RESEARCH

Brazilian Portuguese *in-situ wh*-interrogatives between rhetoric and change

Malte Rosemeyer[1,2]

[1] Albert-Ludwigs-Universität Freiburg, DE

[2] KU Leuven, BE

malte.rosemeyer@gmail.com

Previous studies of the historical development of partial interrogatives have postulated a change from contexts in which the proposition of the interrogative has been explicitly mentioned in the previous discourse, to contexts in which the proposition is discourse-new. The present paper explores whether the historical increase in the usage frequency of Brazilian Portuguese *in-situ wh*-interrogatives represents the same process. Using data from a large corpus of BP theater texts dated between the 19th and 21st century, several discourse functions of InSituWh are identified, the most frequent of which are cataphorical questions, which serve to either open up a question unrelated to the current question under discussion, or raise further questions about the current question under discussion, and rhetorical questions, which question the validity or relevance of a previously mentioned proposition. Rhetorical questions typically do not trigger a response by the interlocutor and are used with psychological verbs and morphologically simple interrogative pronouns. A statistical analysis of the diachronic distribution of InSituWh in the data reveals an increase in the usage frequency of InSituWh especially in contexts in which the proposition is discourse-new. However, the results also indicate that this increase is not due to a grammatical change of InSituWh but rather reflects a consolidation of the rhetorical question function of InSituWh within the genre of theater plays.

Keywords: wh-interrogative; question; in-situ wh; language change; Portuguese; corpus linguistics

1 Introduction

The Portuguese system of partial interrogatives has undergone dramatic changes over time. Whereas until the 19th century, the use of non-clefted *ex-situ wh*-interrogatives (1a) was the norm, the subsequent centuries saw an increase of the usage frequency of clefted *ex-situ* (1b) and *in-situ wh*-interrogatives (1c) (Lopes Rossi 1996). These changes have been documented both for European Portuguese (EP) and Brazilian Portuguese (BP), although Lopes Rossi's (1996) results suggest that the change was implemented to a stronger degree in BP than in EP.

(1)   a.   O que você fez?   [ExSituWh]
                what you do.PST.PFV.3SG
                'What did you do?'

          b.   O que é   que você fez?   [CleftWh]
                what be.PRS.3SG that you do.PST.PFV.3SG
                'What is it you did?'

          c.   Você fez   o quê?   [InSituWh]
                you do.PST.PFV.3SG what
                'You did what?'

Interestingly, similar changes are documented for French, although at different points in time. Elsig (2009) finds an increase of the usage frequency of partial *est-ce que* 'be.PRS.3SG-this that' interrogatives between the 15th and the 17th century. According to Waltereit (2018), this increase went hand in hand with a change regarding to the reference of the pronoun *ce*. *Est-ce que* interrogatives were typically used in Old French when *ce* was anaphoric or deictic, referring to a fact evident in the linguistic or non-linguistic context. Later, speakers started using the construction in contexts in which *ce* no longer had indexical value, indicating a semantic reanalysis persisting into Modern French (cf. also Greive 1974; Kaiser 1980). Regarding InSituWh, Larrivée's (to appear) analysis suggests a slight increase of the usage frequency of French InSituWh between the 20th and 21st century. More importantly, he documents a decrease of the use of InSituWh in contexts involving explicitly mentioned propositions. These results suggest that the path from non-canonical to canonical partial interrogative constructions might be described as a path from contexts in which the proposition is activated to contexts in which it is not necessarily activated.

  The present study tries to determine whether the same type of historical process can be postulated for BP. A typology of the discourse functions of 218 cases of *in-situ wh*-interrogatives in a corpus of almost 300 theater plays dated between 1800 and 2016 is developed and matched to a series of distributional criteria. The two most frequent discourse functions are cataphorical questions, which serve to either open up a "new" question under discussion or raise further questions about an already existing question under discussion, and rhetorical questions, which question the validity or relevance of a previously mentioned proposition. While cataphorical readings can arise both in contexts in which the proposition is discourse-new or discourse-old, rhetorical question readings most typically occur in contexts in which the proposition is discourse-new. This difference is of crucial importance for the interpretation of the historical changes in the distribution of InSituWh attested in the corpus data. Although the analysis reveals an increase in the usage frequency of InSituWh especially in contexts in which the proposition is discourse-new, this result only superficially confirms the hypothesis of a trend paralleling the development of CleftWh and InSituWh in French. A more fine-grained statistical analysis demonstrates that the increase is not due to the fact that BP InSituWh interrogatives have become more information question-like, i.e. serve to advance discourse by asking for a fact unknown to the speaker. Rather, in the specific genre of theater texts, a process of conventionalization of the rhetorical question function has taken place. This genre-internal conventionalization process is not likely to reflect the historical development of InSituWh in spoken BP, which is why the results demonstrate the importance of considering genre-specific tendencies in language use when analyzing language change.

2 The diachrony of BP InSituWh

This section of the paper describes the results from previous studies on the diachrony of BP InSituWh (2.1) and highlights the relevance of the parameters of the degree of activation of the proposition, as well as conditional relevance, for the description of this change (2.2).

2.1 Syntactic change in BP wh-interrogatives

Since the 1990s, a number of studies on the history of Brazilian and European Portuguese *wh*-interrogatives have evinced syntactic change regarding (a) the expression and placement of the subject constituent, (b) the usage frequency of clefting strategies and (c) the preferred position of the *wh* constituent (Duarte 1992; Lopes-Rossi 1993; Lopes Rossi 1996; Kato & Mioto 2005; Kato & Ribeiro 2005; 2009; Fontes 2012a; b; Pinheiro & Marins

2012; Kato 2014; De Paula 2015; 2016; 2017). Regarding (a), several studies demonstrate that over time, null subjects (2a) and postposed subjects (2b) came to be replaced with preposed subjects (2c). For instance, Pinheiro & Marins (2012: 172) find that in a corpus of theater texts dated between the 19th and 20th century, SV word order rose from a relative frequency of 7 to 76 percent, to the detriment of the two other patterns of subject expression.

(2)   a.   O que quer?   [*wh*V]
            what want.PRS.3SG

       b.   O que quer   ela?   [*wh*VS]
            what want.PRS.3SG she

       c.   O que ela quer?   [*wh*SV]
            what she want.PRS.3SG
            'What does she want?'

At the same time, there is a marked increase in the use of clefted *wh*-interrogatives (see 1b) in their data, from 0 to 74 percent. There thus appears to be a relationship between the changes in subject use and clefting. As already observed by Duarte (1992), early tokens of CleftWh are much more likely to display SV word order than non-clefted *wh*-interrogatives, which is why she hypothesized that CleftWh served as a catalyst for the spread of SV from declarative sentences to *wh*-interrogatives.

   This study focuses on the change in (c), i.e. the rise in the usage frequency of InSituWh in Portuguese, documented in several of the cited studies. Lopes Rossi (1996: 68) finds for BP that the relative usage frequency of InSituWh increased from 0 percent in the first half of the 19th century to over 28 percent in the second half of the 20th century. For EP, she only documents an increase from 0 to 2.9 percent in the same period. It has to be noted, however, that Lopes Rossi only analyzed one theater play per 50-year period and dialect, which very much weakens the representativeness of her results. De Paula (2016), analyzing a much bigger sample of theater texts, confirms Lopes Rossi's results for BP and EP. The author documents an increase of the relative usage frequency of InSituWh in $n = 7$ BP theater texts from 0 percent in 1845 to 15.9 percent in 1992. In her much larger sample of EP theater texts ($n = 40$), the relative usage frequency of InSituWh oscillates between 1 and 3 percent (2016: 73). In De Paula (2017), the author also analyzes the diachronic distribution of *wh*-interrogatives in sociolinguistic interviews in EP and BP. She finds an increase in the relative usage frequency of InSituWh in the interviews in BP conducted in the 1970s and the 2010s, from 4 to 10 percent. For EP, she actually documents a decrease in the relative usage frequency of InSituWh, from 17 percent in the interviews conducted in the 1980s to 5 percent in the interviews conducted in the 2010s. In summary, these results suggest that the use of BP InSituWh has become more frequent in theater texts and possibly even in spoken language.

   Although there is thus solid evidence for the existence of this change, to my knowledge no study has yet offered an explanation for it. Likewise, the previous studies do not investigate the possibility of changes in the usage contexts of InSituWh. However, as we shall see in the next section, the possibility of such changes is the key to explaining the changes in the usage frequency of InSituWh.

2.2 Degrees of activation and conditional relevance

Descriptions of the semantics of *wh*-interrogatives distinguish between the proposition of the interrogative (henceforth P) and the referent of the interrogative pronoun or adverb (henceforth X). Following Hamblin (1973), the meaning of a *wh*-interrogative like

Where did John go? can be described as the sum of all possible answers (e.g., 'John went to London', 'John went to Barcelona' etc.). The use of a *wh*-interrogative presupposes the validity of P. Thus, while *Where did John go?* can be uttered in a context in which the proposition has been mentioned in or pragmatically inferred from the previous context, the questioning speech act can also be felicitous in out-of-the blue contexts. In such contexts, the speaker can exploit the fact that the hearer will accommodate the presupposition, i.e. integrate the assumption that John left into her belief system on the basis of the inference that the speaker of the interrogatives appears to believe so.

In Romance languages such as French (Coveney 1990; Obenauer 1994; Chang 1997; Mathieu 1999; Cheng & Rooryck 2000; Adli 2006; Myers 2007; Boucher 2010; Cheng 2013), Spanish (Dumitrescu 1992; 1993; Escandell-Vidal 1999; 2002; Dumitrescu 2008; Rosemeyer 2018a) and Portuguese (Pires & Taylor 2007; Oushiro 2010; 2011a; b; De Paula 2016: 105), InSituWh is most frequently used in contexts in which the proposition has been mentioned in the previous context and can consequently be described as activated. Consider, for instance, example (3), taken from the C-ORAL BRASIL, a corpus of informal spoken conversation between Brazilians from Minas Gerais (Raso & Mello 2012).[1] FLA did not understand EMM's utterance in line 1, which is why she uses the InSituWh interrogative *aparece o quê?* in line 3 in order to ask FLA to repeat the utterance. Such echo question readings exist in all of the mentioned Romance languages.

(3)   *Adriano* (*apud* C-ORAL BRASIL, pubcv01)

```
    01 EMM  <aparece>    mais o  AdRIAno;
            appear.PRS.3SG more  the adriano
            'Adriano appears more'

    02           [((laughs))]

--> 03 FLA  [aparece   ] o QUÊ?
            appear.PRS.3SG   what
            'appears what?'

    04 EMM  o  Adriano que aparece     MAIS;
            the adriano   that  appear.PRS.3SG more
            '(it is) Adriano who appears more'
```

However, there are great differences between the Romance languages with respect to whether or not the use of an InSituWh interrogative *requires* the proposition to be activated. In particular, BP InSituWh can be used with non-activated propositions, as demonstrated by examples such as (4). While tidying up the room, LAO and MBA find Julio's trousers and subsequently discuss how much he loves these trousers and how difficult it was for Ana – supposedly his partner – to convince him to let go of them (lines 1–10). In line 11, MBA discovers another item belonging to Julio, a sandal, and asks LAO how many days this sandal has been lying on the floor. The possibility of such out-of-the blue uses is also documented for French (Adli 2006: 184; Boucher 2010: 109; Larrivée to appear), but is mostly unacceptable in Spanish (Rosemeyer 2018a).

---

[1] The transcript was transcribed using the GAT 2 system for the transcription of talk-in-interaction; a list of the transcription conventions can be found at the end of the paper.

(4)   *a calça* (*apud* C-ORAL BRASIL, bfamdl26)

```
01 LAO  Ana custou           a    ficar livre <<:-)>da CALça lá,>
        ana cost.PST.PFV.3SG to come free             of.the   trouser there
        'It was hard for Ana to get rid of these trousers'

02      ((laughs))

03      chegou          aQUI,
        arrive.PST.PFV.3SG here
        'He came here'

04      outro dia levou        até ela pa CamPInas;
        other day take.PST.PFV.3SG even it   to campinas
        'The other day he even took with him it to Campinas'

05 MBA  é           MESmo?
        be.PRS.3SG really
        'Really?'

06 LAO  É.
        be.PRS.3SG
        'Yes'

07      pa dorMIR né;
        for sleep.INF TAG
        'To sleep, you know'

08      (2.0)

09 LAO  pelo menos ele não sAI        na    rua   com
        at   least he  not leave.PRS.3SG in.the street with
        esse neGÓcio;
        this  thing
        'At least he doesn't go out onto the street with it'

10 MBA  é <<creaky> melHOR;>
        be.PRS.3SG      better
        'That's something'

11      (1.5)
```

-->
```
12 MBA  nOssa essa sandália dele tem        quantos dias que
        wow   this sandal   his  have.PRS.3SG how.many days that
        tá         aqui no   CHÃO;
        have.PRS.3SG here on.the floor
        'My goodness, how many days has this sandal of his been lying here
        on the floor?

13 LAO  por QUÊ,=
        why
        'Why?'
        =já   tava (0.2)  quando cê VEio;
        already be.PST.IPFV.3SG when  you come.PST.IPFV.3SG
        'It was already there when you came here'
```

Note that whereas in (3), the interrogative *aparece o quê?* is used with an ascending into-
nation (indicated by ? in the transcript), in (4) the interrogative receives a falling intona-
tion (indicated by; in line 12). This observation conforms to the generalization established
in Kato (2013) for BP that echoic InSituWh usually receives a rising intonation, whereas
non-echoic InSituWh is typically pronounced with a falling intonation.

   Crucially for our purposes, the parameter of the activation of P is not only related to
intonation but also to the degree of to which the interrogative even needs to be answered.
This idea can be couched in terms of the notion of conditional relevance (Schegloff 1968).
Wh-interrogatives that express information questions occur as the first part of an adja-
cency pair, making the second part (i.e. the answer) highly relevant. Both (3) and (4) can
be described as such contexts, since in each example the speaker expects the hearer to
answer the question. However, sometimes *wh*-interrogatives are not used to ask for infor-
mation. Consider, for instance, example (5), in which the speaker uses the InSituWh con-
struction as a rhetorical question. VER clearly assumes that everyone knows the answer
to his question, which is why he does not even wait for the interlocutors to provide this
answer; rather, he provides it himself. The answer consequently has a very low degree of
conditional relevance and VER could have easily left the answer out.

(5)   *projeto* (*apud* C-ORAL BRASIL, bfamcv12)

    01 VER: mas eu preciso   que ele faça   o   proJEto,
             but I   need.PRS.1SG that he   do.PRS.SBJ.3SG the project
             'but I need him to do the project'

--> 02   final de conta ele é   DOno (1.0) pra QUÊ;
             end   of count he be.PRS.3SG owner   for what
             'after all, what is he the owner for?'

    03   pra GEnte::;
             for us
             'for us'

    04   (2.0)

    05   pra eu ir   fazendo   o   orçaMENto-=
             for I go.INF make.PROG the cost.estimate
             'so that I can make the cost estimate'

    06   =ir   comprando os materiais aos POUcos-=
             go.INF buy.PROG   the materials to.the few
             'buy the materials bit by bit'

    07   =e TAL;
             and so
             'and so on'

The three examples in (3–5) thus suggest an intricate relationship between the param-
eters of activation of P, conditional relevance and intonation in the use of BP InSituWh.
Whereas echoic uses of InSituWh typically receive an ascending intonational contour,
non-echoic uses receive a falling contour. The non-echoic uses can be either information
questions or rhetorical questions, and may have propositions that are discourse-new.

   These considerations are relevant to the analysis of the changes in the use of InSituWh in
BP because some authors contend that the parameter of activation of P played an impor-
tant role in the history of French InSituWh. For instance, Larrivée (to appear) analyzes

the use of InSituWh in a comparative corpus of French sociolinguistic semi-directed interviews conducted between 1969 and 1974, and in 2014. Following Dryer (1996), the author annotates the data according to the parameter of activation of P, distinguishing between propositions that are explicitly activated and those that are discourse-new. He documents an increase of the relative frequency of InSituWh in contexts in which the proposition is discourse-new, from 46 percent ($n = 6/13$) in the 20th century data to 86 percent ($n = 60/70$) in the 21st century data. However, as acknowledged by Larrivée himself, follow-up studies analyzing a greater quantity of data are necessary in order to confirm this result.

In line with Larrivée's (to appear) analysis of the diachrony of French InSituWh, this paper explores the question of whether the increase in the usage frequency of BP InSituWh was accompanied by a change in the activation of P. In particular, it tests the hypothesis that InSituWh became less strongly bound to contexts in which P is activated.

3 Data

The data for this study were extracted from a self-compiled corpus of BP theater plays dated between 1800 and 2016 (Rosemeyer 2018c).[2] The corpus contains almost 300 theater plays (see Table 1 for a summary), mostly from Southern Brazil (Rio de Janeiro and São Paulo), with a total of about 2,5 million words. The theater plays were collected either from existing linguistic databases of BP theater plays (such as the corpus of the Grupo de Morfologia Histórica do Português and the Grupo Sujeito em Peças Teatrais at the Universidade Federal do Rio de Janeiro) or from homepages on which modern BP playwrights make their plays freely available (for instance, oficinadeteatro.com (last access 3 May 2019)). Only plays for which metadata was given — especially date of publication and author information — were taken into consideration; in dubious cases the authors were contacted in order to verify the information.

The decision to use the corpus of theater plays was taken on the basis of the fact that whereas direct documentation of spoken BP before the 20th century is scarce, the corpus of theater texts arguably has sufficient time depth and volume in order to study the changes in the system of *wh*-interrogatives. A crucial premise of this approach, shared with studies in historical pragmatics such as Jacobs & Jucker (1995) or Culpeper & Kytö (2010), is that although such texts "cannot be expected to have preserved speech with the accuracy that modern audio-recording devices do" (Kytö 2011: 432), theater plays at least approximate contemporary spoken language. They are more reliable representations of spoken language than prose or specialized texts. As was shown in Section 2, the use of *wh*-interrogatives is extremely sensitive to discourse pragmatics. Unsurprisingly, previous studies have shown that the use of *wh*-interrogatives in prose or specialized texts massively differs from spoken language both in terms of usage frequency and discourse functions (see Kato & Mioto [2005]; Oushiro [2011: 33, 35]; Ehmer & Rosemeyer [2018]). For this reason, the use of a "conventional" big corpus of digital texts such as the Corpus do português (Davies 2006) was dispreferred.

Table 1: Summary statistics for the corpus of Brazilian Portuguese theater plays.

	19th c.	20th c.	21st c.	Total
n_{words}	787,015	740,389	947,900	2,482,610
n_{plays}	82	63	153	298

---

[2] More information on the corpus, including a list of the plays, can be found on the author's homepage.

All tokens of *wh*-interrogatives in sentences followed by a question mark were extracted using regular expressions. A list of the extracted interrogative pronouns and adverbs can be found in (7).

(7)   *aonde* 'to.where', *cadê* 'where.is', *como* 'how', *onde* 'where', *porque/porquê* 'why', *quais* 'which ones', *qual* 'which one', *quando* 'when', *quanta* 'how.much.F.SG', *quantas* 'how.much.F.PL', *quanto* 'how.much.M.SG', *quantos* 'how.much.M.PL', (*o*) *que/quê* 'what', *quem* 'who'

Of the $n = 18{,}903$ tokens of direct *wh*-interrogatives encountered in the data, a total of $n = 390$ tokens were coded as InSituWh. However, not all of these tokens can be said to represent InSituWh in a strict sense, i.e. *wh*-interrogatives in which the *wh*-element is placed after a finite verb. In particular, the verb before the interrogative element does not always have to be finite (see 8) and can even be non-verbal (see 9).

(8)   *O quarta do casal*, Jorge Gomes de Oliveira, 1984

FLORA -   […] O   "Wesley" já   chegou???
              the wesley   already arrive.PST.PFV.3SG
              'Did Wesley already arrive?'

LILY -   Já   sim, está   tranc ad o   no   Banheiro…
          already yes be.PRS.3SG lock.PTCP.M in.the bathroom
          'Yes he did, he is locked in the bathroom…'

MARIZA – **Fazendo o   quê???**
          do.PROG   what
          'Doing what?'

(9)   *O demônio familiar,* José de Alencar, 1857
PEDRO: Disse   a sr. moço   Eduardo […] que
        say.PST.PFV.3SG to mr. young.man eduardo   that
        Sr. Vasconelos [sic] é_____ um… nome muito ruim!
        mr. vasconcelos   be.PRS.3SG a   name very bad
        'He said to the young mister Eduardo […] that Mr. Vasconcelos is
        a… very bad thing!'

VASCONCELOS: **Um que, moleque?**
              a   what rascal
              'A what, you rascal?'

PEDRO: Um pinga!
           a   pinga
           'A pinga!'

Although to some degree there is free variation between the different types of InSituWh (for instance, in (9) Vasconcelos could have also asked *É um quê?* 'Is a what?'), the interrogatives without a finite verb differ from "true" InSituWh in that they are more dependent on the previous context syntactically and pragmatically, and are consequently much less likely to be used in contexts in which the proposition is not activated (see Rosemeyer to appear a; b). All tokens of InSituWh that either do not have a verb or a non-finite verb were therefore excluded from the dataset, leading to a final corpus of $n = 218$ examples of InSituWh in the strict sense.

4 InSituWh in discourse

A closer look at the $n = 218$ examples of InSituWh in the corpus demonstrates that these cases of InSituWh realize very different functions in discourse, which emerge as situated meanings (Linell 2009) from the contextual properties of the utterance. In broad terms, we can distinguish between cataphorical, anaphorical and rhetorical discourse functions (cf. also Fiengo 2009 and Rosemeyer 2018a for similar classifications).

Cataphorical discourse functions are "true" information questions in the sense that they serve to elicit a propositional answer by the hearer. This answer advances the discourse by adding new information into the interlocutors' Common Ground, defined as the set of propositions that the interlocutors assume to be presupposed (see, e.g., Stalnaker 2002). By eliciting this answer, the utterer of the interrogative thus actively manages discourse progression, which can for instance be modeled in terms of the notion of Question Under Discussion (Klein & von Stutterheim 1987; Ginzburg 1996; Roberts 1996, among others). Two subtypes of cataphorical discourse functions can be distinguished. First, in contexts in which the proposition is discourse-new and the answer to the interrogative is unknown to the speaker, InSituWh is used to establish a new discourse topic, which thus answers a less specific question under discussion. In out-of-the-blue contexts, this will be the "Big Question" *What is the way things are?* (see Roberts 1996; Riester et al. 2018: 415). Consider, for instance, example (10), the beginning of a recent theater play. Ping and Pong have been calling for each other from offstage. First Pong enters and leaves the stage searching for Ping. Then Ping enters the stage and asks the audience if they have seen Pong; the scripted reaction *Ah!* suggests an answer by the audience. The proposition of Ping's subsequent interrogative *Ele foi pra onde?* is pragmatically inferred from her observation that Ping was on stage but is not anymore. The proposition is thus discourse-new in the sense that it is not derived from the previous discourse. As a result, the interrogative expresses an information question and necessitates an answer, which the playwright supposes the audience will indeed give (cf. *Pra la?*). In doing so, it introduces a new discourse topic, i.e. a question under discussion not closely related to the question under discussion in the previous co-text. This function of InSituWh will therefore be called NewTopic.

(10)   *Um curumim no sítio do pica pau amarelo*, Marcondys França, 2012

PONG: –   (Entra)   Ping? Ping? Oi! Vocês viram   a   Ping?
enter.PRS.3SG ping ping hi   you   see.PST.PFV.3PL the ping
Lá?   Vocês estão   tentando me enganar! Então vou
there you   be.PRS.3PL try.PROG me deceive   then go.PRS.1SG
pra lá. (Sai)
to   there
'(enters) Ping? Ping? Hi there! Did you see Ping? There? You're trying to deceive me! So I'll go over there'

PING: –   (Entra) Pong? Pong? Vocês viram   o   Pong? Ah! **Ele**
enters pong pong you   see.PST.PFV.3.PL the pong ah he
foi   pra onde? Pra lá?   (Sai)
go.PST.PFV.3SG to   where to   there leave.PRS.3SG
'(enters) Pong? Pong? Did you guys see Pong? Ah! Where did he go? There? (leaves)'

Second, in many cataphorical instances of InSituWh the interrogative is used to elaborate on an already established question under discussion; this function will be called Elaboration. It thus serves to clarify a question related to an issue raised in the previous

co-text. The Elaboration function typically arises when the proposition of the interrogative has either already been mentioned (see 11) or is derived via inference from a piece of information in the previous context (see 12). Although like NewTopic questions, they can be classified as a "real" information questions, Elaboration questions advance the discourse to a much smaller degree than NewTopic questions and cannot arise in thetic contexts.

(11)   *Propostas*, Humberto Mello, 2005

   M.b. –   Dona   eu tenho   um pobrema que eu queria
             mistress I have.PRS.1SG a   problem that I   want.PST.IPFV.1SG
             falar pra senhora, é   um pobrema meu, particular de minha
             talk to mrs   be.PRS.3SG a   problem my   particular of my
             pessoa que eu tenho   desde pequeno.
             person that I   have.PRS.1SG since little
             'Mistress, I have a problem that I wanted to talk to you about, it's a
             personal problem of mine that I have had since I was little'

   Medusa – **Você tem   o que querido?**
             you have.PRS.3SG what beloved
             'What (problem) do you have, my dear?'

(12)   *Desejo secreto*, Miguel Oniga, 1991

   Lucas –   Vai   jantar?
             go.PRS.3SG eat.lunch
             'Will you have lunch?'

   Laura –   Não. Jantei   na   legação australiana.
             no   have.lunch.PST.PFV.1SG in.the embassy australian
             'No. I already had lunch in the Australian embassy'

   Lucas –   **Comeu   o que?**
             eat.PST.PFV.3SG what
             'What did you eat?'

   Laura –   Carneiro com pêras.
             mutton   with pears
             'Mutton with pears'

Anaphorical discourse functions do not aim at eliciting a piece of information from the hearer that advances the discourse, but rather to negotiate the meaning or significance of a previous utterance. Consequently, these meanings are bound to contexts in which the proposition of interrogative is discourse-old. Repeat readings (see 13, also 3) emerge when there is reason to believe that the utterer of the interrogative does not know the answer to her question; she asks for repetition of a recent word or phrase either because she did not hear it or it was omitted, or because it is unknown to her.

(13)   *Mês de cachorro louco*, Isaias Edson Sidney, 1997

   VÍTOR:-   Pai! Mãe! Não dá   mais… acabou…
             father mother not give.PRS.3SG more   end.PST.PFV.3SG
             'Father! Mother! It doesn't matter anymore… it's over…'

   APARECIDA:- **'Cabou…   'cabou   o quê, meu filho?**
                end.PST.PFV.3SG end.PST.PFV.3SG what   my   son
                'It's over… What is over, my son?'

In contexts in which both the proposition and the asked-for-element are discourse-old and the speaker thus clearly knows the answer to the interrogative, a challenge reading arises. This pragmatic effect can be modeled via the concept of implicature; by ostensively asking for a piece of information that is both known to the speaker and known by the hearer to be known to the speaker, the speaker implicates that a plain answer is not the preferred reaction. Rather, the interrogative serves to challenge the previous move by the interlocutor. To illustrate this effect, consider example (14). Julieta has clearly understood what Romeu said; her InSituWh interrogative expresses her indignation resulting from Romeu's previous utterance and can therefore be characterized as an exclamative (cf. also Chernova 2015: 166). Note that the exclamative challenge function is also indicated by the simultaneous use of question and exclamation marks in the text. As demonstrated by Rosemeyer (2018a), even though challenges are clearly not information questions they do typically still prompt a reaction by the hearer, namely a justification for her or his previous move.

(14)   *Romeu & Julieta,* Marcondys França, 2012
        ROMEU – É   que agora você ta   meio fora de forma né…
                 be.PRS.3SG that now   you   be.PRS.3SG half out   of form   TAG
                 'It's that you are a bit off form now, aren't you…'

        JULIETA – (levantando o   braço e   avançando   para bater em Romeu)
                   raise.PROG   the   arm   and advance.PROG to   hit   in romeu
                   To   meio o que Romeu?!?!
                   be.PRS.1SG half what romeu
                   '(raising the arm and advancing in order to hit Romeu) I am a bit
                   what, Romeu?'

The last group of situated meanings can be said to realize rhetorical functions, i.e. they do not require an answer by the hearer. First, outloud questions are questions the speaker asks herself viz. no one in particular (Stivers & Enfield 2016: 2623). In spoken language, they are typically uttered in a low voice. In the theater plays, outloud questions sometimes merely serve to express the speaker's train of thought, as in example (15).

(15)   *O poeta e a inquisição,* Gonçalves de Magalhâes, 1870
        LÚCIA – É   um pouquinho d'água   com vinagre.
                 be.PRS.3SG a   bit   of.water with vinegar
                 Molha-se   o   lenço… assim… É   coisa santa;
                 wetten.PRS.3SG-REFL the cloth   so   be.PRS.3SG thing holy
                 Não tenha   medo; aplique-o   sobre as
                 not   have.PRS.SBJ.3SG fear   apply.PRS.SBJ.3SG-it over   the
                 fontes. **Ensinou-me   quem mesmo?…** nem   me
                 temples teach.PST.PFV.3SG-me who   again   not.even REFL
                 lembro
                 remember.PRS.1SG
                 'That's a bit of water with vinegar. The cloth has to be wetted; like
                 that. It's a holy thing; don't be afraid; apply it to the temples. Who
                 taught this to me again? I don't even remember'

Lastly, there are uses of InSituWh that can be characterized as rhetorical questions in a stricter sense. According to Rohde (2006: 135), there are three felicity conditions in order for rhetorical readings to arise: (a) there is an obvious answer to the question, (b) both speaker and hearer would give the same answer and (c) if an answer is given, it

is consequently uninformative. Fiengo (2009: 61–63) distinguishes between rhetorical questions with an obvious answer and rhetorical questions that are unanswerable. However, unanswerable rhetorical questions like *What has John ever done to help?* are really only rhetorical questions with *one* obvious answer (*Nothing*). I will therefore consider the most important definitional criterion for rhetorical questions to be the presence of an obvious answer to the question.

The presence of an obvious answer is frequently indicated in the constructed dialogues by the fact that the speaker does not wait for the hearer to give an answer but continues her turn. Typically, she gives her own answer to her "question", but using an interrogative intonation contour. This indicates the low conditional relevance of the question. Consider, for instance, example (16), in which the interrogative simply expresses André's dismay that there is nothing to drink around.

(16)   *Coração rasgado,* Aloisio Villar, 2012
JULIANA–   Não!! Isso não é   pra beber,   é   pra
(interrompendo) no   this not be.PRS.3SG for drink.INF be.PRS.3SG for
limpar   os   dedos.
clean.INF the fingers
'No! This is not for drinking, it's for cleaning your fingers'

ANDRÉ –   Esse líquido é   pra limpar   os   dedos? **Eu bebo**
this liquid be.PRS.3SG for clean.INF the fingers I   drink.PRS.1SG
o que   então? Guardanapo?
what then then   napkin
'This liquid is for cleaning fingers? Then what should I drink? The napkin?'

The rhetorical question reading frequently arises with psychological predicates, such as 'think', 'believe' or 'want'. For instance, in example (17) the father's question expresses that there really is nothing that he could have done to meet his daughter's standards. Crucially for the analysis developed in this paper, example (17) also demonstrates that rhetorical questions are used rather freely with respect to the parameter of activation of the interrogative proposition. Thus, in (17) the proposition is clearly discourse-new. The same applies to outloud questions (see example 15).

(17)   *Pai por acaso, filha por acidente,* Romário Machado, 2003
FILHA –   Você só   tem   revista   de mulher nua?
daughter you only have.PRS.3SG magazine of woman naked
'Do you only have magazines with naked women?'

PAI –   **E   você queria   o quê?** Que eu tivesse
father   and you want.PST.IPFV.3SG what   that I   have.PST.IPFV.SBJ.3SG
ai   um monte de Revistas   repletas de homem nu?
there a   heap of magazines full   of man   naked
'And what did you want? That I had an overflowing heap of magazines with naked men there?'

In addition, the examples in (16) and (17) illustrate three other interesting characteristics of rhetorical questions. First, note the absence of address terms in (16) and (17), whereas in many of the preceding examples address terms are used in order to reinforce the conditional relevance of the interrogative (cf. *querido* 'beloved' in example (11), *meu filho* 'my son' in (13) and *Romeu* in (14)). By directly addressing the interlocutor, the speaker

thus implies that an answer or other reaction to the interrogative is very important to her. Since this is not the case in rhetorical questions, address terms are expected to be less frequent in such contexts.

Second, it appears that a rhetorical reading is more likely to arise in the case of morphologically non-complex *wh* pronouns referring to arguments of the interrogative, such as *quem* 'who', *o que* 'what' or *onde* 'where'. This might be due to the pragmatic mechanism that is responsible for the emergence of rhetorical question readings, i.e. the use of an interrogative in a context in which the answer is known to the speaker and the addressee. For instance, the question *[What]$_{wh}$ could I do?* usually implicates that there is no acceptable referent of the interrogative pronoun *what*. A *wh*-interrogative with a morphologically more complex *wh*-element, such as *[How many years]$_{wh}$ did he work there?* might be less well suited for the expression of a rhetorical question because the *wh*-element carries more semantic information, which to some degree invalidates the implicature that no satisfactory referent for the *wh*-element exists.

Third, rhetorical InSituWh interrogatives typically contain elements that explicitly mark the question as advancing the discourse, such as the conjunction *e* 'and' (example 17) and the adverb *então* 'then' (example 16). However, as has been shown in previous research (Rosemeyer 2018b), this characteristic might not be distinctive for rhetorical questions, as the formulation of cataphorical information questions frequently involves the use of such markers, too.

Table 2 summarizes the typology of situated meanings established in this section according to the parameters of activation and knowledge of the interrogative proposition P and the asked-for-element X, as well a conditional relevance, and gives the usage frequencies of every one of these meanings in the data.

5 The development of InSituWh in BP theater plays

On the basis of the typology of situated meanings established in the last section, a descriptive and inferential statistical analysis of the changes in the usage of BP InSituWh was carried out.

5.1 Frequency changes over time

Figure 1 illustrates the changes in the usage frequency of InSituWh in the corpus of BP theater texts. Each point in the graph represents the log-transformed normalized usage frequency of InSituWh for one year (in turn representing one or more plays published that year), whereas the line represents the result of a local polynomial regression analysis indicating the trend. Figure 1 demonstrates that InSituWh interrogatives have become more frequent in BP theater plays, especially after the beginning of the 20th century, confirming the results from previous historical studies (see Section 2).

Table 2: Summary of the typology of situated meanings of BP InSituWh.

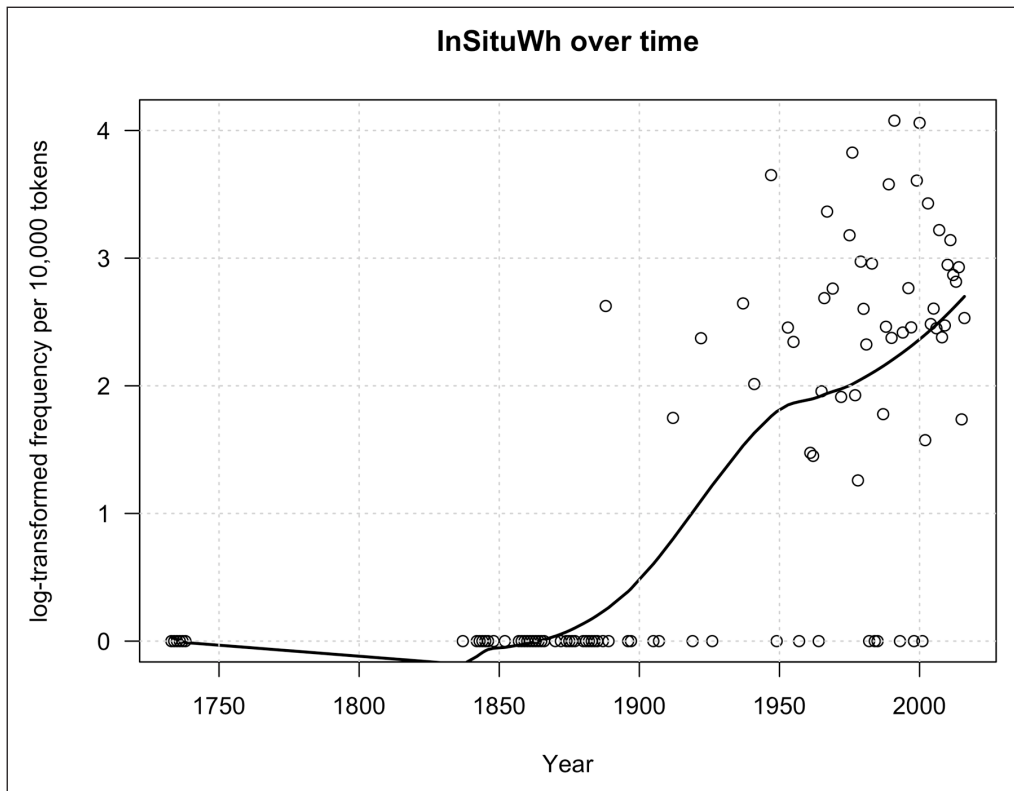|	Activation of proposition	Activation of X	Speaker's knowledge of X	Conditional relevance	Frequency
NewTopic	New	New	Unknown	High	65
Elaboration	Old	New	Unknown	High	52
Repeat	Old	Old	Unknown	High	17
Challenge	Old	Old	Known	High	8
Outloud	New/Old	New	Known	Low	6
RhetoricalQuestion	New/Old	New	Known	Low	70

Figure 1: Normalized usage frequency of InSituWh by year in the corpus of BP theater texts.

Table 3: Usage frequency of InSituWh by 25-year period in the corpus of BP theater texts.

1870–1899	1900–1924	1925–1949	1950–1974	1975–1999	2000–2016
2	3	8	17	57	131

Table 3 gives the non-normalized usage frequencies of InSituWh per 25-year time period, again demonstrating that InSituWh is very infrequent before the second half of the 20th century.

The increase in the usage frequency of InSituWh goes hand in hand with distributional changes. First, as hypothesized in Section 2, there are changes with respect to whether or not InSituWh can be used with discourse-new propositions. Figure 2 visualizes this change, masking the scarce ($n = 5$) number of tokens of InSituWh before 1925. Whereas in the first half of the 19th century, a majority of InSituWh tokens occur in discourse-old contexts (i.e. contexts in which the proposition has either been mentioned or is logically dependent on a previous proposition), after 1950 InSituWh gradually advanced into discourse-new contexts. In the 21st century data, almost 75 percent of the InSituWh tokens occur in discourse-new contexts.

In line with the description of the typology of discourse functions of InSituWh from Section 4, one might interpret this result as a change from anaphorical to cataphorical discourse functions. If this were the case, the change could be interpreted as an intrusion of the use of InSituWh into information question functions previously reserved to ExSituWh interrogatives.

Crucially however, it was shown in Section 4 that rhetorical InSituWh tokens can also appear in discourse-new contexts. As it turns out, the increase of InSituWh in discourse-new contexts is largely due to a change from non-rhetorical to rhetorical discourse functions. Figure 3 visualizes this change, collapsing the three groups of discourse functions (ANA = anaphorical, CATA = cataphorical and RHET = rhetorical) and the parameter of whether or not the proposition is activated (NewP vs. OldP).

Figure 3 clearly illustrates that the change from usage contexts with activated propositions to usage contexts with non-activated propositions is moderated by discourse function. Most importantly, there does not appear to be a strong increase in the relative usage frequency of non-rhetorical cataphorical discourse functions in contexts in which
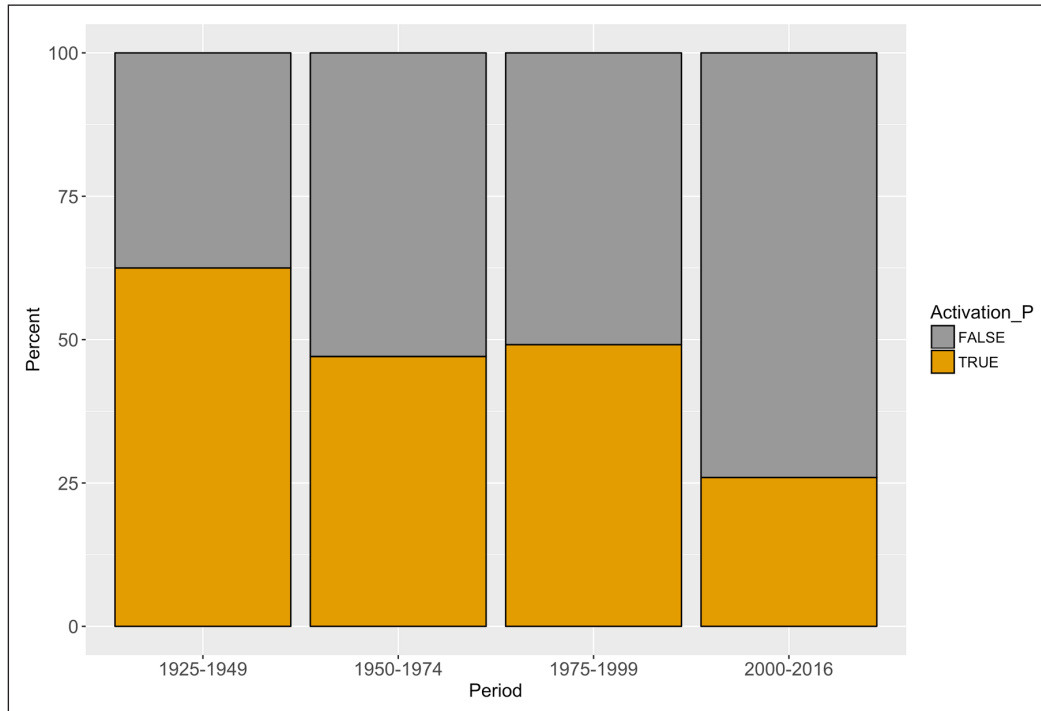
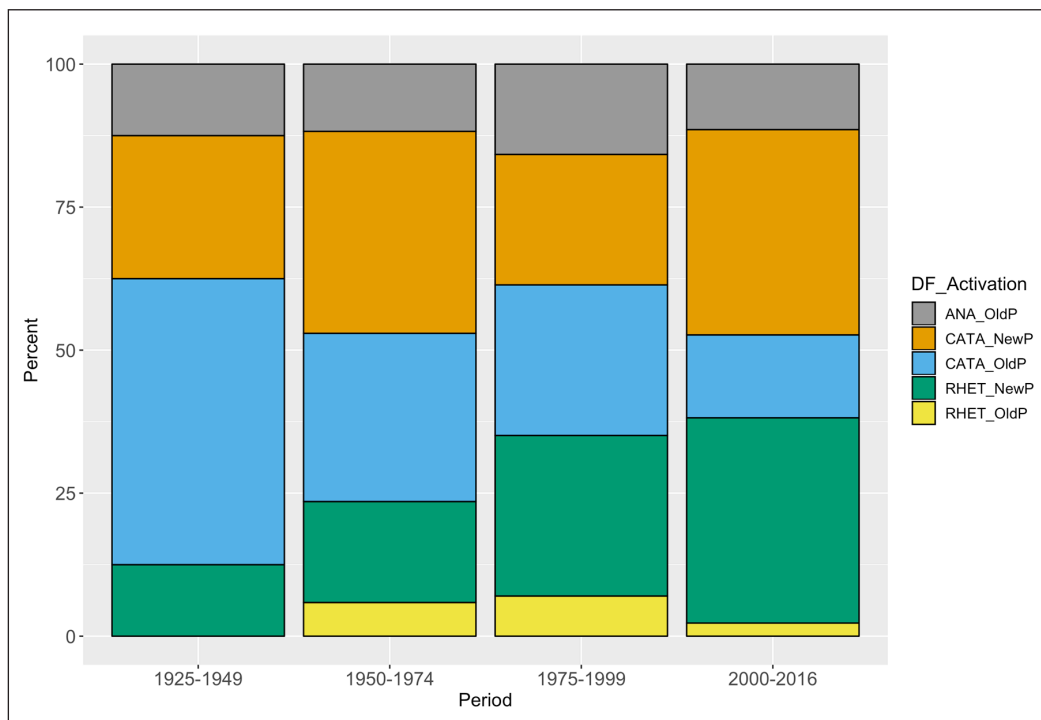**Figure 2:** Activation of proposition in InSituWh by 25-year time period (after 1925).

Figure 3: Relative frequencies of discourse functions (DF) by activation and 25-year time period (after 1925).

the proposition is discourse-new (orange color). The relative usage frequency of these contexts only increases from 25 percent in 1925–1949 to about 36 percent in the 21st century. Note however that due to the fact that most of the InSituWh tokens are found in texts dated between 1974 and 2016, the increase in the frequency of these contexts between the two last time periods might still be significant.

We document a stronger and more consistent increase in the use of rhetorical discourse functions, especially with discourse-new propositions (green color), from about 13 percent in 1925–1949 to about 36 percent in the 21st century. This increase coincides with a decrease in the usage frequency of the use of InSituWh in non-rhetorical cataphorical contexts with a non-activated proposition, i.e. with the discourse function of Elaboration (blue color). These results do not suggest that the historical trend of BP InSituWh towards a higher usage frequency in contexts in which the interrogative proposition is discourse-new illustrated in Figure 2 is symptomatic of a functional change of InSituWh towards information question use. Rather than intruding into information question contexts, the construction becomes more frequent with rhetorical question readings.

5.2 Predicting the rhetorical discourse function

The next steps in the analysis aim at confirming the results from the last section using inferential statistical methodology. In a first step, a logistic regression model was calculated in order to validate the coding of discourse functions by evincing correlations between the interpretation of the data and more objective contextual criteria.

Based on the results from the last section, a binary dependent variable RHET_NEWP was constructed that distinguished between rhetorical uses of InSituWh in contexts with non-activated propositions (RHETNEWP = 'True') and all other discourse functions of InSituWh (RHETNEWP = 'False'). The regression model thus measured the probability for an InSituWh token to be used in contexts classified as rhetorical with non-activated propositions. Table 4 summarizes the predictor variables used in the regression model.

Table 4: Predictor variables in the regression model predicting the use of InSituWh in contexts classified as RHETNEWP = 'True'.

Predictor	Type	Levels or range	Description	*n*
LOGORALITY	Numerical	0.7–5.9	Logarithmic normalized usage frequency of orality indicators in the play	–
NEXTMOVE BYUTTERER	Categorical	None (RL)	No next move by the utterer, i.e. the interlocutor takes the turn	72
		Answer	Speaker herself gives an answer	8
		Answer-question	Speaker herself gives an answer, in a questioning stance	64
		Explanation	Speaker explains her question	24
		Question	Speaker reformulates the question	24
		Other	Speaker utters another type of speech act	26
CONNECTOR	Categorical	False (RL)	No connector (*e, então*) used	194
		True	Connector used	24
ADDRESS	Binary	False (RL)	No address term used	191
		True	Address term used	27
SIMPLEWH	Binary	False (RL)	Complex *wh*-element	94
		True	Simple *wh*-element	124
VERBCLASS	Categorical	Event/State (RL)	Non-psychological predicates	156
		Psychological	Psychological predicates	62

These predictor variables were hypothesized to have an influence on the likelihood for an *in-situ wh*-interrogative to realize a rhetorical reading on the basis of the discussion of the properties of rhetorical questions in section 4 and the descriptive findings in section 5.1, with the exception of the variable LOGORALITY.

The predictor LOGORALITY was included in order to test the intuition that rhetorical uses of *wh*-interrogatives are more likely to occur in formal than in informal registers. Although the corpus used for the analysis can be characterized as homogeneous in terms of genre, the theater plays still represent orality to a greater or lesser degree, depending on the degree to which the authors heed the norms of written BP. Consequently, it was hypothesized that in plays that represent spoken language more accurately, the use of rhetorical questions is less likely. The predictor LOGORALITY establishes a measurement of the degree to which the plays in the Brazilian Portuguese corpus represent orality, by using Biber & Finegan's (2004 [1987]: 68) dimension of "involvement" of the oral/literate dimensions of variation.[3] Five linguistic variables, listed in Table 5, were extracted from the corpus of BP theater texts. These variables can be said to represent orality due to the fact that their use is contingent on temporal, spatial or discourse deixis (present progressive, demonstrative neuter pronouns, time and place adverbs and discourse markers) or because they represent intellectual states prone to expression in orality (the type of verbs that Biber & Finegan call "private verbs"). Both realizations typical for EP (e.g., *estar* + infinitive progressives) and BP (e.g., *estar* + gerund progressives) were included in order to capture all variants in all temporal periods. As proposed in Biber and Finegan's study, the frequencies of the five variables were aggregated for each text. Afterwards, they were normalized and log-transformed in order to ensure comparability of the numerical variable across the different texts. The higher the value of a text for the resulting variable LOGORALITY, the more likely it is to represent spoken language more accurately. As summarized in Rosemeyer (to appear a), one interesting analytical result from this procedure was that there is a significant increase in the degree of orality of theater plays over time, reflecting a gradual weakening of the norms of written BP.

Having described the operationalization of the dependent and predictor variables, we can now go on to discuss the results from the logistic regression model, calculated in R (R Development Core Team 2019), which predicted the likelihood for an InSituWh token to be used as a rhetorical question in contexts in which the proposition is discourse-new dependent on the predictor variables listed in Table 5. With a c index of concordance

Table 5: Linguistic variables used to measure the degree of orality in the corpus of Brazilian Portuguese plays.

Variable	Description	*n*
Private verbs in present tense singular	The verbs *achar* 'mean', *pensar* 'think', *acreditar* 'believe', *crer* 'believe'	3260
Present progressive	*estar* + gerund (e.g., *est-á diz-endo* 'be-PR.3SG say-GERUND) and *estar* + *a* + infinitive (e.g., *est-á a diz-er* 'be-PR.3SG to say-INF) constructions	4860
Demonstrative neuter pronouns	*isso* and *isto* 'this'	8655
Time and place adverbs	*aqui* 'here' and *agora* 'now'	9836
Discourse markers	*né* 'isn't it?', *bom* 'well', *pois* 'so', *então* 'so', *olha* 'listen'	6141
Total		32752

---

[3] See Culpeper and Kytö (2000) and Rosemeyer (to appear a) for a similar approach in historical linguistics.

of 0.87, the model explains about 83 percent of the variation in the data regarding the dependent variable. The results from the regression analysis are summarized in Table 6.

The regression analysis evinces correlations between the dependent and the predictor variables, most of which reach statistical significance. The strongest of these effects is the variable NextMoveByUtterer. Figure 4 visualizes this effect. In comparison to contexts in which the speaker stops her turn after uttering the interrogative (level 'None'),[4] the likelihood for the interrogative to receive a rhetorical interpretation is highest in contexts in which the speaker goes on to deliver an answer, whether using an assertive or questioning intonational contour ('Answer' and 'Answer-question'), followed by explanations of her motive to utter the interrogative ('Explanation'). Rephrasing the question after uttering the interrogative (level 'Question') does not lead to a significantly higher likelihood for the interrogative to receive a rhetorical interpretation.

The analysis also found significant effects for the predictor variables VerbClass, SimpleWh and LogOrality, as well as a marginally significant effect for Address. As predicted, InSituWh is more likely to be used in rhetorical function with a discourse-new proposition with psychological verbs (the most frequent of which were *querer* 'want', and *pensar* 'think'), as well as with simple as opposed to complex *wh*-elements (cf., e.g., *quem* 'who' vs. *de onde* 'from where'). In addition, InSituWh is significantly less likely to be used in texts that scored high on the variable LogOrality, thus representing less formal writing styles that represent spoken discourse more accurately. Although with a *p* value of .0578 this effect does not reach statistical significance, it might be that InSituWh is less

Table 6: Results from the regression model predicting the use of InSituWh in contexts classified as RhetNewP = 'True'*.

Variable	Level	Beta	OR	SE	Z	p
(Intercept)	(numerical variable)	−1.332	0.264	1.241	−1.073	>.05
LogOrality	(numerical variable)	−0.487	0.615	0.244	−1.997	<.05*
NextMove ByUtterer	None	(RL)				
	Answer	3.672	39.336	0.904	4.062	<.001***
	Answer-question	2.983	19.748	0.611	4.884	<.001***
	Explanation	2.712	15.054	0.659	4.112	<.001***
	Question	0.995	2.706	0.853	1.167	>.05
Connector	False	(RL)				
	True	0.419	1.521	0.581	0.722	>.05
Address	False	(RL)				
	True	−1.522	0.218	0.802	−1.897	<.01
SimpleWh	False	(RL)				
	True	0.931	2.536	0.421	2.211	<.05*
VerbClass	Event/State	(RL)				
	Psychological	1.026	2.79	0.405	2.534	<.05*

* *Note*: RL = Reference level, OR = odds ratio, SE = standard error, Z = z-values, p = p-values. Model evaluation: c index of concordance = 0.87, somers' dxy = 0.74, AIC = 198.92.

---

[4] In this category cases were included in which the interrogative was followed by an order to answer the question, as in (i), repeated below as (18).

(i) Vado – Pensa   que estou   aqui por quê? Anda, responde!
         think.PRS.3SG that be.PRS.1SG here why   go.IMP answer.IMP
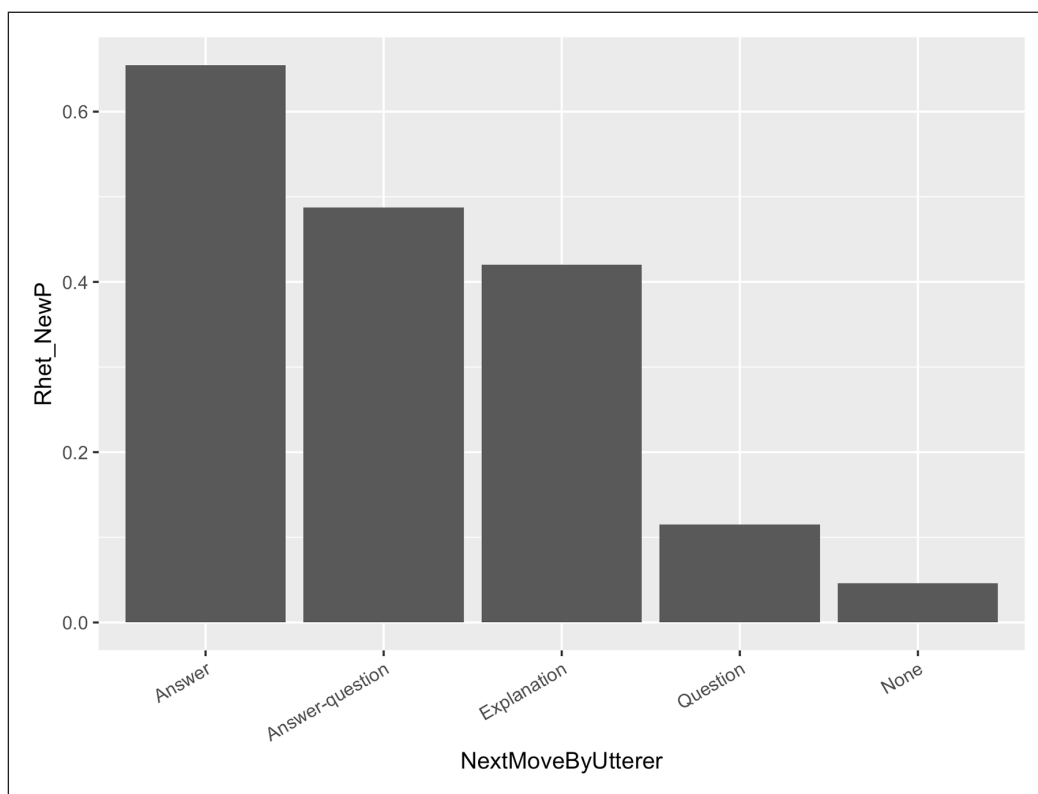   'Why do you think that I am here? Come on, answer me!'

Figure 4: The effect of the predictor variable NextMoveByUtterer on the dependent variable
Rhet_NewP. The values on the y axis represent the conditional probability (0 = minimum,
1 = maximum) that Rhet_NewP receives the value 'True'.

likely to be used in rhetorical function with a discourse-new proposition when there is an
address term present. The analysis did not find a significant effect for the predictor vari-
able Connector, measuring the use of the discourse-connecting conjunction *e* 'and' and
adverb *então* 'then'.

By uncovering the correlations between the rhetorical interpretation of InSituWh and
contextual parameters it is finally possible to verify the findings regarding the change
from non-rhetorical to rhetorical meanings on the basis of these objective quantifiable
contextual criteria. Figure 5 plots the diachronic changes in the distribution of InSituWh
with respect to NextMoveByUtterer.

Surprisingly, Figure 5 demonstrates an increase in the frequency of InSituWh tokens in
contexts in which the speaker does *not* give an answer to her or his own interrogative,
from 25 percent in 1925–1949 to almost 40 percent in the 21st century. This result, which
a logistic regression analysis shows to be statistically significant,[5] appears to contradict
the finding of an increase of rhetorical readings of InSituWh because not an increase,
but a decrease of the use of InSituWh in contexts with high conditional relevance would
be expected. However, a closer look at the data demonstrates that the two historical
trends are not incompatible. Consider Figure 6, a heat plot of the diachronic relationship
between discourse function and the next move by the speaker. The colors of the tiles in
Figure 6 indicate which context-meaning association is strongest in each time period. For

---

[5] A logistic regression analysis was conducted that predicted whether the speaker gives an answer on the
basis of two contextual predictors: (a) the discourse function of the interrogative and (b) the year of
publication of the play. Both predictors had a statistically significant effect on the dependent variable
(AnswerByUtterer, modelled as true/false). Most importantly, the predictor Year was found to have
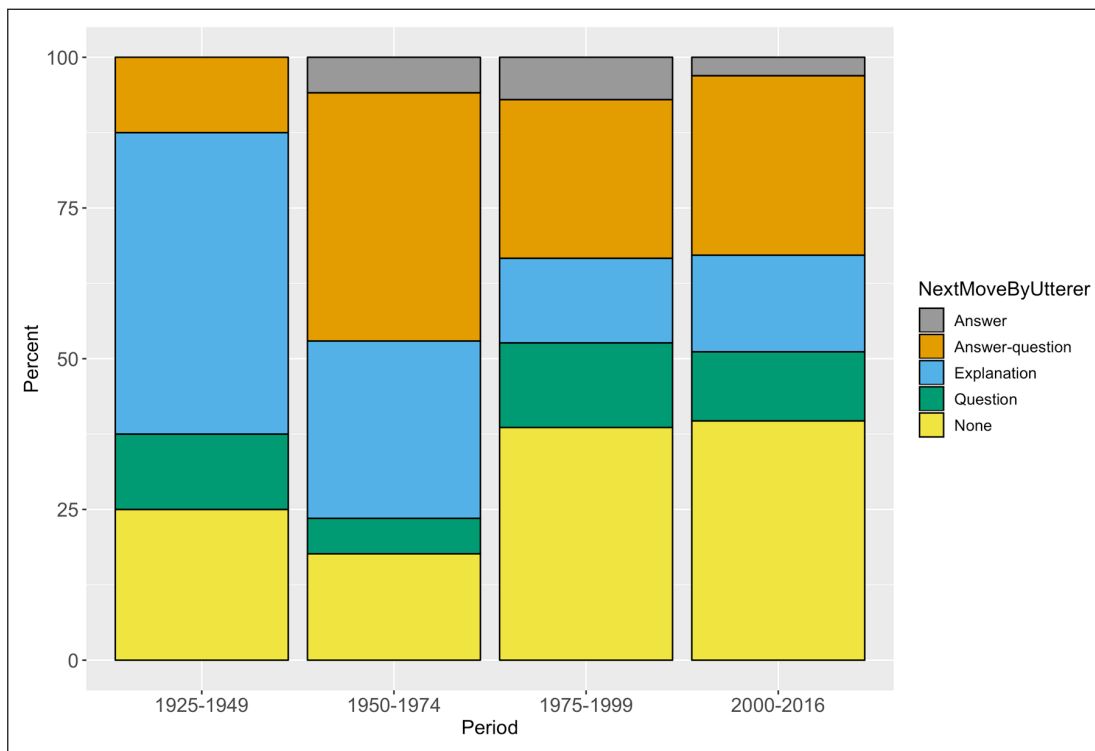a negative effect on AnswerByUtterer, at a significance level of $p < .05^*$.

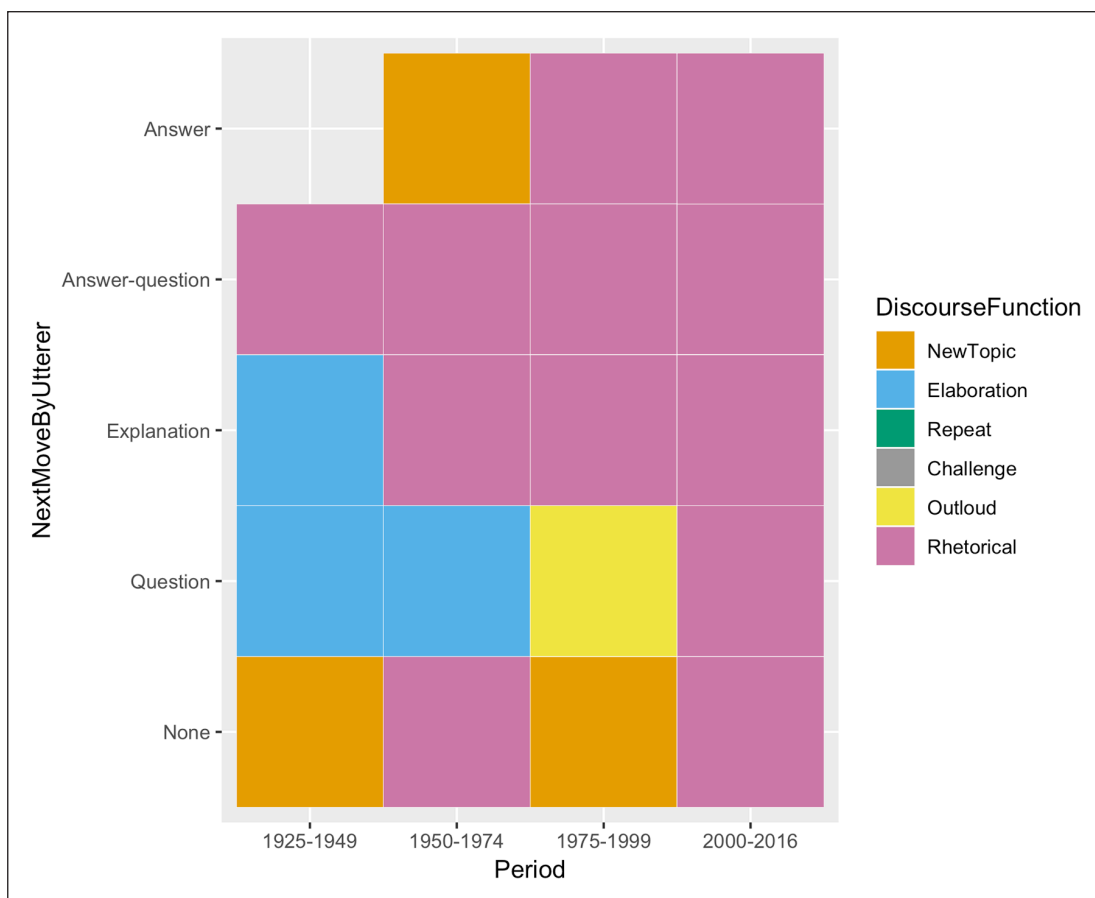**Figure 5:** Changes in the distribution of InSituWh with respect to the next move by the speaker.

Figure 6: Heat plot of the distribution of NextMoveByUtterer by discourse function of the InSituWh interrogative and 25-year period.

instance, in 1950–1974, elaboration uses of InSituWh are most frequently followed by another question.

Figure 6 demonstrates that between 1925 and 1974, the rhetorical readings of InSituWh typically arise in contexts in which the speaker follows up on her or his interrogative with a "questioning answer" (as in the examples (16) and (17)) or an explanation of her or his motivation to utter the interrogative. However, especially in the 21st century texts InSituWh also frequently acquires a rhetorical reading in contexts which are atypical for rhetorical question interpretations, i.e. contexts in which the speaker does not give an answer to the interrogative (18) or follows up on the interrogative with a second (rhetorical) question (19).[6]

(18)   *Navalha na carne,* Plínio Marcos, 1967
       Vado —   Eu estou   duro! Estou   a nenhum! Eu estou   a
                I   be.PRS.1SG broke be.PRS.1SG to nobody   I   be.PRS.1SG to
                zero! A   zero, sua vaca!
                zero to zero   you cow
                'I am broke! I have become nobody! I have become nothing! Nothing,
                you cow!'

       N. Sueli —   E   a   culpa é   minha?
                    and the guilt be.PRS.3SG mine

       Vado —   Vagabunda, miserável! Sua puta sem-calça!   Quem tu
                vagabond   idiot   you slut without-pants who   you
                pensa   que é? **Pensa   que estou**
                think.PRS.3SG that be.PRS.3SG think.PRS.3SG that be.PRS.1SG
                aqui por quê? Anda,   responde!
                here why   go.IMP answer.IMP
                'You vagabond, idiot, pantless slut! Who do you think you are? Why
                do you think I'm here? Come on, answer me!'

(19)   *Romeu & Julieta,* Marcondys França, 2012
       ROMEU – Mas não precisa   Paulão, você dormira   aqui na
                but not need.PRS.3SG paulão you sleep.FUT.3SG here in.the
                sala   mesmo…
                room itself
                'But you don't need [to pack your bags], Paulão, you can sleep right
                here in the living room'

       PAULÃO – O que Romeu? Na   sala? Mas claro   que não, já
                 what romeu in.the room but certain that not already
                 vou   levando   as malas para o   seu quarto onde
                 go.PRS.1SG carry.PROG the bags to   the your room where
                 dormirei,   **ta   pensando o quê?** Isso é
                 sleep.FUT.1SG be.PRS.3SG think.PROG what   this be.PRS.3SG
                 jeito de se   tratar uma visita pô?!?!
                 form of REFL treat a   visitor
                 'What, Romeu? In the living room? Certainly not, I am already
                 carrying the bags to your room where I will sleep, what are you
                 thinking? Is this how you treat a visitor?'

---

[6] Incidentally, example (19) demonstrates that ExSituWh interrogatives, such as *Quem tu pensa que é?*, can also express rhetorical questions. The analysis proposed in this paper would predict for the corpus of theater texts that while both InSituWh and ExSituWh can express this discourse function, it is more likely for InSituWh than for ExSituWh.

5.3 Is the change in activation of the proposition explained by discourse function?

The last step in the analysis aimed at corroborating in an inferential statistical analysis that the change from discourse-old to discourse-new propositions documented in Figure 2 does not result from the fact that InSituWh became more frequent in information question functions, but rather reflects the increase in the frequency of InSituWh in rhetorical question functions documented in Figure 3.

A generalized additive regression model (GAM, cf. Wieling et al. 2011) was calculated in which the parameter of activation of the proposition (variable Activation_P, with the levels 'False' and 'True') was predicted from the year of publication of the play (numerical predictor variable Year). This baseline model (α) was then compared to a second GAM (β), which additionally included the random effect Rhetorical referring to whether or not the InSituWh token expressed a rhetorical question.[7] In doing so, the analysis allowed evaluating to which degree the effect of Year on Activation_P is contingent on the predictor Rhetorical. The assumption was that if the effect of Year disappeared when controlling for Rhetorical, it could be assumed that Rhetorical is a better predictor of Activation_P than Year, implying that the observed change in activation is only due to the simultaneous increase in rhetorical uses of InSituWh.

The results from the two regression models are summarized in Figure 7. Both in the baseline model α (left plot) and the controlled model β (right plot) there is a negative correlation between the parameter of activation of the proposition and the year of publication. The baseline model judges this correlation to be stronger in the baseline model than in the controlled model; in the left plot, the probability of activation of P descends from 0.5 to –1.1 (a decrease of 1.6), whereas in the right plot, it merely descends from –0.1 to –0.5 (a decrease of 0.4) when taking into account the whole time span, which would indicate that the change is almost entirely explained by the use of InSituWh in rhetorical function. However, the controlled model also differs from the baseline model in that it displays a more informative curve; according to the controlled model, the decrease in the
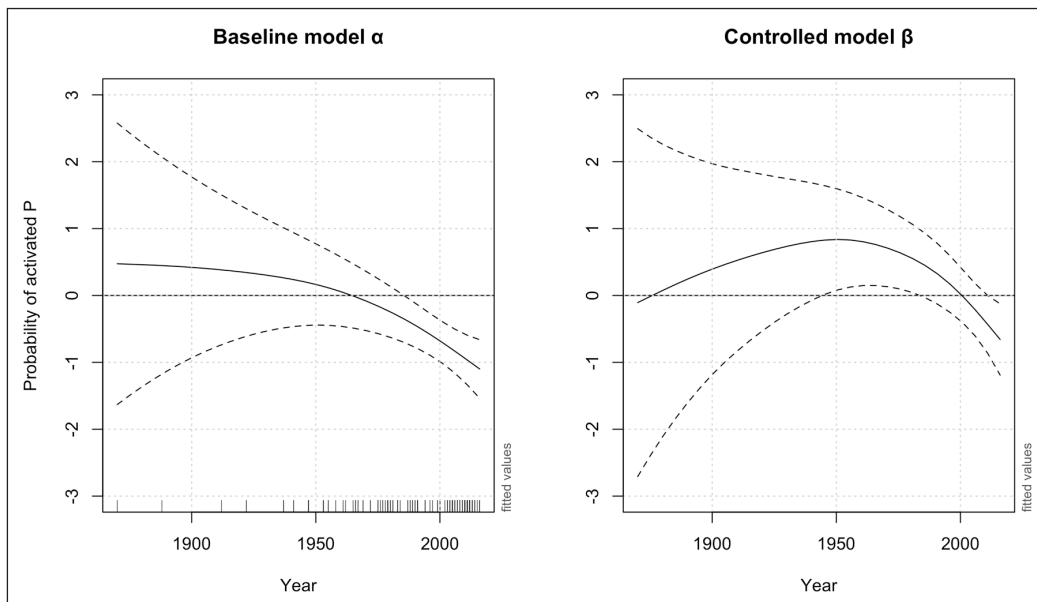

Figure 7: Probability of activation of proposition in InSituWh by year as predicted by two GAMs, a baseline model α (left) and a model β controlling for rhetorical uses of InSituWh.

---

[7] The GAM analysis was calculated and plotted using the packages mgcv (Wood 2018) and itsadug (Van Rij et al. 2017) in R.

probability that the proposition is activated is restricted to the time period between 1950 and 2016. In this time period, the trend is also relatively strong in the controlled model; the probability that the proposition is activated descends from about 0.8 in 1950 to –0.5 in 2016 (a decrease of 1.3).

Although due to the limited amount of data, the analysis is not very reliable (note the huge error bars indicated by the dotted lines in Figure 7), it does suggest that the decrease in the probability for the proposition of InSituWh tokens to be activated is at least partially due to the fact that InSituWh became more frequent in rhetorical question uses.

6 Discussion and conclusions

This paper set out to explore the historical development of InSituWh in a corpus of Brazilian Portuguese theater plays, testing the hypothesis that over time InSituWh became more common in usage contexts in which the interrogative proposition is discourse-new. This original hypothesis was found to be only superficially tenable. A closer inspection of the correlation between usage contexts and the discourse functions of InSituWh revealed that whereas discourse-old propositions typically lead to anaphorical discourse functions (i.e. asking for repetition or challenging a previous statement), both information and rhetorical question readings can arise when the proposition of the interrogative is discourse-new. The difference between information and rhetorical questions is not so much governed by the informational status of the interrogative proposition but by conditional relevance, i.e. whether the question necessitates an answer. A statistical analysis of the data demonstrated that this difference can however be measured in terms of a number of contextual predictors, namely whether or not the utterer of the interrogative allows turn-taking after the interrogative, the morphological complexity of the *wh*-element, whether or not the interrogative has a psychological predicate, the degree of orality of the text and possibly the use of address terms.

The result that the informational status of the proposition of the *wh*-interrogative is not a sufficient predictor of the difference between information and rhetorical question readings had important repercussions for the diachronic analysis of the data. The analysis indeed demonstrated a historical trend for BP InSituWh with respect to the informational status of the proposition from the use of InSituWh, from discourse-old to discourse-new propositions. However, this change is at least partially due to the rising usage frequency of rhetorical questions with discourse-new propositions, to the detriment of Elaboration questions, which represent the most frequent discourse function in the earliest plays.

The fact that the change to discourse-new interrogative propositions is contingent on rhetorical discourse functions can be taken to imply that genre-specific conventions played a crucial role for the attested historical development. Rhetorical questions represent highly expressive and stylized forms of discourse management, which explains why our analysis found rhetorical questions to be more frequent in low-orality, i.e. more formal, theater plays. It therefore seems probable that rhetorical questions are more frequent in all constructed dialogue than in natural spoken language. Consequently, it is doubtful to which degree the historical trend documented in the corpus of BP theater plays represents an actual change in spoken language (cf. also Rosemeyer to appear a). Although the more fine-grained statistical analysis using generalized additive mixed-effects modeling in Section 5.3 demonstrated the existence of a trend towards the use of non-activated proposition that is independent from the rhetorical question uses of InSituWh and which could consequently reflect a change in spoken language, the paucity of data makes it impossible to carry out a definitive assessment on this question. Further research on an even bigger corpus of BP theater texts is therefore necessary in order to reach a clearer evaluation in this matter.

The discourse functions of InSituWh described in this paper arise as situated meanings from the combination of the identified contextual parameters. There is no evidence that the use of InSituWh was barred from certain pragmatic usage contexts at any time in the corpus of theater plays. Consequently, the documented change in the distribution of InSituWh appears to rather represent probabilistic change than the actual creation or loss of a specific discourse function. In addition, the rhetorical question function is not restricted to InSituWh, but can also arise with other types of *wh*-interrogatives (see footnote 6). By applying diachronic variationist methodology to the constructional variation in Brazilian Portuguese *wh*-interrogatives, future studies could analyze whether the conventionalization of the rhetorical question reading in theater texts was restricted to InSituWh or can be found also in other types of *wh*-interrogatives.

To conclude, the description of the historical development of InSituWh in Brazilian Portuguese adds to the existing literature by demonstrating that the historical trajectory of *wh*-interrogatives does not always involve a change towards information question functions. The results from this case study rather suggest that in theater plays, BP InSituWh has gradually conventionalized a rhetorical discourse function. This finding is concordant with recent studies demonstrating that in specific communicative settings, *wh*-interrogatives can come to acquire exclamative discourse functions (see Auer 2016; Ehmer & Rosemeyer 2018).

Abbreviations

Constructional types of *wh*-interrogatives
CleftWh = clefted *ex-situ wh*-interrogative
ExSituWh = non-clefted *ex-situ wh*-interrogative
InSituWh = *in-situ wh*-interrogative

Linguistic glosses
PRS = present tense
PST = past tense
PFV = perfective
IPFV = imperfective
SBJ = subjunctive
INF = infinitive
PROG = progressive
1,2,3 = 1st, 2nd, 3rd person
SG = singular
PL = plural

Additional File
The additional file for this article can be found as follows:

- **Appendix.** Summary of the most important GAT 2 transcription conventions (Selting et al. 2011). DOI: https://doi.org/10.5334/gjgl.900.s1

Acknowledgements

References

Adli, Aria. 2006. French wh-in-situ questions and syntactic optionality: Evidence from three data types. *Zeitschrift für Sprachwissenschaft* 25. 163–203. DOI: https://doi.org/10.1515/ZFS.2006.007

Auer, Peter. 2016. "Wie geil ist das denn?". Eine neue Konstruktion im Netzwerk ihrer Nachbarn. *Zeitschrift für Germanistische Linguistik* 44(1). 69–92. DOI: https://doi.org/10.1515/zgl-2016-0003

Biber, Douglas & Edward Finegan. 2004 [1987]. Historical drift in three English genres. In Sampson, Geoffrey & Diana McCarthy (eds.), *Corpus linguistics: Readings in a widening discipline*, 67–77. London: Continuum.

Boucher, Paul. 2010. Wh-questions in French and English. Mapping syntax to information structure. In Breul, Carsten & Edward Göbbel (eds.), *Comparative and contrastive studies of information structure*, 101–138. Amsterdam, Philadelphia: Benjamins. DOI: https://doi.org/10.1075/la.165.05bou

Chang, Lisa. 1997. *WH-in-situ phenomena in French*. Vancouver: University of British Columbia dissertation.

Cheng, Lisa Lai-Shen. 2013. Wh-in-situ. *Glot International* 7(4). 103–109.

Cheng, Lisa Lai-Shen & Johan Rooryck. 2000. Licensing wh-in-situ. *Syntax* 3. 1–19. DOI: https://doi.org/10.1111/1467-9612.00022

Chernova, Ekaterina. 2015. *The syntax of wh-movement in multiple (true and echo) questions: A Q-particle approach*. Doctoral dissertation, Universitat de Girona.

Coveney, Aidan. 1990. Variation of interrogatives in spoken French. In Green, John N. & Wendy Ayres-Bennett (eds.), *Variation and change in French*, 116–133. London: Routledge.

Culpeper, Jonathan & Merja Kytö. 2000. Data in historical pragmatics: Spoken interaction (re)cast as writing. *Journal of Historical Pragmatics* 1(2). 175–199. DOI: https://doi.org/10.1075/jhp.1.2.03cul

Davies, Mark. 2006. *O corpus do português*. https://www.corpusdoportugues.org. Last access 2 April 2018.

De Paula, Mayara Nicolau. 2015. A ordem VS/SV em interrogativas-Q: um estudo diacrônico em peças teatrais brasileiras e portuguesas. In Baalbaki, Angela, Janaína Cardoso, Poliana Arantes & Sandra Bernardo (eds.), *Linguagem: Teoria, análise e aplicações*, 585–595. Rio de Janeiro: Programa de Pós-graduação em Letras.

De Paula, Mayara Nicolau. 2016. *A ordem VS/SV e as interrogativas-Q no PE e no PB: uma análise diacrônica*. Rio de Janeiro: Universidade Federal do Rio de Janeiro dissertation.

De Paula, Mayara Nicolau. 2017. A comparative diachronic analysis of wh-questions in Brazilian and European Portuguese. *Diadorim* 19. 173–196.

Dryer, Matthew S. 1996. Focus, pragmatic presuppositions, and activated propositions. *Journal of Pragmatics* 26(4). 475–523. DOI: https://doi.org/10.1016/0378-2166(95)00059-3

Duarte, Maria Eugênia L. 1992. A perda da ordem V(erbo) S(ujeito) em interrogativas qu-no português do Brasil. *D.E.L.T.A.* 8. 37–52.

Dumitrescu, Domnia. 1992. Estructura y función de las preguntas retóricas repetitivas en español. In Vilanova, Antonio (ed.), *Actas del X Congreso de la Asociación Internacional de Hispanistas (Barcelona, 21–26 de agosto, 1989)* IV. 1323–1338. Barcelona: Promociones y Publicaciones Universitarias.

Dumitrescu, Domnia. 1993. Función pragma-discursiva de la interrogación ecoica usada como respuesta en español. *Diálogos Hispánicos* 12. 51–86.

Dumitrescu, Domnia. 2008. Alo-repeticiones interrogativas en el español mexicano: funciones discursivas y estrategias de (des)cortesía. *Pragmatics* 18(4). 659–680.

Ehmer, Oliver & Malte Rosemeyer. 2018. When "questions" are not questions. Inferences and conventionalization in Spanish *but*-prefaced partial interrogatives. *Open Linguistics* 4. 70–100. DOI: https://doi.org/10.1515/opli-2018-0005

Elsig, Martin. 2009. *Grammatical variation across space and time. The French interrogative system.* Amsterdam, Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/silv.3

Escandell-Vidal, Victoria. 1999. Los enunciados interrogativos. Aspectos semánticos y pragmáticos. In Bosque, Ignacio & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española* 3. 3930–3991. Madrid: Espasa.

Escandell-Vidal, Victoria. 2002. Echo-syntax and metarepresentations. *Lingua* 112. 871–900. DOI: https://doi.org/10.1016/S0024-3841(02)00051-7

Fiengo, Robert. 2009. *Asking questions. Using meaningful structure to imply ignorance.* Oxford: Oxford University Press.

Fontes, Michel Gustavo. 2012a. *As interrogativas de conteúdo na história do português brasileiro: uma abordagem discursivo-funcional.* São Paulo: Universidade Estadual Paulista "Júlio de Mesquita Filho" dissertation.

Fontes, Michel Gustavo. 2012b. A clivagem do constituinte interrogativo em sentenças interrogativas do português brasileiro: uma abordagem diacrônica. *Estudos Linguísticos* 15(3). 149–170. DOI: https://doi.org/10.5433/2237-4876.2012v15n3p149

Ginzburg, Jonathan. 1996. Interrogatives: questions, facts and dialogue. In Lappin, Shalom (ed.), *The handbook of contemporary semantic theory*, 385–422. Oxford: Basil Blackwell. DOI: https://doi.org/10.1111/b.9780631207498.1997.00018.x

Greive, Artur. 1974. *Neufranzösische Formen der Satzfrage im Kontext.* Wiesbaden: Steiner.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10. 41–53.

Jacobs, Andreas & Andreas H. Jucker. 1995. The historical perspective in pragmatics. In Andreas H. Jucker (ed.), *Historical pragmatics. Pragmatic developments in the history of English*, 3–33. Amsterdam, Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/pbns.35.04jac

Kaiser, Egbert. 1980. *Strukturen der Frage im Französischen: Synchronische und diachronische Untersuchungen zur direkten Frage im Französischen des 15. Jahrhunderts (1450–1500).* Aachen: Narr.

Kato, Mary A. 2013. Deriving wh-in-situ through movement. In Camacho-Taboada, Victoria, Ángel L. Giménez Fernández, Javier Martín-González & Mariano Reyes-Tejedor (eds.), *Information structure and agreement*, 175–191. Philadelphia, Amsterdam: John Benjamins.

Kato, Mary A. 2014. Focus and wh-questions in Brazilian Portuguese. In Torres Cacoullos, Rena, Nathalie Dion & André Lapierre (eds.), *Linguistic variation. Confronting fact and theory*, 111–130. London: Routledge.

Kato, Mary A. & Carlos Mioto. 2005. A multi-evidence study of European and Brazilian Portuguese wh-questions. In Reis, Marga & Stephan Kepser (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives*, 307–328. Berlin, New York: De Gruyter. DOI: https://doi.org/10.1515/9783110197549.307

Kato, Mary A. & Ilza Ribeiro. 2005. *Cleft sentences and wh-questions in Brazilian Portuguese: A diachronic analysis*. Campinas: Editora da Unicamp.

Kato, Mary A. & Ilza Ribeiro. 2009. Cleft sentences from Old Portuguese to Modern Portuguese. In Dufter, Andreas & Daniel Jacob (eds.), *Focus and background in Romance languages*, 123–154. Amsterdam, Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/slcs.112.06kat

Klein, Wolfgang & Christiane von Stutterheim. 1987. Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte* 109. 163–183.

Kytö, Merja. 2011. Corpora and historical linguistics. *Revista Brasileira de Linguística Aplicada* 11(2). 417–457. DOI: https://doi.org/10.1590/S1984-63982011000200007

Larrivée, Pierre. To appear. Historical pragmatics, explicit activation and wh in situ in French. In Feldhausen, Ingo, Martin Elsig, Imme Kuchenbrandt & Mareike Neuhaus (eds.), *Romance languages and linguistic theory 15: Selected papers from 'Going Romance' 30, Frankfurt*. Amsterdam, Philadelphia: John Benjamins.

Linnell, Per. 2009. *Rethinking language, mind and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte, NY: Information Age Publishing.

Lopes-Rossi, Maria A. G. 1993. Estudo diacrônico sobre as interrogativas do português do Brasil. In Roberts, Ian & Mary A. Kato (eds.), *Português Brasileiro: uma viagem diacrônica. Homenagem a Fernando Tarallo*, 307–342. Campinas: Editora da Unicamp.

Lopes Rossi, Maria Aparecida. 1996. *A sintaxe diacrônica das interrogativas-Q do português*. Campinas: Universidade Estadual de Campinas dissertation.

Mathieu, Eric. 1999. WH in situ and the intervention effect. *UCL Working Papers in Linguistics* 11. 441–474.

Myers, Lindsy Lee. 2007. *WH-Interrogatives in spoken French: A corpus-based analysis of their form and function*. Austin: University of Texas at Austin dissertation.

Obenauer, Hans-Georg. 1994. *Aspect de la syntaxe A-barre – Effets d'intervention et mouvement des quantifieurs*. Paris: Université de Paris VIII dissertation.

Oushiro, Livia. 2010. Condicionamento discursivo-pragmático no uso variável de Interrogativas-Q. *Estudos Linguísticos* 39. 628–639.

Oushiro, Livia. 2011a. *Um análise variacionista para as Interrogativas-Q*. São Paulo: Universidade de São Paulo dissertation.

Oushiro, Livia. 2011b. Wh-interrogatives in Brazilian Portuguese: The influence of Common Ground. *University of Pennsylvania Working Papers in Linguistics* 17(2). 145–154.

Pinheiro, Diogo & Juliana Marins. 2012. A trajetória das interrogativas QU- clivadas e não clivadas no Português Brasileiro. In Duarte, Maria Eugênia L. (ed.), *O sujeito em peças de teatro (1833–1992): Estudos diacrônicos*, 161–179. São Paulo: Parábola.

Pires, Acrisio & Heather L. Taylor. 2007. The syntax of wh-in-situ and Common Ground. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 43(2). 201–215.

R Development Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Accessible online at https://www.R-project.org. Last access 17 January 2019.

Raso, Tommaso & Heliana Mello. 2012. The C-ORAL BRASIL I: reference corpus for informal spoken Brazilian Portuguese. In Caseli, Helena, Aline Villavicencio, António Teixeira & Fernando Perdigão (eds.), *Computational processing of the Portuguese lan-*

guage, 362–367. Dordrecht, Kluiwer: Springer. DOI: https://doi.org/10.1007/978-3-642-28885-2_40

Riester, Arndt, Lisa Brunetti & Kordula De Kuthy. 2018. Annotation guidelines for Question under Discussion and information structure. In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in lesser-described languages. Studies in prosody and syntax*, 403–443. Amsterdam, Philadelphia: John Benjamins.

Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. In Jae Haak Yoon & Andreas Kathol (eds.), *OSUWPL volume 49: Papers in semantics*, 91–136. Columbus, OH: The Ohio State University Department of Linguistics.

Rohde, Hannah. 2006. Rhetorical questions as redundant interrogatives. *San Diego Linguistics Papers* 2. 134–168.

Rosemeyer, Malte. 2018a. The pragmatics of Spanish postposed-wh-interrogatives. *Folia Linguistica* 52(2). 283–317. DOI: https://doi.org/10.1515/flin-2018-0007

Rosemeyer, Malte. 2018b. Las funciones discursivas de las interrogativas parciales en castellano. *Romanistisches Jahrbuch* 69. 277–314. DOI: https://doi.org/10.1515/roja-2018-0017

Rosemeyer, Malte. 2018c. PorThea: a historical corpus of Portuguese theater plays. Available online at http://www.romanistik.uni-freiburg.de/rosemeyer/05corpus.html. Last access 27 May 2019.

Rosemeyer, Malte. To appear a. Actual and apparent change in Brazilian Portuguese *wh*-interrogatives. *Language Variation and Change*.

Rosemeyer, Malte. To appear b. A história das interrogativas parciais *in-situ* no português brasileiro em uma perspectiva baseada no uso. *LaborHistórico*.

Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist* 70. 1075–1095. DOI: https://doi.org/10.1525/aa.1968.70.6.02a00030

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. 2011. A system for transcribing talk-in-interaction: GAT 2. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 12. 1–51.

Stalnaker, Robert. 2002. Common Ground. *Linguistics and Philosophy* 25. 701–721. DOI: https://doi.org/10.1023/A:1020867916902

Stivers, Tanya & Nicholas J. Enfield. 2016. A coding scheme for question–response sequences in conversation. *Journal of Pragmatics* 42. 2620–2626. DOI: https://doi.org/10.1016/j.pragma.2010.04.002

Van Rij, Jacolien, Martijn Wieling, Harald Baayen & Hedderik Van Rijn. 2017. itsadug: Interpreting time series and autocorrelated data using GAMMs. *R package version 2.3*, accessible online at https://CRAN.R-project.org/package = itsadug. Last access 4 April 2019.

Waltereit, Richard. 2018. Inferencing, reanalysis, and the history of the French *est-ce que* question. *Open Linguistics* 4. 35–48. DOI: https://doi.org/10.1515/opli-2018-0004

Wieling, Martijn, John Nerbonne & Harald Baayen. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, 6(9). e23613. DOI: https://doi.org/10.1371/journal.pone.0023613

Wood, Simon. 2018. mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. *R package version 1.8-28*, accessible online at https://CRAN.R-project.org/package = mgcv. Last access 4 April 2019.