



UNIVERSITY OF HELSINKI



<https://helda.helsinki.fi>

Helda

---

## On the emergence of quotative bueno in Spanish : a dialectal view

Rosemeyer, Malte

2023-09-05

---

Rosemeyer, M & Posio, P 2023, On the emergence of quotative bueno in Spanish : a dialectal view. in D Casartelli, S Cruschina, P Posio & S Spronck (eds), The grammar of thinking : From reported speech to reported thought in the languages of the world. Trends in Linguistics. Studies and Monographs (TiLSM), vol. 379, De Gruyter Mouton, Berlin. <https://doi.org/10.1515/9783111065830-005>

---

<http://hdl.handle.net/10138/585347>

10.1515/9783111065830-005

---

unspecified

acceptedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

## On the emergence of quotative *bueno* in Spanish: a dialectal view

Malte Rosemeyer (FU Berlin)

Pekka Posio (University of Helsinki)

### Abstract

The Spanish discourse marker *bueno*, literally ‘good’, is a notoriously polyfunctional item that has been argued to serve several seemingly divergent functions, ranging from the expression of agreement to the expression of disagreement. It has also been observed that, as *bueno* frequently occurs at the beginning of reported speech, it may be grammaticalizing into a quotative marker, perhaps replacing other, more canonical markers like the verb *decir* ‘say’ (Borreguero Zuloaga 2017). In this paper we adopt a cross-dialectal view to the use of *bueno* as a discourse marker, mapping it with the expression of reported discourse (a notion subsuming both reported speech and thought) in a multi-dialect corpus of spoken Spanish. Our analysis provides quantitative and qualitative evidence for the assumption that *bueno* is grammaticalizing into a marker of reported discourse. We also show that reported speech and reported thought are not routinely distinguished from each other in Spanish: rather, reported thought is conceptualized as reported speech, recurring to the metaphor of thinking as speaking. Nevertheless, we find evidence for the assumption that in those dialects in which the use of *bueno* is particularly productive, some speakers seem to routinely associate the use of *bueno* with reported thought rather than speech.

Keywords: Discourse marker; Reported speech; Reported thought; Polyphony; Spanish; Pragmaticalization; Grammaticalization; Dialect

### 1. Introduction

In Spanish, the polyfunctional discourse marker *bueno*, grammaticalized from the adjective ‘good’, has several functions that range from the expression of agreement or disagreement to turn organization in conversation. For instance, *bueno* is frequently used at the beginning of a turn, as in example (1), taken from the PRESEEA corpus consisting of semi-structured sociolinguistic interviews.<sup>1</sup> Here, the use of *bueno* serves an interactional function that can be described as attenuation (the speaker I does not want to fully subscribe to the statement that she had a bad time with her first child). The closest equivalent of this use of *bueno* in English would be *well*. At the same time, the first occurrence of *bueno* can be considered to serve a turn-taking function (see Section 2).

- (1) [I and E are talking about I's children. I explains that she is enjoying the time with her second child more than she did with her first time.]

E: *con la primera lo pasaste peo:r ¿no?*

I: *bueno yo soy muy tranquila pero y tampoco lo pasé mal lo que pasa que bueno pues la inexperiencia te hace: ir aprendiendo en el día a día y en el... la hora a hora y en el minuto a minuto ¿no?*

E: ‘with the first one you had a worse time, right?’

I: ‘BUENO, I am very calm and I did not really have a bad time, it's rather that BUENO the lack of experience makes you learn every day and... from hour to hour and from minute to minute,

---

<sup>1</sup> The PRESEEA corpus will be introduced in Section 3.

right?’  
(ALCA\_M22\_028)<sup>2</sup>

As observed by Borreguero Zuloaga (2017), *bueno* also occurs with certain regularity in sequences of direct reported speech (see example (2)).

- (2) *y yo decía pues bueno-- me decía la gente → / ya te enterarás de cómo es ¿no? y yo decía bueno ya me enteraré*

‘and I said “okay BUENO”... people would say to me “eventually you will find out how it is”, right? and I said “BUENO yes, I will find out”’  
Borreguero Zuloaga (2017: 70, mark-up in the original, translation ours)

On the basis of an interactional analysis of *bueno* and other discourse markers in Peninsular Spanish, Borreguero Zuloaga (2017) claims that one of the functions of these discourse markers is to signal transition to reported speech. In other words, the use of *bueno* in examples such as (2) does not necessarily reflect an actual earlier use of *bueno* that is being reported by the speaker, replicating its original interactional function, but rather serves to mark the segment as reported speech. Borreguero Zuloaga’s (2017) analysis furthermore implies that this new function of *bueno* has arisen over time on the basis of the original interactional functions of *bueno* in discourse, in line with recent proposals such as Detges and Waltreit (2011), who argue that turn-taking mechanisms in interaction can lead to grammaticalization processes.

In this paper, we aim at testing two hypotheses, based on Borreguero Zuloaga’s observations from Peninsular Spanish: (1) *bueno* has developed new functions as marker of reported speech and discourse polyphony based on its earlier dialogical discourse marker uses and (2) these functions arise in those dialects with highest overall frequency of *bueno* as a discourse marker. In addition, we aim at teasing out any eventual differences in the use of *bueno* associated with reported speech (RS) and reported thought (RT). Since the distinction between RS and RT is not usually made in previous literature, we use the term reported discourse (RD) to subsume both. Due to the difficulty of tracing changes in the pragmatic functions of linguistic elements typically associated to spoken language in diachrony (see Rosemeyer 2019a), we adopt a cross-dialectal perspective. Standard theories of grammaticalization (e.g., Hopper and Traugott 2003) assume a correlation between the acquisition of new functions by a linguistic element and its usage frequency, such that an expansion of the functional domain of the element is usually accompanied by a general increase in its usage frequency. If *bueno* has indeed acquired the new function of marking polyphony, we should be able to document a cross-dialectal correlation between the proportion of discourse-marker uses of *bueno* (relative to lexical uses as an adjective) and the likelihood that speakers use *bueno* at the beginning of RD.

This chapter is structured as follows: in Section 2 we present an overview of the functions of *bueno* in addition to the “new” quotative function discussed here, in the light of both previous studies and a scrutiny of our data, which is presented in Section 3. Section 4 focuses on the qualitative analysis of *bueno* in RD in the data, and Section 5 presents a

---

<sup>2</sup> The examples follow the transcription conventions of the PRESEEA corpus; however, in the interest of perspicuity, they have been simplified with regard to the marking of overlapping turns. ‘E’ refers to the interviewer and ‘I’ to the informant. The sign : has been adopted to mark lengthening of the previous vowel and ... is used instead of the tag *vacilación* ‘hesitation’. A single (/) and double (//) slash refer to shorter and longer pauses, respectively. Since our analysis is not morphosyntactic but pragmatic, we do not provide glosses but a translation as close to the original as possible, leaving the discourse marker BUENO untranslated.

qualitative analysis of the quotative contexts across dialects displaying different usage patterns. Our findings are summarized in the concluding Section 6.

## 2. *Bueno* between argumentative, metadiscursive and polyphonic functions

There is a considerable body of research on discourse particles in Spanish (see, e.g., Martín Zorraquino and Portolés 1999; Briz, Pons, and Portolés 2008; López Serena and Borreguero Zuloaga 2010, to name but a few). As pointed out by Martín Zorraquino and Portolés (1999), the discourse particle *bueno* is extremely versatile and can express several pragmatic functions. The main functions identified in previous studies are the following:

- i. expression of agreement with the interlocutor, possibly with hedges or modifications of the propositional content being agreed with;
- ii. expression of disagreement with the interlocutor while protecting the positive face of the speaker through mitigation;
- iii. metadiscursive uses, e.g., beginning, ending or changing a conversational topic (Martín Zorraquino and Portolés 1999); reformulation of previous speech (Pons 2003), presenting what is being said as continuation of previous topic (Briz, Pons, and Portolés 2008).

It should be noted that these classifications are based on the uses of *bueno* in spontaneous conversations, while our data consist of sociolinguistic interviews (see Section 3), where it is difficult to find clear instances of agreement and disagreement with the interlocutor due to the asymmetric roles of interviewer and informant. However, our data present several contexts of question-answer turn pairs where the first member projects an affirmative or a negative answer: *bueno* typically occurs in the second members of these adjacency pairs when the answer is not simply affirmative or negative but, rather, presents a modification or elaboration of the projected response. This is illustrated by examples (3) and (4). In (3), the speaker affirms the assumption expressed by the interviewer but also modifies it: she had to quit studies for various reasons, not only family-related ones, as implied by the interviewer. In (4), the informant begins both of her turns by *bueno*: the first is a negative answer to the interviewer's question, providing an explanation of why she does not like participating in manifestations, and the second one is an evasive answer that can be interpreted as neither affirmative nor negative.

### (3) positive answer with a modification of the presupposition of the question

I: [...] *yo podía haber sido un buen estudiante lo que pasa es que lo dejé / lo tuve que dejar*

E: *por problemas familiares ¿no?*

I: *bueno sí // problemas familiares problemas económicos // ahí ayudó: / todo un poquito*

I: '[...] I could have been a good student but I quit / I had to quit'

E: 'due to family problems, right?'

I: 'BUENO yes // family problems economic problems // everything summed up there a little bit'

(MALA\_H11\_114)

### (4) negative answer with an explanation / evasive answer

E: *señora G. / y ¿le gusta ir para las marchas y eso? ¿o no? /*

I: *bueno no me gusta ir porque como me duelen las piernas / no puedo caminar mucho /*

E: *¿y tocar cacerola? //*

I: *bueno / en vez en cuando //*

E: 'Ms G. / and do you like to go to the marches and all that? or don't you?'

I: 'BUENO I don't like to go because my feet hurt / I can't walk much /'  
E: 'and bang the pots? //'  
I: 'BUENO / from time to time //'  
(CARA\_M32\_067)

In the type of question-answer pairs represented by our data, *bueno* signals that the answer is not just positive or negative – in which case the speakers tend to use the polarity items *sí* 'yes' and *no* 'no' – but rather adds an explanation or a modification of the presupposition included in the question.

As for the metadiscursive function, our data present several examples of contexts where *bueno* is not a reaction to something said by the interlocutor but rather serves topic management functions. The *bueno* in the last line of example (5), may serve simultaneously as a marker of acceptance of the answer 'it's a friendship' given by the informant and a marker of closing this topic and moving to the next one (the informant's children).<sup>3</sup>

(5) **metadiscursive marker (=non-preferred change of topic; acknowledgement: closing one conversational topic and moving to the next one)**

E: *eh / ¿eres casada actualmente? /*  
I: *no / estoy separada del papá del niño //*  
E: *eeh / ¿estás con otro muchacho?*  
I: *mmm*  
E: *bueno no / no*  
I: *no / no / no // no / no vamos a / no vamos a / a mentir // no / no es una relación*  
E: *ya*  
I: *es una amistad*  
E: *una amistad // bueno / mmm // me dijiste que tienes dos niños*  
  
E: 'eh / are you married at the moment? /'  
I: 'no / I'm separated from the father of the child //'  
E: 'eeh / are you with another guy?'  
I: 'mmm'  
E: 'BUENO no / no'  
I: 'no / no / no // let's not / let's not / lie // no / it's not a relationship'  
E: 'ok'  
I: 'it's a friendship'  
E: 'a friendship // BUENO / mmm // you told me you have two children'  
(LHAB\_M21\_019)

The fourth function of *bueno*, namely marking a sequence of speech as reported, was identified by Borreguero Zuloaga (2017), who examines the occurrence of several discourse particles in stretches of reported speech in spoken, colloquial Spanish. Her conclusion is that *bueno* has acquired a new discourse function as a marker of reported speech and discourse polyphony along with other markers such as prosodic cues and quotative uses of the verb *decir* 'say' (see Posio and Pešková 2020). Although Borreguero Zuloaga (2017: 81) does not systematically distinguish between reported speech (RS) and reported thought (RT), she remarks that *bueno* and other discourse markers with quotative functions may introduce the voice of another person as well as the speaker themselves or their thoughts (cf. Benavent Payá 2015 on the quotative uses of *decir* 'say').

---

<sup>3</sup> The first occurrence of *bueno*, produced by the interviewer in the fifth line, can be understood as a mitigation marker, as she seems to start an apology for having asked a potentially face-threatening question.

*Bueno* is not the only quotative marker in Spanish. Typically, reported speech is introduced using verbal markers such as *digo* 'I say' and nominal markers such as the *y* 'and' + PRO(NOUN) construction (e.g., *y yo* 'and I') (cf. Posio and Pešková 2020). Although *bueno* often co-occurs with these other quotative markers, this is not always the case (Borreguero Zuloaga 2017). As we will show in Section 5, the quotative function of *bueno* presents different degrees of detachment from its other discourse functions, and that the quotative function emerges in those dialects displaying the most extensive use of *bueno* for other functions.

### 3. Data

Our data come from the PRESEEA corpus of spoken Spanish (PRESEEA 2014) and include the subcorpora currently available from nine countries (Chile, Guatemala, Colombia, Cuba, Mexico, Peru, Spain, Uruguay, and Venezuela). Table 1 lists the number of interviews, total number of words (including tags and metadata), and places and years of recordings by country.

**Table 1. Summary statistics of the PRESEEA corpus**

Country	<i>n</i> of interviews	<i>n</i> of words	Place of recording	Date of recording
Chile	18	235,069	Santiago de Chile	2007-2009
Guatemala	18	152,611	Guatemala	2003-2005
Colombia	71	615,407	Barranquilla, Cali, Medellín, Pereira	2001-2019
Cuba	17	161,535	La Habana	2010-2011
Mexico	70	919,240	Guadalajara, Mexicali, México D.F., Monterrey	2001-2018
Peru	18	177,004	Lima	2009-2010
Spain	88	916,331	Alcalá de Henares, Granada, Madrid, Málaga, Valencia	1988-2011
Uruguay	18	168,424	Montevideo	2007-2010
Venezuela	18	218,462	Caracas	2004-2008
<b>Total</b>	<b>336</b>	<b>3,564,083</b>		

As already mentioned in Section 2, the type of discourse represented by these data is semi-structured sociolinguistic interview. Consequently, each conversation has two participants with highly asymmetric discourse roles, as the interviewer is expected to ask questions and the informant is expected to answer them. Discourse phenomena like agreement and disagreement, and to a certain extent also turn management, are thus underrepresented in our data. However, the interviews offer plenty of examples of narratives where the informants tell the interviewers about their experiences, leading to occurrences of reported discourse (RD).

For our quantitative analysis, we extracted all instances of *bueno* ( $n = 12,625$ ), as well as all instances of RD ( $n = 12,143$ ) from the data. The extraction of RD was based on the

coding with the tag *cita* 'quote' that was included in the corpus by the transcribers. Although the description of the corpus does not specify whether the code *cita* is applied to both RS and RT, it is evident from the examples that both types are included, and we will therefore use the hyperonym RD for all sequences marked as *cita*. This procedure thus allowed us to not only analyze the usage frequency of *bueno* in the interviews, but also the number of utterances that represent RD. Once these two data sets were extracted, all subsequent data annotation was carried out manually, starting with the separation of the discourse marker uses ( $n = 11,343$ ) and the adjectival<sup>4</sup> uses ( $n = 1,182$ ) of the word.

In the following quantitative analysis (Section 4), we analyze the contextual variation in the use of *bueno* on the level of the individual interviews. For instance, we first measure the normalized usage frequency of *bueno* in each interview and then group these results by country. The advantage of this analytical procedure is that it allows for capturing idiosyncratic variation displayed in each interview, which in turn reflects the speakers' individual usage patterns and preferences. We are thus able to gauge with much higher precision the range of variation when considering cross-dialectal differences in the usage of *bueno* than if we were relying on normalized frequencies on country level alone.

#### 4. Cross-dialectal differences in the use of *bueno*

This section presents a quantitative analysis of cross-dialectal differences in the use of *bueno* in our data. First, we analyze overall cross-dialectal differences in the usage frequency and productivity of *bueno* (4.1). In Section 4.2, we test the hypothesis that quotative uses emerge most often in those dialects with highest overall frequency of *bueno* marker by inspection of the correlation between the usage frequencies of *bueno* in RD contexts and its overall usage frequencies. In Section 4.3, we test the hypothesis that this change leads to a lower usage of quotative markers such as *digo* 'I say'. Sections 4.2 and 4.3 employ both descriptive statistical methods and inferential statistical methods from variationist sociolinguistics, in particular linear mixed-effects regression models, in order to test for statistical significance. The results from these analyses are discussed in Section 4.4.

The analyses presented in this section are based on two crucial premises. First, we assume that cross-linguistic and cross-dialectal differences can be used as synchronic evidence for the existence of historical processes of linguistic change, common in typological approaches to the study of language (see, e.g., Croft 2016). The simplest case of such an approach is to analyze the use of an element in various languages or dialects in terms of distributional features. When significant differences regarding the use of that element in the distinct languages or dialects are found, a conceptual space (Croft 2016: 591–592) is posited for this range of functions. From this conceptual space, a historical change in terms of the relative succession of these functions can be inferred. To give an example from Romance, Carlier, De Mulder and Lamiroy (2012: 288–290) order Modern Spanish, Italian and French in terms of the degree of grammaticalization of verbal mood and, specifically, the productivity of the subjunctive mood. Such an approach can thus be used to infer a historical change when there is no access to real-time historical data. This is particularly useful for the description of linguistic change in elements such as discourse markers because discourse markers typically grammaticalize in interaction (cf. Heine 2013) and historical data usually lacks the richness required in order to conduct longitudinal diachronic analyses that model changes in interactional routines (Rosemeyer 2019a). As was shown in Rosemeyer (2019b), even when historical texts approximate spoken interaction in narrated discourse, this representation of spoken interaction is influenced by textual models imposed by genre. This typically leads to biases in terms of the represented discourse functions. In addition, the

---

<sup>4</sup> Note that *bueno* is the masculine singular form of the adjective: thus, the frequency of the adjective *bueno* in our data does not reflect the overall frequency of the lexeme, but only the frequency of this particular form.

present case is even more complex, as a diachronic analysis of the usage of *bueno* in historical written texts would have to trace the development of reported speech markers within narrated discourse, in itself reported speech. In the light of these methodological obstacles, it seems to us that the hypothesis of an emergence of quotative functions of *bueno* is extremely difficult to prove in a longitudinal analysis.

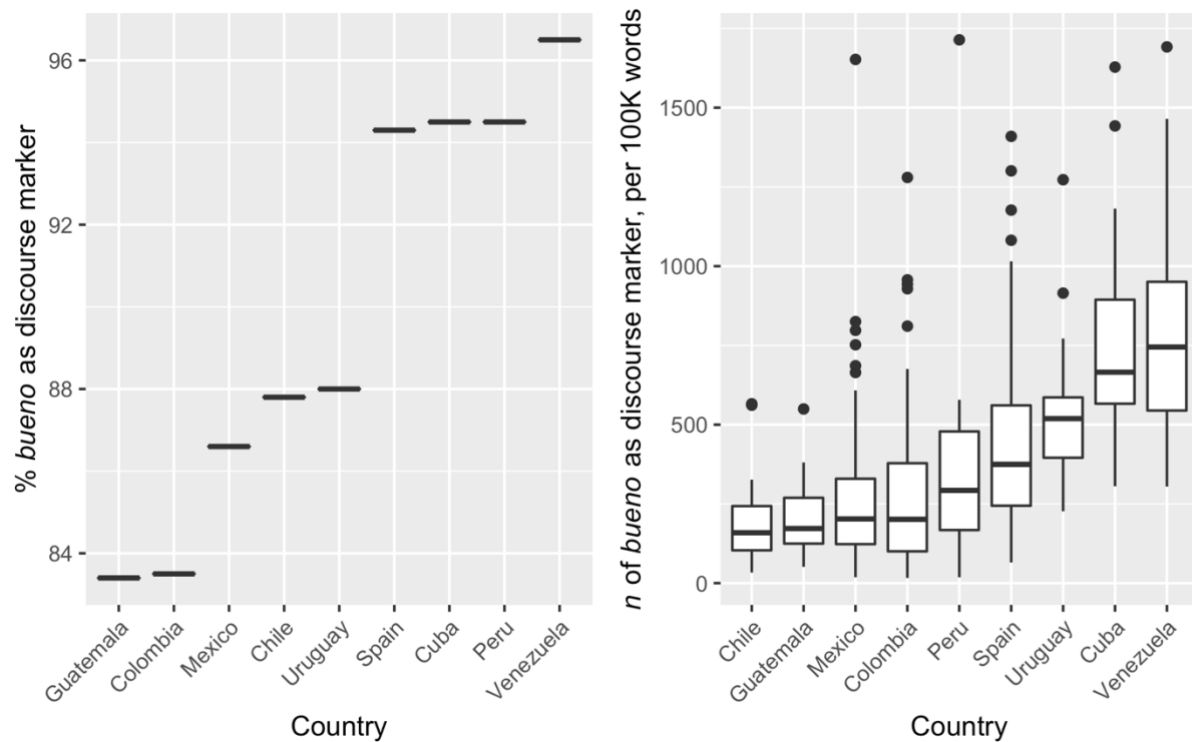
Second, the analyses in this section employ the usage frequency of *bueno* as a crucial predictor of the assumed historical change towards polyphonic functions. This premise rests on a solid basis of historical evidence for a correlation between changes in usage frequency and grammatical change. Grammaticalization processes are usually accompanied by an increase in the usage frequency of the grammaticalizing element, reflecting the acquisition of new functions of the element. In particular, the “repetition of forms may lead to their ‘liberation’, or ‘emancipation’ (Haiman 1994), from their earlier discourse contexts and to increased freedom to associate with a wider variety of other forms” (Hopper and Traugott 2003: 127; cf. also Bybee and Torres Cacoullós 2009). For instance, Hopper and Traugott (2003: 194–196) explain the grammaticalization of Akkadian *enma*, a marker possibly derived from a verb of saying, towards a quotative discourse marker as the result of the frequent and conventionalized co-occurrence of *enma* with reported speech. A similar process appears to have taken place for the verb *decir* ‘to say’ in Peninsular Spanish (cf. Posio and Pešková 2020).

#### 4.1 Usage frequency and productivity of *bueno* as a discourse marker

Figure 1 visualizes the usage frequency of *bueno* as a discourse marker ( $n = 11,343$ ), both relative to the usage frequency of *bueno* in general (= including discourse both discourse marker uses and adjectival uses;  $n = 1,182$ , left plot) and in absolute usage frequency (right plot), by country.<sup>5</sup> We take the relative frequency of *bueno* as a discourse marker to be indicative of its degree of grammaticalization and consequently, its productivity: the higher the usage frequency of *bueno* as a discourse marker relative to adverbial uses in a specific dialect or idiolect, the more likelier speakers are to assign the new grammatical meaning to that element (see, e.g., Hay 2001). While the results differ slightly in terms of the relative order of the dialects, they suggest that the discourse marker use of *bueno* is most productive in (1) Venezuelan, Peruvian, Cuban, Uruguayan and European Spanish and least productive in (2) Guatemalan, Colombian, Mexican, and Chilean Spanish.

---

<sup>5</sup> All plots were created in R (R Core Development Team 2021), using the ggplot2 package (Wickham 2016).



**Figure 1. Usage frequency of *bueno* as a discourse marker relative to overall usage frequency of *bueno* (including adverbial uses), by country**

**Left plot: Mean percentage of the usage of *bueno* as a discourse marker by country**

**Right plot: Normalized usage frequency of *bueno* as a discourse marker, grouped by interview and country**

Having assessed the usage frequency of *bueno* as a discourse marker relative to the usage frequency of *bueno* as an adjective in Figure 1, we continued to eliminate the adjectival uses of *bueno* from the data, leaving us with a final dataset of  $n = 11,443$  occurrences of *bueno* as a discourse marker. In the remainder of this paper, we will use the term *bueno* to refer exclusively to the discourse marker uses.

As was discussed at the beginning of Section 4, we assume that a greater productivity is mirrored in the freedom to be combined with a wider variety of other forms. This effect is well-known in grammaticalization studies, where it was observed that grammaticalization leads to an increase in the type frequency of grammaticalizing elements. For instance, Bybee (2003) analyzes the grammaticalization of the Old English verb *cunnan* ‘to know’, leading to Modern English *can*. She finds that initial modal uses of *cunnan* were restricted to a limited set of complement verbs (most importantly, communication verbs). The modal meaning was then gradually transferred to uses of *cunnan* with other complement verbs by virtue of an actualization process. Consequently, the high grammatical productivity of Modern English *can* can be measured in terms of its high type frequency, i.e. its freedom to combine with virtually all types of complement verbs.

In this paper, we assume that the productivity of discourse markers such as *bueno* can be measured in very much the same way. *Bueno* frequently co-occurs with prefaces such as *ah* ‘ah’. Prefaces are formally defined in Conversation Analysis as elements that occur in turn-initial position and that are “followed by additional elements within the same intonation

contour” (Heritage 1998: 292), meaning that no prosodic break between the preface and *bueno* may occur. Table 2 provides an inventory of the prefaces preceding *bueno* in our data.<sup>6</sup>

**Table 2. Preface types occurring before *bueno* in the PRESEEA data, sorted by usage frequency**

Preface	Translation	Frequency
<i>y</i>	‘and’	1,129
<i>ah</i>	‘ah’	339
<i>ee(h)(m)</i>	‘eh’	210
<i>pues/pos</i>	‘then/well’	182
<i>entonces</i>	‘then’	100
<i>porque</i>	‘because’	90
<i>este</i>	‘eh’	80
<i>mm</i>	‘mm’	78
<i>ya</i>	‘okay’	40
<i>aha/aja</i>	‘aha’	16
<i>ay</i>	‘oh’	14
<i>así</i>	‘so’	12
<i>aunque</i>	‘although’	9
<i>ahor(it)a</i>	‘now’	8
<i>además</i>	‘also’	7
<i>hm</i>	‘hm’	7
<i>uhm</i>	‘uhm’	7
<i>hh</i>	‘hh’	5
<i>okay</i>	‘okay’	4
<i>oh/oy</i>	‘oh’	2
<i>uy</i>	‘uh’	1
<i>bah</i>	‘bah’	1

These preface-*bueno* combinations have specific interactional functions, which cannot be described in detail here. As a case in point, however, consider example (6), in which the informant, who is a schoolteacher, has just told the interviewer that the kids call her by her first name (anonymized as X in the transcript). In the first line of the transcript, E asks I for confirmation. I repeats her assertion in the next line and gives an example from everyday life in order to give greater credibility to her claim. In the last line, E uses the assertion *sí* ‘yes’ to signal that she believes I now. E then goes on to introduce a new topic, namely the question whether they should finish the interview. This topic change is signaled by the sequence *y bueno*. Arguably, the discourse function of *y bueno* in this context cannot be derived in terms of its compositional meaning (‘and’ + ‘well’); rather, it appears that the hedged agreement expressed by *bueno* serves to attenuate the face-threatening potential of the topic transition (see Pons 2003), which is why *y bueno* seems to have conventionalized the function of topic transition.

<sup>6</sup> Note that the term “preface” in Conversation Analysis is defined in terms of its function in discourse (indeed, just like the term “discourse marker”). As a result, the prefaces listed in Table 2 belong to very different morphosyntactic categories (conjunctions, interjections, adverbs etc.). As correctly noted by one of the anonymous reviewers, a previous classification of these prefaces could offer a more accurate picture of the combination possibilities of *bueno*. Since this approach transcends the scope of the present paper, we leave this analysis to future research.

(6) E: *te llaman X / ni señorita ni profesora*

I: *nada no / X / el otro día precisamente ayer o antesdeayer encontré a uno pero muy mayor... que ya está en carrera y todo "hola X"*

E: *sí... y BUENO // pue:::s yo creo que más o menos / no sé ya ha terminado*

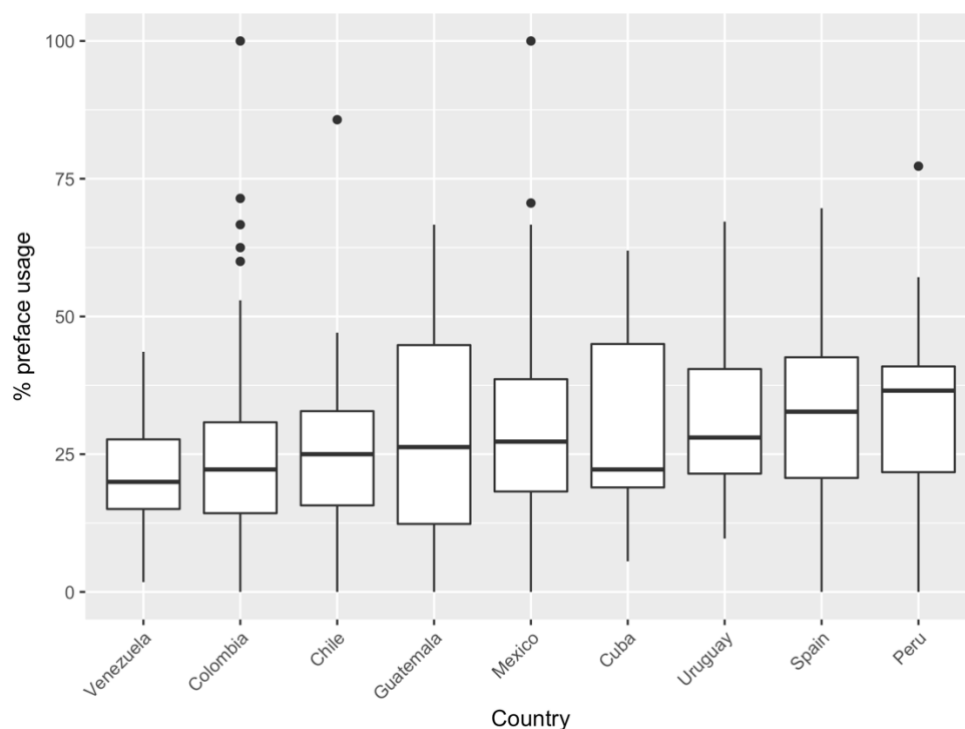
E: 'everybody calls you X / neither madam nor professor'

I: 'no no / [only] X / the other day, yesterday or the day before yesterday, to be precise, I met one who was already old... he was already on the street and [says] "Hi X"'

E: 'yes, and BUENO // so... I think that was it more or less / it's over, I think'

(ALCA\_M23\_10)

Crucially for our purposes, it seems plausible to assume that the creation of conventionalized preface-*bueno* sequences is only possible in a situation in which *bueno* is already a grammaticalized discourse marker. Consequently, we would expect *bueno* to occur more frequently with prefaces in those dialects where it has a high usage frequency. The results from Figure 2, which visualizes the mean percentage of preface usage before *bueno* by interview and country, demonstrates that this prediction is mostly borne out by our data. In particular, four of the five dialects in which the use of *bueno* was shown to be most productive also show the highest rates of mean preface usage (Peru, Spain, Uruguay, Cuba). However, Venezuela is a notable exception, as mean preface usage is lowest in this dialect despite the high frequency of *bueno*. Despite this exception, the positive correlation between the usage frequency of *bueno* and mean rate of preface usage reaches high statistical significance.<sup>7</sup>



**Figure 2. Mean percentage of preface usage before *bueno*, by interview and country**

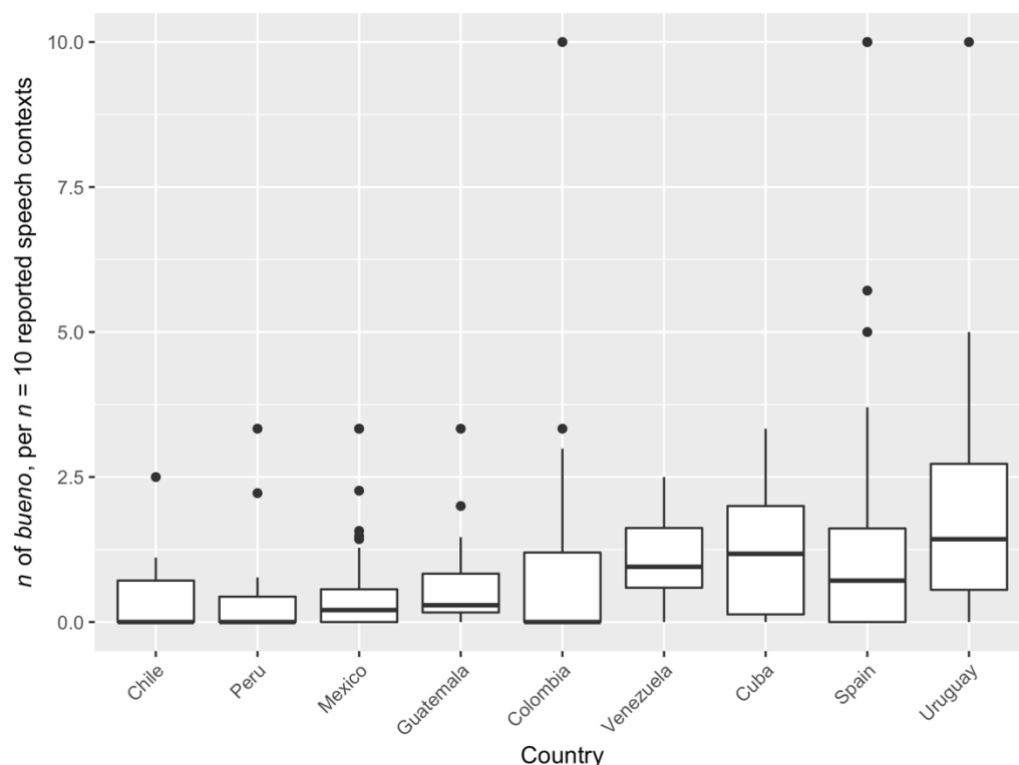
<sup>7</sup> Since the data violated the assumption of normality, a U-test was computed (Gries 2009: 209-210). The U-test showed that the positive correlation between the usage frequency of *bueno* and mean rate of preface usage is significant ( $W = 2770.5$ ,  $p_{\text{two-tailed}} < .001$ ). Here and in the remaining statistical analyses, we assumed a significance threshold of  $p < .05$ .

#### 4.2 Usage frequency of *bueno* in reported discourse (RD)

Having established the overall distinction between dialects with more vs. less productive use of *bueno* as a discourse marker, we can now examine the hypothesis that the quotative uses emerge most often in those dialects with highest overall frequency of *bueno* as a discourse particle. We predict that the more frequently *bueno* is used in a specific interview and a specific dialect, the more likely it is to be found in the beginning of a stretch of RD. Such a correlation might be indicative of a grammaticalization of *bueno* towards a quotative marker.

We calculated the usage frequency of *bueno* at the beginning of RD contexts (i.e., as the first element of the stretch of RD, eventually preceded only by one of the ‘prefaces’ listed in Section 4.1) relative to the overall number of RD contexts, which, as described in Section 3, was obtained by extracting all instances of the *cita* ‘quote’ tags from the data. In the process of this analysis, we realized that the tag *cita* had not been used in 62 interviews that nevertheless contained reported speech. Consequently, we eliminated these files from our analysis, leading to a total of  $n = 274$  interviews.

Figure 3 visualizes the frequency of *bueno* at the beginning of RD relative to all occurrences of RD, by interview and country. It demonstrates that, in line with our expectations, *bueno* is relatively more frequent at the beginning of RD contexts in those dialects that were identified as displaying a high productivity of *bueno* (i.e., Uruguay, Spain, Cuba, and Venezuela).



**Figure 3. Frequency of *bueno* at the beginning of RD relative to all occurrences of reported speech, by interview and country**

In order to assess the statistical significance of this finding, we calculated a linear mixed-effects regression model predicting the usage frequency of *bueno* in RD relative to all occurrences of RD from the independent variables summarized in Table 3. Most importantly,

we included the usage frequency of *bueno* per file as a predictor to test the hypothesis that the speakers who show a particularly productive use of *bueno* are more likely to use *bueno* at the beginning of RD, as well. We also included a number of sociolinguistic variables in order to control for social characteristics of the speakers. Testing such sociolinguistic predictors is a standard approach in variationist sociolinguistics (see, e.g. Tagliamonte 2012) and did not reflect any hypotheses on our part. As random effects, we included the variables COUNTRY (the dialect in question), and YEAR (the year of recording of the interview), thus also controlling for this variation. The model was calculated in R (R Core Development Team 2021) using the lmer (Bates et al. 2015) and lmerTest (Kuznetsova, Brockhoff, and Rune 2017) packages.

**Table 3. Operationalization of predictor variables for the linear mixed-effects regression model predicting the usage frequency of *bueno* in RD relative to all occurrences of RD**

Variable	Description
BUENO_FREQUENCY	Usage frequency of <i>bueno</i> per 100,000 words in the file, z-standardized
PERCENT_SEX_FEMININE	Percent of speakers of feminine sex per file, z-standardized
MEAN_AGE	Mean age of speakers per file, z-standardized
PERCENT_EDUCATION_HIGH	Percent of speakers with university education per file, z-standardized
PERCENT_SPEAKER_INTERVIEWER	Percent of turns attributed to the interviewer per file, z-standardized

Table 4 summarizes the results from the regression model. As predicted, the model found a statistically significant positive correlation between (a) the usage frequency of *bueno* in RD relative to all occurrences of RD and (b) the overall, i.e. RD-independent, usage frequency of *bueno* in the interviews. While none of the social variables were shown to correlate significantly with the usage frequency of *bueno* in RD relative to all occurrences of RD, the model found marginally statistically significant effects for the variables PERCENT\_SEX\_FEMININE and PERCENT\_EDUCATION\_HIGH.

**Table 4. Results from the linear mixed-effects regression model predicting the usage frequency of *bueno* in reported discourse relative to all occurrences of reported discourse in  $n = 274$  sociolinguistic interviews from the PRESEEA corpus**

Variable	Log odds	Odds ratio	Standard error	z value	p value
(Intercept)	0.068	1.070	0.102	0.662	0.508
BUENO_FREQUENCY	0.197	1.217	0.066	2.964	0.003
PERCENT_SEX_FEMININE	-0.116	0.891	0.060	-1.94	0.052
MEAN_AGE	0.013	1.014	0.063	0.214	0.830
PERCENT_EDUCATION_HIGH	0.114	1.120	0.065	1.756	0.079
PERCENT_SPEAKER_INTERVIEWER	-0.065	0.937	0.068	-0.955	0.340

#### 4.3 Competition with quotative markers

Although the results from Section 4.2 correspond to expectations, it might be argued that the positive correlation between the probability of usage of *bueno* at the beginning of RD

contexts and the overall usage frequency of *bueno* is a simple frequency effect, i.e., the higher probability of usage of *bueno* at the beginning of RD contexts for interviews in which *bueno* is used particularly productively is not higher **than would be expected** on the basis of the simple fact that *bueno* is generally more productive in those dialects. Consequently, in order to demonstrate that high-frequency *bueno* is indeed more likely to develop into a quotative marker than low-frequency *bueno*, additional analyses are necessary.

One further possibility to assess the relevance of the grammaticalization hypothesis is to analyze to what extent *bueno* competes with other quotative expressions such as the verb *decir* 'to say'. In Borreguero Zuloaga's (2017) view, the existence of instances of RD not introduced by *decir* or other quotative markers, but rather *bueno* alone, would be an indicator that *bueno* has undergone the grammaticalization process towards expression of polyphonic functions. This proposal leads to the prediction that dialects in which the use of *bueno* is particularly productive should demonstrate lower rates of overt quotative expressions before *bueno*.

Consequently, we restricted our dataset to all tokens of *bueno* in RD ( $n = 905$ ) and established a new variable QUOTATIVE, which described the type of immediately preceding quotative marker. Table 5 describes the operationalization of this variable.

**Table 5. Operationalization of the variable QUOTATIVE describing the type of quotative marker preceding *bueno* in RD**

Level	Description
say	immediately preceding <i>verbum dicendi</i> such as <i>digo</i> 'I say' or <i>dije</i> 'I said'
think	immediately preceding <i>verbum cogitandi</i> such as <i>pienso</i> 'I think'
noun	immediately preceding pronoun or noun referring to the speaker of the RD, such as <i>yo</i> 'I', <i>mi tía</i> 'my aunt'
adverbial	immediately preceding adverbial expression such as <i>total</i> 'thus', <i>así como</i> 'so how', <i>entonces</i> 'then'
none	no immediately preceding quotative marker

It is crucial to note that the coding of QUOTATIVE only refers to the immediately preceding quotative marker. In fact, we did not find a single clear-cut case in which *bueno* is used at the **beginning** of a stretch of RD without being preceded by a quotative marker in the extended context. Rather, tokens annotated as QUOTATIVE = 'none' are typically structured as the example in (7), where we have annotated the turns in the stretch of RD by indentation:

- (7) *y me dijo / no pues es que no piense [...] se pone a trabajar y todo / y ya viene y ya se ha olvidado de la novia*  
*/ le dije no acá es que / yo estaba ya muy enamorado /*  
*yo dije no / no acá no / no puedo hacer eso / yo la quiero mucho a ella /*  
*¡ah! / bueno haga lo que a usted le provoque*

'and he **said** to me "no well you don't think [...] you will start working and everything / and then you come and you'll have forgotten about the girlfriend"  
*/ I said* to him "no dad it's like" / I was very much in love /  
**I said** "no / no dad no / I can't do that / I love her very much" /  
**"ah! / BUENO** you do whatever you want" (MEDE\_H31\_002)

We find that in the first three stretches of RD in (5), the verb *decir* 'to say' is used as a quotative. In the last reported turn, attributed to the speaker's interlocutor, *bueno* is used

without a preceding quotative marker. However, the characterization of the turn as RD is clearly projected by the use of the previous quotative markers, and possibly also by specific intonational features (which are not reflected in the corpus).

A very similar discourse configuration is typical of the usage of (pro)nominal quotative markers, as evinced in the last line of example (8). The use of a communication verb *pregunté* 'I asked' activates the quotative frame in which *y yo* 'and I' receives a quotative interpretation. Note that the stretch of RD consists solely of *ah bueno* which might be interpreted here as an expression of disbelief (cf. English *and I was like "okay"*). This example also illustrates a case where it is difficult to say whether the RD is intended to represent reported speech or reported thought.

- (8) *entonces le pregunté "señor no me ha dicho nada de B"*  
"no: no a él le ha ido bien ta ta / ni desordenado ni nada / bien bien que le ha ido"  
*y yo "ah bueno"*  
'then I asked him "sir you haven't said to me anything about B"  
"no, no he's been good, ok? / not disoriented or anything / all good, all good he's been"  
**and I "ah BUENO"**  
(BARR\_H23\_015)

These considerations give rise to an alternative hypothesis regarding the use of quotative expressions. In particular, examples such as (7) and (8) might not be indicative of a change of *bueno* (or *y yo*, for that matter) towards the expression of polyphony, but rather a process by which the scope of quotative verbs has extended, such that these quotative verbs can project over longer stretches of discourse through the activation of a **quotative frame** in which the subsequent turns are also interpreted as RD.

Indeed, in line with studies on the grammaticalization of quotative markers (see the beginning of Section 4), the assumed change of *bueno* towards the expression of polyphony would lead us to expect not a decrease in the overall usage frequency of quotative markers such as *digo* 'I say', but rather the inverse effect. Thus, we would predict an increased strength of collocation between such quotative markers and *bueno*, in the sense that the usage of *bueno* becomes more frequent after quotative markers than would be expected on the basis of the usage frequency of *bueno* alone. In line with usage-based approaches to the description of language, such a routinization process should be mirrored in an increased, i.e., conventionalized, syntagmatic association between the quotative marker and *bueno*, leading to the creation of a prefab (see, e.g., Bybee 2006: 713).<sup>8</sup> In particular, we would predict a decrease in the frequency of prosodic boundaries between the quotative marker and *bueno*. Crucially, the routinization of the association between quotative markers and *bueno* would result in an increased predictability of *bueno* from quotative marker usage (i.e., upon hearing a quotative marker, the hearer will be more likely to predict the use of *bueno* as the next word uttered by the speaker). This higher predictability of *bueno* in sequences marked as RD might lead to a higher-based reanalysis (see Detges and Waltereit 2002; Eckardt 2009; Rosemeyer and Grossman 2017), such that *bueno* comes to be associated with RD irrespectively of the

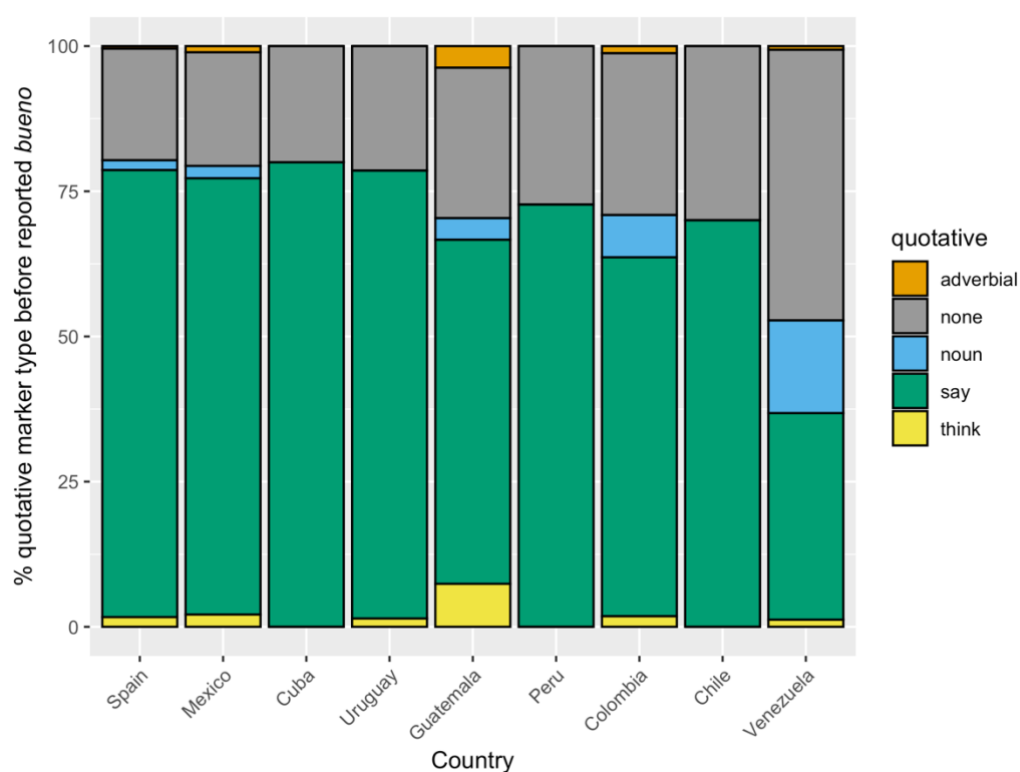
---

<sup>8</sup> Bybee (2010: 35) defines prefabs (=prefabricated expression) as "any conventionalized multi-word expression" that results from the historical process of "chunking". A prefab such as Engl. *pull strings* differs from its compositional counterpart *pull strings* in that its form is immutable (e.g., saying I pulled a string does not lead to the meaning 'to secretly use the influence you have over important people in order to get something or to help someone' (<https://dictionary.cambridge.org/de/worterbuch/englisch/pull-strings>, accessed 7 July 2021), suggesting a holistic representation of the prefab in the language user's mind. Likewise, the meaning of the prefab is stored holistically, and is consequently not established compositionally (i.e., by combining the meaning of *pull* and the meaning of *strings*), due to a historical conventionalization process.

usage of a previous quotative marker. Only such a change would enable an eventual omission of other quotative markers with *bueno*.

Essentially, the scenario sketched above is not so much a contradiction to the negative correlation between quotative marker usage and usage frequency of *bueno* in RD, but rather a more detailed description of the change that might lead to the existence of such a negative correlation. In other words, if this latter scenario were found to be consistent with the data, it would suggest that although there is no conclusive evidence for the assumption that *bueno* is starting to replace quotative markers, such a change might well be incipient.

Figure 4 visualizes the mean percentage of quotative marker usage before *bueno*, by country, in RD ( $n = 905$ ). It demonstrates that omission of quotative markers before *bueno* is less likely in those dialects in which the usage of *bueno* is particularly frequent and productive (especially Spain, Cuba, and Mexico).



**Figure 4. Mean percentage of quotative marker usage before *bueno*, by country**

However, this generalization appears to be rather weak, given that not only the Venezuelan, but also the Mexican and, to a smaller extent, Peruvian data do not seem to conform to the interpretation. In order to tackle this question, we again calculated a linear mixed-effects regression model to confirm the significance of these findings. This regression model predicts the mean percentage of quotative marker usage in each interview from the predictors already used in the previous regression model, summarized again in Table 6. It thus tests the hypothesis that speakers who show a particularly productive use of *bueno* are more likely to introduce *bueno* with a quotative marker. In keeping with standard sociolinguistic practices, we again tested for various social variables (sex, age, education, role in the conversation).

**Table 6. Operationalization of predictor variables for the linear mixed-effects regression model predicting percentage of quotative marker usage in RD introduced by *bueno* in the PRESEEA corpus**

Variable	Description
BUENO_REPORTED_FREQUENCY	Log-transformed usage frequency of reported <i>bueno</i> per file, normalized by words per file, z-standardized
PERCENT_SEX_FEMININE	Percent of speakers of feminine sex per file, z-standardized
MEAN_AGE	Mean age of speakers per file, z-standardized
PERCENT_EDUCATION_HIGH	Percent of speakers with university education per file, z-standardized
PERCENT_SPEAKER_INTERVIEWER	Percent of turns attributed to the interviewer per file, z-standardized

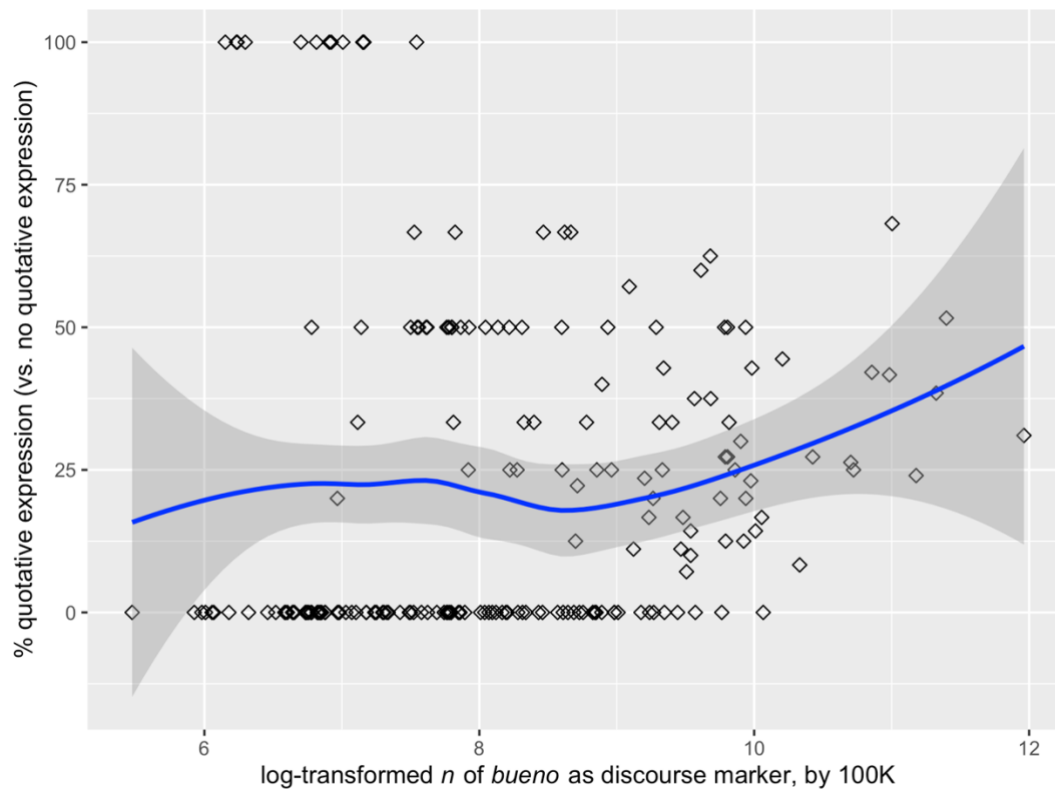
As random effects, we again included the variables COUNTRY (the dialect in question), and YEAR (the year of recording of the interview), thus controlling for this variation. Crucially, we also included a third random effect, namely REPORTED\_FREQUENCY, which represents the mean normalized and z-standardized usage frequency of all occurrences of RD (irrespective of whether *bueno* is used in these stretches of reported speech) within a file. This random effect was nested within the random effects for COUNTRY and YEAR, in order to rule out the possibility that a higher frequency of quotative expressions is simply due more frequent RD and/or differences in the rigor of coding for RD in the different interviews.<sup>9</sup> Table 7 reports the results from the regression analysis.

**Table 7. Results from the linear mixed-effects regression model predicting the percentage of quotative marker usage in RD introduced by *bueno* in the PRESEEA corpus**

Variable	Log odds	Odds ratio	Standard error	z value	p value
(Intercept)	0.129	1.137	0.142	0.907	0.365
BUENO_REPORTED_FREQUENCY	0.095	1.100	0.032	2.938	0.003
PERCENT_SEX_FEMININE	-0.089	0.915	0.066	-1.358	0.174
MEAN_AGE	0.118	1.125	0.066	1.796	0.073
PERCENT_EDUCATION_HIGH	0.016	1.016	0.064	0.243	0.808
PERCENT_SPEAKER_INTERVIEWER	0.075	1.078	0.077	0.98	0.327

Table 7 demonstrates a significant positive correlation between the usage frequency of *bueno* in RD and the mean percentage of quotative marker use. This effect is visualized in Figure 5, demonstrating that especially quotative marker usage before *bueno* is more likely for interviews that display a particularly high usage frequency of reported *bueno*.

<sup>9</sup> See Gries (2015) for a description of the relevance of using nested random effect structures and instructions for their implementation.



**Figure 5. Correlation between the usage frequency of *bueno* in RD contexts and the mean percentage of quotative marker use**

These results contradict Borreguero Zuloaga's (2017) assumption of a negative correlation between quotative marker usage and usage frequency of *bueno* in RD. Rather, we find the inverse effect, in that probability of quotative marker usage before *bueno* in RD increases the more *bueno* is used in RD.

However, our finding is consistent with the alternative scenario proposed above, namely, an increased routinized association between quotative markers and the usage of *bueno* in RD, which is likewise indicative of the historical change of *bueno* towards the expression of polyphony. As was suggested above, such a scenario can be tested by exploring the degree to which quotative markers and *bueno* have become a prefab. In order to do so, we analyzed the correlation between the frequency with which speakers insert a prosodic break between immediately preceding quotative markers and reported *bueno* in our data. Given that quotatives based on *verbi dicendi* such as *decir* 'to say' constitute by far the most frequent group of quotative expressions in our data ( $n = 595$  of 905, see also Section 5 below), we restricted our analysis to contexts in which reported *bueno* is immediately preceded by such a speech act verb.

The PRESEEA distinguishes between three types of prosodic breaks, i.e. none (9a), weak (9b) and strong (9c) prosodic breaks. A weak prosodic break is conceptualized as a short pause, a strong prosodic break as a longer pause.<sup>10</sup>

- (9) a. *dice bueno pues ahora lo apagaré*  
 '(s/he) says "BUENO well now I'll turn it off"' (GRAN\_M23\_010)

<sup>10</sup> See [https://preseea.linguas.net/Portals/0/Metodologia/Marcas\\_etiquetas\\_minimas\\_obligatorias\\_1\\_2.pdf](https://preseea.linguas.net/Portals/0/Metodologia/Marcas_etiquetas_minimas_obligatorias_1_2.pdf) (p. 7, accessed 11 February 2021).

b. *entonces dice: / bueno mijo pero no vayas*

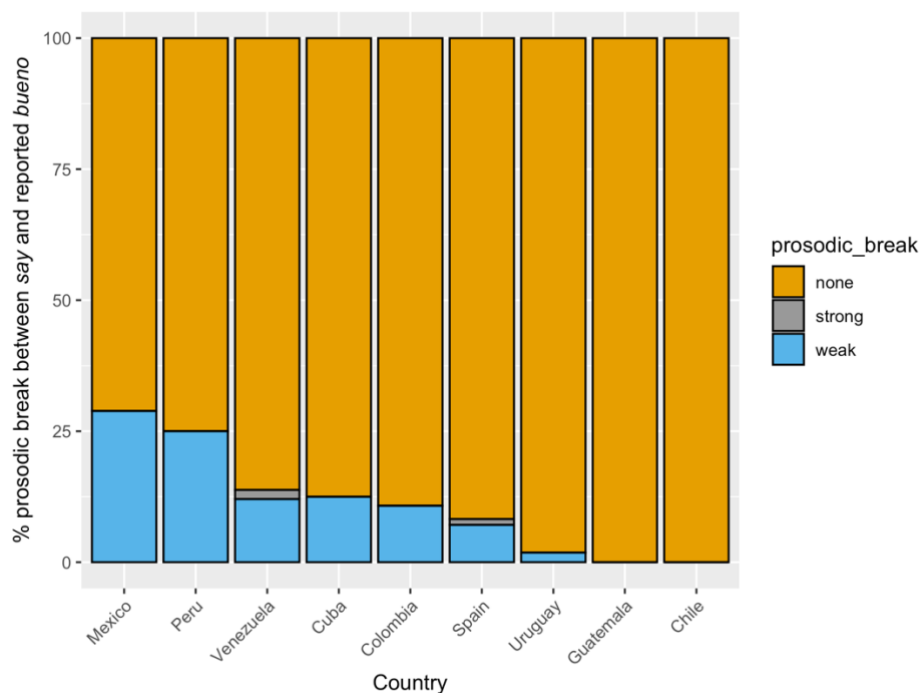
‘then (s/he) says / “BUENO kid but don’t go there”’ (BARR\_H23\_015)

c. *entonces mi esposa me dice // bueno no se te olvide que nosotros vamos a jugar [...]*

‘then my wife says to me // “BUENO don’t forget that we are going to play...”’

(CARA\_H21\_013)

Figure 6 visualizes the mean percentage of prosodic breaks between quotative ‘say’ verbs and *bueno*, by country. It seems to suggest that indeed, those dialects in which the usage of *bueno* is particularly frequent and productive display fewer prosodic breaks between the quotative ‘say’ verbs and *bueno*, in line with the assumption of a higher degree of conventional association between the two expression types.



**Figure 6. Mean percentage of prosodic breaks between quotative ‘say’ verbs and *bueno*, by country**

However, once again, several dialects in Figure 6 do not seem to conform to this proposed pattern. Contrary to expectations, Mexican Spanish shows the highest mean frequency of prosodic breaks, whereas Chile and Guatemala have the lowest mean frequency of prosodic breaks. This might be due to relative differences in sample size, as indicated in Table 8. If we consider only those dialects with a minimum usage frequency of reported *bueno* following a quotative ‘say’ expression of  $n = 50$ , a much clearer picture emerges, where Uruguay and Spain show low rates of prosodic breaks between *bueno* and the ‘say’ quotative, whereas Mexico, Venezuela and Colombia show relatively high rates of prosodic breaks.

**Table 8. Usage frequency of reported *bueno* following a quotative ‘say’ expression**

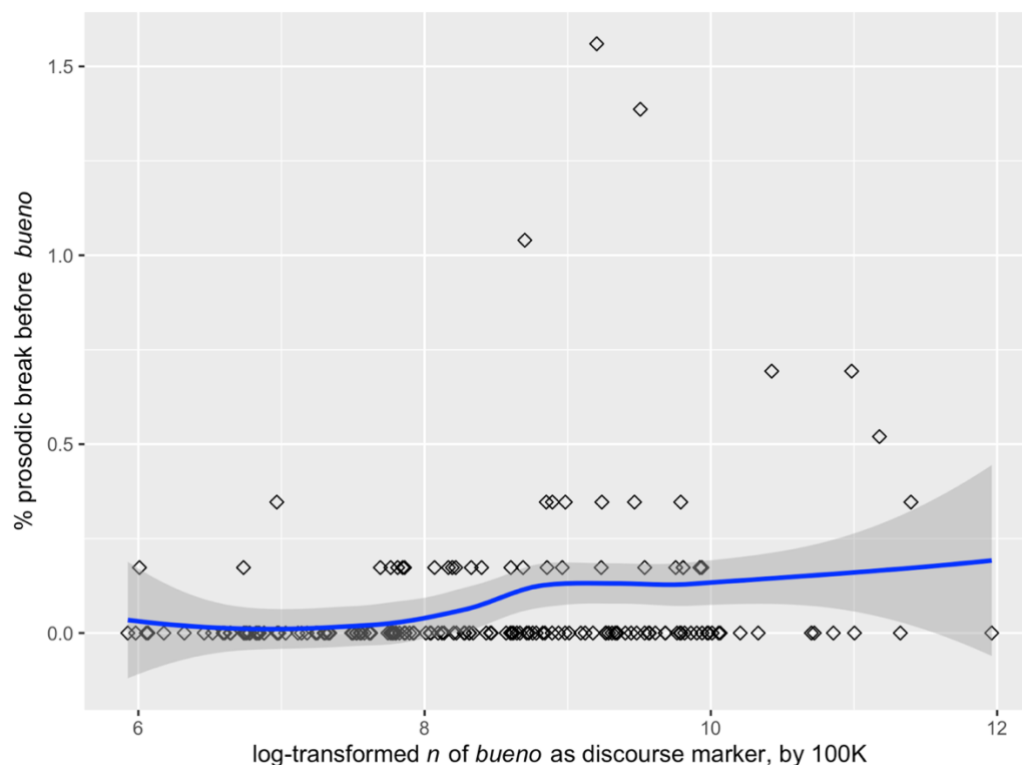
Spain	Mexico	Colombia	Venezuela	Uruguay	Cuba	Guatemala	Peru	Chile
182	142	102	58	54	16	16	16	7

This interpretation is backed by a simple linear mixed-effects regression model predicting the mean percentage of prosodic breaks between *bueno* and quotative ‘say’ in the interviews from the usage frequency of *bueno* in RD introduced by quotative ‘say’, controlling for YEAR and COUNTRY as random effects. The model, summarized in Table 9, calculates a significant positive correlation between the mean percentage of prosodic breaks between *bueno* and quotative ‘say’ and the usage frequency of *bueno* in RD introduced by quotative ‘say’.

**Table 9. Results from the linear mixed-effects regression model predicting the mean percentage of prosodic breaks between *bueno* and quotative ‘say’ in the interviews from the usage frequency of *bueno* in RD introduced by quotative ‘say’**

Variable	Log odds	Odds ratio	Standard error	z value	p value
(Intercept)	-0.350	0.704	0.105	-3.348	0.001
BUENO_REPORTED_FREQUENCY	0.053	1.055	0.013	4.165	0.000

Figure 7 visualizes this positive correlation.



**Figure 7. Correlation between the mean percentage of prosodic breaks between *bueno* and quotative ‘say’, and the usage frequency of reported *bueno* in the interviews**

#### 4.4 Discussion

In summary, our quantitative analysis of the usage of *bueno* in the PRESEEA supports the hypothesis based on Borreguero Zuloaga (2017) that *bueno* is evolving towards a quotative marker used to express polyphony in discourse. In concrete, we have shown that (a) *bueno* is more likely to be used in RD contexts in those interviews that show a particularly productive and frequent usage of *bueno*, indicating that higher productivity of *bueno* in general can lead to a more frequent usage of *bueno* in RD, and (b) that this change seems to be bound to a

routinization of the collocation between *bueno* and preceding quotative markers (especially quotative *decir* 'say'), as mirrored in a decreasing frequency of prosodic breaks between *bueno* and *decir*. Although we did not find evidence for the assumption that *bueno* has started to replace other quotative expressions in those dialects in which it is particularly productive and hence grammaticalized, our findings are consistent with proposals of mechanisms of grammaticalization based on the notion of hearer-based reanalysis. Consequently, our analysis suggests that while *bueno* has not developed into a quotative marker itself even in those dialects in which it is most productive, it may do so in the future.

## 5. *Bueno* in reported speech and thought

In this section, we examine the eventual distinctions between reported speech (RS) and reported thought (RT) in our data, focusing especially on those contexts where *bueno* occurs as part of the stretch of reported discourse (RD).

A simple quantitative examination of the lexical items most frequently used to introduce RD reveals a strong predominance of the verb *decir* 'say', while *pensar* 'think' is marginal in this function. The verb *decir* precedes the sequence of reported discourse in 6,450 of the 12,143 occurrences of the tag *cita* 'quote' in the totality of our data consisting of all occurrences of the tag, i.e., over half of the instances of RD are introduced using this verb. The verb *pensar* occurs only 27 times before RD. This comparison confirms Posio and Pešková's (2020) observation that *decir* has clearly specialized in the quotative function while other communicative verbs are marginal in the quotative function. It also shows that RT is very rarely distinguished from RS by the choice of the verb in the quotative expression. Indeed, speakers recur to the verb *decir* in contexts that are clearly instances of RT, as in example (10). In this example, the speaker does not refer to any actual dialogue but rather describes what he is thinking at the moment or in general (i.e., the quotative is not episodic but generic; see Posio and Pešková (2020) for a discussion of the two types).

- (10) *ahora lo pienso y digo qué bárbaro*  
'now I think about it and I say "how amazing"'  
(GUAD\_H22\_002)

Since the same lexical verb can be used to introduce both RS and RT, let us now turn into a qualitative scrutiny of the data in order to examine whether other usage patterns distinguishing between reported speech and thought emerge, focusing on those dialects where *bueno* is most frequently associated with RD.

The qualitative analysis of these data lends further support to our hypothesis that quotative *bueno* does not necessarily display any of the other discourse functions of *bueno* such as expressing agreement, disagreement or mitigation (see Section 2), but can rather just mark the beginning of the reported sequence. This can be observed in examples (11) and (12) from the Uruguayan data – one of the dialects with most frequent and productive use of *bueno*. We find frequent use of the quotative *bueno* in combination with *decir* 'say' in contexts that clearly pertain to RS, i.e., stretches of dialogue that the speaker presents as episodic, having occurred and being anchored in a point in time. Such a case of RS is found in example (11), where the speaker attributes the words to a third person (medical personnel in a hospital). There is nothing in the context supporting an interpretation of *bueno* (in this particular example) as serving other discourse marker functions in addition to marking the sequence as RS.

- (11) [...] *sentía dolor muy fuerte en el pecho y... ta fui al médico y me dijeron bueno / andá a hacerte una placa / y ahí fue que descubrieron eso / [...]*

‘[...] I felt a very strong pain in my chest and... so I went to the doctor and **they said to me** “**BUENO** / go and let them do an X-ray” / and it was there they noticed that / [...]’  
(MONV\_H11\_035)

The same combination of *decir* and *bueno* also occurs before RT, i.e., stretches of dialogue that the speaker presents as inner speech or something that they or anyone in their position might say or think in a given scenario that are not episodic (anchored to a specific point in time and space) but rather generic or irrealis (see Posio and Pešková 2020). A clear example of reported thought is given in (12), where the interviewer asks if the informant has ever experienced a near-death experience – an essential part of the classic Labovian sociolinguistic interview.

(12) [...] *y ¿te ha pasado alguna vez de tener... / aparte de esta operación ¿no... peligro de muerte? de que vos decís bueno... me me muero o sea...*

‘[...] and has it happened to you some time that you’d have... / in addition to that surgery, like... danger of death? like **you say** “**BUENO**... I, I’m going to die” like...’  
(MONV\_H11\_035)

As can be observed in examples (11) and (12), the Uruguayan data does not present any formal differences in the marking of RS and RT. However, looking at other dialects with frequent and productive *bueno* usage, the Cuban data seems to show some evidence to the effect that speakers may formally distinguish between RT and RS. An analysis of the Cuban interviews suggests that the quotative *bueno* typically occurs in sequences intended to represent what the speaker (or anyone, typically expressed as a generic *tú* ‘you’) may think or say (to themselves) in a given scenario, with no reference to actual dialogue. This is illustrated by example (13) where the speaker describes the old times and introduces what we interpret as RT, with no reference to actual, episodic, dialogue.

(13) [...] *tocaban grupos musicales en la playa / eeh / aquí en el malecón / o sea / que había una diversidad / para tú poder escoger / una salida tuya un fin de semana para desconectar que era inmensa / tú podías darte el lujo de decir “bueno / voy aquí” / los cines / la ciudad estaba llena de cines / de películas de estreno [...]*

‘[...] there were bands playing on the beach / eeh / here on the pier / like / there was diversity / for you to choose / where you want to go out on a weekend to disconnect that was immense / you had the luxury of **saying** “**BUENO** / **I’m going here**” / the movies / the city was full of movie theaters / movie premieres [...]’  
(LHAB\_H22\_049)

Example (14) from the Cuban data illustrates a complex stretch of RD where the quotative verb *decir* is used in the beginning, in combination with *bueno* (lines a, b) while *bueno* alone functions as a quotative marker once the frame of reported discourse has been established (line e). In this example, the speaker is explaining why she decided to quit her job as an informatician. Our interpretation is that the example contains both RT (the speaker saying – to herself – that she must change jobs: lines a, b, e) and RS by the speaker herself and two other persons (lines c, d, f). Interestingly, *bueno* only occurs here in those instances of RD that seem to be construed as RT (lines a, b, e) but not in RS (lines a, b, e). This interpretation is supported by the fact that in the former cases there is no recipient (i.e., the speaker does not direct her words to anyone in particular) while in the latter cases the recipient is marked (*a mi hermana* ‘to my sister’, *me* ‘to me’) which construes a reference to an actual dialogue rather than thought or ‘inner speech’ without an external recipient.

- (14) [...] *por un tiempo fui la informática como un año y pico dos años / hasta que llegaron... muchachos recién graduados los pusieron ahí y llegamos a ser como once personas... pero...eh //*  
**[a] dije “bueno voy a cambiar ya porque”**  
*además que ahí trabajaba mucho mucho mucho y con la universidad no me daba casi tiempo / y hacía lo que me tocaba y lo que no me tocaba bueno llegué hasta hacer nómina y todo... eh*  
**[b] dije “bueno me voy a cambiar”**  
**[c] le dije a mi hermana un día “oye eeh averíguame mmh algún trabajo que haya por allá [...]” / [...] y entonces**  
**[d] Adriana me dijo “no ahí está hay una plaza componedor... pero tienes que aprender un programa hay que no sé que” //**  
**[e] “bueno yo voy a ir a probar” // vine / acababan de inaugurar la editorial //**  
**[f] Aidé habló conmigo me dijo “¡no! para que empieces” no sé que [...]**

‘[...] for some time I was the informatician like a year and a bit two years / until there arrived... some recently graduated guys they were placed there and we went up to like 11 people... but... eh //

**[a] I said “BUENO I will change [job] already because”**

in addition to working a lot a lot a lot there and with the university I had hardly any time / and I was doing my duties and other stuff well I started receiving a salary and everything.. eh

**[b] I said “BUENO I will change”**

**[c] I said to my sister** one day “listen eeh find me mmh a job that’s available there [...]” and then

**[d] Adriana said to me** “no there’s a place for a technician... but you have to learn a program you have to I don’t know what” //

**[e] “BUENO I’m going to try” // I came / they had just opened the agency //**

**[f] Aidé talked with me she said to me** “no! why would you start” I don’t know what [...]

(LHAB\_M12\_043)

Example (14) also illustrates the distribution of *bueno* and other quotative markers already discussed in Section 4.3: in longer stretches of RD, *decir* occurs in the beginning either alone or with other markers, while *bueno* and other markers alone function as quotatives only after the quotative frame has been established by the use of *decir*.

Contrasting with the Uruguayan and Cuban examples discussed above, in the varieties with least occurrences of *bueno* in RD, not only is *bueno* less frequently used but it also seems to convey other discourse marker functions in addition to marking a stretch of dialogue as RD. Rather, it seems to maintain the ‘original’ discourse marker function, for instance, marking a dispreferred turn in a conversation. This is exemplified by (15) from the Guatemalan data. This excerpt is part of a lengthy monologue where the informant refers to a situation where his father was offering money to his son, which caused him to feel both upset and ashamed and to respond ‘*bueno* then I’ll just leave’.

- (15) [...] *le dio diez pesos y yo pues dije yo ganas de // “dáselos nene” le dije // me dieron ganas de decirle así pero bueno / [teeth-sucking sound] / aquel es un niño veá / me enojó mucho y // y / con él me dio vergüenza / “bueno entonces ya me voy” le dije yo*

‘he gave him ten pesos and I said wanted to [say] “**give them to him kid**” **I said to him** // I wanted to say that to him but well [teeth-sucking sound] / it is a child, look / I became very angry and // and / I was ashamed for him / “**BUENO then I’ll go already**” **I said to him**’  
 [...](GUAT\_H11\_011)

Since the speaker in example (15) is reproducing a conflictive situation where *bueno* may have been used as a discourse marker to signal, e.g., the unexpectedness of the verbal reaction or used as a hedge, it can be interpreted as being part of the RS. Thus, this example differs from (11)-(14), where the presence of *bueno* is not motivated by contextual clues from the original discourse context where the stretch of RD is nested.

Although examples like (13) and (14) do suggest that local discourse patterns distinguishing between RS and RT emerge in dialects with most productive usage of *bueno*, it is difficult to say to what extent such patterns are generalized or, rather, depend on the local discourse context – or even reflect the idiolectal preferences of individual speakers. Thus, regarding the eventual distinction between RS and RT, the main finding of our analysis is that the two are generally not distinguished in Spanish but, rather, RT is represented similarly to RS. In other words, thinking – an invisible, mental activity – is metaphorically represented as speaking when reported to an interlocutor.

## 6. Discussion and conclusions

This paper has examined the emergence of the polyphonic functions of the discourse particle *bueno* across dialects of Spanish, testing the hypothesis, formulated on the basis of Borreguero Zuloaga's (2017) observation that *bueno* and other discourse particles frequently occur in the beginning of reported discourse, that *bueno* may grammaticalize into a marker of polyphony and that it may even replace quotative markers such as *decir* 'say'.

A quantitative analysis of our data identified dialects with very productive use of *bueno* (i.e., our Venezuelan, Peruvian, Cuban, Uruguayan and European samples) and another group of dialects with less productive use of *bueno* (i.e., our Mexican, Guatemalan, Colombian, and Chilean samples). All dialects of the first group except Venezuela also showed frequent combination of *bueno* with other discourse particles as prefaces. As expected, the dialects with most frequent and productive uses of *bueno* were also the ones with most frequent use of *bueno* at the beginning of reported discourse. In order to exclude the possibility that this finding is due to a simple frequency effect, we moved on to assess the degree to which *bueno* has started competing with quotative markers in dialects in which it is particularly productive. Contrary to expectations, in these dialects reported *bueno* is more likely to co-occur with quotative markers (again, with the exception of Venezuela). Finally, we identified a positive correlation between the appearance of prosodic breaks between the quotative verb *decir* and the quotative marker *bueno*, and the usage frequency of the quotative *bueno*. We interpreted these findings in line with a scenario in which the grammaticalization of *bueno* towards expression of a quotative function is due to a hearer-based reanalysis on the basis of the frequent collocation of *bueno* with quotative markers. Thus, while we did not find direct evidence of *bueno* replacing other quotative markers (as predicted by Borreguero Zuloaga 2017), the creation of *decir* + *bueno* prefabs might lead to a situation in which the use of *decir* is no longer necessary in order to mark reported discourse.

A more detailed qualitative analysis provided further evidence for this proposal by analyzing the use of reported *bueno* in two dialects in which *bueno* is particularly productive, namely Uruguayan and Cuban Spanish. Thus, it was shown that in *decir* + *bueno* collocations from these dialects, *bueno* no longer necessarily displays any of the interactional functions typical for *bueno* in direct, non-reported, discourse, suggesting that *bueno* is indeed conventionally used to (co-)mark RD.

Our analysis also shows that reported speech and reported thought are not routinely distinguished by speakers of Spanish: the use of mental verbs to introduce reported discourse is marginal, and typically thoughts are reported as if they were spoken words, i.e., using the verb *decir* 'say' and, occasionally, the quotative *bueno*. However, inspection of the Cuban

data suggested that in dialects in which *bueno* is particularly productive, *decir + bueno* collocations might even be recruited as markers of reported thought. Such uses of *bueno* can be characterized as even more innovative, as the function of reporting thought is even more dissociated from the original interactional functions of *bueno* (cf. also Rosemeyer and Sansiñena 2019 for a similar approach to *que* 'that'-prefaces in Spanish). In particular, when using *bueno* in a stretch of RT, it is highly unlikely that the speaker has indeed mentally uttered the word *bueno* in the situation they are describing. However, since operationalizing the distinction between reported speech and reported thought for a quantitative analysis seems extremely difficult due to non-distinct marking of the two types of reported discourse, this observation is based only on qualitative scrutiny of the data. Subsequent quantitative corroboration of these tendencies seems necessary, an issue we leave to future research.

## References

- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67 (1). 1–48.
- Benavent Payá, Elisa. 2015. *Decir y discurso directo en los relatos de la conversación coloquial*. València: Universitat de València.
- Borreguero Zuloaga, Margarita. 2017. Los relatos coloquiales: partículas discursivas y polifonía. *Pragmalingüística* 25. 62–88.
- Briz, Antonio, Salvador Pons & José Portolés (eds.). 2008. *Diccionario de partículas discursivas del español*. [www.dpde.es](http://www.dpde.es) (accessed 21 September 2020).
- Bybee, Joan L. 2003. Mechanisms of change in grammaticization: the role of frequency. In Richard Janda & Brian Joseph (eds.), *The Handbook of Historical Linguistics*, 624–647. Oxford: Blackwell.
- Bybee, Joan L. 2006. From usage to grammar: the mind's response to repetition. *Language* 82 (4). 711–733.
- Bybee, Joan L. 2010. *Language, Usage, and Cognition*. Cambridge & New York: Cambridge University Press.
- Bybee, Joan L. & Rena Torres Cacoullous. 2009. The role of prefabs in grammaticization: how the particular and the general interact in language change. In Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen M. Wheatley (eds.), *Formulaic Language, Volume I: Distribution and Historical Change*, 187–217. Amsterdam & Philadelphia: Benjamins.
- Carlier, Anne, Walter De Mulder & Béatrice Lamiroy. 2012. Introduction: The pace of grammaticalization in a typological perspective. *Folia Linguistica* 46 (2). 287–301 [Special issue *The Pace of Grammaticalization in a Typological Perspective*, eds. Anne Carlier, Walter De Mulder & Béatrice Lamiroy].
- Croft, William. 2016. Typology and the Future of Cognitive Linguistics. *Cognitive Linguistics* 27 (4). 587–602.

Accepted for publication (2021) in Spronck, S., Casartelli, D., Cruschina, S. & Posio, P. (eds.). *The grammar of thinking - From reported speech to reported thought in the languages of the world*. De Gruyter Mouton, Trends in Linguistics. Studies and Monographs (TiLSM). (Author's accepted manuscript)

Detges, Ulrich & Richard Waltereit. 2002. Grammaticalization vs. reanalysis: a semantic-pragmatic account of functional change in grammar. *Zeitschrift für Sprachwissenschaft* 21 (2). 151–195.

Detges, Ulrich & Richard Waltereit. 2011. Turn-taking as a trigger for language change. In Sara Dessì Schmid, Ulrich Detges, Paul Gévaudan, Wiltrud Mihatsch & Richard Waltereit (eds.), *Rahmen des Sprechens. Beiträge zur Valenztheorie, Varietätenlinguistik, Kreolistik, Kognitiver und Historischer Semantik. Peter Koch zum 60. Geburtstag*, 175–190. Tübingen: Narr.

Eckardt, Regine. 2009. APO: Avoid Pragmatic Overload. In May-Britt Mosegaard Hansen & Jaqueline Visconti (eds.), *Current Trends in Diachronic Semantics and Pragmatics*, 21–41. Bingley: Emerald.

Gries, Stefan Th. 2009. *Quantitative Corpus Linguistics with R. A Practical Introduction*. New York & London: Routledge.

Gries, Stefan Th. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10 (1). 95–125.

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39. 1041–1070.

Heine, Bert. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51 (6). 1205–1247.

Heritage, John. 1984. A change-of-state token and aspects of its sequential placement. In J. Maxwell Atkinson & John Heritage (eds.), *Structures of Social Action: Studies in Conversation Analysis*, 299–345. Cambridge: Cambridge University Press.

Heritage, John. 1998. *Oh*-prefaced responses to inquiry. *Language in Society* 27. 291–334.

Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press.

Kuznetsova, Alexandra, Per B. Brockhoff & Rube H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82 (13). 1–26.

López Serena, Araceli & Margarita Borreguero Zuloaga. 2010. Los marcadores del discurso y la variación lengua hablada vs. lengua escrita. In Óscar Loureda & Esperanza Acín Villa (eds.), *Los estudios sobre marcadores del discurso en español, hoy*, 415–495. Madrid: Arco Libros.

Martín Zorraquino, María & José Portolés. 1999. Los marcadores del discurso. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua Española*, 4051–4213. Madrid: Espasa Calpe.

PRESEEA. 2014. *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. <http://preseea.linguas.net> (accessed 9 September 2020).

Accepted for publication (2021) in Spronck, S., Casartelli, D., Cruschina, S. & Posio, P. (eds.). *The grammar of thinking - From reported speech to reported thought in the languages of the world*. De Gruyter Mouton, Trends in Linguistics. Studies and Monographs (TiLSM). (Author's accepted manuscript)

Pons, Salvador. 2003. From agreement to stressing and hedging: Spanish *bueno* and *claro*. In Gudrun Held (ed.), *Partikeln und Höflichkeit*, 219–236. Frankfurt am Main: Peter Lang.

Posio, Pekka & Andrea Pešková. 2020. Construccionalización de los introductores cuotativos con el verbo *decir* en español peninsular y argentino. *Spanish in Context* 17 (3). 391–414.

R Development Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org> (accessed 9 February 2021).

Rosemeyer, Malte. 2019a. Actual and apparent change in Brazilian Portuguese *wh*-interrogatives. *Language Variation and Change* 31 (2). 165–191.

Rosemeyer, Malte. 2019b. Brazilian Portuguese *in-situ wh*-interrogatives between rhetoric and change. *Glossa* 4 (1). 1–29.

Rosemeyer, Malte & Eitan Grossman. 2017. The road to auxiliariness revisited: the grammaticalization of FINISH anteriors in Spanish. *Diachronica* 34 (4). 516–558.

Rosemeyer, Malte & María Sol Sansiñena. 2019. The discourse functions of insubordination in Spanish *wh*-interrogatives. *Papers of the Linguistic Society of Belgium* 13. <https://sites.uclouvain.be/bkl-cbl/en/journals/pap> (accessed 3 March 2021).

Tagliamonte, Sali. 2012. *Variationist Sociolinguistics. Change, Observation, Interpretation*. Malden, MA: Wiley-Blackwell.

Wickham, Harald. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.