

**Preprint.** To appear in: Cristina Sanz & Alfonso Morales-Front (eds.). *Issues in Hispanic and Lusophone Linguistics. Selected Proceedings of HLS 2016*. Johns Benjamins.

**The Progressive-to-Imperfective shift: contextually determined variation in Rioplatense, Iberian, and Mexican Altiplano Spanish.**

---

**Martín Fuchs<sup>1</sup>, Ashwini Deo<sup>2</sup> & María Mercedes Piñango<sup>1</sup>.**

<sup>1</sup> Department of Linguistics, Yale University.

<sup>2</sup> Department of Linguistics, The Ohio State University.

---

**Keywords:** Grammaticalization, Variation, Progressive, Imperfective, Context Modulation.

**Abstract:** Spanish has two markers (claimed to be in free alternation) to convey that an event is in progress at reference time: the Simple Present (e.g., *canta*, ‘sings’) and the Present Progressive (e.g., *está cantando*, ‘is singing’). Based on evidence from sentence acceptability studies in three different Spanish dialects, we show that the distribution of the two markers is not random, but sensitive to contextual modulation. Specifically, results show that the (ambiguous) Simple Present is more acceptable in contexts where interlocutors share perceptual access to the event at issue. Otherwise, participants favor the (unambiguous) Present Progressive. We conclude that this variation reflects and is constrained by the well-attested *grammaticalization path* in which progressive markers gradually generalize to become imperfective markers.

---

## **1. Introduction**

Crosslinguistically it has been observed that the link between specific functional meanings and their corresponding linguistic markers changes systematically across time (Bybee et al., 1994; Dahl, 1985; Traugott, 1989; *i.a.*). In light of their systematicity, such changes are viewed as manifestations of *unidirectional grammaticalization paths* that manifest cyclically. Although the empirical observations that give rise to these characterizations are robustly attested, the representations and processes that might support them are still not well understood. Here, we seek

*Corresponding author:*

Martín Fuchs.

370 Temple St., New Haven, CT, USA

[martin.fuchs@yale.edu](mailto:martin.fuchs@yale.edu)

to bridge that gap by examining the Progressive-to-Imperfective shift from the perspective of synchronic variation.

One prerequisite for analyzing any particular grammaticalization path is a characterization of: (a) its *structural component*, including its participating meanings, (b) its *dynamic component*, including the series of sub-steps that comprise the shift, and (c) the factors that enable the transition from one stage to the following. These are explained below.

The Progressive-to-Imperfective shift has been attested in different languages and language families, such as Turkish (Göksel & Kerslake, 2005), Tigre (Raz, 1983), and some Indo-Aryan languages (Deo, 2006). The observation is that progressive marking is innovated in an existing linguistic system, co-exists in variation with priorly existing imperfective marking, and gradually “becomes” the general imperfective marking, ousting out the older imperfective (Bybee et al., 1994; Comrie, 1976; *i.a.*).

Deo (2015) provides the first comprehensive proposal that captures the *structural* and the *dynamic* components of this particular change. The model, however, neither makes predictions regarding the variation between the progressive and the imperfective forms at various synchronic stages over the course of the change, nor does it specify the contextual factors that ultimately support the transition between the sub-stages of the change. Here, we address these issues by conceptually grounding the *structural component*. This allows us to make behavioral predictions for synchronic variation, which is the basis for the possible steps within the larger pathway of change. To this end we investigate the Spanish tense-aspect system, which contains both a progressive and an imperfective marker (i.e., the Present Progressive and the Simple Present respectively). These have been observed as markers of the Progressive meaning, as in sentences (1a) and (1b):

- (1) a. Ana est-á fuma-ndo.  
 Ana be-PRS.3.SG smoke-PROG.  
 ‘Ana is smoking’
- b. Ana fum-a.  
 Ana smoke-PRS.IPFV.3.SG  
 ‘Ana smokes/is smoking’

One traditional implicit assumption about the Spanish system (or other Spanish-like) systems is that these two forms are in *free alternation*, so that a speaker who intends to convey the Progressive meaning may use either (1a) or (1b) randomly (e.g., Comrie, 1976). Here we claim that the assumption that the two variants alternate freely has not been empirically justified. Instead, it seems to be a consequence of the difficulty to identify the factors that might underlie a possible variation pattern; factors that may ultimately underpin a process of change.

By contrast, we propose that the use of these markers is subject to contextual constraints – a possibility that suggests that variation is rooted in fundamental properties of meaning structure that ultimately guide diachronic change. Specifically, we propose that when conveying information that some event is in progress, a speaker must assess whether she and her interlocutor have shared perceptual access to the event described by the predicate. We argue that the choice between the *ambiguous* Simple Present and the *unambiguous* Present Progressive depends on whether the interlocutors can rely on shared perceptual access to disambiguate the intended meaning or whether in the absence of sufficient extra-linguistic context they must use the unambiguous form. This is the **Shared Perceptual Access Hypothesis**.

The remainder of the paper is structured as follows. **Section 2** provides morphosemantic considerations regarding both the synchronic characterization of the Progressive and Imperfective

meanings – the *structural component* – and the description of the series of steps involved in the Progressive-to-Imperfective shift – the *dynamic component* – (Deo, 2009, 2015). **Section 3** presents a description of the distribution of PROG-marking and IMPF-marking in Modern Spanish, which motivates our study. **Sections 4** and **5** present the methods for data acquisition and results. Finally, **Section 6** discusses the results, provides some conclusions, and indicates directions for further research.

## 2. Morphosemantic Considerations of the Progressive-to-Imperfective Shift

### 2.1 A synchronic characterization of the Progressive and Imperfective meanings

From a descriptive perspective, there are several indications that the Progressive and Imperfective meanings share a semantic core. On the one hand, IMPF-marking is usually associated with three main readings: (a) the *event-in-progress* reading, (b) the *characterizing* reading, and (c) the *continuous* reading with lexically stative predicates, illustrated in (2):

- (2) a. Ana fum-a afuera ahora (event-in-progress)  
 Ana smoke-PRS.IPFV.3.SG outside now.  
 ‘Ana is smoking outside now’.
- b. Ana fum-a diez cigarrillos por día. (characterizing)  
 Ana smoke-PRS.IPFV.3.SG ten cigarettes a day.  
 ‘Ana smokes ten cigarettes a day’.
- c. Ana viv-e en Bogotá. (continuous)  
 Ana live-PRS.IPFV.3.SG in Bogotá.  
 ‘Ana lives in Bogotá’.

On the other hand, PROG-marking is most saliently associated with the *event-in-progress* reading in (3):<sup>1</sup>

---

<sup>1</sup> An anonymous reviewer points out that the progressive marker can also present a *habitual* reading (*Ana está fumando diez cigarrillos por día* ‘Ana is smoking ten cigarettes a day’) or a *characterizing* reading with

- (3) Ana est-á fuma-ndo afuera. (event-in-progress)  
 Ana be- PRS.3.SG smoke-PROG outside  
 ‘Ana is smoking outside’

This overlap in readings has led to the proposal that the Progressive is a subtype of Imperfective, having a more specific meaning that can be included within the more general imperfective one. This characterization of the relation between the two meanings is also supported by *semantic blocking* facts: languages that present two distinct morphological markers often show that the IMPF-marker is restricted to the *characterizing and continuous* readings.

Seeking to account for the range of imperfective readings, Deo (2009) provides a unified account of the two meanings. Under this account, the Progressive and the Imperfective meanings are characterized as operators. Both operators, labeled PROG and IMPF respectively, apply to a predicate of eventualities and modify its aspectual nature. They are similar in that both involve a universal quantifier whose domain of quantification is a *regular partition* of some interval; i.e., a set against which the instantiation of a given predicate is evaluated regarding its distribution over time. The idea of a *regular partition* captures the intuition that events described by the Imperfective are understood to be regularly *distributed over time*. The *partition-measure*, which determines the value of each cell of the regular partition, is a free variable with a value determined by context. The distinct readings associated with the markers arise as a result of different values of the partition-measure in different contexts.

The above characterization further specifies that the contrast between PROG and IMPF is that in

---

stative predicates (*Ana está viviendo en Bogotá* ‘Ana is living in Bogotá’). These cases, however, license the inference that the situation described by the predicate is temporally contingent and subject to change (e.g., Comrie, 1976). We take cases like these to be driving the vanguard side of the change (i.e., the encroachment of the progressive marker into the domain of the imperfective). Space limitations force us to leave discussion of this possibility and its implications to future reports.

the case of PROG the domain of the quantifier is *a regular partition of the reference interval*, whereas in the case of IMPF it is a *regular partition of some superinterval of the reference interval*.

In what follows, we describe how the process of composition takes place.

When the PROG operator combines with a predicate, it returns the proposition that there is some reference interval in which every cell of a regular partition coincides with the predicate. Moreover, the *partition-measure* is set by context to an infinitesimal length, which gives rise to the *event-in-progress* reading. So, in a PROG-marked sentence such as (3), repeated here as (4), the ‘smoking’ event is said to coincide with every cell of a regular partition of the reference interval; that is, its *regular distribution* over time resides in that it extends over every small part of the reference interval:

- (4) Ana est-á fuma-ndo afuera.  
Ana be- PRS.3.SG smoke-PROG outside  
‘Ana is smoking outside’.

By contrast, when the IMPF operator combines with a predicate, it returns the proposition that there is a superinterval of the reference interval in which every cell of a regular partition coincides with the predicate. When the *partition-measure* is set by context to a larger length, it gives rise to the *characterizing* reading. So, in an IMPF-marked sentence such as (2b), repeated here as (5), the ‘smoking’ event coincides with every cell of a regular partition of a superinterval of the reference interval. Its *regular distribution* over time resides in that it is instantiated at every cell given by the partition of that superinterval (i.e., *ten times a day*):

- (5) Ana fum-a diez cigarrillos por día  
Ana smoke-PRS.IPFV.3.SG ten cigarettes a day.  
‘Ana smokes ten cigarettes a day’.

Further, when the IMPF operator applies to a predicate and the context provides an infinitesimal *partition-measure* for the regular partition, the reading obtained is that the event is instantiated at every infinitesimal part of the interval. This is the *event-in-progress* reading that IMPF may generate, as in (2a), repeated here as (6):

- (6) Ana fum-a afuera ahora  
 Ana smoke-PRS.IPFV.3.SG outside now.  
 ‘Ana is smoking outside now’.

The *characterizing* and the *event-in-progress* readings of the IMPF-marker depend thus on the context in which the sentence is uttered. In the case of the *event-in-progress* reading of the IMPF-marker, the *partition-measure* is set by context to an infinitesimally small length (i.e., the predicate is instantiated at infinitesimally small cells of a regular partition of an interval). When the predicate is instantiated at larger cells of a regular partition, the *characterizing* reading obtains.

This analysis also tells us that a progressive-marked sentence *asymmetrically entails* an imperfective-marked sentence: PROG is a semantically narrower version of IMPF. This is because if every cell of the reference interval coincides with some event, it logically follows that there is a superinterval of the reference interval such that every cell in a regular partition of that superinterval coincides with the event. Positing a shared semantic core for the Progressive and the Imperfective in this manner both helps us to understand the cross-linguistic distribution of PROG-marking and IMPF-marking, and sets up the *structural component* in the unidirectional language shift from Progressive-to-Imperfective.

## 2.2. The grammaticalization path from PROG to IMPF

Within the *grammaticalization* literature, it is a well-attested fact that PROG-markers emerge in

languages to realize the PROG operator, get grammaticalized, and gradually evolve into realizations of the IMPF operator (e.g., Bybee et al., 1994). Deo (2015) describes this process as a series of four stages that repeat in cyclic fashion:

- i. A *context-dependent* stage, in which there is only one IMPF-marker that is compatible with both *event-in-progress* and *characterizing* readings, which are disambiguated in context.
- ii. A *partially context-dependent* stage, where a PROG-marker arises in order to express *event-in-progress* readings in some contexts, while the IMPF-marker can still be used in all contexts.
- iii. An *explicit marking* stage, in which PROG-marking becomes obligatory to express *event-in-progress* readings, and thus restricts IMPF-marking to express *characterizing* and *continuous* readings by *semantic blocking* (i.e., IMPF-marking gets associated to contexts in which PROG-marking cannot occur).
- iv. A *context-dependent'* stage, in which PROG-marking generalizes to all contexts and becomes the only marker that expresses *event-in-progress*, *characterizing*, and *continuous* readings. This stage is equivalent to the first stage.

Given these stages, three diachronic transitions are observed: the *recruitment* of a new device in the language to unambiguously express a new functional meaning; the *categorization* of this innovated marker, which begins to be obligatory in some contexts; and the *generalization* of the innovated marker to all contexts and to the wider domain of the Imperfective.

But how precisely do such transitions proceed? Here we follow Deo (2015), who proposes that these transitions are guided by the communicative strategies of speakers and hearers at different

points in time. One may take assertions in the Imperfective domain to refer either to events that occur at reference time, presenting the *event-in-progress* reading (e.g., *Ana is smoking*), or to events that characterize the world at a given reference time, displaying the *characterizing* reading (e.g., *Ana smokes*). This *conceptual* distinction triggers the appearance and development of a *linguistic* contrast between progressive and imperfective marking whereby the emergence of a progressive marker and its gradual grammaticalization is induced and maintained by a functional pressure to unambiguously express the Progressive meaning; the meaning that makes reference to the here and now.

Here is how the process of emergence of the marker takes place. A language in the *context-dependent* stage will disambiguate between these two types of assertions by relying on shared contextual information between speaker and hearer. This strategy, however, does not guarantee communicative success; the ambiguity inherent to a single marker can result in misunderstandings triggered by possible mismatches in the contextual assessment of the speaker and the hearer. Disambiguation in such a system can be achieved either through pragmatic context or through the use of adverbials that explicitly indicate which of the two meanings is the intended one (e.g., *now* vs. *every day*).

At some point – and given that pragmatic context is not always successful in disambiguating meanings – the use of adverbials or other markers that make salient the *event-in-progress* reading will be established as a convention in a speech community. The *partially context-dependent* stage is the result of such conventionalization of a grammatical device to disambiguate between the two possible readings. This new marker serves the purpose of restricting the temporal interpretation of the predicate to an event that occurs at reference time, the *event-in-progress* reading.

The *explicit marking* stage defines the period in which the new progressive marking entrenches

in the grammar – extending to new syntactic contexts, for instance – and becomes obligatory for referring to events that occur at reference time. In turn, imperfective marking gets restricted to events that characterize the world at reference time. Such a system constitutes an optimal solution to the communicational goal of conveying unambiguously the relevant meanings.

However, imperfect learning of the system can lead to a new *context-dependent* stage. We can assume that during the process of language acquisition, children and caregivers engage more frequently in interactions about events in progress, so that PROG-marking becomes more used than IMPF-marking. Increased frequency can lead children to overgeneralize and extend PROG-marking to the broader domain of the Imperfective in their production, triggering the observed *generalization* process.

In what follows we present the facts of Spanish that illustrate the Progressive-to-Imperfective grammaticalization path in the language.

### **3. The Spanish Progressive and Imperfective markers**

As mentioned above, traditional descriptions of Spanish assume that PROG-marked sentences unambiguously express an ongoing event, while IMPF-marked sentences are able to express either a habit (i.e., a *characterizing* reading) or an ongoing event (Yllera, 1999; NGRAE, 2009; *i.a.*). These accounts suggest that the *event-in-progress* reading can be expressed optionally by either PROG-marking or IMPF-marking.

By contrast, other accounts present the intuition that PROG-marking and IMPF-marking are not in *free alternation*, but rather contextually determined. Those accounts, however, do not make clear what the relevant contextual properties should be. For instance, they report that the Spanish PROG-marker *actualizes* an event (Fernández de Castro, 1999) or *enhances* its duration (Roca Pons, 1958), yet they leave the key processes of *actualization* or *enhancement* vaguely defined.

A closer diachronically-informed look reveals evidence for a *partially context-dependent* stage: Spanish shows a clear morphological exponent for PROG, which alternates with the IMPF-marker to express the *event-in-progress* reading. In line with contextual-determination approaches, we expect the existence of contextual factors that make one marker more or less acceptable than the other in expressing the *event-in-progress* reading. If this is correct, the markers would not alternate freely, but their use would be conditioned by additional features that have not yet been factored into the model. This would also lead to a view of the differences between dialects of Spanish as the diachronic shift at work in different ‘synchronic cuts’: different dialects could be at different sub-stages of the same unidirectional shift, just as different languages may be at different stages along the same grammaticalization path (e.g., Bybee et al. 1994).

We capture the differential use of PROG-marking and IMPF-marking in Spanish through the Shared Perceptual Access Hypothesis: the claim that the markers are not *alternating freely*, but that the use of these two variants is *contextually conditioned* by whether the speaker and the hearer share perceptual access to the event described by the predicate. From this hypothesis, it follows that in Spanish IMPF-marking is already a dispreferred form to convey the *event-in-progress* reading and that PROG-marking has increased in frequency taking over most of the instances where the Progressive meaning is to be expressed (e.g., Torres Cacoullos 2012). If this is true, Spanish would be *between* the *partially context-dependent* stage and the *explicit marking* stage – the alternation is not free, but PROG-marking has not yet been made completely obligatory to express the Progressive meaning.

One concrete way to test the hypothesis that progressive marking arises due to communicational pressures to refer to the *event-in-progress* reading unambiguously would be to examine diachronic records of languages that have been documented over long periods of time.

Here, we take a different approach by examining instead the synchronic variation measured through sentence acceptability judgments across three distinct Spanish dialects. The details of this study are presented directly below.

#### **4. An experimental study in three dialects of Spanish**

We test the Shared Perceptual Access Hypothesis through three sentence acceptability studies – one per dialect under consideration. The Shared Perceptual Access Hypothesis states that while PROG-marking is always appropriate for conveying the Progressive meaning in Spanish, the use of IMPF-marking for the same goal is modulated by whether the speaker and the hearer share perceptual access to the event described by the predicate.

We operationalized *context* as a variable by defining two types of contexts based on the amount of information they contain: Rich Experiential Contexts and Poor Experiential Contexts. A Rich Experiential Context places both speaker and hearer in the same perceptual environment such that they both have equal access to the physical environment in which the event described by the utterance takes place. It is this property what allows the hearer to access more easily the *event-in-progress* reading of an IMPF-marked utterance. A Poor Experiential Context lacks this feature. In this context, the physical environment is not shared, and it is this reduction in shared information what affects the hearer's ability to access the intended *event-in-progress* reading from the ambiguous IMPF-marked utterance.

If the two markers are in *free variation* with respect to this parameter, changing the larger situational contexts for a PROG-marked sentence or an IMPF-marked sentence should not yield a change in the degree of acceptability for either marker to convey the *event-in-progress* reading. On the other hand, if the *Shared Perceptual Access Hypothesis* is on the right track, the acceptable use of the imperfective marker with an *event-in-progress* reading will be mediated by the richness

of the context. In Rich Experiential Contexts, both markers should be accepted equally to convey the *event-in-progress* reading. By contrast, in Poor Experiential Contexts, the preferred way to convey this reading would be via the unambiguous PROG-marking, given that, unlike IMPF-marking, it eliminates the possibility of miscommunication resulting from ambiguity.

In summary, the sentence acceptability studies were designed to address three issues: (i) a clear picture of which diachronic sub-stage different dialects of Spanish are in, (ii) a test as to whether the markers are in free variation or have a contextually determined distribution, and if the use of the markers is mediated by context, (iii) a way to operationalize the *richness* of the context (determined by degree of shared perceptual access) as the factor that modulates the choice of marker to convey the *event-in-progress* reading.

#### 4.1. Methods

##### 4.1.1. Materials

A contextual vignette was presented to the participant, which featured a speaker and a hearer. Following the description of the larger context, the speaker made an utterance to the hearer. Those utterances were identical in each contextual condition. The studies had a 2x3 design, which resulted in a total of 6 conditions. The independent variables were the following:

- a. **Context:** The contextual vignettes presented either a *Rich Experiential Context* or a *Poor Experiential Context*, which differed minimally. In Rich Experiential Contexts, there was enough contextual information to assume that the interlocutors shared perceptual access to the event described by the predicate. On the other hand, Poor Experiential Contexts were cases in which speakers could not assume that hearers had perceptual access to the event. Examples are given in (7) and (8):

**Rich Experiential Context**

- (7) *Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación. Golpea la puerta, la abre, y ve al hijo sentado en el escritorio. Antes de que ella diga nada, el hijo le dice:*

‘Anna comes home from work, and goes to her son’s room to look for him. She knocks on the door, opens it, and sees him sitting at his desk. Before she can say anything, her son tells her:’

**Poor Experiential Context**

- (8) *Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación. Golpea la puerta, pero el hijo no contesta. Sin que ella llegue a abrir la puerta, el hijo le dice:*

‘Anna comes home from work, and goes to her son’s room to look for him. She knocks on the door, but her son does not answer. Before she gets to open the door, her son tells her:’

- b. **Aspectual Marker:** The corresponding sentences to the preceding context appeared either with IMPF-marking (Simple Present), PROG-marking (Present Progressive) or the Spanish *pretérito* (Simple Past) as a baseline condition. Examples in the progressive condition (a), the imperfective condition (b), and the *pretérito* condition (c) are given in (9):

- (9) a. Est-oy           hacie-ndo    la    tarea.  
be-PRS.1.SG   do-PROG    the   homework  
‘I am doing homework’
- b. Hag-o                    la    tarea.  
do-PRS.IPFV.1.SG    the   homework.  
‘I am doing homework’
- c. Hi-ce            la    tarea.  
do-PST.1.SG    the   homework  
‘I did homework’

All sentences were *declarative affirmative* sentences<sup>2</sup> that uniformly conveyed a Progressive meaning. The sentences had first, second, and third person singular subjects distributed evenly. The task included 30 items per condition, which resulted in a total of 180 experimental stimuli, which were adapted to each of the Spanish varieties.

#### 4.2. Procedure

Participants were presented with a contextual vignette (a Rich Experiential Context or a Poor Experiential Context) followed by only one sentence out of the three possible ones (i.e., PROG-marked, IMPF-marked, or *pretérito*-marked). They were asked to judge the acceptability of the whole context-sentence pairing. Participants were asked to judge if the choice of the speaker in the contextual vignette was a felicitous one given the setting they were involved in. The context-sentence pair appeared together on the screen, with button choices from 1 to 5 to select a rating. Five practice stimuli were presented before the beginning of the task, together with clear instructions, which asked the participant to perform the task in a single sitting, without distractions, while in a quiet room. The instructions also presented definitions for each point on the scale from 1 to 5, which ranged from “I don’t understand the phrase. A native speaker would not say it” to “I understand the phrase perfectly and I would say it myself”.

The studies were administered online in a survey platform (Qualtrics, 2015) during the months of December 2015 and January 2016. 10 blocks of 39 stimuli each were divided in three sessions. Each complete session lasted approximately 60 minutes. Once they were done with a session, participants received an URL link to the following one. For each session, subjects were

---

<sup>2</sup> See Yllera (1999), Torres Cacoullos (2012), on why negatives and interrogatives contexts favor the appearance of IMPF-marking in the expression of the *event-in-progress* reading.

compensated with 10 US dollars, to a total of 30 US dollars. Yes-no comprehension questions were presented after 75% of the stimuli. For instance, after the context-sentence pair presented in (7/8-9) above, a question such as *¿Ana llegó del trabajo hace unas horas?* ‘Did Ana come back from work a few hours ago?’ would appear on the screen. These questions were designed so that they required comprehension of the whole context-sentence pair, and could not be answered correctly by mere scanning of the vignettes.

Given that the prediction for ratings of the PROG-marker was that they would always be high ratings – because PROG-marking is the unambiguous choice to express the *event-in-progress* reading – 30 fillers were developed, which described a past completed event (unacceptable with the PROG-marker).<sup>3</sup> Conversely, given that the prediction for the *pretérito*-marker was that it would always be rated low – because its meaning is not compatible with describing events in progress –, another 30 filler items were constructed; these described completed past events that were acceptable with the *pretérito*-marker. This was to avoid the possibility that subjects gave high scores just because they encountered the PROG-marker or provided only low ratings when evaluating the *pretérito*-marker. In summary, these fillers were designed to obtain the opposite pattern of ratings than those in the target stimuli. Another 150 fillers from an unrelated study about copula alternation in Spanish completed the experimental stimuli, resulting in a total of 390 items. Items were divided in 10 blocks with 3 stimuli per condition of the present study (18 stimuli per block) plus 21 fillers. There was full items randomization within blocks, while the order of presentation of the blocks to participants was pseudo-randomized by the experimenter, so that each subject saw the blocks in a different order.

---

<sup>3</sup> For instance, a context in which a friend asks another one out to go see a movie, and she answers that she has already seen that movie. In that context, the use of PROG-marking to convey an event in the past would be unacceptable, so participants rated those accordingly.

#### 4.3. Subjects

Speakers of three different Spanish dialects were assessed: 39 Iberian Spanish speakers (20 female) from Madrid, Castilla-La Mancha y Castilla y León; 38 Rioplatense Spanish speakers (21 female) from Buenos Aires, and 37 Mexican Altiplano Spanish speakers (18 female) from Mexico City and the states of Morelos, Tlaxcala, Puebla and the south of Hidalgo. All participants were between 18 and 37 years old, and had at least finished 12 years of schooling. Subjects had not lived in other states or provinces for more than a year.

#### 4.4. Predictions

If the markers are in *free variation*, participants' ratings will not be impacted by changes in the *richness* of the contextual information. Moreover, the acceptability of PROG-marked sentences and IMPF-marked sentences should be not significantly different to express the *event-in-progress* reading. By contrast, if the markers are not in *free variation* and their use is instead driven by *Shared Perceptual Access*, participants' ratings will be impacted by the *richness* of the context and by the presence of different markers when judging the acceptability of context-sentence pairs that are expressing the *event-in-progress* reading. Specifically, while PROG-marked sentences will always be high ranked and *pretérito*-marked sentences will always be low ranked, there could be contextual modulation in the cases of the IMPF-marked sentences. When enough information is provided by the context to help the hearer disambiguate between the available readings of the imperfective marker, participants will judge the choice of IMPF-marker for conveying the *event-in-progress* reading as more acceptable than when the context does not provide such information. If we observe differences of acceptability between dialects, they should only be in the IMPF-marked sentences, and they should point to different sub-stages in the grammaticalization path under scrutiny.

## 5. Results

We performed an analysis that checked participants' accuracy in the comprehension questions, by coding each answer as correct or incorrect. All subjects answered correctly more than 85% of these questions, so no subject was excluded from subsequent analyses. We also checked the ratings in the two filler conditions, as a measure of sensitivity to the task. PROG-marked sentences that expressed an event in the past had low rankings across dialects (mean = 2.18, SE = 0.05).<sup>4</sup> In turn, *pretérito*-marked sentences that expressed an event in the past presented high ratings (mean = 4.51, SE = 0.04).

To analyze the experimental data, we performed a *linear mixed effects analysis* in R (R Core Team, 2015) using the *lme4* package (Bates, Maechler & Bolker, 2015). For model selection, we followed regular recommendations for linguistic analysis (Winter, 2013) and performed Likelihood Ratio Tests of the full model with the effect in question against the model without it. Post hoc tests were run with the *multcomp* package (Hothorn, Brest & Westfall, 2008), and corrected by Tukey. The model analyzed the effects of three variables – dialect, context, and aspectual marker – on our dependent variable, the participants' ratings. As fixed effects, and given our predictions in §4.4., we entered the interaction of context, aspectual marker and dialect. As random effects, the model had random intercepts for subject and item and by-subject random slopes for the effects of aspectual marker, context and grammatical person, which was the maximal random effect structure justified by the data. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

The model with a fixed effect of the interaction between context, aspectual marker and dialect

---

<sup>4</sup> All reported standard errors are calculated not over tokens, but over subject means.

performed significantly better to explain the data than the model without the effect ( $\chi^2(17) = 436.13, p < .001$ ). Post hoc tests with Tukey correction showed a main effect of aspectual marker, which significantly favored the PROG-marker over the IMPF-marker ( $\beta = 1.23, p < .001$ ), and a significant main effect of context that favored Rich Experiential Contexts over Poor Experiential Contexts ( $\beta = .477, p < .001$ ). On the other hand, there was no main effect of dialect (Iberian – Rioplatense,  $\beta = .269, p = .127$ ; Mexican – Rioplatense,  $\beta = .074, p = .857$ ; Iberian – Mexican,  $\beta = .195, p = .342$ ). No effects of gender or age of the participants were found in the data.

The interaction of aspectual marker by context by dialect was a significant fixed effect in the model. To assess the source of the significant differences, we subset the data by dialect and performed a linear mixed effect analysis for each dialect analyzing the fixed effect of context and aspectual marker. Random effects were kept the same as in the *all-dialects* model. For **Rioplatense Spanish**, the model with a fixed effect of the interaction between context and aspectual marker performed significantly better than the null model to explain the data ( $\chi^2(5) = 109.41, p < .001$ ). We then subset the data by aspectual marker and analyzed the effect of context as a fixed effect on each of the aspectual markers. In the case of the PROG-marker, no significant effect of context was found ( $\chi^2(1) = 0.276, p = .599$ ). No context effect was found in the case of the *pretérito*-marker either ( $\chi^2(1) = 0.330, p = .566$ ). However, in the case of the IMPF-marker, a significant difference arose between contexts ( $\chi^2(1) = 11.125, p < .001$ ). A post hoc test with Tukey correction showed a significant difference in favor of the Rich Experiential Context over the Poor Experiential Context ( $\beta = 0.477, p < .001$ ).

**Iberian Spanish** showed the same pattern as Rioplatense Spanish. A linear mixed effects model with the fixed effect of the interaction between context and aspectual marker performed significantly better than the null model ( $\chi^2(5) = 149.01, p < .001$ ). Subsets by aspectual marker

showed no significant effect of context neither on the PROG-marker ( $\chi^2(1) = 0.585$ ,  $p = .445$ ) nor on the *pretérito*-marker ( $\chi^2(1) = 0.004$ ,  $p = .950$ ). However, the IMPF-marker showed significant contextual modulation ( $\chi^2(1) = 16.746$ ,  $p < .001$ ), favoring again Rich Experiential Contexts over Poor Experiential Contexts in a post hoc test corrected by Tukey ( $\beta = 0.478$ ,  $p < .001$ ).

**Mexican Altiplano Spanish** was subjected to the same linear mixed effects analysis. While the model with the fixed effect of the interaction between context and aspectual marker performed significantly better than the null model ( $\chi^2(5) = 92.756$ ,  $p < .001$ ), it did not perform better than a model that only had a main effect of aspectual marker as a fixed effect ( $\chi^2(3) = 4.895$ ,  $p = .18$ ), showing that the effect firstly observed was only due to the main effect of aspectual marker and was not modulated by the type of context. Given our particular interest for the IMPF-marker, we analyzed the effect of context on that particular marker, which showed no significant difference to the null hypothesis ( $\chi^2(1) = 0.185$ ,  $p = .667$ ).

A summary of the results in terms of the participants' ratings means by context, aspectual marker and dialect is given in Table 1. Standard errors are indicated in parentheses. Conditions where we found significant differences are bolded:

	<i>Iberian Spanish</i>		<i>Rioplátense Spanish</i>		<i>Mexican Alt. Spanish</i>	
	<i>Rich</i>	<i>Poor</i>	<i>Rich</i>	<i>Poor</i>	<i>Rich</i>	<i>Poor</i>
PROG	4.78 (.03)	4.74 (.03)	4.68 (.05)	4.66 (.05)	4.51 (.06)	4.46 (.06)
IMPF	<b>4.18 (.11)</b>	<b>3.70 (.09)</b>	<b>3.90 (.11)</b>	<b>3.43 (.08)</b>	3.57 (.12)	3.51 (.12)
PRÉT	2.16 (.08)	2.15 (.09)	2.67 (.08)	2.57 (.08)	2.67 (.09)	2.63 (.08)

TABLE 1: Participants' ratings means and standard errors by condition (Dialect \* Aspectual Marker \* Context).

Figure 1 also presents a summary of the results in terms of means of the ratings – by dialect and by aspectual marker –, and shows the significant differences between the context conditions.

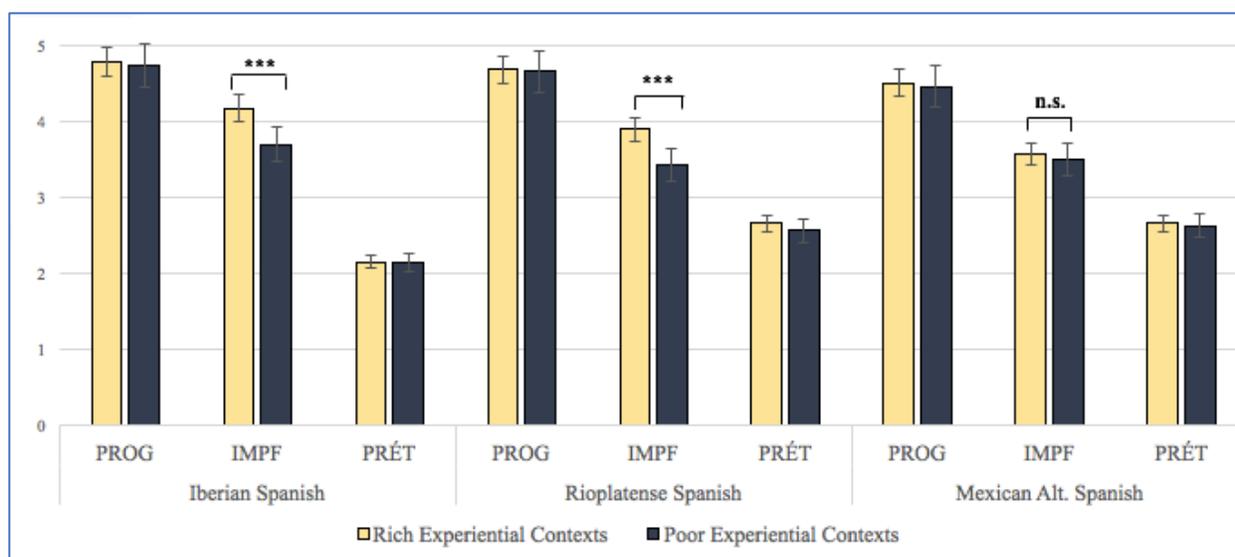


FIGURE 1: Effect of Context on Aspectual Marker by Dialect (Means)

## 6. Discussion and concluding remarks

The first conclusion that can be drawn from the data is that in the three Spanish dialects that were studied PROG-marking and IMPF-marking are not in free alternation to express the *event-in-progress* reading. PROG-marking has reached ceiling ratings in all three dialects, and it is the preferred form to express the *event-in-progress* reading. On the other hand, IMPF-marking is not completely limited to the non-progressive readings (i.e. *characterizing* and *continuous* readings), but its acceptability in conveying the *event-in-progress* reading is significantly lower than that of PROG-marking. In the case of Rioplatense and Iberian Spanish, it seems that this alternation is *contextually determined* – i.e., speakers have a *partially context-dependent* strategy to express the *event-in-progress* reading. In contrast, in Mexican Altiplano Spanish, the acceptability of IMPF-marking to express the *event-in-progress* reading is low regardless of the *richness* of the contextual information – the main strategy to convey the *event-in-progress* reading is an *explicit marking* strategy.

The main finding is that context has an effect on the acceptability of a given marker. In

particular, the acceptability of IMPF-marking for describing events in progress was enhanced in Rioplatense and Iberian Spanish when the context in the vignette assumed shared perceptual access between speaker and hearer (i.e., *Rich Experiential Contexts*); moreover, the acceptability in those cases was significantly higher than when the context did not carry this feature (i.e., *Poor Experiential Contexts*). The cases in which *rich* information was presented allowed the speaker to rely on this contextual information and use a broader imperfective marker. Conversely, in the cases where the information given in the scenario was *poor*, the speaker had to assume that the hearer needed explicit linguistic information to disambiguate the reading at play and thus be able to comprehend the intended reading. Therefore, in those scenarios, the more acceptable marker was the unambiguous PROG-marker, while the IMPF-marker was dispreferred.

This outcome provides support to the *Shared Perceptual Access Hypothesis*: the alternation between the markers is *contextually determined* by whether the speaker and the hearer share perceptual access to the event described by the predicate. The factor that modulates the acceptability of each marker is the speaker's assessment of the richness of the context, which expresses the convergence of contextual information and shared perceptual access between them and the hearer. Thus, when expressing the Progressive meaning in a dialect with two distinct markers whose alternation is contextually determined, the speaker has the choice of relying on contextual information or using an unambiguous marker. On the other hand, in the case of Mexican Altiplano Spanish, IMPF-marking is equally dispreferred in Rich and Poor Experiential Contexts to express the *event-in-progress* reading, showing that contextual information and shared perceptual access no longer modulate the acceptability of the marker. Decreased sensitivity to contextual information correlates with lowering the acceptability of a semantically broader form.

These differences between dialects can be explained within the *diachrony-based* theory of

variation that we propose here. We expect a PROG-marker to emerge in a language to express unambiguously the *event-in-progress* reading in Poor Experiential Contexts. For Spanish, we reason that when there was only one marker (i.e, Simple Present) that expressed all possible imperfective readings, its ambiguity was more salient in Poor than in Rich Experiential Contexts. Poor Experiential Contexts could not help disambiguate between the readings, thus reducing communicative success between interlocutors. Over time, this situation motivated the recruitment of a new marker to unambiguously express the *event-in-progress* reading in those contexts. This new marker would later generalize to Rich Experiential Contexts, while also making IMPF-marking inappropriate for conveying the *event-in-progress* reading in Poor Experiential Contexts. The only situation in which IMPF-marking would then become appropriate to express the *event-in-progress* reading would be when the context was rich enough to disambiguate the intended reading. We suggest that this is the stage in which Rioplatense and Iberian Spanish are at the moment. However, Mexican Altiplano Spanish seems to be one step further in the diachronic path. PROG-marking has also become the widely-preferred option in Rich Experiential Contexts, and IMPF-marking has dropped its acceptability regardless of the richness of the contextual information. The prediction is that it will continue in that direction until the PROG-marker becomes the *only* possibility to express the *event-in-progress* reading. In summary, we can see in all the dialects how the spread of a new marker is a function of the available contextual information, and how the variation that we observe has a diachrony-based explanation, which follows a *unidirectional grammaticalization path*.

This paper presents a novel *methodological approach* to encode the role of context in language variation as a proxy to synchronically observe attested language changes, by obtaining acceptability ratings of different constructions from three samples of distinct but similar

populations; namely, speakers of three different dialects of the same language. We show that when having clear definitions of the *structural* and *dynamic* components involved in a grammaticalization path, a sentence acceptability study that manipulates the role of contextual information in a subtle – but sharp and systematic – way can be designed, and we can start to have a better grasp at how language change operates as a result of the choices that speakers make. A clear way to increase the validity of this methodology would be to extend it to other grammaticalization paths that have been observed across languages and times. The expectation is that context modulation should be found in analogous ways to the one presented in this paper.

## 7. References

- Bates, Douglas; Martin Maechler & Ben Bolker. 2015. Fitting Linear Mixed-Effects Models using *lme4*. *Journal of Statistical Software*, 67, 1-48.
- Bybee, Joan; Revere Perkins, & William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: UCP.
- Comrie, Bernard. 1976. *Aspect. An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: CUP.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Basil Blackwell.
- Deo, Ashwini. 2006. *Tense and Aspect in Indo-Aryan languages: Variation and Diachrony*. PhD thesis, Stanford University.
- Deo, Ashwini. 2009. Unifying the imperfective and the progressive: partitions as quantificational domains. *Linguistics and Philosophy*, 32, 475-521.
- Deo, Ashwini. 2015. The semantic and pragmatic underpinnings of grammaticalization paths: the progressive to imperfective shift. *Semantics and Pragmatics*, 8, 1-52.

- Fernández de Castro, Félix. 1999. *Las perífrasis verbales en el español actual*. Madrid: Gredos.
- Göksel, Asli & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. Routledge.
- Hothorn, Torsten; Frank Bretz & Peter Westfall. 2008. Simultaneous inference in General Parametric Models. *Biometrical Journal*, 50, 346-363.
- NGRAE = Real Academia Española y Asociación de Academias de la Lengua Española. 2009. Las perífrasis verbales. *Nueva Gramática de la Lengua Española*. 2105-2222. Madrid: Espasa.
- Qualtrics Platform. 2015. Qualtrics, Provo, Utah, USA.
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raz, Shiomo. 1983. *Tigre grammar and texts*. Undena Publications.
- Roca Pons, José. 1958. *Estudios sobre perífrasis verbales del español*. Madrid: CSIC.
- Torres Cacoullos, Rena. 2012. Grammaticalization through inherent variability: The development of a progressive in Spanish. *Studies in Language*, 36, 73-122.
- Traugott, Elizabeth. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, 65, 31-55.
- Winter, Bodo. 2013. Linear models and linear mixed effects models in R with linguistic applications, ms.
- Yllera, Alicia. 1999. Las perífrasis verbales de gerundio y participio. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*. 3390-3441. Madrid: Espasa.