**DE GRUYTER**

**Research Article**

Oliver Ehmer, Malte Rosemeyer*

# When "Questions" are not Questions. Inferences and Conventionalization in Spanish *But*-Prefaced Partial Interrogatives

**Abstract:** The present paper analyzes the discourse-pragmatic function of introducing Spanish *qué* 'what'-interrogatives with the concessive connective *pero* 'but'. In some contexts, a *pero*-preface contributes to the interpretation of the interrogative as the realization of an interactional challenge rather than a request for information (e.g. an information question). We explore the inferential processes by which the *pero*-preface leads to an interpretation of the interrogative as an interactional challenge and try to demonstrate that this challenge function of *pero*-prefaced *qué*-interrogatives may not only achieved 'ad hoc' by a local combination of the constitutive elements, but also by conventionalized form-function associations that developed diachronically. In a first step, we analyze *pero*-prefaced *qué*-interrogatives in a corpus of spoken Present Day Spanish. There are three main functions of *pero*-prefaces: to signal that a previous answer to the same interrogative is insufficient, to insist on an answer to a previously unattended request, or to challenge an immediately preceding action by an interlocutor. Using methodology from variationist linguistics, we identify entrenched patterns of *pero*-prefaced *qué*-interrogatives that have conventionalized the challenge function. In a second step, we conduct a diachronic variationist analysis of the development of Spanish *pero*-prefaced *qué*-interrogatives between 1700 and 1975, testing the hypothesis that the challenge reading developed later than the question reading. Our results show that due to their largely monological nature, the same inferential processes cued by *pero* lead to different discourse functions in historical texts. Over time, however, the use of *pero*-prefaced interrogatives started to become more likely in constructed dialogues. We argue that this change reflects an ongoing conventionalization of the challenge function in *pero*-prefaced interrogatives in spoken language.

**Keywords:** interrogative; concession; inference; wh-question

## 1 Introduction

Interrogatives, i.e. constructions such as *What do you think?,* are not always questions. It is well known that apart from requesting information, interrogatives can be used to perform a broad range of other conversational functions, such as performing an offer *'Would you like to have a cup of tea?'* or an evaluation *'How beautiful is that!'* (cf. Stivers and Enfield 2010). However, the precise function of a specific interrogative

---

---

**\*Corresponding author: Malte Rosemeyer,** KU Leuven, Leuven, Belgium , E-mail: malte.rosemeyer@gmail.com
**Oliver Ehmer,** Albert-Ludwigs-Universität Freiburg

construction is often difficult to describe and depends on a variety of linguistic and cognitive factors.

The present paper analyzes the influence of one of these factors, namely the effect of prefacing Spanish partial interrogatives formed with the interrogative pronoun *qué* 'what' with the concessive connective *pero* 'but'.[1] We are specifically interested in the fact that in some contexts, a *pero*-preface changes the interpretation of the interrogative from requesting information to performing an *interactional challenge*, signaling that the speaker considers a previous utterance to be inappropriate. The aims of the paper are twofold. First, we explore the inferential processes by which the *pero*-preface leads to an interpretation of the interrogative as an interactional challenge. Second, we try to demonstrate that this challenge function of *pero*-prefaced *qué*-interrogatives may not only be achieved 'ad hoc' by a local combination of the constitutive elements, but also by conventionalized form-function associations that developed diachronically.

To pursue these aims, we analyze a corpus of spoken Present Day Spanish and a diachronic corpus of written Spanish, using both methodologies from interactional and variationist linguistics. We proceed with our analysis in two steps. In a first step, we analyze the function of *pero*-prefaced *qué*-interrogatives in a corpus of spoken Present Day Spanish drawing on the corpora *cola* (Jørgensen et al.), *callhome* (Canavan/Zipperlen 1996), *cespla* (Ehmer 2010), *c-oral-rom* (Cresti and Moneglia 2005), *hacspa* (Gabriel 2011), and *nccsp* (Torreira/Ernestus 2010). We demonstrate that there are four main functions of *pero*-prefaces in *qué*-interrogatives: to signal that a previous answer to the same interrogative is insufficient, to insist on an answer to a previously unattended request, to challenge an immediately preceding action by an interlocutor and to perform a strong evaluation. First, these functions are dependent on the local context of use of the interrogative. Concrete examples will be discussed in detail in section 4, taking into account various contextual factors that, based on inferences, lead to the different functions. Second, we argue that these three functions reflect a gradual alteration of the basic concessive meaning of *pero*. On the basis of a quantitative analysis using variationist methodology, we identify entrenched patterns of *pero*-prefaced *qué*-interrogatives that have conventionalized the challenge function.

In a second step, we conduct a diachronic variationist analysis of the development of Spanish *pero*-prefaced *qué*-interrogatives between 1700 and 1975, using the CORDE corpus (Real Academia Española 2016). We test the hypothesis that the challenge reading developed later than the question reading, focusing on realizations with the verb *decir* 'to say'. Our results show that due to their largely monological nature, the same inferential processes cued by *pero* lead to different discourse functions in historical texts. Over time, however, the use of *pero*-prefaced interrogatives starts to become more likely in fictional dialogues. We argue that this change reflects an ongoing conventionalization of the challenge function in *pero*-prefaced interrogatives in spoken language.

## 2 Theoretical prerequisites

In this section we establish the theoretical prerequisites for the analysis. We first discuss patterns of functional variation in interrogatives (2.1), then go on to describe the relevance of prefacing for inference management in general and for 'but' in particular (2.2).

### 2.1 Functional variation in interrogatives

It is well known that interrogative constructions can be used not only used to request information, but rather to comment on something that the interlocutor has said (Stivers/Enfield 2010). Consider example (1), taken from the COCA (Davies 2015-2016), in which an enlistment situation is narrated.

---

**1** In this paper, we only consider partial interrogatives formed with the interrogative pronoun *qué* 'what', although we believe that our findings should also be of relevance for interrogatives formed with other interrogative pronouns and adverbs.

**Example (1): enlistment**
(Source: *COCA* (Davies 2015-2016), *Shot in the back*, 2015)

01   *An army sergeant glanced up from the game and saw Alexander standing just*
02   *inside the door.*
03   *"Yes, sir, something we can do for you?" the sergeant asked.*
04   *"Is this where you join the army?" Alexander asked.*
05   *"It sure is. If your grandson is looking to join, why you just bring him right on down*
06   *here, old-timer, and we'll sign him right up."*
       *[...]*
07   *"No, sir, I'm not askin' about my grandson. I'm askin' about me."*
08   *Sergeant Kilbride got a confused look on his face. "I don't understand. What do you*
09   *mean, asking about you?"*
10   *"Joinin' up", Alexander said. "I want to join the army."*
11   *The sergeant laughed. "**You want to fight for Uncle Sam, do you?**".*
12   *"You're damn right I do. And I'll fight as hard for the Yankee government this time*
13   *as I once fought against it".*

Given that the participants are in an enlistment office, the recruiters' standard assumption is that anyone entering the office wants to join the army (l. 1–4). However, given that Alexander is very old, they come up with an alternative explanation of his presence, namely, that he wants to enlist his grandson (l. 5–6). This forces Alexander to clarify twice that he indeed wants to join the army (l. 7 and 10). This information first leads to confusion (l. 8–9) and then disbelief (l. 11). The recruiter expresses his disbelief both by laughing and then asking about something that has already been said twice: whether Alexander wants to join the army. Consequently, the interrogative *You want to fight for Uncle Sam, do you?* is only superficially a question. The interrogative is not used to ask for information, but rather challenges the proposition 'I want to join the army' asserted by Alexander in the preceding context.

Example (1) thus illustrates that "grammar and prosody do not reliably identify questions" (Hayano 2012). Rather, we can assume that the most basic pragmatic function of an interrogative is simply to mobilize a response by the interlocutor (Stivers and Rossano 2010). Which type of response is expected or preferred by the interrogative utterer depends on (a) the way the interrogative is designed (e.g., whether it is a polar or a partial interrogative, or its lexico-syntax and prosody) and (b) the preceding context, specifically the degree to which the information requested by the interrogative is presupposed. Prefaces such as *but* or *and* belong to the first strategy.

## 2.2 *Pero*-prefaces and inferences

Prefacing elements at the beginning of a turn or utterance play an important role in its pragmatic interpretation. To name an example, it has been shown that *well*-prefaces in English serve to "index some kind of departure from expectations for the subsequent turn" (Heritage 2013). Prefaces establish a kind of interpretative 'indexical frame' (Auer 1996) for the understanding of the upcoming turn that guides inferential processes concerning, for instance, the connectedness of the turns (Schegloff and Sacks 1973; Mazeland and Huiskes 2001; Heritage 2013; Li 2016), the relation of the turn to an institutional 'agenda' (Heritage and Sorjonen 1994) or the expectations of the participants. Prefaces can include a whole range of different kinds of elements (Heritage 2013) as for example breathing and bodily visual behavior (Streeck and Hartge 1992), address terms (Clayman 2010; 2012) particles like *oh* and *ah* (Heritage 1984; Küttner 2016; Küttner this SI) as well as discourse markers and connectives (Schiffrin 1987; Hansen 1998a; b) like *but*.

There is a broad range of studies on the concessive connective *but* pointing to the inferential processes involved in its use, especially within the *denial of expectations* approach (Lakoff 1971; Blakemore 1989; Iten 2005). We follow the general assumption of this approach, which claims that *but* signals a deviation from

an expectation that is assumed to hold in conversation. These expectations have been described in terms of presuppositions (Lakoff 1971), argumentative viewpoints (Anscombre and Ducrot 1977) and the relevance of contextual interpretations (Blakemore 1989). Consider the constructed dialogue in example (2).

**Example (2): ring**
(Assumed context: Standing in front of a shop-window)

> 01   A:   *This ring is beautiful.*
> 02   B:   *Oh yes it is!*
> 03        *But it's too expensive.*

A's utterance *This ring is beautiful* (02) can be said to raise the expectation that A would like to buy the ring or even make B buy it. This expectation is cancelled by B's utterance *But it's too expensive* (03). The function of *but* is to signal that while B accepts the truth or validity A's utterance, a likely consequence (in this case to buy the ring) is denied. That the concession actually involves the (partial) acceptance of a contrasting position is evidenced in this example by B's utterance *Oh yes it is!* (02), explicitly approving the position previously taken by A. The example thus also conforms to the *cardinal concessive schema* developed in interactional linguistics (Couper-Kuhlen and Thompson 2000; Barth-Weingarten 2003), where concession in conversation in its basic form is seen to involve three sequential moves: (1) a claim by one participant A, (2) an acknowledging move by an interlocutor B, who subsequently (3) formulates a claim that is possibly incompatible with A's prior claim.

All of the aforementioned cognitive and argumentative approaches assume that the expectation or assumption that is denied or argued against may or not be inferred by the interlocutors. The assumption can be accessible or manifest to the interlocutors to different degrees (cf. also Iten 2005). The notion of manifestness refers to the fact that interlocutors may not only assume that an assumption actually holds but also that there are 'degrees of probability' for assumptions to be shared by interlocutors as part of their common ground (cf. Sperber/Wilson 1996 [1986]: 39-46). For instance, Bell (1998; 2010) assumes that *but* operates on assumptions that are 'derivable' from the prior discourse, which implies that in some cases, they have actually *not* been derived by some or even all interlocutors. As an alternative to *derivable assumption* Bell also uses the term *potential consequence*. The basic function of *but* is then to "both concede[s] the truth of the prior discourse and signal[s] the cancellation of the potential consequence derivable from the prior discourse segment" (Bell 2010).

Crucially for the present study, the inferred consequence can be located on different levels of discourse. In example (2), the inferred consequence can be said to concern the *semantic/ideational* dimension of discourse (Schiffrin 1987), it can also concern the level of the *sequential organization* of discourse (Bell 2010). The only previous study on *but*-prefaced interrogatives, Jol and van der Houwen (2014), demonstrates that in combination with an interrogative, *but* usually serves to cancel an inference that concerns the sequential organization of discourse. The authors analyze Dutch *maar* 'but'-prefaced interrogatives in police interviews with children. In their data the policemen typically use *but*-prefaced interrogatives in order to retrospectively mark the child's answer to a preceding answer as insufficient, thereby prompting the child to refine his/her previous answer. Likewise, our study will demonstrate that in Spanish partial interrogatives, *pero* 'but' typically concerns the sequential organization of discourse.

Crucially however, Jol/van der Houwen do not describe the inferential processes leading to this pragmatic effect. In the remainder of this paper, we will try to model these inferential processes, as well as offer a perspective on the historical development of one type of *but*-prefaced *qué*-interrogatives in Spanish.

# 3 Data

We conducted qualitative and quantitative analyses of *pero*-prefaced interrogatives on the basis of data from two corpora. The corpus of spoken Spanish contains 1.500.000 words and consists of conversations

that took place between 1996 and 2010 (aligned transcriptions with audio).[2] The conversations took place in various Spanish-speaking countries and represent different speech situations (telephone conversations, interviews, spontaneous multi-party conversations). As a corpus of historical Spanish texts, we used the CORDE (Real Academia Española 2016), which contains more than 300.000.000 tokens of written Spanish from the earliest texts onwards to 1975 from different countries and different text genres.

For the qualitative analysis (sections 4.1 and 4.2), we collected all instances of *pero*-prefaced *qué*-interrogatives from the corpus of spoken Spanish (*n*=260). For the quantitative analyses (sections 4.3 and 5), we focused on the specific case of *pero*-prefaced *qué*-interrogatives formed with the verb *decir* 'to say'. We first collected all instances of the pattern *qué* 'what' + DECIR 'to say' from the corpus of spoken Spanish (*n*=914). We eliminated all cases of other prefaced particles (especially *y* 'and') from the data, which left us with *n*=637 tokens. For the historical analysis, we collected all instances of *qué* + DECIR in Peninsular Spanish texts between 1700 and 1975 from the CORDE. After eliminating all cases of other prefacing particles, the resulting database had a size of *n*=3571 tokens.

# 4 The pragmatics of *but*-prefaced *qué*-interrogatives in spoken Spanish

## 4.1 Interactive functions and sequential patterns

As a first step of our analysis, we conducted a qualitative analysis of all 260 instances of *pero*-prefaced *qué*-interrogatives in our spoken corpus. We identified four major sequential patterns of use corresponding to four interactive, sometimes overlapping, functions: (1) realizing a second request for information while specifying the first request, (2) realizing a second request for information in order to insist on an answer to an unattended previous request, (3) performing an interactional challenge and (4) performing a strong evaluation. In this section, we describe all of these functions on the basis of actual corpus data.

*Pero*-prefaced *qué*-interrogatives are frequently used by speakers to realize a second request for information when the first request has been answered by the addressee in an insufficient manner. The *pero*-prefaced request then both repeats and elaborates the first request. In these cases the *pero*-preface functions in a similar way as the Dutch *maar* ('but')-prefaces described by Jol and van der Houwen (2014), i.e. it signals that the previous answer was insufficient and prompts an elaboration. The following sequence is taken from a shop encounter, where a customer wants to buy a pen of the brand "pilot" of which several variants are available. The seller tries to find out which specific pen the seller wants.[3]

**Example (3):** *rojos*
(Source: *cola* (Jørgensen et al.), mabpe2-07, 899-911 sec)

```
01 D:  eh::: qué clase de pilot queRÍas;=
        'eh::: what kind of pilot (pens) did you want'
02      =de Éstos o [de los O]Tro[s.        ]
          'of these or of the others?'
03 B:                [eh::-    ]
                        'eh'
04                                  [<<all>de]_estos de_EStos.>
                                        'of these of these'
05 D:  de EStos.=
        'of these'
```

```
->    06        qué que[RÍas (hijito).]
                'what did you want (my son)'
      07 B:          [e::l       RO]jo,
                        'the red (one)'
      08        (0.5)
      09 D:   [roJO,       ]
                 'red'
      10 B:   [<<p>hm_HM;>]
->    11 D:   pero qué [QUIEres;   ]=
                'but what do you want'
      12 B:          [<<pp>(y:)>]
                           '(and)'
      13 D:   =el rojo: de los fInos: o de lo::s-
                 'the red (one) of the thin (ones) or of the'
      14 B:   de los GORdos.
                'of the thick (ones)'
      15        (0.2)
      16 D:   de_ÉS[tos;]
                'of these'
      17 B:        [del ] Otro [del OTro.]
                     'of the other of the other'
      18 D:                    [no del   ] OTro.
                                 'no, of the other'
      19        (.)
      20 D:   de BOla.
                'with a ball'
      21        (0.8)
```

The identification of the specific pen the customer wants to buy is accomplished in a stepwise fashion. The sequence starts with a seller's request for information of which kind of pilot pen the customer wants (l. 01), while pointing to two alternative groups of pens (l. 02). The buyer indicates the group of pens of which he wants one (l. 04), which is confirmed by the seller (l. 05), signaling that she is satisfied with the response. Until now the exchange seems to work smoothly. The seller then proceeds with a further request for information *qué que[RÍas (hijito).]* 'what did you want (my son)' (l. 06), to which the buyer responds with the information that he wants 'the red one' (l. 07). The seller first partially acknowledges the response by repeating the color (l. 09), but then repeats her request with a *pero*-preface: *pero qué [QUIEres;]=* 'but what do you want' (l. 11). The *pero*-interrogative is clearly designed as a repetition of the immediately preceding request. Note however that the speaker changes the tense of the verb from past to present in order to emphasize the relevance of her request. Instead of a mere repetition, however, the seller now details which particular information she needs, asking whether the buyer wants a 'fine' or a 'thick' red pen (l. 13-14). The customer then delivers the requested information about the thickness (l. 14) and the conversation goes on with a further specification of the pen tip.

The sequential context in which the *pero*-prefaced *qué*-interrogative is used can be characterized as involving four steps. First, a speaker produces a request for information, which is in a second step attended to by the addressee. In a third step, the requester acknowledges the validity of the response and 'accepts' the information into the Common Ground. However the information is only validated in part, since in a fourth step the requester then repeats the initial request, now prefacing it with *pero*. The repeated *pero*-prefaced request thus functions as a kind of repair marker, as the requester signals that the previously produced response, although acknowledged, is pragmatically insufficient. This is corroborated by the fact that the requester now also details which further information is required to render the response sufficient.

The inferential process involved in this use of the *pero*-preface in second requests can be analyzed as follows. First, the use of *pero* signals that the preceding statement is partially "appropriate", i.e., the proposition expressed by the statement is integrated into the Common Ground. Second, at the same time, an inference is invited regarding the pragmatic level of the progressivity (Schegloff 1979; Stivers and Robinson 2006; Heritage 2007) of the interaction. Spoken interaction usually conforms to the principle "When an information request has been answered, discourse can proceed". However, in our second request uses, this principle is cancelled; the speaker apparently does not accept the response as sufficient. While a literal repetition of the information request would allow for the interpretation that the answer has simply not been heard, the use of *pero* indicates to the hearer that the preceding answer is pragmatically 'insufficient' and needs to be repaired. This inference is further enforced by the further specification of the request.

In our data we find a further use of *pero*-prefaced *qué*-interrogatives, also involving second requests that are literal or slightly modified repetitions of a first request for information. However, what distinguishes this sequential pattern from example (3) is that the first request for information has *not* been answered by the addressee. In the following excerpt F is talking about her uncle.

**Example (4): *tío***
(Source: *cola* (Jørgensen et al.), malce2-09, 316-325 sec)

```
     01 F:   qué (.) qué te iba a deCIR, (.)
             'what, what was I going to tell you'
     02      que mi [tío: es muy GUApo eh?]=
             'that my uncle is really good looking, right?'
     03             [((background voices))]
     04      tiene así los [ojos aZU:]le:s,
             'he's got such blue eyes'
->   05 E:                 [qué TÍo. ]
                              'which uncle'
     06 F:   °hh ahora es mo/ [(.) más moREno;       ]
                'now he is ta/ more tanned'
->   07 E:                     [<<f, h>pero: qué TÍO;>]
                                  'but which uncle'
     08      (0.3)
     09 F:   <<f>el que vive en teneRIfe:->
                'the one that lives in Tenerife'
     10      (0.2)
     11 E:   ah sÍ yo no le coNOZco;
             'oh yes, I don't know him'
```

F is describing the appearance of her uncle (l. 01-03), when E realizes a first request for information with *qué TÍO.* 'which uncle' (l. 5). The request presents an information-seeking *qué*-interrogative, aiming at the clarification of personal reference. Note that E realizes her request in overlapping speech, thereby intending to interrupt the talk of F. F, however, continues the description of her uncle in l. 06, not attending to E's request. E now repeats her first request, repeating the previous interrogative: *<<f, h>pero: qué TÍO;>]* 'but which uncle' (l. 07). This second request is not only prefaced with *pero* but also uttered louder and in a higher pitch register than the first request. By using those signals, E performs an upgrade of the importance of her information request, insisting on an answer. Now F responds to E's request, delivering clarifying details about the uncle (l. 09) that arguably allow E to establish the reference, as evidenced in l. 11.

The pragmatic values of the *pero*-prefaced interrogatives in examples (3) and (4) are based on the same inferential processes. The speaker prompts the addressee to draw the inference that his on-going talk needs to be suspended. What distinguishes this sequential pattern in example (4) from the first pattern is the simple fact that no answer to the first request for information has been produced and acknowledged. This

difference influences the basis for the inferential processes at work. While in example (3), the *pero*-prefaced interrogative is used to signal that the answer to the first request was *insufficient*, in example (4) it is used to signal that there has *not been an answer at all*.

Consequently, whereas the sequential development of the interaction in example (3) still conforms to the cardinal concessive schema, in example (4) it deviates from the schema because there is no longer an acknowledging move by the interlocutor. By using the *pero qué* construction in a context in which there is nothing to be acknowledged, its main function is now to signal that the discourse cannot proceed. It is due to these changes in the sequential structure that the discourse-organizational function of the construction seems particularly strong in example (4). As a result, we believe that examples such as (4) are an important intermediary step in the semantic change from an information-seeking to a challenge function.

Speakers have the option to leave out even more sequential moves from the basic pattern. In particular, we find cases of *pero*-prefaced interrogatives in our data in which the first request for information is missing altogether. In some of these cases, the interrogative is merely used to interrupt the current speaker and signal that the progressivity of talk needs to be suspended until the request is sufficiently attended to. This *suspending* function of *but* is reminiscent of Nemo's (2002) analysis, according to which *but* signals that a process should be/could have been "stopped". However, *pero qué*-interrogatives are frequently also used not to request information but to perform a conversational challenge. The following example is taken from a conversation between friends that are currently eating. We do not know what it is that they are eating, but we can judge from the audio recording that there are several dishes available.

**Example (5): *asquerosas***
(Source: *cola* (Jørgensen et al.), malce2-14, 100-115 sec)

```
      01 E:   están muy BUEnas ((name));
              'they are very good'
      02      (0.4) ((voices in the background))
      03 F:   están de puta MAdre.
              'they are fucking great'
      04      dame OTra.
              'give me another one'
->    05 D:   pero qué DIces-=
              'but what do you say'
      06      =te GUStan;
               'you like them'
      07      (0.5)
      08 D:   casi POto yO=tío.
              'I almost puke, dude'
      09      (1.2) ((voices in the background))
      10 D:   <<f>ah que está asqueROsa.>
                  'ah (that) it's disgusting'
      11      (0.3)
      12 D:   están asqueROsas.
              'they're disgusting'
      13      (0.2) ((voices in the background))
      14 F:   están RIcas.
              'they're tasty'
      15      (0.5)
      16 D:   están asqueRO[sas.]
              'they're disgusting'
      17 F:                [yo: ] es la tercEra que me como YA;
                   'I, it's already the third (one) I eat'
      18      (0.2)
```

The sequence starts with a first positive assessment of the food by E (l. 01). F produces a second assessment, performing an upgrade (l. 02), and asks for more of the food (l. 03). Now D joins the assessment sequence, by using a *pero*-prefaced *qué*-interrogative: *pero qué DIces-=* 'but what do you say'. Here the interrogative is not used to perform a request for information, i.e. that F should repeat what he just said. The speaker D actually makes clear in her following utterance that she has indeed understood F's contribution by rephrasing it in the disbelieving question: *=te GUStan;* 'you like them' (l. 06). Rather, the *pero qué*-interrogative is used to challenge the positive assessment of the food that F has just produced. This interpretation is corroborated by D's next utterance. After a lack of uptake by the other speakers (l. 07), D realizes several extremely negative assessments (l. 08, 10, 12), making the challenge of F's prior assessment event more apparent. In the following course of the conversation F accepts the challenge and defends his position (l. 14-18).

In this sequential pattern the *pero qué*-interrogative is not used to perform a request for information, that is to handle a conversational problem at the level of "insufficient information". Indeed, the information that is "asked for" with the *pero qué*-interrogative is already available to the requester. By requesting this presupposed piece of information, the speaker prompts a specific inference on part of the interlocutor. The hearer has to infer that the request does not concern a problem of insufficient information, but rather operates on a different level of discourse, namely that the statement made in the previous turn is considered to be inappropriate. In a nutshell, asking for presupposed information prompts the conversational inference that this information is challenged to be part of the interlocutors' Common Ground. In dialogical settings, such a proposition challenge may lead to an interactional challenge, especially when directed at a proposition that has been (intended to be) introduced to the Common Ground by a co-participant. The challenge function of *pero qué*-interrogatives represents a well-established rhetorical strategy, as evinced by the fact that we routinely find the respective sequential developments after its use, with co-participants either defending their position or backing down from it.

We find a disproportionate number of *pero qué*-interrogatives in reported speech contexts (see the quantitative analysis in 4.3 for details). In these verbal representations of past interactions *pero qué*-interrogatives seem to be exclusively used as interactional challenges. The following example is taken from a conversation between three women that are exchanging opinions about who will win the soccer league in Spain.

**Example (6): *ganar***
(Source: *nccsp* (Torreira and Ernestus 2010), 01-04-2008_1, 2500-2512 sec)

```
      01 A:   <<all>quién va a ganar la liga este aÑo creo que es: OBvio>,
                   'who is going to win La Liga this year, I think it's obvious'
      02 B:   sÍ <<creaky>ESa>.
              'yes, this (one)'
      03      (0.3)
      04 B:   [((laughs))]
      05 A:   [ESa.      ]
               'this (one)'
      06      el real maDRID,=
              'Real Madrid'
      07      =porque los resultados [lo: (.) conFIRman. ]=
               'because the results confirm it'
      08 B:                          [((laughs silently))] [°h  ]
      09 A:                                               =[ESa-]
                                                            'this (one)'
      10 C:                                               [°h  ]
              yo ayer dije el °h BARça;
              'yesterday I said Barça'
      11      y me dijo daNIEL-=
              'and Daniel said to me'
->    12      =pero <<laughing>qué [DICes-> ((laughs))]
               'but what do you say '
```

```
    13 A:                          [((laughs))          ]
    14 B:                          [((laughs))        ]
    15 A:  daniel [estaba también habLANdo.]
           'Daniel was talking too'
    16 B:         [°hh <<:-)>el bArça      ] ya va terCEro.> ((laughs))
                              'Barça is already third'
```

At the beginning of the sequence A and B collaboratively take the stance that it is obvious that the club Real Madrid will win the league (l. 01-09), even giving an account for their opinion (l. 07). Now C reports that the day before she said to Daniel,[4] a common friend, that the FC Barcelona ("Barça") would win the league (l. 08). She then reconstructs in reported speech what her interlocutor Daniel (l. 09) said to her: *=pero <<laughing>qué [DICes->]* 'but what do you say'. As has been pointed out in the research narrations, reported speech should not be interpreted as a verbatim repetition of an original utterance, but rather as a perspectival reconstruction by the narrator (Günthner 2005; Ehmer 2011). Following such an analysis, we may analyze the use of the *pero qué*-interrogative in this reported speech as a rhetorical device to portray the reported situation as being conflictual (cf. Briz 1993). More specifically, C uses the *pero qué*-interrogative to signal that her interlocutor in the narrated situation challenged her position. Actually, this reported challenge is the point of this mini-narration, since it serves C to perform a self-disclosure of having said something foolish, thereby creating a humorous effect (l. 13-14). We take uses in reported speech to portray a situation as being conflictual as an indicator of the conventionalization of the challenge function of *pero qué*-interrogatives.

A last function that we want to mention here is closely related to the one just discussed: *pero*-prefaced *qué*-interrogatives may also be used as exclamatives, expressing strong assessments and evaluations. This function is most commonly found with *pero qué*-interrogatives introducing noun phrases or adjectives, which may but need not be expanded by relative or copula phrases, for example *pero qué cabrón (que eres)*, 'but what (a) goat/bugger (you are)' or *pero qué lindo (es eso)*, 'but what/how beautiful (is this)'. The following example is taken from a conversation between two friends, who are talking about problems in partnerships in a playful and humorous way.

### Example (7): cacho cabrón
(Source: *c-oral-rom* (Cresti/Moneglia 2005), etelef09, 354-374 sec)

```
       01 J:  probablemente lo mejor que es cuando menos
               <<laughing>HAblan>-
               'probably the less they speak the better'
       02      [((laughs))]
       03 T:  [((laughs))]
       04      pero-
               'but'
       05 J:  <<laughing> cuando (.) cuando menos coñazo DAN>
                            'the (.) the less trouble they cause you'
       06      ((laughs)) [((laughs))                                ]
->     07 T:             [<<:-)> pero qué cacho cabrón que ERes>-]
                              'but what a piece of crap (that) you are'

       08      <<:-)>es que (.) es que te lo tengo que deCIR;
                    'you know (.) you know I have to tell you that'
       09 J:  ((laughs))
       10 T:  JO:de:r-
               'damn'
       11 J:  ((laughs))
```

---
**4** The names have been altered to anonymize the transcript.

As part of a mocking activity J states that T prefers her partners not talk so much (l. 01). Both participants laugh (l. 02 and 03) and J produces an exaggerated version of his prior statement (05), attributing to T the attitude to her partners as possible sources of trouble. T, who already has started to formulate a counter position with *pero* in (l. 04) then uses a *pero qué*-interrogative with *cacho cabrón* 'piece of crap' as core of the noun phrase, which is extended by the relative phrase *que eres* 'that you are'. While in Spanish *cabrón* is polysemous and can 'mean' goat as well serve as a swear word ('bugger/crap'), it is already clear by the relative extension *que eres* ('that you are') that the *pero qué*-interrogative is not used to realize an information question ('But which piece of crap are you?') but an assertion and a strong evaluation. The expressive quality is co-signaled by the metalinguistic expression in l. 08 and the swear word in l. 10. The continuous laughs and the use of smile voice provide further support for the interpretation that the interrogative in l. 07 is used to express a strong evaluation, since they serve to soften it. We may conclude here that in terms of the interactional dynamics that a *pero*-preface to an exclamative serves to signal disagreement and co-indexes the challenging character of the utterance. Such uses are not only found in situation of co-presence but also in reported speech in narrations. We will argue below that in such cases the speaker uses *pero* to characterize the reported situation as being conflictual.

## 4.2 Syntactic and lexical variation

Table 1 reports the distribution of constructional types in the data from the oral corpora. *Pero qué* interrogatives are most frequently used in combination with verb phrases (VP 55.8%), although a substantial minority of cases involve a noun phrase (NP 21.2 %) or and adverb or adjective (ADV/ADJ 9.6 %), which frequently express strong assessments or evaluations.

**Table 1.** Syntactic variation in *pero qué* interrogatives

| Constructional type | Examples | *n* |
| --- | --- | --- |
| *pero qué* VP | *pero qué compro* 'but what do I buy'<br>*pero qué hiciste* 'but what did you do'<br>*pero que es esto* 'but what is this' | 55.8% (145/260) |
| *pero qué* NP | *pero qué metro* 'but which tram'<br>*pero qué cosa más rica* 'but what a delicious thing'<br>*pero qué asco* 'but what disgust' | 21.2% (55/260) |
| *pero qué* ADV/ADJ | *pero qué tonto* 'but how stupid'<br>*pero que guay tío* 'but how cool man'<br>*pero qué más* 'but what more' | 9.6% (25/260) |
| *pero qué* | *pero qué* 'but what' | 12.3% (32/260) |
| (unclassifiable) | (unintelligible phrases, broken off phrases, …) | 1.2% (3/260) |

In the following, we will focus on the 145 cases of *pero qué*-interrogatives that involve a verb phrase because they represent the most frequent construction type. A closer look at the lexical realization of these interrogatives (Table 2) suggests that they frequently express a challenge function. We observe a predominant use of the construction with *situational* verbs, i.e. verbs that make reference to an action that is evident in the situational context. In particular, the three verbs *hacer* 'to make' (25 %), *pasar* 'to happen' (17 %) and *decir* 'to say' (14%) account for 55% of *pero qué*-interrogatives with a verb phrase. These verbs are specifically apt for the expression of a challenge function because the interrogative proposition is typically manifest and perceptually accessible in the discourse (something that is happening, something that someone does, something that someone says). This means that typically, the information asked for is already available to the questioner.

Note, however, that it is not only in the case of the three most frequent verbs that we find a situational reading. If we consider the randomly selected examples from the non-situational verbs (e.g., *ser* 'to be'), we see that these tokens too suggest a situational reading in that they represent contexts in which the

proposition is mutually manifest to the interlocutors. For instance, in the utterance *pero qué es eso* 'but what is this' the deictic pronoun *eso* signals that the referent or event that is being discussed is mutually manifest to the discourse participants.

**Table 2.** Lexical variation in *pero qué* + verb phrase interrogatives

| Verb | Examples | *n* |
|---|---|---|
| hacer 'to do/make' | *pero qué estás haciendo* 'but what are you doing' | 24.8% (36/145) |
| pasar 'to happen' | *pero qué pasa* 'but what happens' | 16.6% (24/145) |
| decir 'to say' | *pero qué dices* 'but what do you say' | 14.5% (21/145) |
| ser 'to be' | *pero qué es eso* 'but what is this' | 13.1% (19/145) |
| querer 'to want' | *pero qué quieres comprar* 'but what do you want to buy' | 6.2% (9/145) |
| saber 'to know' | *pero qué sé yo* 'but what do I know' | 4.8% (7/145) |
| tener 'to have' | *pero qué tienes* 'but what do you have' | 2.8% (4/145) |
| hablar 'to speak' | *pero qué hablas tú normalmente* 'but what do you normally say' | 2.1% (3/145) |
| importar 'to matter' | *pero qué me importa* 'but what do I care' | 1.4% (2/145) |
| preguntar 'to ask' | *pero qué te iba a preguntar* 'but what was I going to ask you' | 1.4% (2/145) |
| (hapaxes) | | 11.7% (17/145) |
| (broken off) | | 0.7% (1/145) |

## 4.3 A quantitative analysis of the use of *(pero) qué* + DECIR in spoken interaction

Our qualitative analysis of *pero*-prefaced *qué*-interrogatives has not only described the typical discourse functions of these interrogatives, but also demonstrated correlations between these discourse functions and different contextual factors. In particular, it has shown that in contrast to information question readings, challenge readings typically arise in contexts when an interrogative about mutually manifest information is directed at an interlocutor who is present in the speech situation. Consequently, we expect the likelihood of a challenge reading to be higher in present tense and with second person singular morphology. In addition, the analysis has shown that challenge readings seem to be more likely in contexts of reported speech. In the present section, we aim at confirming the statistical relevance of these observations.

Our descriptive quantitative analysis in Section 4.2 furthermore suggested that three frequent verbs with situational readings – *hacer* 'to make', *pasar* 'to happen' and *decir* 'to say' are more likely to be used with a prefacing *pero* and consequently in challenge function than verbs with non-situational readings. We selected all interrogatives formed with one of these verbs, *decir*, as a representative sample of the alternation between *pero*-prefaced *qué*-interrogatives and non-prefaced *qué*-interrogatives (*n*=914). The reason for selecting precisely this verb was that in contrast to the two other situational verbs, *decir* refers to an action that describes an auditory rather than a visual stimulus. Consequently, it is much easier for the analyst to assess whether the interrogative proposition is manifest in discourse for *decir* than for *hacer* and *pasar*. While this fact is not yet relevant to the synchronic analysis, it will be for the diachronic analysis in Section 5.

In order to confirm the results from the descriptive qualitative analysis, we used a combination of *classification and regression trees* (CARTs) and *logistic mixed-effects regression analysis* over the alternation of the presence of a *pero*-preface in all *qué* + DECIR interrogatives. The dependent variable PERO refers to whether or not the interrogative is prefaced with the conjunction *pero*.

Classification and regression trees (CART) can be described as nonparametric, data-driven regression analyses that are ideal tools for exploratory statistical analyses (Tagliamonte and Baayen 2012). We constructed a CART over our sample of the alternation between *pero*-prefaced *qué*-interrogatives and non-prefaced *qué*-interrogatives, with the dependent variable PERO, by using the function *ctree()* from the

package *party* (Hothorn et al. 2015). Given that CARTs are not susceptible to multicollinearity, we included all variables coded, at various levels of granularity. These variables operationalized tense, person and number morphology of the verb, whether an auxiliary was present and the type of the auxiliary (modal, non-modal), whether the interrogative was repeated from an earlier utterance, whether the interrogative occurred in reported speech, the utterance length, whether or not there were elements interposed between the interrogative pronoun *qué* and the interrogative verb, whether or not there was a term of address used (pronoun, address term…) and the speech situation (see Table 4 in the appendix for a summary).

We plot the results from the CART in Figure 1. The most important predictor of the presence of a *pero*-preface to a *qué* + DECIR interrogative is whether or not the interrogative is used in reported speech (cf. Node [1]). Thus, the relative usage frequency of a *pero*-preface is significantly higher in reported speech contexts (about 40%) than in other speech contexts (between 2 and 10%). The second most important predictor is whether or not the interrogative verb is conjugated in second person singular morphology and present tense (cf. Node [2]). There is a significant increase in the relative usage frequency, from 2 to 10%, of a *pero*-preface in the presence of these morphological markers.
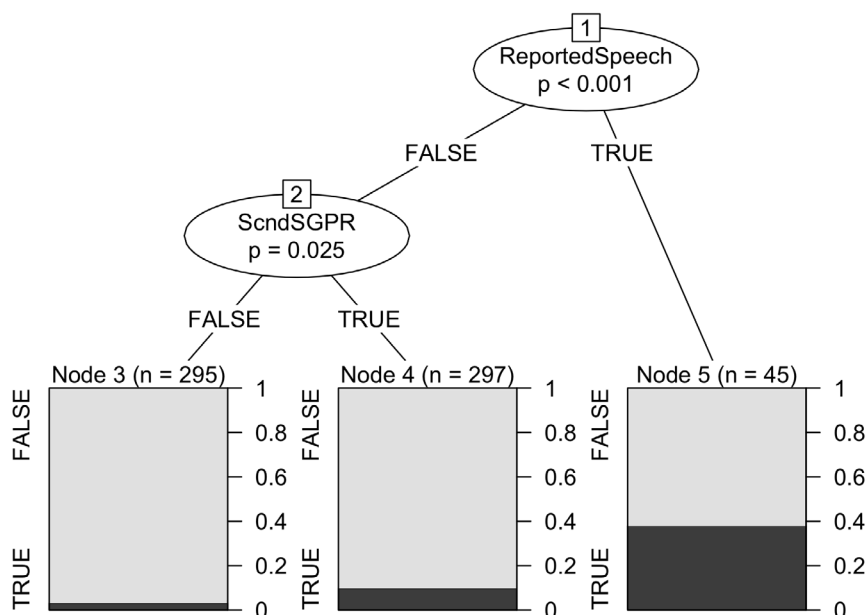


**Fig. 1.** Conditional inference tree for the variation between *pero qué* + DECIR and *qué* + DECIR in the corpus of spoken Spanish (c index of concordance = 0.72, somers' dxy = 0.45)

While the advantage of CARTs is that they can cope with multicollinearity and complex interactions, they cannot be used to test hypotheses. We therefore used a logistic mixed-effects regression analysis to test the findings from the CART, by using the function *glmer()* from the package *lme4* (Bates et al. 2015), including two random effects: GENRE, which refers to the type of speech situation ('spontaneous', 'telephone', and 'non-spontaneous'), and DATAID, which refers to the transcript number. The variable DATAID was used as a proxy for both interpersonal variation and diatopic variation. The analysis confirmed the statistical significance of the variables REPORTEDSPEECH and SCNDSGPR even when controlling for GENRE and DATAID (see Table 5 in the appendix for a summary of the results).

In summary, the quantitative analysis suggests that the most typical function of *pero*-prefaces is to signal a challenge. First, the analysis demonstrates that *pero* is typically used in *qué* + DECIR interrogatives that directly address an interlocutor in the here-and-now (effect of SCNDSGPR), i.e. in situations of

communicative immediacy. This means that these interrogatives are likely to express a challenge function because are typically used to discuss propositions that are manifest to the interlocutors in discourse.

Second, the analysis also suggests a conjoined influence of the parameters of person, number and tense morphology, in that the aggregate variable SCNDSGPR was selected as more significant than the separate variables for these predictors. Consequently, it appears that the influence of each of these predictors increases even further if the other predictors have the correct value. From a linguistic perspective, this interplay of the three predictors could be characterized as a type of conventionalization of a form-function association. In particular, there is a strong conventionalization of the specific function of challenge the content of a previous utterance for the sequence *¿Pero qué dices?* 'But what are you saying?'.[5]

Lastly, the effect of the variable REPORTEDSPEECH suggests that the interlocutors use *¿Pero qué dices?* to present a given reported situation as conflictual. We assume that this metalinguistic use of the construction depends on its conventionalization. Given that reported speech contexts do not take place in the here-and-now, the hearer of the reported interrogative is less likely to be able to have an intuition about why the reported situation is particularly conflictual. In other words, the discourse function of the interrogative cannot be easily derived from context, which is why we consider the fact that *pero*-prefaced *qué* + DECIR interrogatives are frequently used in reported speech contexts an indicator for a strong form-function association in the construction, i.e. its conventionalization.

# 5 The diachronic development of *pero qué* + DECIR

Our analysis of spoken Present Day Spanish suggested that *pero qué* + DECIR constructions are undergoing a process of conventionalization, i.e. a strengthening of the form-function association in specific contexts. Specifically, the use of *pero* is more likely in second person singular interrogatives in present tense because interactional challenge readings mostly arise in contexts in which a speaker directly addresses a co-present interlocutor. Likewise, the fact that the use of *pero* is more likely in reported speech indicates that the speakers exploit this form-function correlation in order to portray the narrated situation as conflictual. Like other entrenchment processes, conventionalization processes are historical processes. If a certain construction is frequently employed to express a specific discourse function, hearers of the construction will increasingly take exactly this discourse function to be the conventionalized meaning of the construction.[6] We would therefore expect that the construction diachronically has expanded into the contexts associated with the challenge function, namely (a) second person singular present tense contexts and (b) contexts of reported speech, i.e. fictional dialogues.

At this point, however, we encounter a serious methodological obstacle, i.e. the lack of documentation of spoken language data in historical texts. Due to the unavailability of records of spoken language, studies in historical linguistics usually have to use written historical texts under the premise that these texts reflect the spoken language of that time with more or less fidelity. However, written texts are different from spoken interaction in that the reader is situationally detached from the linguistic event. As a result writers only infrequently address their readers directly, since these cannot be expected to reply immediately.

Due to the differences in the communicative situations represented by spoken language and historical texts, we would expect the frequency with which *pero qué*-interrogatives are used to perform interactional challenges to be lower in historical texts. The challenge function arises when the challenged proposition is mutually manifest in a given communicative situation of immediacy. Communicative immediacy is

---

**5** In the framework of usage-based linguistic, this type of conventionalization of the form-function association could be defined as *entrenchment* (Langacker 1987; Croft 2013). Crucially however, our analysis reveals that the form-function association for *pero qué* + DECIR constructions evinced in this paper does not hold true for just one speaker but rather the speech community as a whole. Broccias (2013) defines conventionality as the "entrenchment of a linguistic expression within a speaker community". In the remainder of this paper, we will therefore prefer the term *conventionalization* over *entrenchment*.

**6** As an example, consider Auer's (2016) description of the WIE X IST DAS DENN? 'How X is that?' construction in German, where X is an evaluative adjective. Auer argues that while the construction initially was used in an information-request function, it was reanalyzed as an exclamative construction. This change is mirrored in the conventionalization of the construction; for instance, the construction differs from questioning interrogatives in that the nuclear accent is invariably placed on the demonstrative *das*.

therefore a prerequisite for the challenge function. In the absence of such a situation of communicative immediacy, as in historical texts, *pero qué*-interrogatives are unlikely to be used to express a challenge towards the reader. This is especially so for the *pero qué* + DECIR construction, since the writer does not have access to what the reader says.

In summary, the differences between written and spoken texts seem to suggest that our first prediction of an increase of the relative usage frequency of *pero qué* + DECIR interrogatives in second person singular present tense cannot be reliably be assessed in our data. Because written texts represent different communicative situations, we would expect differences in the way writers use the *pero qué* + DECIR construction. In other words, if conventionalization processes can be identified, they are not necessarily identical to the ones identified in our analysis of spoken Spanish.

Crucially however, we do find direct addresses and consequently, use of second person morphology, in historical texts. In fictional dialogues, writers represent the oral speech of their or a previous time in similar way that speakers represent previous interactions in reported speech. Consequently, it is to be expected that if we do find *pero qué* + DECIR interrogatives in second person singular present tense, it is in fictional dialogues. Indeed, these contexts may be the only contexts in which data from written and spoken language might be relatively comparable. Just like speakers exploit the rhetorical potential of *pero qué* + DECIR in reported speech, writers might exploit it in fictional dialogues. In other words, it is in fictional dialogues where we can assess the degree of conventionalization of *pero qué* + DECIR interrogatives. In the light of these considerations, we can indeed assume an increase in the use of second person singular *pero qué* + DECIR interrogatives to be evidence for a historical process of conventionalization.

We conducted qualitative and quantitative analyses over the 3571 tokens of *pero qué* + DECIR from historical texts between the 18th and 20th century (see Section 3 for information on the data gathering procedure). We describe the results from the qualitative analysis in Section 5.1 and the quantitative analysis in Section 5.2.

## 5.1 The functions of *pero qué* + DECIR in historical texts

In this section, we will show that while written *pero qué* + DECIR has similar rhetorical functions as in spoken language, these rhetorical functions are distributed differently. We will demonstrate that the crucial parameter for these differences is person morphology because person morphology restricts the type of discourse function that an interrogative construction can be used for.

In the examples (8) and (9) we give typical cases of second person *pero qué* + DECIR in historical written texts.[7] In both cases, the construction is used to realize an interactional challenge. The examples also demonstrate the kinds of linguistic clues that the writers use to indicate that *pero qué* + DECIR is not used as an information question in these contexts. In example (9), the speaker uses the address term *Manuel*. By calling the name of the interlocutor although no other interlocutors are present, the speaker signals his surprise over Manuel's previous turn. A second strategy, found in example (9), is punctuation, namely the use of exclamation marks over question marks.

**Example (8)**
(Source: CORDE (Real Academia Española 2016), *El reinado de Witiza*, 1968)

- *… Esto tiene pinta de ser asunto que excede las capacidades de un pobre Jefe de la Guardia Municipal de Tomelloso.*
  'This seems to be a problem that exceeds the abilities of a humble chief of the Municipal Guard of Tomelloso.'
- ***¿Pero qué dices, Manuel?*** *Si tú eres el más grande. Nunca has fallado.*
  'But what are you saying, Manuel? You are the greatest. You have never failed.'

---

**7** Note that we classified polite third person forms of *pero qué* + DECIR such as *pero usted qué dice* 'but PRO.POL.SG what say. PRS.3SG', as second person forms.

**Example (9):**
(Source: CORDE (Real Academia Española 2016), *Los árboles mueren de pie*, 1949)

| Abuela: | *Malo es que lo pienses, pero, por Dios, que no lo sepa él o estás perdida. Siempre se ha dicho que el amor es un poco como esos carritos chinos: uno muy cómodo, sentado dentro, y el otro tirando. Por lo visto, esta vez te ha tocado a ti tirar del carrito.* |
|---|---|
| Grandmother: | 'It is bad that you have this thought, but, by God, he [your husband, OE/MR] cannot know this or you are lost. People have always said love is a little bit like those little Chinese carts: one is seated inside very comfortably, the other has to pull. It looks like this time you are the one who has to pull.' |
| Isabel | *¡Y qué importa si es mío lo que va dentro! Ojalá fuera más pesada la carga y más duro el camino para merecerlo mejor a su llegada.* |
| | 'And what does it matter whether the one inside is really mine! I wished that the load was heavier and the way more difficult so that I would deserve him more when we arrive.' |
| Abuela: | ***¡Pero qué estás diciendo!*** *Hablas de tu marido como si no fuera tuyo; como si tuvieras que ganártelo aún.* |
| Grandmother: | 'But what are you saying! You are speaking of your husband as if he was not yours; as if you had to still win him over.' |

The use of first person *pero qué* + DECIR is much more common in the historical texts than in the spoken data. The construction is almost always used with a specific type of challenge function. In particular, *pero qué* + DECIR is used in monological speech situations in which the narrator evaluates and corrects a previous thought process. Consider example (10) below.

**Example (10):**
(Source: CORDE (Real Academia Española 2016), *Voz de la naturaleza*, 1787–1803)

*¡Inicuo! Impostor! Yo te buscaré, te arrancaré esas pérfidas entrañas: no, no podrás huir de mi cólera y venganza, **pero, ¿qué digo?** ¿Acaso su crueldad sincerará mis delitos? ¿Porque él sea un malvado debo yo ser delincuente? [...] No.*
'Wicked man, impostor! I will search for you and rip out these perfidious entrails; no, no, you will not be able to flee from my anger and vengeance, but, what am I saying? Is it so that his cruelty would exculpate my crimes? Do I have to be a delinquent because he is a villain? No.'

The rhetorical strategy employed by the utterer of the interrogative in example (10) is very similar to the one that leads to interactional challenge readings in second person *pero qué* + DECIR interrogatives in the spoken data. In the written texts the narrator exploits the reader's assumption that he knows perfectly well what he has been saying in the previous discourse stretch. In addition, the reader is in his right to assume that the narrator actually knows what he just wrote. By addressing himself with the interrogative, the narrator challenges his own previous statement and suspends progressivity in the text. In other words, he implicates that his previous statement is not sufficient or even unacceptable, and then goes on to deliver a better perspective on the state of affairs. The use of the conjunction *pero* consequently serves to reinforce the contrast between the old, unacceptable, statement, and the new statement. This serves to show that the complex rhetorical strategy underlying cases such as example (10) is based on the same inferential mechanisms we found in spoken Present Day Spanish.

When *pero qué* + DECIR is used with third person morphology, the most common use appears to be that of an information question. Consider example (11), a fictional dialogue between two speakers. When the speaker of the first turn tries to discourage the protagonist from worrying too much about the opinion of

other people, the protagonist uses the *pero qué* + DECIR construction to prompt the interlocutor to reveal information about the exact nature of this opinion. The use of *pero* seems to be motivated by the fact that the protagonist cannot accept this proposal without knowing the opinion. In other words, the speaker uses *pero* to insist on the relevance of the asked-for information for the evaluation of the proposal.

**Example (11)**
(Source: CORDE (Real Academia Española 2016), *La sombra*, 1870)

> – *Quiero decir que te cuidas demasiado de la opinión de las gentes, cosa que se debe despreciar las más de las veces, sobre todo cuando [...] no se funda en nada positivo, sino en esas presunciones vulgares, hijas de una gran decadencia moral.*
> 'I want to tell you that you worry too much about the opinion of other people, an attitude that should usually be disdained, especially when it is not based on something positive but rather on those vulgar presumptions, the daughters of great moral decadence'
> – ***Pero ¿qué dice la opinión de las gentes?*** *–pregunté yo–*
> 'But what is [lit. says] the opinion of the people?' – I asked.

The use of third person *pero qué* + DECIR in interactional challenge function (example 12) is infrequent in our data. In cases such as *¡pero qué dirá el padre!* 'but what will father say!' the challenge is not realized towards a third person, i.e. the father. Rather, the third person is used as a moral authority that is assumed to support the criticism against an action realized by the interlocutor. Such cases seem to reflect an advanced state in the conventionalization of *pero qué* + DECIR, in that the challenge function arises even though the addressee of the interrogative is not identical to its subject referent.

**Example (12):**
(Source: CORDE (Real Academia Española 2016), *Traducción de Las mil y una noches*, c. 1916)

> *Cuando la princesa hubo oído el relato de aquella historia de Maruf, se echó a reír de tal manera, que se cayó de trasero. [...] Y ella le dijo: "En verdad ¡oh Maruf! que estás ducho en astucias, y nadie puede compararse a ti en listeza, en sagacidad, en delicadeza y en buen humor. **Pero qué dirá mi padre, y sobre todo qué dirá su visir,** enemigo tuyo, si llegan a saber la verdad de tu historia y la invención de la caravana?*
> 'When the princess had heard Maruf's story, she started laughing so hard that she fell on her behind. [...] And she told him: "Really, oh Maruf, you are an expert in clever stories and no one compares to you regarding cleverness, sagacity, tact and good humor. But what will my father say, and especially, what will his vizier, your enemy, say, when they learn the truth about your story and how the caravan was invented?'

Our brief qualitative analysis of *pero qué* + DECIR interrogatives in written Spanish texts has demonstrated that their range of discourse functions crucially depends on who is addressed by the interrogative. As in spoken interaction, challenge readings arise in our written data when the writer or fictional speaker uses an interrogative even though the information that s/he seems to be asking for is manifest to both him or herself and the fictional interlocutor viz. reader. With verbs of saying and in present tense, this is almost always the case in first person interrogatives (the writer seems to be asking for something that s/he has just said/wrote). Likewise, interlocutors will usually assume that what they have said has been heard, which is why challenge readings can easily arise in second person contexts. In contrast, challenge readings are less frequent in third person contexts because the subject referent of the verb DECIR is typically absent in the situation and it is impossible to challenge a person in his/her absence.[8]

---

**8** Indeed, we observe that challenge readings in third person *pero qué* + DECIR seem to be more frequent not in present indicative tenses, but rather in future tense or conditionals because the rhetorical strategy involves threatening the interlocutor with a possible negative reaction of the moral authority to his/her actions, as in example (12).

## 5.2 Quantitative analysis

Our claim that *pero* is used to reinforce challenge readings in written historical texts makes the following predictions based on the observations in Section 5.1:

(a)  The probability of prefacing *pero* is higher with first and second person morphology in comparison to third person morphology.
(b)  The probability of prefacing *pero* is higher when the narrator or speaker uses an address term.
(c)  The probability of prefacing *pero* is higher when the writer uses exclamation marks or full stops in comparison to question marks.

An additional prediction concerns the historical development of *pero qué* + DECIR interrogatives. In particular, we would expect an increase in the strength of the form-function association, i.e. conventionalization, for challenging *pero qué* + DECIR interrogatives over time. Given that we expect the best predictor of the challenge reading to be person morphology, this assumption leads to the following additional prediction:

(d)  The probability of prefacing *pero* increases over time at a faster rate for first and second person morphology than for third person morphology.

As is evident in the nature of the predictions we have formulated in (a–d), we adopt an indirect approach to the description of the changes in *pero qué* + DECIR interrogatives, in which we use contextual variables (e.g., person morphology) as indicators of the discourse function of the tokens in our corpus. They reason for this procedure is that we believe that it is problematic to impose interpretations on historical data that are based on intuitions by speakers of Modern Spanish (see, e.g., Rosemeyer 2016a). In our approach, historical changes in the data are measured by interaction effects between a variable measuring time and the other contextual predictors (an approach detailed in Rosemeyer 2016b). If such an interaction reaches statistical significance, this means that the distribution of *pero qué* + DECIR changed significantly over time, which in turns allows an interpretation in terms of the changing discourse functions of *pero qué* + DECIR.

### 5.2.1 Analytical procedure and model selection

In order to evaluate the predictions (a–d), we conducted a mixed-effects logistic regression model over our corpus of *(pero) qué* + DECIR interrogatives in written historical texts. As in the regression analysis for the spoken data, the dependent variable was PERO, referring to whether we find prefacing with *pero*. The included predictor variables – listed in Table 3 below – derived from predictions (a–d). We also included the random effect variable GENRE, thereby controlling for variation due to differences in text genre.

**Table 3.** Predictor variables in the mixed-effects logistic regression model on the corpus of *(pero) qué* + DECIR interrogatives in written historical texts

| Variable name | Type | Levels | Description |
|---|---|---|---|
| YEAR.MEAN.S | numerical | – | Year of publication, centered and z–standardized |
| PERSON | factorial | 1 | Person morphology on the verb |
|  |  | 2 |  |
|  |  | 3 |  |
| QUESTIONMARK | binary | True | Variable referring to punctuation: whether a question mark is |
|  |  | False | used (TRUE) or a full stop / exclamation mark (FALSE) |
| SUBJECT | factorial | None | Whether there is no subject pronoun or address term used |
|  |  | PreposedSubj | (None), there is a subject pronoun expressed (PreposedSubj) or |
|  |  | Address | an address term used (Address) |

In a first step, we included all of the four variables and all interactions between these variables. In order to arrive at the best model for our data, i.e. the model that makes the best predictions while simultaneously being most parsimonious, we eliminated these interaction effects in a stepwise fashion by using the function *pdredge()* of the *MuMIn* package (Bartón 2015). We then selected the regression model with the lowest AIC as the final regression model. We give the formula of this regression model in (11). Note that with a c index of concordance of 0.67 and a somers' dxy of 0.33, the model does not reach a very high degree of fit to the data. This problem seems to be a result of the high degree of inter-author variation. By including a random effect variable AUTHOR, referring to the authors of the texts, the c index of concordance rises to 0.79 and the somers' dxy to 0.59, a much more acceptable result. However, although the results of the analysis remain stable, the model with the AUTHOR random effect does not converge, probably because of the fact that the variables YEAR.MEAN.S and AUTHOR are strongly correlated (many data points in YEAR.MEAN.S refer to one text written by one author). While it is possible that we could improve the model fit by including other predictor variables, it appears that the best possible model with the selected variables is the model described by the formula in (13).

(13)  PERO ~ PERSON + SUBJECT + PERSON + QUESTIONMARK + YEAR.MEAN.S +
      PERSON : QUESTIONMARK + PERSON : YEAR.MEAN.S + (1| GENRE)

### 5.2.2 Results

All of the included predictor variables turned out to have statistically significant main effects on the probability of use of *pero*. Of the interaction effects, only the interactions between PERSON and QUESTIONMARK, as well as PERSON and YEAR.MEAN.S were statistically significant. In the following, we will illustrate these effects by using effect plots. The full results of the regression model are available in the appendix (Table 6).

Figure 2 illustrates the joint effect of the variables PERSON and QUESTIONMARK on the dependent variable PERO. The plot illustrates that the probability of using a prefaced *pero* is highest with first person morphology, followed by second and, in some distance, third person morphology. However, these differences are moderated by punctuation. In *qué* + DECIR interrogatives in which a question mark is used, there is no significant difference in the probability of use of *pero* by person morphology (see the red dotted line). The differences in the probability of use of *pero* by person morphology are basically restricted to *qué* + DECIR interrogatives in which a full stop or an exclamation mark is used (see the black line). In these contexts, the probability of use of *pero* is less than 2% in third person morphology, about 8.5% with first person morphology and 16% with second person morphology.

Figure 3 illustrates the effect of the variable SUBJECT on the probability of use of *pero*. While there is no statistically significant difference between interrogatives with or without subject pronoun ('None' – 'PreposedSubj'), the use of *pero* is significantly more likely in interrogatives in which an address term is used than in interrogatives without an address term.

Lastly, although (a) the model does not find a significant effect of the variable YEAR.MEAN.S as a main effect and (b) the interaction effects QUESTIONMARK:YEAR.MEAN.S and SUBJECT:YEAR.MEAN.S do not reach statistical significance, the interaction effect between PERSON and YEAR.MEAN.S does reach statistical significance. This finding is relevant in that it suggests a historical change in the distribution of the use of *pero*. Figure 4 illustrates this change. The probability of use of *pero* with third person morphology (black line) increases moderately over time. In contrast, there is a strong increase in the probability of use of *pero* in second person contexts (green dotted line), and a significant decrease in the probability of use of *pero* in first person contexts (red dotted line). Note that because we are comparing relative frequencies (the use of *qué* + *decir* against *pero qué* + *decir*) and we additionally control for GENRE by using the random effect, these changes cannot be attributed to shifts in the distribution of the texts in the various genres.
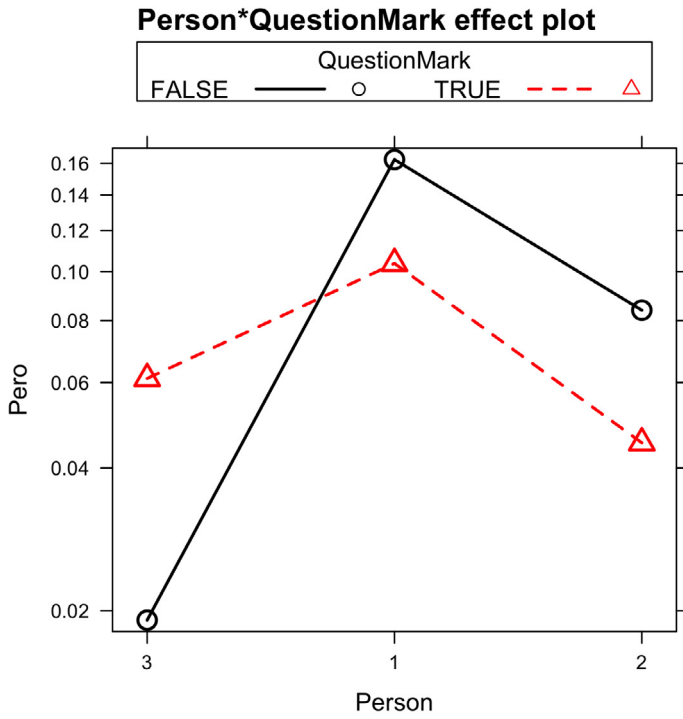
## Person*QuestionMark effect plot



**Fig. 2.** Plot of the interaction effect of PERSON and QUESTION on the dependent variable PERO
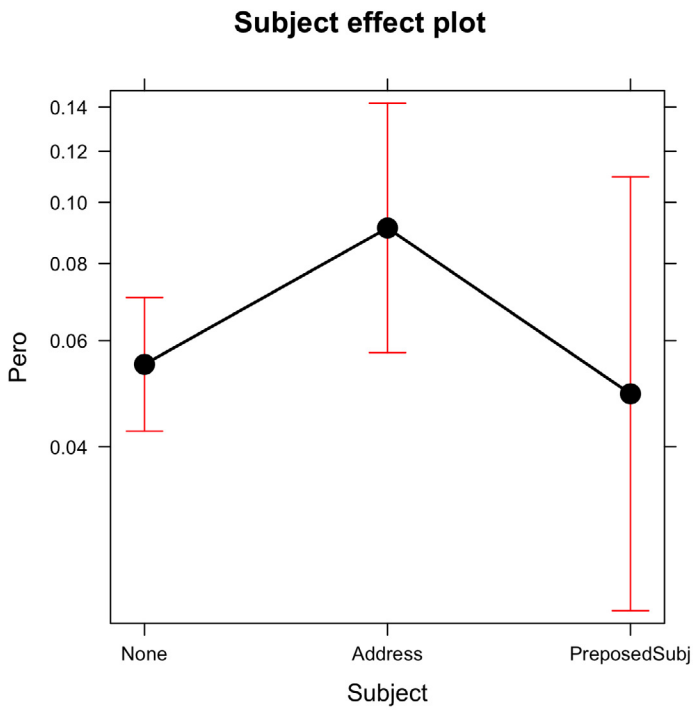
## Subject effect plot



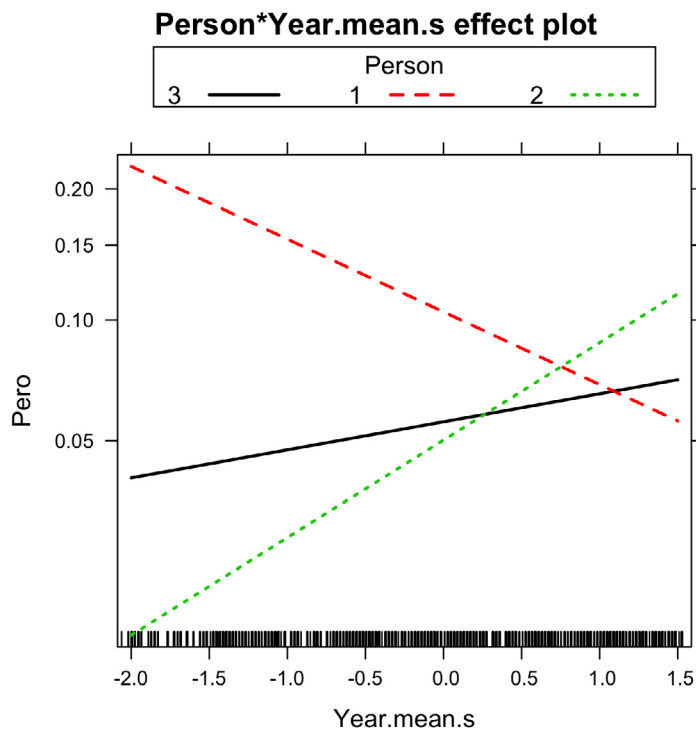**Fig. 3.** Plot of the effect of SUBJECT on the dependent variable PERO

**Fig. 4.** Plot of the interaction effect of Person and Year.mean.s on the dependent variable Pero

## 5.3 Discussion of the results

Our results suggest that as in spoken Present Day Spanish, the authors of historical written Spanish texts use *pero*-prefaced *qué* + decir interrogatives in order to either insist on the relevance of the interrogative or to challenge a previous utterance. The inferential statistical analysis suggests that the last of these two functions is particularly relevant for the distinction between simple *qué* + decir and *pero qué* + decir interrogatives. Authors are significantly more likely to use *pero*-prefaced *qué* + decir than simple *qué* + decir with first and second person morphology. This correlation is even stronger when a full stop or an exclamation mark is used instead of a question mark. In addition, the analysis demonstrated that authors are significantly more likely to use *pero*-prefaced *qué* + decir interrogatives when an address term is used. This means that authors do not use *pero* and address terms as alternatives but rather cumulate those cues to index the challenge function.

The statistical analysis also showed that this distribution underwent historical changes. While, as predicted, there is a significant increase in the probability of use of *pero* with second person morphology, there is also a significant decrease in its use with first person morphology. In contrast, its probability of use with third person morphology remains relatively constant. Since this finding only partially corresponds to our prediction (d), these diverging developments are worth considering in greater detail. Crucially, we have to interpret them with respect to more general diachronic change, namely the fact that there was an overall decrease of the usage frequency of first and second person morphology in *qué* + decir interrogatives, accompanied by a rise in the usage frequency of third person *qué* + decir interrogatives. We illustrate this change in Fig. 5 below.
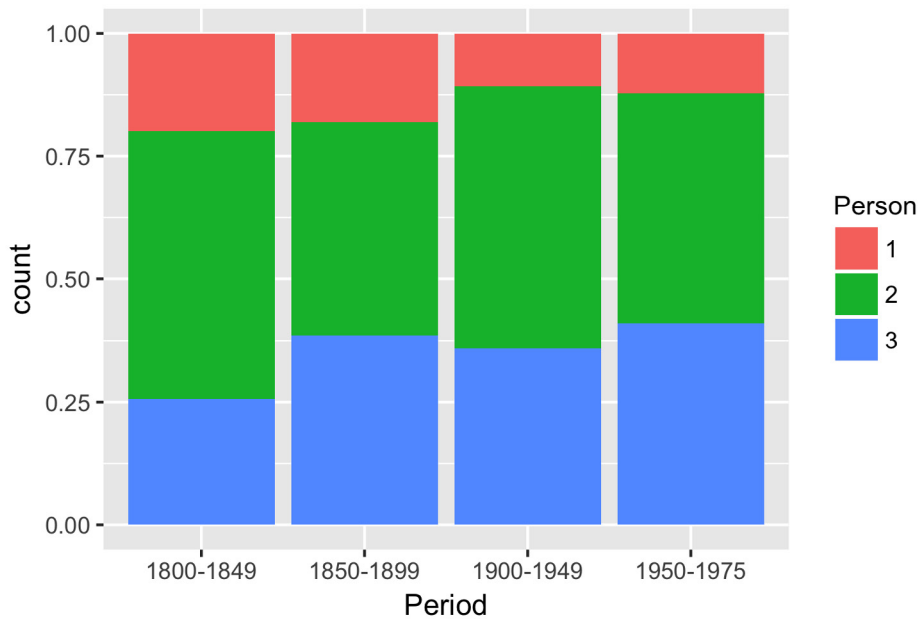
**Fig. 5.** Development of the relative frequencies of person morphology in the (*pero*) *qué* + DECIR data over time

The decline of the challenge function of first person *pero qué* + DECIR might be due to the simple fact that authors increasingly dislike to use first person *qué* + DECIR in the texts. This development also affects the use of *pero qué*-interrogatives: writers disprefer to challenge themselves with interrogatives. In other words, there is a decline of the rhetorical strategy of self-correction dependent on a general trend in writing style.

In contrast, the increase of the challenge function of second person *pero qué* + DECIR cannot be attributed to a similar process. Rather, it can be described as an instance of a specialization process. While we also observe a general decrease of the use of second person morphology in writing, there is an increase in the strength of association of *pero* with second person morphology. This specialization of *pero qué* + DECIR as a marker of interactional challenge in second person contexts might reflect an increasing conventionalization of this construction in the spoken language at the time. The crucial premise for this assumption is the idea of a relative equivalence of reported speech in spoken language and fictional discourse in written texts. While the communicative setup of primary spoken language and written texts is radically different, reported speech is similar to fictional discourse in that in both modalities, a narrator reports spoken language. Although there are obviously differences in the degree of elaboration of reported speech and fictional dialogue, these differences do not influence the fact that the two modalities are comparable with respect to the parameter of who is talking.

In our corpus of written texts the use of second person *pero qué* + DECIR interrogatives is almost exclusively restricted to fictional dialogues. Consequently, the fact that we find an increasing form-function correlation between the use of prefaced *pero* and the challenge function for second person *qué* + DECIR interrogatives suggests that in written texts, the conventionalization of *pero qué* + DECIR took place mostly in fictional dialogues. We can thus assume that like the speakers in our corpus of spoken Spanish, authors used *pero qué* + DECIR interrogatives in fictional dialogues to support their representation of a situation as particularly conflictual. This exploitation of the rhetorical potential of the *pero qué* + DECIR construction was only possible due to the conventionalization of second person *pero qué* + DECIR in the spoken language of that time. We therefore assume that the progressive specialization of second person *pero qué* + DECIR in our written data reflects an ongoing conventionalization process in the spoken language of that time, confirming prediction (d).

# 6 Summary and discussion

In a first step, our analysis has demonstrated that in spoken Spanish, prefaced *pero* 'but' has a predictable influence on the discourse function of *qué*-interrogatives. When added to an interrogative, the concessive semantics of *pero* is no longer interpreted on the semantic or ideational level, but rather taken to concern the level of organization of discourse. Specifically, while *pero* is used to acknowledge a previous utterance by the interlocutor, it signals at the same time that this utterance is insufficient in one of various ways; as a second request, the *pero qué* interrogative signals that the previous utterance was not a sufficient answer to the first request. As a first request, the interrogative either upgrades the importance of the request or challenges a previous utterance.

In a second step, we used variationist methodology to demonstrate that there are quantitative reflexes of the specific function of *pero qué* interrogatives. Our comparison of *qué* + DECIR and *pero qué* + DECIR interrogatives has shown that the likelihood of prefacing *pero* increases in reported speech contexts and in first person singular present tense. Both of these results suggest a conventionalization of the challenge function in *pero qué* interrogatives.

We tested this hypothesis of a conventionalization of *pero qué* + DECIR interrogatives in the diachronic analysis of written texts. Although we found different conventionalization processes – particularly the conventionalization of first person singular *pero qué* + DECIR interrogatives – the challenge function of these uses of the construction arises by the same inferential mechanism as in second person singular *pero qué* + DECIR. Our quantitative analysis demonstrated that in the historical texts, the use of *pero* is (a) more likely in first and second person *qué*-interrogatives, especially when a punctuation mark other than a question mark is used, and (b) more likely with an address term. These results suggest that *pero qué* + DECIR is more likely to have a challenge function in first and second person than in third person contexts. Crucially, our analysis also revealed a significant increase in the likelihood of use of second person *pero qué* + DECIR over time. We argued that this increase can be interpreted as a reflex of the conventionalization of *pero* as a marker of interactional challenges in interrogatives in the language of that time; as in our spoken data, writers used *pero qué dices* in order to mark a narrated situation as particularly conflictual, a rhetorical strategy that is only viable when *pero* is already typically associated with the challenge function.

At this point we want to emphasize that the quantitative diachronic study in Section 5 has focused only one the conventionalization of the *pero qué*-construction with one verb, namely *decir* ,say'. We argued that verbs with a situational meaning referring to the current communicative situation are more likely to conventionalize the challenge function than other verbs. In fact, we have seen in section 4.2 that *pero qué*-interrogatives are typically used with situational verbs, namely *decir* 'to say', *hacer* 'to do/make' and *pasar* 'to happen'. We assume that a similar diachronic process of conventionalization could be traced for constructions with those verbs too, as *pero qué estas haciendo* 'what are you doing' and *pero qué (te) pasa* 'but what is happening' seem to have conventionalized a challenge function, as well. We need further studies to analyze the development of those constructions and relate their development to the more schematic *pero qué*-construction. Another promising perspective seems to be the inclusion of other, non-situational verbs in the construction. While we assume these verbs not to have conventionalized challenge readings, there are interesting exceptions, such as the frequent use of *(pero) qué sé yo*, '(but) what do I know', which like the cases of *pero que digo* 'but what am I saying' in our corpus of historical texts seems to have conventionalized first person usage.

To conclude, we believe that our paper makes three contributions to the study of inferences in interaction and language change. First, it gives an explicit account of the inferences raised by the use of prefaced *pero* in *qué*-interrogatives and explains how the discourse functions of *pero qué*-interrogatives arise on the basis of these inferences. While we regard the basic functions of both *pero* and *qu*é-interrogatives as invariant, these functions change according to the discourse context in which the *pero qué*-interrogative is used. Consequently, the *pero qué*-interrogative can be used as an information request (in second or first position) or even as an interactional challenge. This gradual change conforms nicely to models of *actualization* in historical linguistics (cf. De Smet 2012), in that a new discourse function is created by small steps of reanalysis in slightly changed contexts (e.g., unacceptable response vs. missing response).

Second, our paper demonstrates how to successfully combine methodologies from interactional and variationist linguistics. In particular, it has shown that by using statistical tools such as classification and regression trees and mixed-effects logistic regression modeling, it is possible to find quantitative reflexes of the rhetorical strategies evinced by our in-depth qualitative analysis. This finding suggests that these rhetorical strategies are indeed pervasive in our populations of Spanish speakers and writers, and not idiosyncratic.

This leads us to the third important point, namely the fact that such rhetorical strategies can be conventionalized in a speaker community over time. By complementing the qualitative and quantitative analysis of spoken Spanish with an analysis of the alternation between *qué* + DECIR and *pero qué* + DECIR interrogatives in historical texts, we have demonstrated that it is possible to relate interactional analyses of synchronic spoken data with variationist analyses over written historical texts. Our analysis has not only shown that the conventionalization of *pero qué*-interrogatives in the spoken data might be the result of a historical process, but also suggests that the one usage context in which spoken synchronic data and written diachronic data turn out to be relatively similar is represented speech (be it reported speech in spoken language or fictional discourse in written texts). Both speakers and writers exploit the rhetorical potential of particularly conventionalized constructions in order to portray a narrated situation as being conflictual and thus also to increase the expressivity of their narrations of spoken language. This finding strengthens the common assumption in historical linguistics that the existence of a given conventionalized form-function association in fictional discourse is an important indicator of its conventionalization in the spoken language at that time. We consequently propose that the frequent use of a construction with a specific discourse function in either reported speech in spoken discourse or in written fictional discourse is a good indicator of its conventionalization in spoken language.

# References

Anscombre, Jean-Claude & Oswald Ducrot. 1977. Deux *mais* en français? *Lingua* 43(1). 23–40.

Auer, Peter. 1996. The pre-front field in spoken German and its relevance as a grammaticalization position. *Pragmatics* 6(3). 295– 322.

Auer, Peter. 2016. 'Wie geil ist das denn?' Eine neue Konstruktion im Netzwerk ihrer Nachbarn. *Zeitschrift für Germanistische Linguistik* 44(1). 69–92.

Barth-Weingarten, Dagmar. 2003. *Concession in spoken English. On the realisation of a discourse-pragmatic relation*. Tübingen: Narr.

Bartón, Kamil. 2015. *MuMIn: Model selection and model averaging based on in- formation criteria (AICc and alike)*. Available online at https://cran.r-project.org/web/ packages/MuMIn/index.html. Last access 3 May 2017.

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai & Gabor Grothendiekt. 2015. *lme4: Linear Mixed-Effects Models using 'Eigen' and S4*. Available online at https://cran.r-project.org/web/packages/lme4/index.html. Last access 23 March 2017.

Bell, David M. 1998. Cancellative discourse markers: a core/periphery approach. *Pragmatics* 8. 515–541.

Bell, David M. 2010. *Nevertheless*, *still* and *yet*: Concessive cancellative discourse markers. *Journal of Pragmatics* 42. 1912–1927.

Blakemore, Diane. 1989. Denial and contrast: A relevance theoretic analysis of *but*. *Linguistics and Philosophy* 12(1). 15–37.

Briz, Antonio. 1993. Los conectores pragmáticos en español coloquial (I): su papel argumentativo. *Contextos* XI/21-22. 145-188.

Broccias, Cristiano. 2013. Cognitive Grammar. In Graeme Trousdale & Thomas Hoffman (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Canavan, Alexandra & George Zipperlen. 1996. *CALLHOME Spanish Speech*. Philadelphia: Linguistic Data Consortium.

Clayman, Steven E. 2010. Address terms in the service of other actions: the case of news interview talk. *Discourse & Communication* 4. 161-183.

Clayman, Steven E. 2012. Address terms in the organization of turns at talk: the case of pivotal turn extensions. *Journal of Pragmatics* 4. 1853-1867.

Couper-Kuhlen, Elizabeth & Dagmar Barth-Weingarten. 2011. A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2. *Gesprächs-forschung – Online-Zeitschrift zur verbalen Interaktion* 12. 1–51.

Couper-Kuhlen, Elizabeth & Sandra A. Thompson. 2000. Concessive patterns in conversation. In Elizabeth Couper-Kuhlen & Bernd Kortmann (eds.), *Cause – Condition – Concession – Contrast. Cognitive and Discourse Perspectives*, 33, 381–410. Berlin: de Gruyter.

Cresti, Emanuela & Massimo Moneglia (eds.). 2005. *C-ORAL-ROM. Integrated reference corpora for spoken Romance languages*. Amsterdam: Benjamins.

Croft, William. 2013. Radical Construction Grammar. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 211–232. Oxford: Oxford University Press.

Davies, Mark. 2015-2016. *Corpus del español, web/dialects*. Available at http://www.corpusdelespanol.org/. Last access 8 March 2017.

De Smet, Hendrik. 2012. The course of actualization. *Language* 88(3). 601-633.

Ehmer, Oliver. 2010. *cespla – Corpus de Conversaciones ESpontáneas PLAtenses*. http://www.cespla.de. Last access 5 January 2018.

Ehmer, Oliver. 2011. *Imagination und Animation. Die Herstellung mentaler Räume durch animierte Rede*. Berlin/New York: de Gruyter.

Gabriel, Christoph. 2011. Hamburg Corpus of Argentinean Spanish (HaCASpa). https://corpora.uni-hamburg.de/hzsk/de/islandora/object/spoken-corpus:hacaspa. Last access 5 January 2018.

Günthner, Susanne. 2005. Narrative reconstructions of past experiences. Adjustments and modifications in the process of recontextualizing past experience. In Uta M. Quasthoff & Tabea Becker (eds.), *Narrative Interaction*, 285–301. Amsterdam: John Benjamins Publishing Group.

Hansen, Maj-Britt Mosegaard. 1998a. Discourse Markers. *Lingua* 104. 235–260.

Hansen, Maj-Britt Mosegaard. 1998b. *The function of discourse particles. A study with special reference to spoken French*. Amsterdam: Benjamins.

Hayano, Kaoru. 2012. Question Design in Conversation. In Tanya Stivers & Jack Sidnell (eds.), *The Handbook of Conversation Analysis*, 395–414. Malden: Wiley-Blackwell.

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*, 299–345. Cambridge: Cambridge University Press.

Heritage, John. 2007. Intersubjectivity and progressivity in person (and place) reference. In Nick J. Enfield & Tanya Stivers (eds.), *Person Reference in Interaction: Linguistic, Cultural and Social Perspectives*, 255-280. Cambridge: Cambridge University Press.

Heritage, John. 2013. Turn-initial position and some of its occupants. *Journal of Pragmatics* 57. 331-337.

Heritage, John & Marja-Leena Sorjonen. 1994. Constituting and Maintaining Activities across Sequences: And-Prefacing as a Feature of Question Design. *Language in Society* 23(1). 1-29.

Hothorn, Torsten, Kurt Hornik, Carolin Strobl & Achim Zeileis. 2015. *party: A Laboratory for Recursive Partytioning*. Available online at http://cran.r-project.org/ web/packages/party/index.html. Last access 23 March 2017.

Iten, Corinne. 2005. *Linguistic Meaning, Truth Conditions and Relevance. The Case of Concessives*. New York. Palgrave.

Jol, Guusje & Fleur van der Houwen. 2014. Police interviews with child witnesses: pursuing a response with maar (= Dutch but )- prefaced questions. *International Journal of Speech, Language & the Law* 21(1).

Jørgensen, Annette Myre, Esperanza Eguía Padilla, Anna-Brita Stenström, Juan Antonio Martínez López, Eli Marie Drange Danbolt, Mariano Reyes Tejedor, Anna Acevedo, Giovanna Angela Mura, Stine Huseby, Lise Holmvik, Solfrid Hernes, Evert Jakobsen, Kristine Eide & Marie Espeland. proyecto COLA. Corpus Oral de Languaje Adolescente. http://www.colam.org/. Last access 5 January 2018.

Küttner, Uwe. This SI. Investigating inferences in sequences of action: The case of claiming "just-now" recollection with *oh that's right*.

Küttner, Uwe-A. 2016. *That*-initial turns in English conversation. Doctoral dissertation. In. Potsdam, Germany: University of Potsdam.

Lakoff, Robin. 1971. If's, and's and but's about conjunction. In Charles J. Fillmore & Donald Terence Langendoen (eds.), *Studies in Linguistic Semantics*, 114–149. Irvington: New York, Holt, Rinehart & Winston.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Volume 1. Theoretical prerequisites*. Stanford: Stanford University Press.

Li, Xiaoting. 2016. Some discourse-interactional uses of *yinwei* 'because' and its multimodal production in Mandarin conversation. *Language Sciences* 58(Special issue: "Adverbial patterns in interaction"). 51-78.

Mazeland, Harrie & Mike Huiskes. 2001. Dutch 'but' as a sequential conjunction: Its use as a resumption marker. In Margret Selting & Elizabeth Couper-Kuhlen (eds.), *Studies in Interactional Linguistics*, 141–169. Amsterdam: Benjamins.

Nemo, François. 2002. *But* (and *mais)* as morpheme(s). *Delta* 18(2).

Real Academia Española. 2016. *Banco de datos (CORDE) [en línea]. Corpus diacrónico del español*. Available online at http://www.rae.es. Last access 20 April 2012.

Rosemeyer, Malte. 2016a. The development of iterative verbal periphrases in Romance. *Linguistics* 54(2). 235-272.

Rosemeyer, Malte. 2016b. Modeling frequency effects in language change. In Heike Behrens & Stefan Pfänder (eds.), *Experience Counts: Frequency Effects in Language*, 175-207. Berlin, New York: De Gruyter.

Schegloff, Emanuel A. 1979. Identification and Recognition in Telephone Conversation Openings in Everyday Language. In George Psathas (ed.), *Studies in Ethnomethodology*. New York: Irvington.

Schegloff, Emanuel A. & Harvey Sacks. 1973. Opening up closings. *Semiotica* 8. 289–327.

Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.

Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock & Susanne Uhmann. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10. 353–402.

Sperber, Dan & Wilson, Deirdre. 1996 [1986]. Relevance. Communication and Cognition, Oxford: Blackwell.

Stivers, Tanya & Nick J. Enfield. 2010. A coding scheme for question–response sequences in conversation. *Journal of Pragmatics* 42(10).

Stivers, Tanya & Jeffrey Robinson. 2006. A preference for progressivity in interaction. *Language in Society* 35(3). 367-392.

Stivers, Tanya & Federico Rossano. 2010. Mobilizing response. *Research on Language and Social Interaction* 43(1). 3-31.

Streeck, Jürgen & Ulrike Hartge. 1992. Previews: gestures at the transition place. In Peter Auer & Aldo Di Luzio (eds.), *The Contextualization of Language*, 135-157. Amsterdam: Benjamins.

Tagliamonte, Sali & Harald Baayen. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135-178.

Torreira, Francisco & Mirjam Ernestus. 2010. The Nijmegen Corpus of Casual Spanish. *Proceedings of LREC 2010*. http://www.lrec-conf.org/proceedings/lrec2010/. Last access 5 January 2018.

# Appendix

**Summary of the most important GAT 2 transcription conventions**
cf. Couper-Kuhlen and Barth-Weingarten (2011) and Selting et al. (2009)

Sequential structure

| | |
|---|---|
| `[  ]`<br>`[  ]` | overlap and simultaneous talk |
| `=` | fast, immediate continuation with a new turn or segment (latching) |

In- and outbreaths

| | |
|---|---|
| `°h / h°` | in- / outbreaths of appr. 0.2-0.5 sec. duration |
| `°hh / hh°` | in- / outbreaths of appr. 0.5-0.8 sec. duration |
| `°hhh / hhh°` | in- / outbreaths of appr. 0.8-1.0 sec. duration |

Pauses

| | |
|---|---|
| `(.)` | micro pause, estimated, up to 0.2 sec. duration appr. |
| `(-)` | short estimated pause of appr. 0.2-0.5 sec. duration |
| `(--)` | intermediary estimated pause of appr. 0.5-0.8 sec. duration |
| `(---)` | longer estimated pause of appr. 0.8-1.0 sec. duration |
| `(0.5) / (2.0)` | measured pause of appr. 0.5 / 2.0 sec. duration (to tenth of a second) |

Other segmental conventions

| | |
|---|---|
| `:` | lengthening, by about 0.2-0.5 sec. |
| `::` | lengthening, by about 0.5-0.8 sec. |
| `:::` | lengthening, by about 0.8-1.0 sec. |
| `ʔ` | cut-off by glottal closure |
| `and_uh` | cliticizations within units |
| `uh, uhm, etc.` | hesitation markers, so-called "filled pauses" |

Laughter and crying

| | |
|---|---|
| `haha, hehe, hihi` | syllabic laughter |
| `((laughs)), ((cries))` | description of laughter and crying |
| `<<laughing> >` | laughter particles accompanying speech with indication of scope |
| `<<:-)> so>` | smile voice |

Continuers

| | |
|---|---|
| `hm, yes, no, yeah` | monosyllabic tokens |
| `hm_hm, ye_es, no_o` | bi-syllabic tokens |
| `ʔhmʔhm` | with glottal closure, often negating |

Accentuation

| | |
|---|---|
| `SYLlable` | focus accent |
| `sYllable` | secondary accent |
| `!SYL!lable` | extra strong accent |

Final pitch movements of intonation phrases

| | |
|---|---|
| ? | rising to high |
| , | rising to mid |
| – | level |
| ; | falling to mid |
| . | falling to low |

Pitch jumps

| | |
|---|---|
| ↑ | smaller pitch upstep |
| ↓ | smaller pitch downstep |
| ↑↑ | larger pitch upstep |
| ↓↓ | larger pitch downstep |

Changes in pitch register

| | | |
|---|---|---|
| `<<l>` | `>` | lower pitch register |
| `<<h>` | `>` | higher pitch register |

Intralinear notation of accent pitch movements

| | |
|---|---|
| `` `SO `` | falling |
| `´SO` | rising |
| `¯SO` | level |
| `^SO` | rising-falling |
| `ˇSO` | falling-rising |

| | |
|---|---|
| `↑`` | small pitch upstep to the peak of the accented syllable |
| `↓´` | small pitch downstep to the valley of the accented syllable |
| `↑¯SO bzw. ↓¯SO` | pitch jumps to higher or lower level accented syllables |
| `↑↑`SO bzw. ↓↓´SO` | larger pitch upsteps or downsteps to the peak or valley of the accented syllable |

Loudness and tempo changes, with scope

| | | |
|---|---|---|
| `<<f>` | `>` | forte, loud |
| `<<ff>` | `>` | fortissimo, very loud |
| `<<p>` | `>` | piano, soft |

| | | |
|---|---|---|
| `<<pp>` | `>` | pianissimo, very soft |
| `<<all>` | `>` | allegro, fast |
| `<<len>` | `>` | lento, slow |
| `<<cresc>` | `>` | crescendo, increasingly louder |
| `<<dim>` | `>` | diminuendo, increasingly softer |
| `<<acc>` | `>` | accelerando, increasingly faster |
| `<<rall>` | `>` | rallentando, increasingly slower |

Changes in voice quality and articulation, with scope

| | | |
|---|---|---|
| `<<creaky>` | `>` | glottalized |
| `<<whispery>` | `>` | change in voice quality as stated |

Other conventions

| | | |
|---|---|---|
| `<<surprised>` | `>` | interpretive comment with indication of scope |
| `((coughs))` | | non-verbal vocal actions and events |
| `<<coughing>` | `>` | ...with indication of scope |
| `(      )` | | unintelligible passage |
| `(xxx), (xxx xxx)` | | one or two unintelligible syllables |
| `(may i)` | | assumed wording |
| `(may i say/let us say)` | | possible alternatives |
| `((u n i n t e l l i g i b l e, appr. 3 sec))` | | unintelligible passage with indication of duration |
| `((...))` | | omission in transcript |
| `-->` | | refers to a line of transcript relevant in the argument |

**Table 4.** List of predictor variables for the CART over PERO for spoken Spanish

| Variable name | Type | Levels | Description |
|---|---|---|---|
| TENSE | factorial | Present, Future, ImperfectivePast, PerfectivePast, Perfect | Verb tense |
| TENSEPR | binary | True, False | Verb tense = present? |
| SCNDSGPR | binary | True, False | second person singular vs. others |
| PRESENTRELEVANCE | binary | True, False | Present/Perfect vs. others |
| PERSON | factorial | 1, 2, 3 | Person morphology on the verb |
| NUMBER | factorial | sg, pl | Number morphology on the verb |
| PERSONNUMBER | factorial | 1sg, 2sg, 3sg, 1pl, 2pl, 3pl | Person and number morphology on the verb |
| PERSONNUMBERSNDSG | binary | True, False | Whether or not Person/number morphology is 2sg |
| V_PERSON_detail | factorial | él, ellos, nosotros… | Exact type of personal morphology |
| INTERROGATIVEBEFORE | binary | True, False | Interrogative used before or not |
| INTERROGATIVEBEFORE_same | binary | True, False | Same interrogative used before or not |
| INTERROGATIVEBEFORE_pero | binary | True, False | *Pero*-interrogative used before or not |
| INTERROGATIVEBEFORE_factor | factorial | none, different, same | Whether or not an interrogative was used before, and whether or not the same |
| PERIPHRASIS | factorial | Decir, estar decir, haber decir, ir a decir… | Type of periphrasis |
| REPORTEDSPEECH | binary | True, False | Use in reported speech vs. not |
| afterDecirScaled | continuous | – | Number of words after *decir* until utterance end, z-standardized |
| UTTERANCE length | continuous | – | Number of words of the question |
| BETWEENQUEANDDECIRBIN | binary | True, False | Whether or not an element occurs between *qué* and *decir* |
| AUXILIATED | binary | True, False | Whether or not the construction is auxiliated |
| MODALAUXILIARY | binary | True, False | Whether or not the verb is a modal auxiliary |
| ADDRESSTERM | factorial | None, after, before | Presence and position of an address term |
| ADDRESSTERM_before_bin | binary | True, False | Whether or not there is a preposed address term |
| ADDRESSTERM_after_bin | binary | True, False | Whether or not there is a postposed address term |
| SUBJECT | factorial | None, AddressTerm, PreposedSubj | Presence and type of a linguistic element referring to the logical subject |
| GENRE_class | factorial | Non-spontaneous, spontaneous, telephone | Type of interaction |
| GENRE | factorial | Interview, Law, … | Type of interaction (more specific) |

**Table 5.** Full results of the mixed-effects logistic regression model over PERO for spoken Spanish

| Variable | Level | BETA | SE | Z | p |
|---|---|---|---|---|---|
| (Intercept) | | -4.232 | 0.605 | -6.999 | 0.000 |
| SCNDSGPR | False | *Reference level* | | | |
| | True | 1.775 | 0.614 | 2.891 | 0.004 |
| REPORTEDSPEECH | False | *Reference level* | | | |
| | True | 2.226 | 0.362 | 6.153 | 0.000 |
| Model evaluation | C index of concordance = 0.93 Somers' dxy = 0.85 AIC = 338.6 | | | | |

**Table 6.** Full results of the mixed-effects logistic regression model over Pero for Spanish written historical texts

| Variable | Level | Beta | SE | Z | p |
|---|---|---|---|---|---|
| (Intercept) | | -3.965 | 0.728 | -5.446 | 0.000 |
| Person | 3 | *Reference level* | | | |
| | 1 | 2.247 | 0.790 | 2.845 | 0.004 |
| | 2 | 1.584 | 0.823 | 1.925 | 0.054 |
| Year.mean.s | | 0.173 | 0.148 | 1.176 | 0.240 |
| Subject | None | *Reference level* | | | |
| | AddressTerm | 0.547 | 0.241 | 2.268 | 0.023 |
| | PreposedSubj | -0.118 | 0.443 | -0.267 | 0.789 |
| QuestionMark | False | *Reference level* | | | |
| | True | 1.206 | 0.730 | 1.653 | 0.098 |
| Person1 : Year.mean.s | | -0.623 | 0.216 | -2.887 | 0.004 |
| Person2 : Year.mean.s | | 0.430 | 0.200 | 2.148 | 0.032 |
| Person1 : QuestionTRUE | | -1.725 | 0.816 | -2.113 | 0.035 |
| Person2 : QuestionTRUE | | -1.868 | 0.844 | -2.215 | 0.027 |
| Model evaluation | | C index of concordance = 0.67 Somers' dxy = 0.33 AIC = 1323.2 | | | |