico City and its surroundings.

5 Rioplatense Spanish, in this paper, indicates the Spanish spoken in Buenos Aires, Argentina.

6 We are very much aware that there is more diversity in the Spanish spoken within these regions. For instance, the Spanish spoken in Chile is lexically very different from the Spanish spoken in Argentina, but these are the idealized divisions made by the publishing house, with the intention to appeal to the readers of these (broad) geographical areas.

7 The edition intended for Spain was published in Barcelona because the headquarters of the publishing house are in that city. However, the translation was done into the standard norm of the Spanish spoken in Spain, which reflects the Spanish spoken in the Central regions: Castilian Spanish.

8 One anonymous reviewer points out that dialect zones are usually divided on the basis of spoken language, so that a notion like 'Rioplatense Spanish' normally refers to an urban spoken vernacular, the Spanish spoken in Buenos Aires. Written registers, on the other hand, would follow some sort of standard norm. We agree with this point, but still consider that describing *consistent grammatical patterns* found in different written standards of Spanish can shed light into native speakers' grammars from the same geographical areas where these translations were done.

9 Sundell (2010) works with the fifth book in the HP series, *Harry Potter and the Order of the Phoenix*, and its translations to Spanish. In that case, there was an original translation into Castilian Spanish that later was adapted for the Southern Cone by an Argentinian translator, and to the rest of Latin America by a Mexican translator. For details about those translations, see Sundell (2010, chap. 2). It seems that a similar process (but from Rioplatense Spanish to the other two varieties) occurred when the first book was published, since the original publishing house of the first book was in Buenos Aires, Argentina.

10 A reviewer asks how the dimensions in the figures should be interpreted. In MDS, dimensions do not have an inherent linguistic meaning, but are the result of applying the method to the data, which compares similarities across contexts based on the forms chosen by each language to express a given meaning (van der Klis and Tellings 2022). While some studies (e.g., Wälchli and Cysouw 2012) assign linguistic features to the different axes in their maps, our approach is based not on axis interpretation, but on the analysis of clusters of datapoints in the maps. We argue that every cluster of points can point to a relevant linguisic distinction with respect to the distribution of the markers under comparison, since different clusters indicate that at least one language/dialect has changed the grammatical form used to express that context/meaning.

11 One reviewer asks about the rationale for annotating the corpus in the English original and not in one (or more) of the Spanish translations. We assume that staying as close as possible to the meaning intended in the original allows us to control for any artifact introduced by the translators (i.e., 'translation-induced variation', as opposed to grammatical variation).

References

Alarcos Llorach, Emilio. 1947. Perfecto simple y compuesto en español. *Revista de Filología Española* 31: 108–39.

Alarcos Llorach, Emilio. 1970. *Estudios de Gramática Funcional del Español*. Madrid: Gredos.

Azpiazu, Susana. 2013. Antepresente y pretérito en el español peninsular: Revisión de la norma a partir de evidencias empíricas. *Anuario de Estudios Filológicos* 36: 19–31.

Azpiazu, Susana. 2014. Del perfecto al aoristo en el antepresente peninsular: Un fenómeno discursivo. In *Formas Simaples y Compuestas de Pasado en el Verbo Español*. Edited by Susana Azpiazu. Lugo: Axac, pp. 17–30.

Azpiazu, Susana. 2015. La variación antepresente/pretérito en dos áreas del español peninsular. *Verba* 42: 269–92. [CrossRef]

Azpiazu, Susana. 2021. Uso evidencial del perfecto compuesto en el español de Ecuador. In *La Interconexión de las Categorías Semántico-Funcionales en Algunas Variedades del Español*. Edited by Verónica Böhm and Anja Hennemann. Berlin: Peter Lang, pp. 237–62.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Berschin, Helmut. 1975. A propósito de la teoría de los tiempos verbales: Perfecto simple y perfecto compuesto en el español peninsular y colombiano. *Thesaurus* 30: 539–56.

Berschin, Helmut. 1976. *Präteritum und Perfektgebrauch im Heutigen Spanisch*. Tübingen: Max Niemeyer Verlag.

Bertrand, Anne, Yurika Aonuki, Sihwei Chen, Joash Gambarage, Laura Griffin, Marianne Huijsmans, Lisa Matthewson, Daniel Reisinger, Hotze Rullmann, Raiane Salles, and et al. 2022. Nobody's Perfect. *Languages* 7: 148. [CrossRef]

Bogaards, Maarten. 2022. The Discovery of Aspect: A heuristic parallel corpus study of ingressive, continuative and resumptive viewpoint aspect. *Languages*, *in press*.
Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: The Grammaticalization of Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.
Camus Bergareche, Bruno. 2008. El perfecto compuesto (y otros tiempos compuestos) en las lenguas románicas: Formas y valores. In *Tiempos Compuestos y Formas Verbales Complejas*. Edited by Ángeles Carrasco Gutiérrez. Madrid: Iberoamericana, pp. 65–102.
Condoravdi, Cleo, and Ashwini Deo. 2014. Aspect shifts in Indo-Aryan and trajectories of semantic change. In *Language change at the Syntax-Semantics Interface*. Edited by Chiara Gianollo, Agnes Jäger and Doris Penka. Berlin: Mouton de Gruyter, pp. 261–92.
Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Basil Blackwell.
De Jonge, Bob. 1999. El tiempo de todos los tiempos: El uso del presente perfecto en el español bonaerense. In *Actas del XI Congreso Internacional de la ALFAL*. Edited by José A. Samper Padilla and Magnolia T. Déniz. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, pp. 297–304.
De Jonge, Bob. 2001. El valor del presente perfecto y su desarrollo histórico en el español americano. In *Estudios Sobre el Español de América, Actas del V Congreso Internacional de El Español de América*. Edited by Hermógenes Perdiguero and Antonio Álvarez. Burgos: Universidad de Burgos, pp. 838–48.
De Kock, Josse. 1989. Pretéritos perfectos simples y compuestos en España y América. In *El Español de América: Actas del III Congreso Internacional del Español de América*. Valladolid: Junta de Castilla y León, pp. 481–94.
de Swart, Henriëtte. 2007. A cross-linguistic discourse analysis of the Perfect. *Journal of Pragmatics* 39: 2273–307. [CrossRef]
Fleischman, Suzanne. 1983. From pragmatics to grammar: Diachronic reflections on complex pasts and futures in Romance. *Lingua* 60: 183–214. [CrossRef]
Fontanella de Weinberg, María Beatriz. 1992. La estandarización del español bonaerense. In *Scripta Philologica: In Honorem Juan M. Lope Blanch*. Edited by Elisabeth Luna Traill. Mexico: Universidad Nacional Autónoma de México, pp. 425–37.
Fuchs, Martín. 2020. On the Synchrony and Diachrony of the Spanish Imperfective Domain: Contextual Modulation and Semantic Change. Ph.D. thesis, Yale University, New Haven, CT, USA.
González, Paz. 2003. Aspects on Aspect: Theory and Applications of Grammatical Aspect in Spanish. Ph.D. thesis, Utrecht Institute of Linguistics—OTS, Utrecht, The Netherlands.
González, Paz, and Carmen Kleinherenbrink. 2021. Target variation as a contributing factor in TAML2 production. *Círculo de Lingüística Aplicada a la Comunicación* 87: 39–51. [CrossRef]
González, Paz, and Henk J. Verkuyl. 2017. A binary approach to Spanish tense and aspect: On the tense battle about the past. *Borealis—An International Journal of Hispanic Linguistics* 6: 97–138. [CrossRef]
González, Paz, Margarita Jara Yupanqui, and Carmen Kleinherenbrink. 2019. The microvariation of the Spanish Perfect in three varieties. *Isogloss* 4: 115–33. [CrossRef]
Harris, Martin. 1982. The 'past simple' and the present perfect in Romance. In *Studies in the Romance Verb: Essays Offered to Joe Cremona on the Occasion of His 60th Birthday*. Edited by Nigel Vincent and Martin Harris. London: Croom Helm, pp. 42–70.
Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50: 346–63. [CrossRef] [PubMed]
Howe, Chad. 2006. Cross-Dialectal Features of the Spanish Present Perfect: A Typological Analysis of Form and Function. Ph.D. thesis, The Ohio State University, Columbus, OH, USA.
Iatridou, Sabine, Elena Anagnostopoulou, and Roumyana Izvorski. 2001. Observations about the form and meaning of the perfect. In *Ken Hale: A Life in Language*. Edited by Michael Kenstowicz. Cambridge and London: MIT Press, pp. 189–238.
Jara Yupanqui, Margarita. 2012. Peruvian Amazonian Spanish: Linguistic variation, language ideologies, and identities. *Sociolinguistic Studies* 6: 445–69.
Kany, Charles. 1945. *American-Spanish Syntax*. Chicago: University of Chicago Press.
Klein, Wolfgang. 1992. The Present Perfect puzzle. *Language* 68: 525–51. [CrossRef]
Kubarth, Hugo. 1992. El uso del pretérito simple y compuesto en el español hablado de Buenos Aires. In *Scripta Philologica: In Honorem Juan M. Lope Blanch*. Edited by Elisabeth Luna Traill. Mexico: Universidad Nacional Autónoma de México, pp. 553–66.
Le Bruyn, Bert, Martijn van der Klis, and Henriëtte de Swart. 2019. The perfect in dialogue: Evidence from Dutch. *Linguistics in the Netherlands* 36: 162–75. [CrossRef]
Löbner, Sebastian. 2002. Is the German Perfekt a Perfect Perfect? In *More Than Words: A Festschrift for Dieter Wunderlich*. Edited by Ingrid Kaufmann and Barbara Stiebels. Berlin: Akademie Verlag, pp. 369–91.
McCawley, James D. 1971. Tense and time reference in English. In *Studies in Linguistic Semantics*. Edited by Charles J. Fillmore and Terence Langendoen. New York: Holt, Rinehart and Winston, pp. 96–113.
McCawley, James D. 1981. Notes on the English present perfect. *Australian Journal of Linguistics* 1: 81–90. [CrossRef]
McCoard, Robert William. 1978. *The English Perfect: Tense-Choice and Pragmatic Inferences*. Amsterdam: North-Holland (Elsevier).
Michaelis, Laura A. 1994. The ambiguity of the English present perfect. *Journal of Linguistics* 30: 111–57. [CrossRef]
Moreno de Alba, José. 1978. *Valores de las Formas Verbales en el Español de México*. México: Universidad Nacional Autónoma de México.
Mulder, Gijs, Gert-Jan Schoenmakers, Olaf Hoenselaar, and Helen de Hoop. 2022. Tense and aspect in a Spanish literary work and its translations. *Languages*, *in press*.

NGRAE. 2009. Real Academia Española y Asociación de Academias de la Lengua Española. In *Nueva Gramática de la Lengua Española*. Madrid: Espasa.
Nishiyama, Atsuko, and Jean-Pierre Koening. 2010. What is a Perfect State? *Language* 86: 611–46. [CrossRef]
Portner, Paul. 2003. The (Temporal) semantics and (Modal) pragmatics of the perfect. *Linguistics and Philosophy* 26: 459–510. [CrossRef]
Potterglott. 2022. Harry Potter and the Spanish Tykes. Available online: https://www.potterglot.net/harry-potter-and-the-spanish-tykes/ (accessed on 31 January 2022).
R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: Dover Publications.
Rodríguez Louro, Celeste. 2009. Perfect Evolution and Change: A Sociolinguistic Study of the Preterit and Present Perfect Usage in Contemporary and Earlier Argentina. Ph.D. thesis, University of Melbourne, Melbourne, Australia.
Rodríguez Louro, Celeste. 2012. Los tiempos de pasado y los complementos adverbiales en el español rioplatense argentino: Del siglo XIX al presente. *Signo and Seña* 22: 215–34.
Rodríguez Louro, Celeste, and Margarita Jara Yupanqui. 2011. Otra mirada a los procesos de gramaticalización del presente perfecto en español: Perú y Argentina. *Studies in Hispanic and Lusophone Linguistics* 4: 55–80. [CrossRef]
Schaden, Gehrard. 2009. Present Perfects compete. *Linguistics and Philosophy* 32: 115–41. [CrossRef]
Schaden, Gehrard. 2012. Modelling the 'aoristic drift of the Present Perfect' as Inflation. An Essay in Historical Pragmatics. *International Review of Pragmatics* 4: 261–92. [CrossRef]
Schwenter, Scott. 1994. The grammaticalization of an anterior in progress: Evidence from a Castilian Spanish dialect. *Studies in Language* 18: 71–111. [CrossRef]
Schwenter, Scott, and Rena Torres Cacoullos. 2008. Defaults and indeterminacy in temporal grammaticalization: The 'perfect' road to perfective. *Language Variation and Change* 20: 1–39. [CrossRef]
Soto, Guillermo. 2015. El pretérito perfecto compuesto en el español estándar de nueve capitales americanas: Frecuencia, subjetivización y deriva aorística. In *Formas Simples y Compuestas de Pasado en el Verbo Español*. Edited by Susana Azpiazu. Lugo: Axac, pp. 131–46.
Squartini, Mario, and Pier Marco Bertinetto. 2000. The simple and compound past in Romance languages. In *Tense and aspect in the languages of Europe*. Edited by Östen Dahl. Berlin: Mouton de Gruyter, vol. 6, pp. 385–402.
Sundell, David. 2010. El Español Neutro en la Traducción Intralingüística. Un Estudio Sobre el uso del Español Neutro en las Traducciones Intralingüísticas de Harry Potter y la Orden del Fénix. Master's thesis, University of Oslo, Oslo, Norway.
Tarantino, Patricio. 2018. *Historia Secreta del Mundo Mágico*. Buenos Aires: Numeral.
van der Klis, Martijn, Bert Le Bruyn, and Henriëtte de Swart. 2017. Mapping the Perfect via Translation Mining. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2—Short Papers*. Edited by Mirella Lapata, Phil Blunsom and Alexander Koller. Valencia: Association for Computational Linguistics, pp. 497–502.
van der Klis, Martijn, Bert Le Bruyn, and Henriëtte de Swart. 2021. A multilingual corpus study of the competition between PAST and PERFECT in narrative discourse. *Journal of Linguistics* 58: 423–57. [CrossRef]
van der Klis, Martijn, and Jos Tellings. 2022. Multidimensional scaling and linguistic theory. *Corpus Linguistics and Linguistic Theory*, *advance online version*.
Verkuyl, Henk J. 1972. *On the Compositional Nature of the Aspects*. Dordrecht: D. Reidel Publishing Company.
Verkuyl, Henk J. 1993. *A Theory of Aspectuality. The Interaction between Temporal and Atemporal Structure*. Cambridge: Cambridge University Press.
Verkuyl, Henk J. 1999. *Aspectual Issues. Studies of Time and Quantity*. Stanford: CLSI Publications.
Vet, Co. 1992. Le passé composé: Contextes d'emploi et interprétation. *Cahiers de Praxématique* 19: 37–59.
Wälchli, Bernhard, and Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50: 671–710. [CrossRef]
Westmoreland, Maurice. 1988. The distribution and the use of the present perfect and the past perfect forms in American Spanish. *Hispania* 71: 379–84. [CrossRef]
Winter, Bodo. 2019. *Statistics for Linguistics. An Introduction Using R*. New York and London: Taylor and Francis.