Language Variation and Change, **31** (2019), 165–191.

doi:10.1017/S0954394519000097

Actual and apparent change in Brazilian Portuguese *wh*-interrogatives

MALTE ROSEMEYER

KU Leuven (Belgium) / Albert-Ludwigs-Universität Freiburg (Germany)

ABSTRACT

Previous studies on the diachrony of *wh*-interrogation in Brazilian Portuguese have observed a replacement process of ex-situ-*wh* interrogatives by cleft-*wh* and in-situ-*wh* interrogatives in the twentieth century. The present study analyzes almost 19,000 *wh*-interrogatives from a corpus of theater plays dated between 1800 and 2016, demonstrating that not all of these frequency changes constitute actual change. The increase in the usage frequency of several types of *wh*-interrogatives is partially or entirely due to changes in the degree of orality of theater plays, or changes in word order. Moreover, only some of these changes can be characterized as changes from below, that is, changes in which high-orality texts are affected by the frequency increase first. This notion is also relevant for functional change in *wh*-interrogatives. Over time, the use of cleft-*wh* and in-situ-*wh* interrogatives spread from contexts in which the proposition is highly accessible to low-accessibility contexts. For cleft-*wh*, this change is moderated by orality, again indicating change from below.

Present-day Brazilian Portuguese (BP) possesses several *wh*-interrogative constructions. In the correct pragmatic context, a sentence like ‘Where did you go?’ can be expressed in at least five ways (1).[1]

(1) a. *Onde você foi?* [EXSITUWH]
where you go.PST.PFV.3SG

b. *Onde é que você foi?* [CLEFTWH]
where be.PRS.3SG that you go.PST.PFV.3SG

c. *Onde que você foi?* [REDUCEDCLEFTWH]
where that you go.PST.PFV.3SG

d. *Você foi (pra) onde?* [INSITUWH]
you go.PST.PFV.3SG (to) where

e. *Onde?* [BAREWH]
where

I am grateful to the audience at the workshop “Information Structure and Language Change” in Caen, Scott Schwenter, Freek Van de Velde, and the two anonymous reviewers at LVC for their valuable comments on a previous version of this paper. I would also like to express my profound gratitude to Célia Regina dos Santos Lopes and Maria Eugênia Lammoglia Duarte for very helpful discussions and sharing their corpora with me. This research was funded by the Research Foundation-Flanders (FWO) in the context of the research project “Variation and change in Spanish and Portuguese partial interrogatives” (12N1916N).

Previous studies have demonstrated that the usage frequency of CleftWh (1b) and InSituWh (1d) has increased over time in BP, to the detriment of ExSituWh (1a). Likewise, since the second half of the twentieth century ReducedCleftWh, that is, reduced cleft constructions (1c), are attested and extremely frequent in spoken language.[2]

However, changes in text frequencies can be due to apparent change reflecting environmental changes in the text genre (cf., for example, Szmrecsanyi [2016]). These considerations are nontrivial to the study of the changes in the Portuguese system of *wh*-interrogatives, because the spoken-written dimension plays a crucial role for the variation in the use of *wh*-interrogatives. It is well known that, in Indo-European languages such as French, CleftWh and InSituWh constructions are more frequent in spoken than in written language and also display greater pragmatic flexibility (Armstrong, 2001; Elsig, 2009; Kaiser & Quaglia, 2015; Mathieu, 2004). Although almost all of the previous studies on changes in Portuguese *wh*-interrogatives analyze theater texts, a genre that might represent spoken language more accurately than, for example, prose, *a priori* we cannot exclude the possibility that the increase in the usage frequencies of CleftWh, InSituWh, and ReducedCleftWh is due to genre change in these theater texts.

The aim of this paper is to answer the question whether actual grammatical change has taken place in the Portuguese system of *wh*-interrogatives. I analyze almost 19,000 *wh*-interrogatives from a corpus of theater plays dated between 1800 and 2016. After a discussion of the problem of actual and apparent change in Portuguese *wh*-interrogatives and a description of the data used, I provide an overview of the overall changes in usage frequency of these interrogatives. The subsequent analysis demonstrates that not all of these changes constitute actual change. By controlling for the degree of orality of the texts, three types of change are identified: (i) apparent change, that is, change that is due to the rising degree of orality in BP theater plays; (ii) actual change "from below", i.e., reflecting social conventionalization processes in the speaker community; and (iii) genre change that is independent from orality. The analysis also demonstrates that word order had an important influence on the development of the distribution of ExSituWh and CleftWh. In a further step, the changes in the usage contexts of CleftWh and InSituWh interrogatives are analyzed, demonstrating that (a) there is an increase in the probability of CleftWh and InSituWh to be used in contexts in which the proposition has a low degree of accessibility, and (b) for CleftWh, this increase is moderated by orality.

THE PROBLEM OF ACTUAL AND APPARENT CHANGE IN PORTUGUESE *WH*-INTERROGATIVES

Several previous diachronic studies document changes in the BP and European Portuguese (EP) system of partial interrogatives (De Paula, 2015, 2016, 2017; Duarte, 1992; Fontes, 2012a, 2012b; Kato, 2014; Kato & Mioto, 2005; Kato &

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

Ribeiro, 2009; Lopes Rossi, 1996; Pinheiro & Marins, 2012). There are changes regarding (a) the availability and usage frequency of construction types and (b) the expression and placement of subject constituents within these *wh*-interrogatives, although many of these studies conflate the two factors because (b) is taken to be the cause of (a). Lopes Rossi (1996:44–48; 68) documents an increase of CLEFTWH constructions for EP and BP in the twentieth century. Lopes Rossi's results also suggest a strong increase in the use of INSITUWH in BP from zero attestations in the first half of the nineteenth century to a relative usage frequency of 38 percent in the second half of the twentieth century, but a much less pronounced increase in EP to three percent in the second half of the twentieth century. Unsurprisingly, this increase in the usage frequency of "marked" *wh*-interrogative constructions coincided with a decrease of the relative usage frequency of EXSITUWH. While partially reproducing Lopes Rossi's results, De Paula (2016) finds a stronger increase in the usage frequency of CLEFTWH in EP (documenting an increase to 32 percent in the second half of the twentieth century), while Pinheiro and Marins (2012) finds a less strong increase of INSITUWH in BP.

The great majority of these studies are not so much interested in the development of the competition between these constructional types as in the expression and placement of subject constituents within *wh*-interrogatives, demonstrating an increase in both the expression of subjects (i.e., a loss of null subjects) and in SV word order compared to VS word order. What is more, the majority of these analyses are based on datasets of relatively limited size. Given the low numbers of tokens, well under 1,000 per language variety, and the fact that entire periods are frequently represented by one or two texts, the fluctuations in the results of these studies are not surprising. Crucially for my argumentation, this problem is exacerbated by the fact that *wh*-interrogatives essentially represent a pragmatic phenomenon that is very much governed by the rules of spoken interaction. As a result, there are great differences between spoken and written texts in (a) the distribution of constructional types of *wh*-interrogatives and, relatedly, (b) the functions that *wh*-interrogatives are used for. Regarding the first point, consider Oushiro's (2011) comparison of the use of *wh*-interrogatives in spoken (sociolinguistic interviews) and written texts (newspaper articles and student essays) in São Paulo, summarized in Table 1.[3] The use of all marked types of *wh*-interrogatives is vastly more frequent in spoken language than in the written texts, in which as much as 96 percent of the *wh*-interrogatives correspond to the EXSITUWH type. Although theater plays, used by all of the diachronic studies mentioned above, doubtlessly represent spoken language better than other types of written texts, they are still written texts and consequently more affected by standardization processes than spoken language. None of the diachronic studies based on corpora of theater plays mentioned above report a similarly high usage frequency of REDUCEDCLEFTWH constructions in theater plays after the 1980s.

Given that the distribution of types of *wh*-interrogatives is strongly dependent on the distinction between spoken and written language, any change in the register of theater plays affecting the degree to which these texts obey current linguistic norms

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

TABLE 1. *Distribution of wh-interrogative types in spoken and written Present-day BP (data from Oushiro [2011:33, 35])*

	Sociolinguistic interviews		Newspaper articles and student essays	
	n	%	*n*	%
ExSituWh	721	44	1168	96
ReducedCleftWh	579	35	14	1
CleftWh	121	7	16	1
InSituWh	227	14	19	2
Total	1648	100	1217	100

is bound to have had a profound influence on the distribution of types of *wh*-interrogatives in these texts.

This problem can be framed in terms of the difference between actual and apparent change, proposed in Szmrecsanyi (2016). Szmrecsanyi argues that frequency changes in historical corpora do not always reflect actual grammar change (in his definition, change in either the repertoire of structural units or probabilistic constraints on the use of these structural units) but may reflect apparent, that is, environmental, change. The author analyzes the development of the genitive alternation in English. Like other studies, he observes a decrease in the usage frequency of the *s*-genitive (relative to the *of*-genitive) between 1675 and 1825, followed by an increase until 1970 to higher levels than at the beginning of the change. The alternation between the *s*- and *of*-genitive is governed by the animacy of the possessor. Szmrecsanyi demonstrates that the curious drop in relative frequency of the *s*-genitive is in part due to changes in the overall frequency of animate noun phrases.

In line with Szmrecsanyi's proposal, *prima facie* there is no way of knowing whether the changes in the distribution of *wh*-interrogatives constitute actual grammar changes as long as we do not rule out the possibility that the increase in the usage frequency of non-canonical types of *wh*-interrogatives is due to changes in the degree of formality of theater plays, that is, environmental change.

A second way in which the notion of actual and apparent change might be relevant concerns the discourse function of these *wh*-interrogatives. Consider again Oushiro's (2011) study of the variation between the different constructional types of *wh*-interrogatives in Present-Day BP. Using multivariate statistical analysis, Oushiro demonstrates that the use of InSituWh interrogatives (see 1c) is favored in a so-called "discourse-continuing" function in which the speaker himself or herself gives an answer to the question, as in example (2), in contrast to information questions and rhetorical questions in which no answer is required (cf., also Kato [2013] for a syntactic motivation of the different discourse functions of InSituWh).

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

(2) Informal sociolinguistic interview (between 2003 and 2008), *apud* Oushiro (2011:101)

Marco: *então quer dizer… **isso daí prejudica quem?** …não prejudica o professor… ela tá lá ganhando o dinheiro dele… prejudica você que é o aluno… entendeu?*

'So you want to say… this is bad for who? It is not bad for the professor… she is earning his money… It is bad for you who is the student, you understand?'

In theater plays, the distribution of the discourse functions of *wh*-interrogatives might be expected to depend on the level of formality. For instance, more formal theater genres typically rely more on monologues than on dialogues, which is why one would expect more rhetorical questions and possibly discourse-continuing questions in these types of plays. While many of the studies mentioned above try to control for this problem by only including comedic plays, it stands to reason that such genre changes affect these corpora as well. In parallel to Srmrecsanyi's analysis of the genitive alternation in English, it is thus in principle possible that changes in the distribution of the different constructional types of *wh*-interrogatives are due to the frequency with which certain discourse functions are expressed in the plays.

In summary, there is a lacuna in the research on the development of the system of Portuguese *wh*-interrogatives, in that previous studies (a) are based on datasets of rather limited size and have not addressed the problem of actual and apparent change; and (b) have not studied whether the overall changes in the distribution of the *wh*-interrogative constructions were accompanied by changes in the functions of these constructions. The present study addresses exactly these points, and, in doing so, proposes a principled way of distinguishing between actual and apparent change that can also be applied to other phenomena and languages.

DATA

Corpus construction

As in the previous studies mentioned, the analyses reported here were conducted on a self-compiled corpus of Portuguese theater plays (Rosemeyer, 2018b). This is because theater plays are the only text type with time depth in which representations of direct speech are frequent enough to allow for quantitative analyses. Given that existing historical corpora such as the *Corpus do português* (Davies, 2006) and the *Tycho Brahe* corpus (Galves, De Andrade, & Faria, 2017) do not contain a sufficient number of theater plays, a new corpus of theater plays was constructed on the basis of texts, dated between 1800 and 2016, available from existing corpora, as well as electronic databases of modern Portuguese plays. Table 2 summarizes the distribution of the data across the three centuries. Although the BP section of the corpus is almost five times as

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

TABLE 2. *Summary statistics for the corpus of Portuguese theater plays*

		Nineteenth c.	Twentieth c.	Twenty-first c.	Total
Brazil	n_{words}	787015	740389	947900	2482610
	n_{plays}	82	63	153	298
Portugal	n_{words}	269338	127604	140188	537130
	n_{plays}	21	15	22	58

large as the EP section, in no century is the total number of words lower than 120,000 words. The asymmetry in the sizes of the BP and EP corpora means that the results will be much more reliable for the BP than for the EP data, though with a total of 58 plays, the EP section of this corpus is bigger than the EP corpora in any previous study.

Search queries and data elimination procedures

In a first step, all tokens of *wh*-interrogatives were extracted using regular expressions. The query identified all instances of the interrogative pronouns or adverbs in (3) followed by a question mark before encountering a full stop (i.e., "." or "!"). Because, as in other Romance languages, most of the Portuguese interrogative pronouns or adverbs can also be used as complementizers, the overall number of 140,000 tokens returned from the queries without the restriction to sentences marked as questions was too high to allow for manual coding. The restricted query still led to an extraction of more than 34,000 cases.

(3) *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'

In a second step, I manually eliminated all of the tokens in which the pronoun was in fact a complementizer (for instance, CLEFTWH constructions such as *o que é que você quer?* 'what is it that you want?' include the form *que* 'what/that' twice, as an interrogative pronoun and a complementizer).[4] Thirdly, I eliminated a number of contexts in which the use of one or more types of *wh*-interrogatives is impossible for syntactic reasons; these contexts are indirect interrogatives and syntactic islands (as proposed in Oushiro [2011:56–67]).[5] The result of the extraction process was a total number of $n = 18{,}903$ tokens of direct *wh*-interrogatives ($n_{\text{BP}} = 15{,}783$ [83,5%], $n_{\text{EP}} = 3120$ [16.5%]).

OVERALL DEVELOPMENT OF THE DISTRIBUTION OF VARIANTS

Before describing the development of the distribution of variants, it is necessary to introduce a further type of *wh*-interrogatives, not included in the previous list in

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

(1a–e) and undescribed in previous historical studies, which I encountered in the process of data collection. I give three early examples of this type, which I call BareXWh, in (4–6).

(4) *As casadas solteiras*, Martins Pena, 1845
NARCISO - Sim, sim, e podereis então casar-vos de novo com quem quiserdes.
'Yes, yes, and then you will be able to re-marry whoever you like.'
VIRGÍNIA - Casarmo-nos de novo?
'Remarry?'
NARCISO - **E por que não?**
'And why [should you] not?'

(5) *O cigano*, Martins Pena, 1845
BÁRBARA [e] SILVÉRIA - Ah! (Caem desmaiadas nos braços dos amantes.)
'Ah! (They fall unconscious into the arms of their lovers)
ANSELMO - **O que isto**, está a morrer?
'What [is] this, is she dying?'

(6) *A falecida*, Nelson Rodrigues, 1953
TIMBIRA (pigarreando) - Mas é casada?!
'(clears throat)' 'But are you married?'
ZULMIRA - Sou, sim!
'Yes I am!'
TIMBIRA - **Cadê a aliança?**
'Where [is] your wedding ring?'
ZULMIRA - Não uso.
'I don't use it.'

With $n = 744$ tokens, BareXWh interrogatives are more frequent than InSituWh and ReducedCleftWh interrogatives. They can be described as a subtype of BareWh interrogatives in that they do not involve a verb phrase and their interpretation depends on an inferred proposition, indicated in the glosses of the examples with square brackets.[6] However, they differ from BareWh in that they do involve a constituent, such as *não* (4), *isto* (5), or *a aliança* (6), over which the interrogative pronoun has scope.

Figure 1 summarizes the development of the log-transformed normalized usage frequencies of all of the relevant types of *wh*-interrogatives in the BP corpus. The gray dots represent frequency by year (in turn representing one or more plays from that year), whereas the thick lines illustrating the general trends in the data represent estimated values from local polynomial regressions fitted using the function loess() in R.[7] The scale of the y-axis has been adjusted to the range of the frequencies for each of the constructional types, which is why the scales on the y-axes differ. Consequently, one has to bear in mind that, for example, the increase (and fall) in usage frequency is much stronger for CleftWh than for InSituWh.

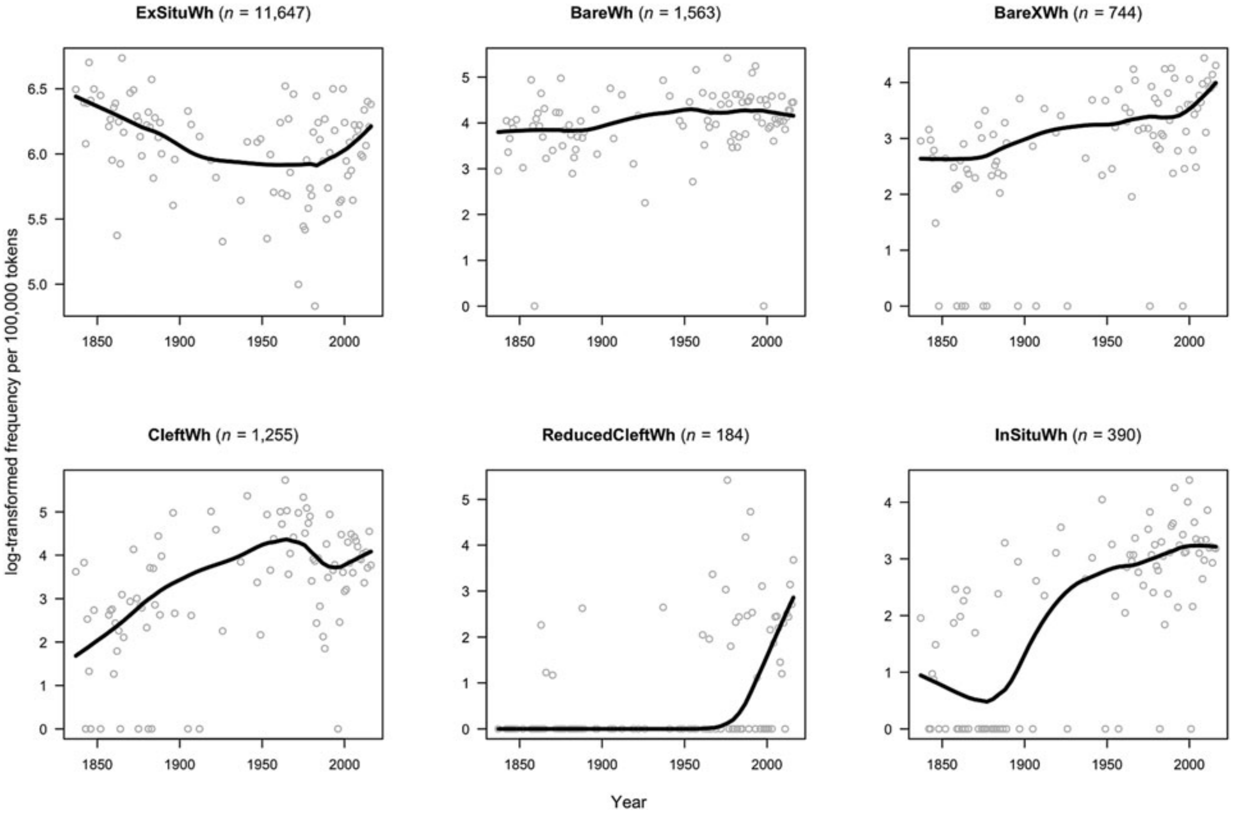

FIGURE 1. Log-transformed normalized frequencies of *wh*-interrogative constructions in BP theater plays by time.

The results illustrated in Figure 1 can be described as follows. ExSituWh interrogatives constitute the default type of *wh*-interrogative in all time periods, despite a strong decrease in their usage frequency between 1800 and 1970. The use of BareWh, the *wh*-interrogative construction with the second highest frequency, remains relatively constant until the beginning of the twentieth century, when it starts to increase, only to remain constant at this plateau until 2016. Regarding BareXWh interrogatives, there is a strong and steady increase from 1900 to 2016. With respect to CleftWh, its use is marginal in the nineteenth century. It is only at the beginning of the twentieth century that we witness a strong increase in its frequency—until about 1970 after which it experiences a slight drop in usage frequency. Another trend starting in the decade of the 1970s is the increased frequency of ReducedCleftWh constructions, virtually nonexistent in the corpus until then. The use of InSituWh is also marginal in the nineteenth century but starts to increase after the beginning of the twentieth century. All of these frequency changes are statistically significant.[8]

To summarize, there seems to have been an increase in the usage frequency of BareXWh, ReducedCleftWh, InSituWh and, to a lesser extent, BareWh constructions in the twentieth century, as well as somewhat curious developments for ExSituWh and CleftWh constructions, which follow u-shaped curves. Two time periods appear to be crucial for the development of *wh*-interrogatives in BP in that apparently, the frequency trajectories of several *wh*-interrogative constructions are correlated. First, in the first half of the twentieth century, we witness a rise in the use of BareXWh, InSituWh, CleftWh, and BareWh, as well as a fall in the use of ExSituWh. Second, after the 1960s, we document the creation of ReducedCleftWh and a simultaneous decrease in the use of CleftWh constructions as well as a recuperation of the use of ExSituWh.

PREDICTORS OF THE CHANGES IN USAGE FREQUENCY

Let us begin by examining the joint rise in usage frequency of InSituWh, CleftWh, BareXWh, and BareWh in the first half of the twentieth century. While the first two changes have already been observed in previous studies, the latter two are undescribed. As it turns out, the change in the usage frequency of BareWh interrogatives is an important hint regarding the question of whether or not actual change has taken place.

BareWh interrogatives such as *Onde?* 'Where?' differ from other types of *wh*-interrogatives in that they have neither a verb phrase nor a subject. This syntactic fact has repercussions for their pragmatics. The use of BareWh interrogatives can be said to rely on inference or maybe structural latency (Auer, 2014:14–18), in that a full interpretation is only possible when the proposition of the interrogative is recoverable from a previous utterance. Consider the simple

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

example in (7). Here, *Para quem?* ‘At who?’ actually receives the interpretation ‘Who was she looking at?’.

(7) *A mulher sem pecado*, Nelson Rodrigues, 1941

UMBERTO (com intenção) - ‘(with hidden agenda)’	Ela estava olhando de vez em quando… ‘She was looking from time to time…’
OLEGÁRIO -	**Para quem?** Diga! ‘At who? Tell me!’
UMBERTO (com descaramento) - ‘(with insolence)’	Para mim. ‘At me.’

Due to their syntactic simplicity, BareWh interrogatives are extremely limited regarding possible usage contexts, being virtually impossible in contexts in which the proposition is not accessible in the immediately previous co-text. Their syntactic limitations prohibit change whereby there would be a spread from contexts in which the proposition is more accessible to contexts in which it is less accessible, a change that we document for CleftWh and InSituWh (see the section Changes in the usage contexts of CleftWh and InSituWh below).

Thus, there is no reason to assume that in a language like Portuguese the use of BareWh became more frequent over time. It seems unlikely that in informal spoken language, nineteenth century speakers of BP used BareWh less frequently than twenty-first century speakers. Rather, I would like to propose that the documented significant increase in the usage frequency of BareWh interrogatives is due to environmental change in the corpus, that is, genre change as BP plays decreased in formality. Given that the frequency increases for CleftWh and InSituWh in the first half of the twentieth century coincided with the frequency increase of BareWh, one might suspect that the rise of CleftWh and InSituWh is also due to genre change.

In order to assess this assumption, I established a measurement of the degree to which the plays in the BP corpus represent orality by using Biber and Finegan’s (2004 [1987]:68) dimension of “involvement” of the oral/literate dimensions of variation, a measure with five linguistic variables (listed in Table 3) that apply to Portuguese and that are easy to extract in a summary fashion. These five linguistic variables represent orality because their use is dependent 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 and Finegan call private verbs). Both realizations typical for EP (for example, *estar* + infinitive progressives) and BP (for example, *estar* + gerund progressives) were included in order to capture all variants in all temporal periods.

As proposed in Biber and Finegan’s study, I aggregated the frequencies of the five variables for each text in a variable “Orality.” Figure 2 illustrates the development of the log-transformed normalized frequency of this variable in the corpus of BP theater plays. As in Figure 1, each point in the plot represents a year.

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

TABLE 3. *Linguistic variables used to measure the degree of orality in 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

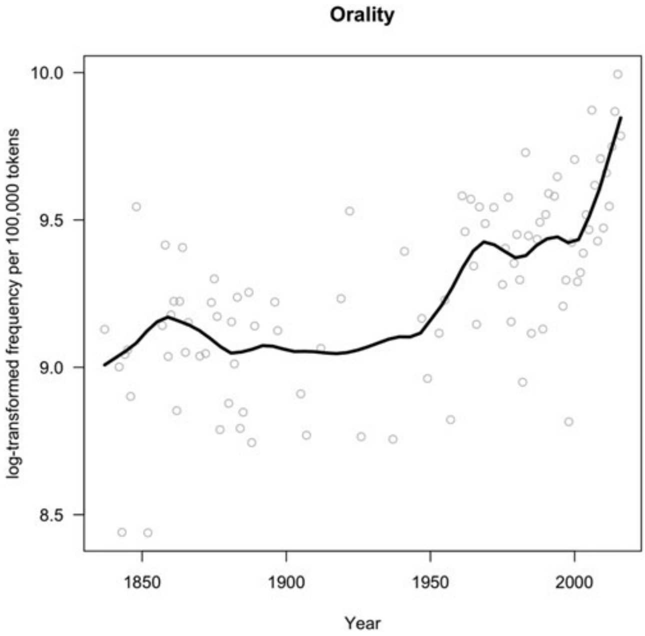

FIGURE 2. Aggregated log-transformed normalized frequencies of five linguistic variables representing orality in the corpus of BP theater plays by time.

As evident in Figure 2, there is a significant increase in the degree of orality as represented by the aggregated usage frequencies of the five linguistic variables.[9] Specifically, there is a small increase in the mean degree of orality between 1830 and 1950. After 1950, this trend picks up considerable speed, reaching the highest levels of orality in the twenty-first century plays. The change in the orality dimension strongly suggests that a genre change has taken place;

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

Brazilian playwrights have come to represent oral speech more accurately over time.

There are strong correlations between orality and the usage frequencies of the *wh*-interrogative types. Figure 3 plots the usage frequencies of the six types of *wh*-interrogatives (y-axis) against the usage frequency of Orality (x-axis). Each point represents one text in the corpus of BP theater plays. Whereas the correlation is not significant for ExSituWh, all other types of *wh*-interrogatives are more frequent in texts scoring high on the Orality variable (as indicated by the regression lines in the plots).[10]

Given that (a) the use of the less frequent *wh*-interrogative types is more frequent in texts scoring high on the Orality variable, and (b) there is an overall increase of texts scoring high on the Orality variable, it stands to reason that the documented overall increase of the usage frequencies of the marked *wh*-interrogative constructions is at least partially due to genre change. For Figure 4, I divided the corpus into a subcorpus of high orality texts and one of low orality texts (that is, the score of a text on the Orality variable was higher and lower, respectively, than the mean of the Orality variable).

The figure demonstrates a clear influence of the orality dimension on the development of most *wh*-interrogative constructions. At least three different types of change can be discerned. First, orality seems to "cushion" the decrease in frequency of ExSituWh, in that the overall decrease is much less strong in high-orality texts than in low-orality texts. Second, for BareXWh, CleftWh, and ReducedCleftWh, we observe a "hump" distribution that, in fact, corresponds to successive s-curves; the frequency increases in low-orality texts are preceded by frequency increases in high-orality texts. Third, although the use of both InSituWh and BareWh interrogatives is more frequent in high-orality texts, the frequency changes in low-orality and high-orality texts mostly run parallel.

Let us begin by discussing the most salient of these distributions, the "hump" distribution changes experienced by BareXWh, CleftWh, and ReducedCleftWh. It seems reasonable to assume that such hump-like changes represent social conventionalization, that is, the diffusion or propagation of an innovation in a speaker community (see, for example, Croft, 2000: chapter 7; Labov, 1994; Schmid, 2015; Weinreich, Labov, & Herzog, 1968). In other words, these results suggest that, in a first step, a spread of these constructions occurred in spoken interactions as represented in higher orality texts at the beginning (BareXWh and CleftWh) or in the second half of the twentieth century (ReducedCleftWh). With the successive diffusion of the innovative *wh*-interrogative constructions they came to be gradually accepted also in more stylized texts scoring lower on the orality dimension. Whereas the first process represents a reflection of co-adaptation in spoken language, that is, "the phenomenon that speakers show a certain tendency to take over and repeat linguistic material produced by their interlocutors earlier on in a given talk exchange" (Schmid, 2015:17), the diffusion of the innovative forms to more formal texts rather represents a change in the writing norms. Consequently, this

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

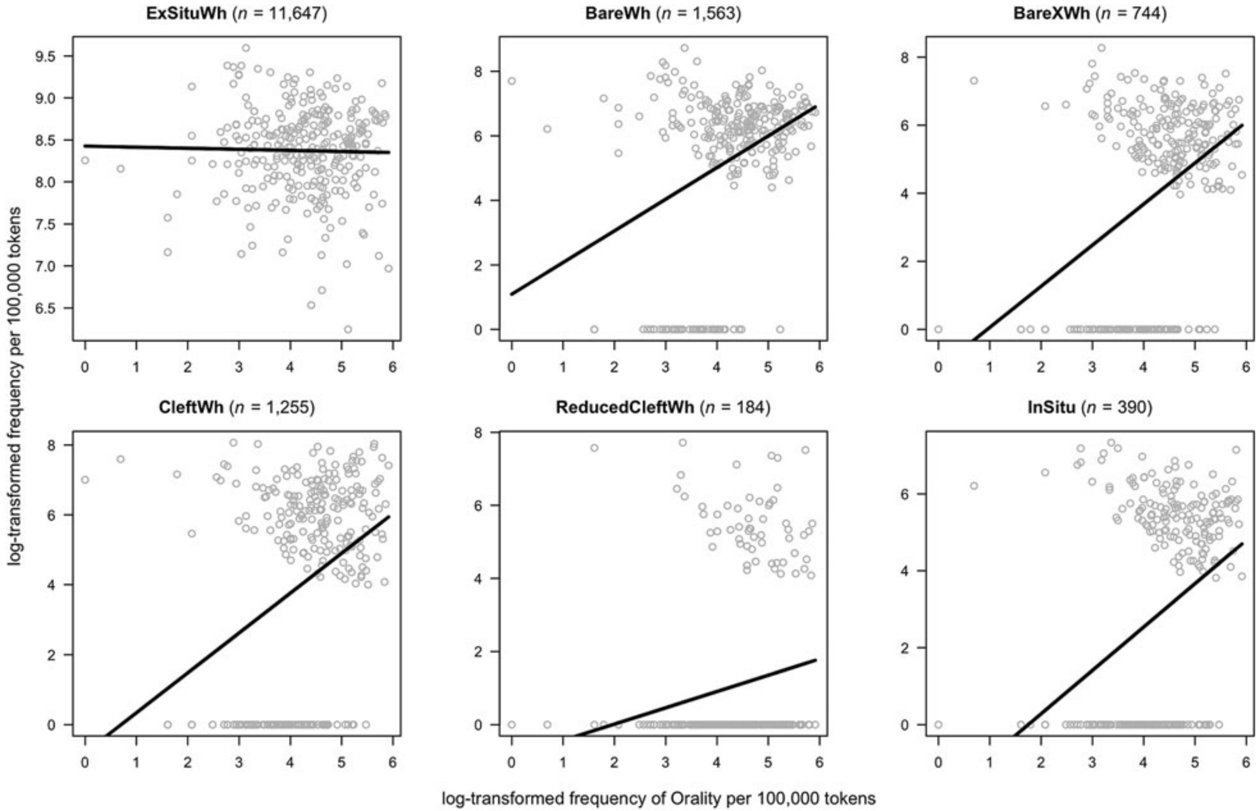

FIGURE 3. Usage frequencies of *wh*-interrogative constructions in BP theater plays by orality.

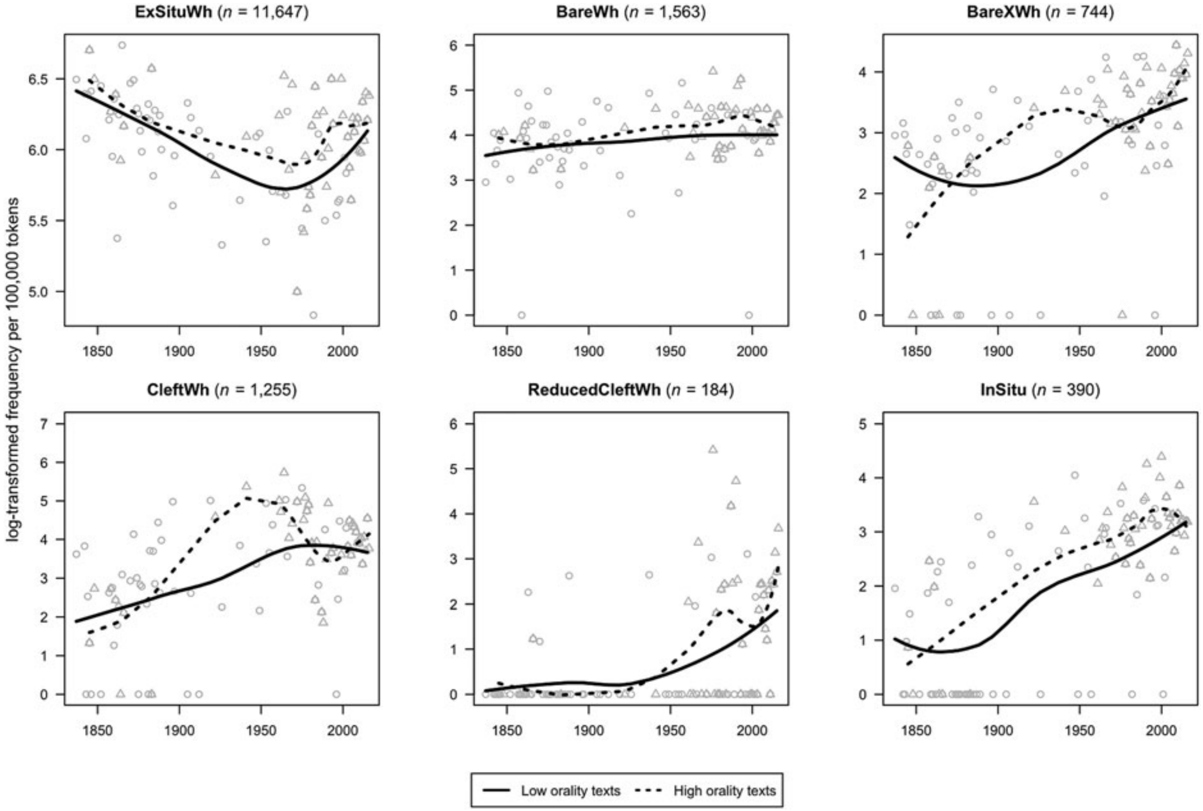

FIGURE 4. Usage frequencies of *wh*-interrogative constructions in BP theater plays by time and orality. (Note: the results per year for high orality texts are represented with triangles, whereas those for low orality texts are represented with circles.)

second part of the diffusion process is based on more deliberation on part of the writer than the first process, which may in many cases be an unconscious choice. This means that changes from below evinced by hump-like change patterns involve the semiconscious adaptation of innovative variants that the writers experience in their everyday life.

As to BareWh and InSituWh interrogatives, the overall increases in usage frequencies are unlikely to be changes from below. There is no evidence that the frequency changes in low-orality texts were preceded by frequency changes in high-orality texts. The fact that the usage frequencies of BareWh and InSituWh interrogatives rise at roughly the same rates rather suggests genre-internal change as the ultimate cause of the frequency increases. Such genre-internal change might simply represent a weakening in writing norms, that is, apparent change. It can also represent an innovation that arose in this specific genre and thus is not necessarily related to change in spoken interaction.

In order to tease apart these two types of change for BareWh and InSituWh, predicted frequencies of InSituWh and BareWh for each year by the mean score on the Orality variable of that year's plays from regression models are compared to the actually observed frequencies.[11] This way it is possible to evaluate how much of the attested change is due to change in the degree of orality of the theater texts. In Figure 5, for BareWh (right plot), when controlling for orality, no statistically significant increase in usage frequency can be documented, which suggests that the increase in the use of BareWh is due to a general relaxation of the writing norms. For InSituWh (left plot), the predicted values show a much lower increase over time than the observed values (from 1.75 to 3 versus 0.9 to 3). This is mostly because, according to the statistical model, the frequency of InSituWh was higher in nineteenth century texts than one would suspect on the basis of the observed frequency, which, in turn, results from the overall lower Orality scores of the nineteenth century plays. However, the predicted values do increase significantly between the 1940s and the 2010s, suggesting that actual, but genre-internal, change has occurred.

In summary, for both InSituWh and BareWh, the increases in usage frequency are much less pronounced than suggested by the changes in their overall distributions in Figure 1. For BareWh no actual change has occurred. For InSituWh, we do document orality-independent change, but later (only after the 1940s or 1950s) and weaker than expected. Since the comparison of low-orality and high-orality texts showed no social conventionalization process, it appears that this orality-independent change of InSituWh was genre-internal and does not reflect actual change in spoken language.

A further point from the discussion of Figure 4 is the "cushioning" effect of orality on the development of ExSituWh interrogatives, suggesting that the frequency decrease was weaker in certain usage contexts bound to high-orality texts. As mentioned in the discussion of the previous research on this topic, there have been changes in subject expression and placement in BP

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

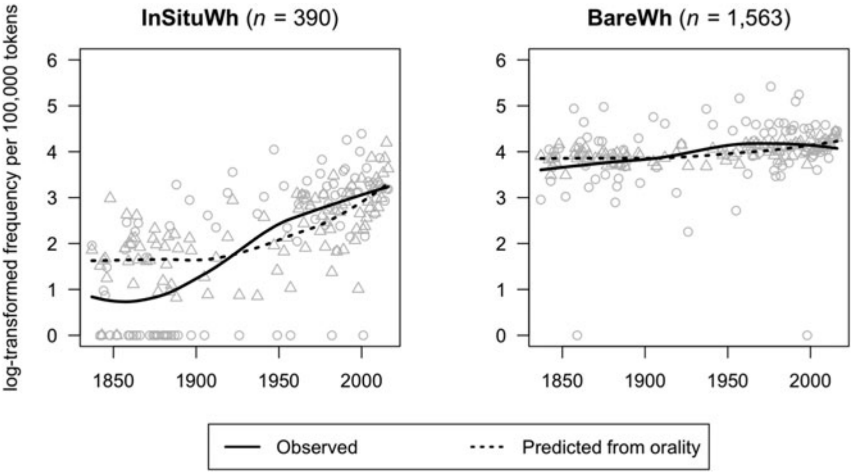

FIGURE 5. Observed versus predicted usage frequencies of InSituWh and BareWh in BP theater plays by time. (Note: gray circles correspond to observed frequencies per year, gray triangles to frequencies predicted by the orality model; the solid regression curve represents the observed values, the dotted regression curve the predicted values.)

wh-interrogatives, namely an increase in the use of overt (versus null) subjects, as well as SV word order in ExSituWh interrogatives. The "cushioning" effect of orality on the development of ExSituWh is thus likely bound to this more general grammatical change in BP.

Figure 6 again illustrates the development of the usage frequencies of ExSituWh and CleftWh interrogatives, but this time distinguishes between the three main types of realization of the subject in these interrogatives: null subject, VS word order, and SV word order.[12]

In line with previous studies, Figure 6 demonstrates that word order had an important influence on the usage frequencies of BP *wh*-interrogatives. The resurgence of ExSituWh after the 1970s is actually entirely due to the fact that SV word order in ExSituWh started to rise in the second half of the nineteenth century, while in null subject and VS word order contexts, there is no significant increase of ExSituWh after 1970. The increase of SV-order ExSituWh interrogatives clearly follows a "hump" distribution in that it first took place in high-orality texts and after the 1950s in low-orality texts, suggesting actual change from below.

It is interesting to contrast this development with the changes in usage frequency for CleftWh interrogatives. Figure 6 demonstrates social conventionalization processes in the development of CleftWh in all three word order configurations; in each, the frequency increase of CleftWh in low-orality texts was preceded by an increase of CleftWh in high-orality texts. Crucially, however, SV word order influenced the development of the construction in that the overall increase in the use of CleftWh is more strongly bound to the increase of SV CleftWh than

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

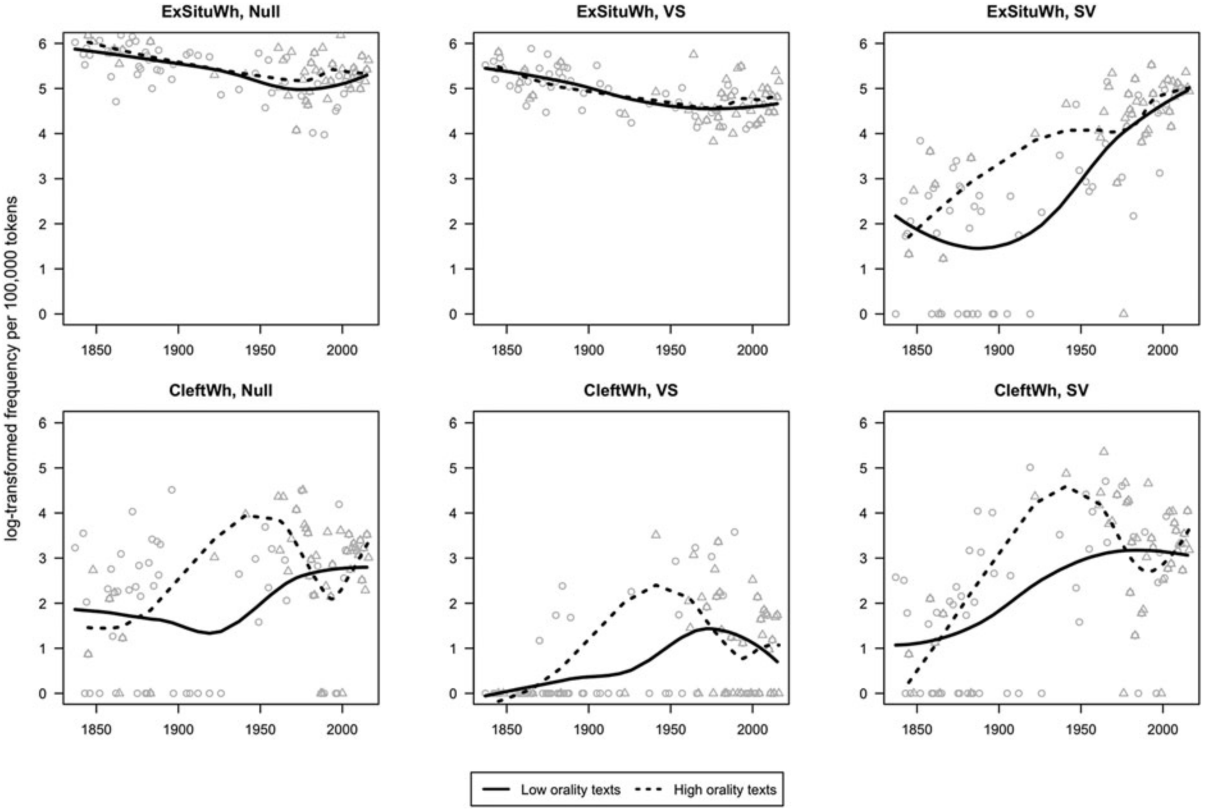

FIGURE 6. Usage frequencies of word order constellations in ExSituWh and CleftWh interrogatives in BP theater plays by time and orality.

null subject CleftWh or VS CleftWh. It is with SV word order that CleftWh experienced by far the strongest rise in usage frequency between the beginning of the nineteenth century and the 1950s. This result confirms claims from previous studies that the rise in BP interrogative and declarative clefts was related to the overall increase in SV word order (Kato & Ribeiro, 2009; Lopes Rossi, 1996).

A last interesting issue is the marked decrease in CleftWh interrogatives after the 1970s. This change might be explained by the parallel rise of ReducedCleftWh interrogatives, which came to replace CleftWh interrogatives as the unmarked type of clefted *wh*-interrogatives in spoken BP (recall Oushiro's [2011] results for spoken BP summarized in Table 1). Two observations from the data support this interpretation. First, both Figure 4 and Figure 6 demonstrate that the decrease of the use of CleftWh is restricted to high-orality texts after the 1950s. In low-orality texts, there is a mostly unbroken increase of the use of most CleftWh constructions in that period. It is in high-orality texts that Brazilian playwrights started to use ReducedCleftWh interrogatives, which led to a competition between these two types of clefted *wh*-interrogatives.

A second argument for this interpretation comes from the comparison of the development of the log-transformed normalized frequencies of CleftWh in BP and EP (see Figure 7). CleftWh interrogatives are less frequent in EP than in BP theater plays until the end of the twentieth century. This difference is mostly due to the fact that the use of CleftWh is already more frequent in the earliest texts of the BP corpus. However, whereas the use of CleftWh starts to decrease after the 1970s in BP, it continues to increase in the EP theater plays. It is well known that the use of ReducedCleftWh is virtually nonexistent in EP (Kato & Ribeiro, 2009), and my results confirm this fact. In the entire EP corpus, only two occurrences of ReducedCleftWh constructions were found, in contrast to $n = 581$ occurrences of CleftWh interrogatives. The unbroken increase in the use of CleftWh interrogatives in EP might thus be due to the fact that no competing clefted *wh*-interrogative arose in EP.

CHANGES IN THE USAGE CONTEXTS OF CLEFTWH AND INSITUWH

The preceding section has demonstrated actual change in the use of CleftWh and InSituWh. The diachronic increase in the usage frequency of a construction is typically correlated with an expansion of the usage contexts of that construction. In the domain of *wh*-interrogatives, evidence for this correlation comes from previous studies on French. Waltereit (2018) analyzes the historical development of the French *que est-ce que* ‘what be.prs.3sg-it that’ interrogative, showing that the earliest attestations occur in contexts in which the pronoun *ce* is anaphoric. Such contexts imply a high degree of cognitive accessibility (Dryer, 1996) of the interrogative proposition. In (8), for example, the proposition ‘she has done something’ is based on a piece of evidence from the situational or discourse context and, consequently, has a high degree of accessibility. In such contexts,

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

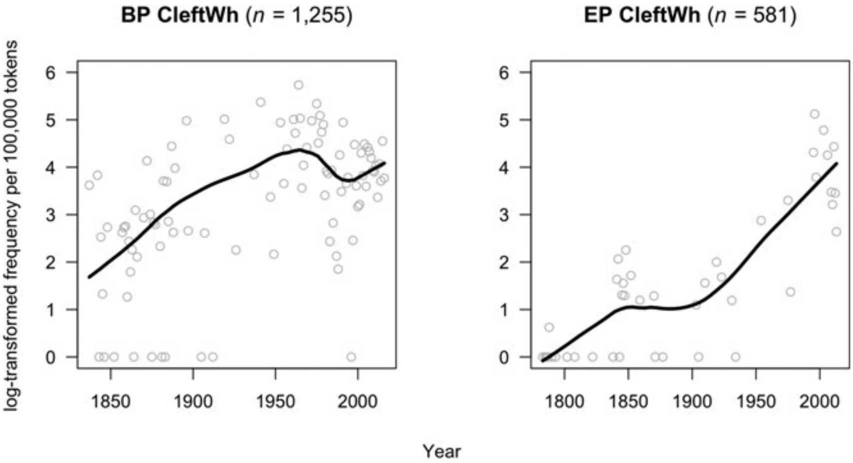

FIGURE 7. Usage frequencies of CLEFTWH interrogatives in Brazilian Portuguese and European Portuguese theater plays by time.

wh-interrogatives typically express disbelief or pretense of disbelief (Rosemeyer, 2018a). They have a low degree of answerability, as no answer is expected.

(8) *Vie de St. Benoit*, end of 12th c., *apud* Waltereit (2018:63)
Suer, li tot poissanz deus espargnet a toi, **ke est ce ke tu as fait**?
'Sister, the almighty God has saved you, what is it that you have done?'

Waltereit documents an expansion of *que est-ce que* to contexts in which the speaker actually expects an answer to her or his question. In present-day French, *que est-ce que* interrogatives can also be used in contexts in which the proposition has a low degree of accessibility, such as thetic contexts. They can thus be regarded as information questions.

Given the actual change documented for CLEFTWH and INSITUWH in the corpus of BP theater plays, one might expect these constructions to have become more frequent in low accessibility contexts. All CLEFTWH and INSITUWH tokens in the data were coded for the degree of accessibility of their proposition. The accessibility variable was coded as "Given" when there was evidence for the proposition on the basis of the previous co-text, as in (9). It was coded as "Inferred" when the proposition could be inferred by logical deduction from something said in the previous co-text, as in (10). It was coded as "New" when neither situation applied, as in (11). Cases coded as "New" thus involve a proposition derived from general world knowledge (e.g., 'Physical entities occupy a place in the world'). Note that the proposition in (11) is also derived from co-text in the sense that the speaker has inferred that she is not in the house. However, there is no strict logical relationship between this inference and the fact that 'she' is necessarily somewhere.

(9) *Comédia sem título*, Martins Pena, 1848
ANA - Então dizei ao Sr. Francisco que aceito.
'So tell Mr. Francisco that I accept.'
CARLOS - **Que é que aceitais?**
'What is it that you accept?'

(10) *Lanterna de fogo*, Qorpo Santo, 1866
MENINA - (para a mulher) Titia… Vovó!… (Puxa-lhe os vestidos com alguma ansiedade.) Titia! Vovó, olha!
'(towards the woman)' 'Auntie… Granny!… (pulls at her clothes with some anxiety.) Auntie! Granny, look!'
A MULHER - (voltando-se para esta) Estás hoje muito incomodativa, muito importuna! **O que é que tu queres?**
'(turning towards her)' 'You are very cumbersome today, very importunate! What is it that you want?'

(11) *Pigmaleoa*, Millôr Fernandes, 1965
EVANDRO: Não tem perigo. Insisti pra que ela entrasse, mas ela disse que prefere a morte.
'There is no danger. I insisted that she enter, but she said that she preferred death'
ISMÊNIA: (Olha na janela) **Onde é que ela está?**
'(looks through window)' 'Where is she?' (lit. 'Where is it that she is?')

Figure 8 illustrates the changes in the distribution of CleftWh ($n = 1255$) and InSituWh ($n = 390$) in terms of the accessibility of the interrogative proposition. The earliest uses of CleftWh and InSituWh are in low-answerability contexts in which the proposition has a high degree of accessibility. The increase in the usage frequencies of the two constructions is correlated with a change from these high-accessibility to low-accessibility contexts.

Given the demonstration above that the usage frequency increases of CleftWh and InSituWh depend on the degree of orality of the texts, it is necessary to control for degree of orality when evaluating the changes in Accessibility summarized in Figure 8. The change towards low-accessibility contexts may be likewise related to the genre change in BP plays from low- to high-orality texts. The previous analysis also demonstrated that CleftWh and InSituWh followed different pathways of change (see Figure 4); whereas the frequency increase of CleftWh was a change from below, the increase in the usage frequency of InSituWh appears to have been a genre-internal change. This leads to different predictions for the influence of orality on the changes in the distribution of accessibility for the two constructions. For CleftWh, one would expect that the change toward low-accessibility contexts first manifested in high-orality texts. For InSituWh, one would not expect orality to influence the change.

In order to test these predictions, I calculated two logistic regression models, one for CleftWh and one for InSituWh, which measured the correlation between

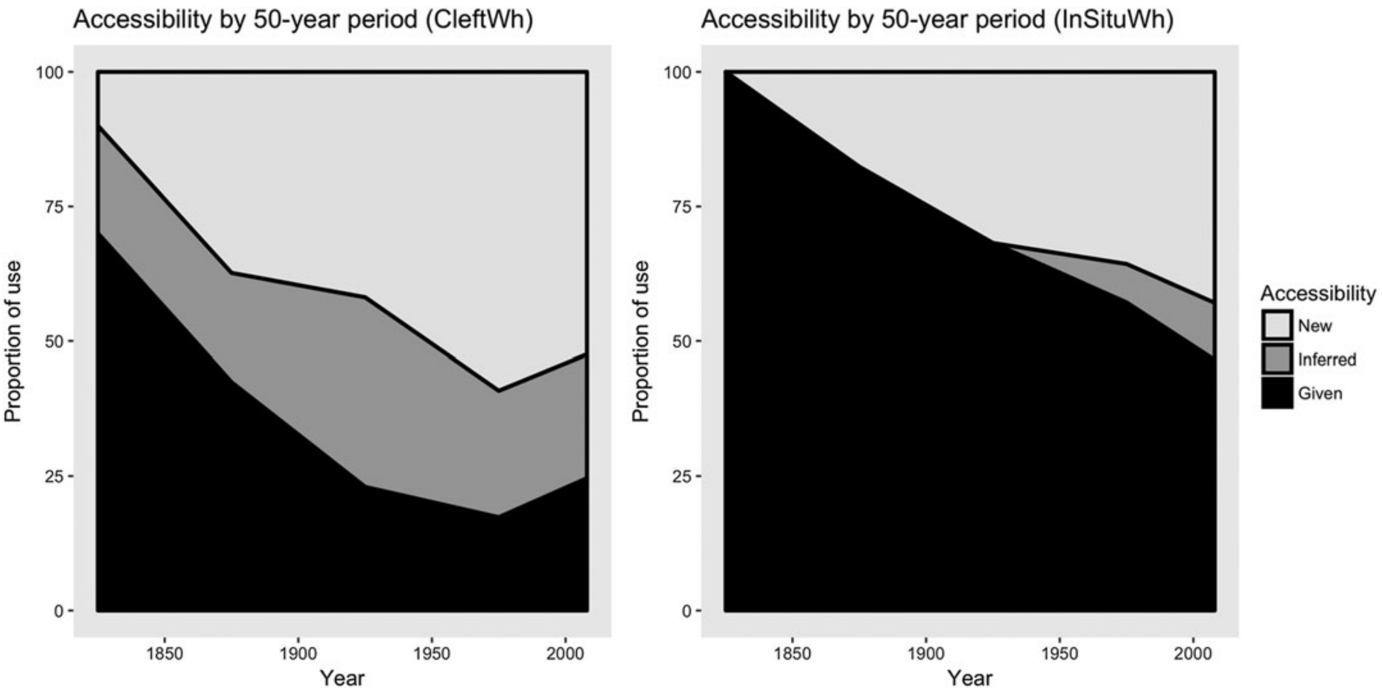

FIGURE 8. Distribution of Accessibility of CLEFTWH and INSITUWH interrogatives in the corpus of BP theater plays by time.

TABLE 4. *Results from the binary logistic regression models (probit link) predicting the use of* CleftWh *and* InSituWh *in low-accessibility contexts in Brazilian Portuguese theater play*s

	CleftWh (*n* = 1,255)				InSituWh (*n* = 390)			
	OR	SE	z	*p*	OR	SE	z	*p*
Year	0.997	2.88	−2.01	<.05*	0.995	0.00	−2.76	<.01**
OralityRes	0.003	0.00	−3.28	<.01**	0.010	6.48	−0.71	>.05
Year : OralityRes	1.003	0.00	2.0	<.05*	1.002	0.00	0.73	>.05
	AIC: 1719.4				AIC: 514.65			
	C index of concordance = 0.47				C index of concordance = 0.43			

(*Note:* OralityRes = residualized values from the regression analysis predicting the log-transformed normalized frequency of Orality from Year. OR = Odds ratio, SE = standard error, z = z value, *p* = *p* value.)

Accessibility on the one hand, and the numerical predictors Year and Orality, as well as their interaction, on the other hand. Accessibility was modeled as a binary variable, collapsing the levels “Given” and “Inferred” into the level “Old” (versus “New”). The statistical modeling was complicated by the strong correlation between Year and Orality (see the discussion of Figure 2) because one prerequisite of regression modeling is that the predictors not be correlated. I therefore created a new variable, OralityRes, which represents the residualized values from the regression analysis predicting the log-transformed normalized frequency of Orality from Year. OralityRes thus represents the score of the texts on the variable Orality that cannot be predicted from time. Table 4 summarizes the results from these models.[13]

According to the regression models, both CleftWh and InSituWh tokens are less likely to occur in contexts in which their proposition is of low accessibility over time. This result confirms the descriptive findings summarized in Figure 8. However, CleftWh and InSituWh differ in that only for the former interrogative type a significant interaction effect between OralityRes and Year is found. Figure 9 visualizes this interaction effect in the regression models for CleftWh and InSituWh. Each line in the plot represents a different mean value of OralityRes, where lower values (e.g., -5) represent low-orality texts and higher values (e.g., 0) represent high-orality texts.

Let us start by reviewing the changes in the usage contexts of CleftWh (left plot). In the earliest texts, the probability for CleftWh to be used in high- or low-accessibility contexts is mediated by the score of the texts on the Orality variable. The probability of use of CleftWh in contexts in which the proposition is old information is highest in low-orality texts (e.g., the line representing the mean value -5) and lowest in high-orality texts (e.g., the line representing the mean value 0). Over time, the probability of use of CleftWh in contexts in which the proposition is old information increased in all texts, irrespective of the degree of orality, thus leveling out the effect of orality in the latest texts. This

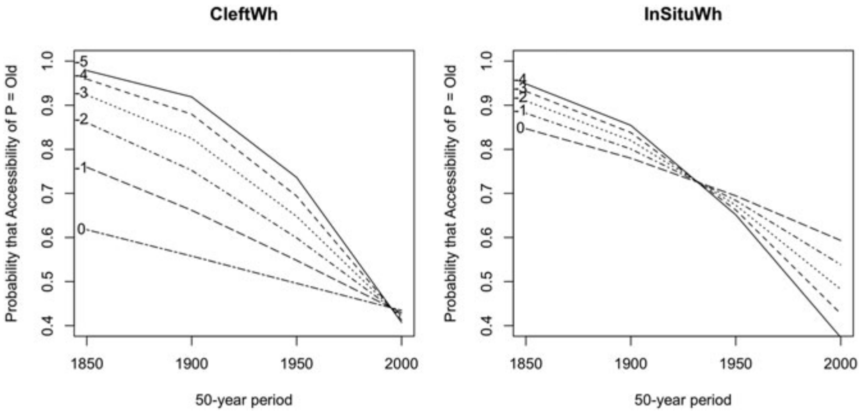

FIGURE 9. Distribution of Accessibility of CleftWh and InSituWh interrogatives in the corpus of BP theater plays by time and orality as predicted by the logistic regression models (lines represent the different degrees of orality).

finding is consistent with the interpretation that the social conventionalization of CleftWh was correlated with the change in the usage contexts of CleftWh. In other words, not only was the frequency increase of CleftWh in low-orality texts preceded by an increase in high-orality texts, but the increase in the use of CleftWh in high-accessibility contexts was bound to high-orality texts, with low-orality texts following in its wake.

In contrast, the interaction between Year and OralityRes does not reach statistical significance for InSituWh (right plot), which is why one cannot rule out the possibility that the changes illustrated in the right plot of Figure 9 are due to random variation. This finding is coherent with the interpretation that the actual change in the use of InSituWh in the corpus of BP theater texts is a genre-internal change.

SUMMARY AND CONCLUSION

The analyses conducted in this paper have demonstrated that the observed changes in the system of BP *wh*-interrogatives represent at least three different types of change. I summarize these changes in Table 5 below. First, it was possible to disentangle actual from apparent, that is, environmental, change. When controlling for orality, the increase in the usage frequency of BareWh interrogatives turned out to be spurious. In other words, there is no evidence that speakers of nineteenth century BP used BareWh interrogatives less frequently than speakers of present-day BP. For InSituWh, controlling for orality did not completely eliminate the frequency increase. Second, the comparison of the development of *wh*-interrogatives in low-orality and high-orality texts demonstrated that, for certain constructions (BareXWh, CleftWh,

TABLE 5. *Types of change in the BP system of wh-interrogative*s

		Frequency changes disappear when controlling for orality	
		No	Yes
High-orality texts display the change before low-orality texts	No	Genre-internal change InSituWh	Apparent change BareWh
	Yes	Change from below BareXWh, CleftWh, ReducedCleftWh, SV ExSituWh	

ReducedCleftWh, and ExSituWh with SV word order), the change affected high-orality texts first and low-orality texts later. Such a constellation is indicative of a social conventionalization process and, consequently, actual change originating in spoken interaction. Third, the analysis identified a change that is neither due to environmental change nor can be characterized as a change from below. The increase in the usage frequency of InSituWh appears to be a genre-internal change independent of the increase of the degree of orality of the theater texts. Further analyses are necessary in order to establish the exact nature of this change (see Rosemeyer [forthcoming]).

The analyses also illustrated that both CleftWh and InSituWh were initially used in contexts in which the interrogative proposition had a high degree of cognitive accessibility. Over time, the use of both *wh*-interrogative constructions expanded to low-accessibility contexts. For CleftWh, this change is documented first in high-orality and only later in low-orality texts, again indicating a change from below and, consequently, social conventionalization. In contrast, the analysis did not evince an influence of orality on the change for InSituWh interrogatives.

Lastly, the results suggest a relationship between word order change and the increase in the usage frequencies of BP CleftWh and ReducedCleftWh interrogatives. In line with the results from previous studies, the analysis demonstrated that SV word order has become more common in ExSituWh and CleftWh interrogatives. This change was a change from below; it affected high-orality texts first and low-orality texts second. There appears to have been a correlation between the increase in SV word order and the increase in the use of CleftWh interrogatives; the analysis has shown that the use of CleftWh first increased in SV word order contexts and later in VS and null-subject contexts. The fact that there has not been a similar increase in SV word order in EP might thus explain why, at least in the beginning, the rise of CleftWh constructions was much stronger in BP than in EP. It was only after the introduction of ReducedCleft constructions and, consequently, the rise of a competing clefted *wh*-interrogative construction, that the historical trend towards the use of CleftWh was broken in BP.

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

NOTES

1. In many present-day Brazilian Portuguese dialects, the pronoun *você* has generalized to the unmarked second person pronoun (Lopes, 2015:204–206; Lopes & Rumeu, 2015). This change coincided with a rise in the overall frequency of use of personal pronouns, frequently explained as a loss of the pro-drop parameter (Duarte, 1992, 1993, 2000). Consequently, in European Portuguese, in these sentences the verb would probably be inflected for second person, although the corresponding personal pronoun *tu* 'you' would probably not be used.

2. The denomination of this type of cleft-*wh* interrogatives as "reduced" cleft-*wh* interrogatives was proposed in Kato and Mioto (2005) and Kato (2014). According to these authors, ReducedCleftWh constructions are derived by ellipsis of the copula from Copula + *wh* + *que* cleft interrogatives, such as *É quem que tá tocando o violão?,* literally, 'Is who that is playing the guitar?' (Kato, 2014:116). Copula + *wh* + *que* cleft interrogatives did not occur in my corpus.

3. Similar figures can be found in Kato and Mioto (2005).

4. An interesting question that, to my knowledge, has not been studied in detail is the gradual replacement of the interrogative pronoun *que* with the reinforced form *o que*, a change that seems to be correlated to the general restructuration of the system of partial interrogatives described in this paper. The development of the usage frequency of *o que* relative to *que* in my data is as follows (only ExSituWh): 1700–1749: 1%; 1750–1799: 4%; 1800–1849: 26%; 1850–1899: 26%; 1900–1949: 11%; 1950–1999: 41%; 2000–2016: 55%. The development of this alternation may of course also depend on the degree of orality of the texts.

5. However, as correctly commented by one of the reviewers of the paper, the elimination of these syntactic contexts does not ensure complete comparability of the different types of *wh*-interrogatives. As will be noted in the discussion of BareWh and BareXWh interrogatives in the later sections of the article, the distribution of these types of *wh*-interrogatives depends on the preceding context more strongly than, for example, ExSituWh interrogatives. Given that the analysis does not work with relative frequencies (that is, percentages) but absolute usage frequencies, this fact does not, however, invalidate the results of this paper.

6. *Cadê* 'where is (it)?' is actually an entrenched and amalgamated form of the sentence *O que é de?* 'What be.prs.3sg of?,' which does have a verb phrase. However, due to the entrenchment process, it is doubtful whether speakers parse *cadê* as involving the verb *é*.

7. For instance, the formula for ExSituWh had the form loess (logExSituWh ~ Year, span = 0.40), where the parameter span controls the degree of smoothing. Local polynomial regressions differ from linear regression models in that they do not make assumptions about the kind of trend encountered in the data, essentially allowing for non-linearity. They are therefore frequently employed to create smoother lines as in Figure 1 (see, for example, Baayen, 2008:94).

8. Statistical testing was done using Kendall's τ because the time variable is not normally distributed (see Gries, 2009:212–213). ExSituWh: $\tau = -0.20$, $z = -2.90$, $p_{\text{two-sided}} < .01$**; BareWh: $\tau = 0.19$, $z = 2.70$, $\mathrm{p}_{\text{two-sided}} < .01$**; BareXWh: $\tau = 0.41$, $z = 5.82$, $p_{\text{two-sided}} < .001$***; CleftWh: Kendall's $\tau = 0.29$, $z = 4.09$, $p_{\text{two-sided}} < .001$***; ReducedCleftWh: Kendall's $\tau = 0.39$, $z = 4.99$, $p_{\text{two-sided}} < .001$***; InsituWh: Kendall's $\tau = 0.47$, $z = 6.56$, $p_{\text{two-sided}} < .001$***. The trends were also tested for autocorrelation using Durbin-Watson tests, none of which showed autocorrelation to be a problem. The concept of autocorrelation describes the fact that, in a historical change, the frequency value of a temporally prior data point will typically be highly correlated with the frequency value of a subsequent data point (see Van de Velde and Petré [forthcoming] for details).

9. Statistical testing was done using Kendall's τ because Orality is not normally distributed, with the following result: $\tau = 0.46$, $z = 6.59$, $p_{\text{two-sided}} < .001$***.

10. Statistical testing was done using Kendall's τ because Orality is not normally distributed, with the following results. ExSituWh: $\tau = 0.00$, $z = -0.04$, $p_{\text{two-sided}} > .05$; BareWh: $\tau = 0.17$, $z = 4.22$, $p_{\text{two-sided}} < .001$***; BareXWh: $\tau = 0.20$, $z = 4.86$, $p_{\text{two-sided}} < .001$***; CleftWh: $\tau = 0.22$, $z = 5.46$, $p_{\text{two-sided}} < .001$***; ReducedCleftWh: $\tau = 0.17$, $z = 3.78$, $p_{\text{two-sided}} < .001$***; InSituWh: $\tau = 0.23$, $z = 5.58$, $p_{\text{two-sided}} < .001$***.

11. Quantile regression was used (Koenker, 2005) because Orality is not normally distributed. Basically, quantile regression works like linear regression, with the difference that it does not estimate the mean of y at each point of x. Rather, it estimates a quantile of the distribution, which is why it can make decent estimates of the quantile for increasing values *of x* despite the increasing variability. In this case, the quantile was set to the median (tau = 0.5), the default setting of the rq() function used for quantile regression in R.

12. A fourth type of word order not included in the graph is SwhV word order, as in *Você o que quer?* 'You what want.prs.3sg?' This word order type was excluded from the graph, because in comparison to

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press

null subject ($n = 8959$), whVS ($n = 3926$), and whSV ($n = 2786$) word order, SwhV word order is marginal ($n = 112$).

13. The c index of concordance is a measure of the goodness of fit of a model to the data, ranging between 0 (no fit) to 1 (perfect fit) (Baayen, 2008:281; Levshina, 2015:259). Typically, a fit above 0.7 is taken to be an adequate fit to the data. With c indexes of concordance of 0.47 viz. 0.43, both of the models thus explain very little variation in the data. Undoubtedly, there are many other parameters that would have to be taken into account to elaborate a full and more explanatory model of the change in the use of CleftWh and InSituWh over time. However, this study does not aim at establishing such a complete model but rather at confirming the hypothesis of an interaction between the functional change and the change in orality in the texts, which is why the low statistical resolution of the models is not a problem for the argument presented here.

REFERENCES

Armstrong, Nigel. (2001). *Social and Stylistic Variation in Spoken French. A Comparative Approach*. Amsterdam: John Benjamins.

Auer, Peter. (2014). The temporality of language in interaction: projection and latency. *Interaction and Linguistic Structures* 54. Available online at http://www.inlist.uni-bayreuth.de/papers/byissue/index.htm. Last access December 20, 2018.

Baayen, Harald. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Biber, Douglas, & Finegan, Edward. (2004 [1987]). Historical drift in three English genres. In G. Sampson, Geoffrey, & D. McCarthy (Eds.), *Corpus Linguistics: Readings in a Widening Discipline*. London: Continuum. 67–77.

Croft, William. (2000). *Explaining Language Change. An Evolutionary Approach*. London: Longman.

Davies, Mark. (2006). O corpus do português. Available online at http://www.corpusdoportugues.org. Last access April 2, 2018.

De Paula, Mayara N. (2015). A ordem VS/SV em interrogativas-Q: um estudo diacrônico em peças teatrais brasileiras e portuguesas. In A. Baalbaki, J. Cardoso, P. Arantes, & S. Bernardo (Eds.), *Linguagem: Teoria, Análise e Aplicações*. Rio de Janeiro: Programa de Pós-graduação em Letras. 585–595.

_______. (2016). *A ordem VS/SV e as interrogativas-Q no PE e no PB: uma análise diacrônica*. Doctoral dissertation, Universidade Federal do Rio de Janeiro.

_______. (2017). A comparative diachronic analysis of wh-questions in Brazilian and European Portuguese. *Diadorim* 19:173–196.

Dryer, Matthew S. (1996). Focus, pragmatics presuppositions, and activated propositions. *Journal of Pragmatics* 26(4):475–523.

Duarte, Maria E. 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.

_______. (1993). Do pronome nulo ao pronome pleno. A trajetória do sujeito do português do Brasil. In I. Roberts, & M. A. Kato (Eds.), *Português Brasileiro: uma viagem diacrônica*. Campinas: UNICAMP. 107–128.

_______. (2000). The loss of the ‘Avoid Pronoun’ principle in Brazilian Portuguese. In M. A. Kato, & E. V. Negrão (Eds.), *The Null Subject Parameter in Brazilian Portuguese*. Frankfurt, Madrid: Vervuert/Iberoamericana. 17–36.

Elsig, Martin. (2009). *Grammatical Variation Across Space and Time. The French Interrogative System*. Amsterdam: John Benjamins.

Fontes, Michel G. (2012a). *As interrogativas de conteúdo na história do português brasileiro: uma abordagem discursivo-funcional*. Doctoral dissertation, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

_______. (2012b). A clivagem do constituiente interrogativo em sentenças interrogativas do português brasileiro: uma abordagem diacrônica. *Estudos Linguísticos* 15(3):149–170.

Galves, Charlotte, De Andrade, Aroldo L., & Faria, Pablo (2017). Tycho Brahe Parsed Corpus of Historical Portuguese. Available online at http://www.tycho.iel.unicamp.br/~tycho/corpus/texts/psd.zip. Last access 2 April 2018.

Gries, Stefan T. (2009). *Quantitative Corpus Linguistics with R. A Practical Introduction*. New York: Routledge.

Kaiser, Georg, & Quaglia, Stefano. (2015). In search of wh-in-situ in Romance: An investigation in detective stories. In E. Brandner, A. Czypionka, C. Freitag, & A. Trotzke (eds.), *Charting the*

Landscape of Linguistics. On the Scope of Josef Bayer's work. Konstanz: Konstanzer Online-Publikations-System (KOPS). 92–103.

Kato, Mary A. (2013). Deriving wh-in-situ through movement. In V. Camacho-Taboada, Á. L. Giménez Fernández, J. Martín-González, & M. Reyes-Tejedor (Eds.), *Information Structure and Agreement*. Amsterdam: John Benjamins. 175–191.

______. (2014). Focus and wh-questions in Brazilian Portuguese. In R. Torres Cacoullos, N. Dion, & A. Lapierre (Eds.), *Linguistic Variation. Confronting Fact and Theory*. London: Routledge. 111–130.

Kato, Mary A., & Mioto, Carlos. (2005). A multi-evidence study of European and Brazilian Portuguese wh-questions. In M. Reis, & S. Kepser (Eds.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. Berlin: De Gruyter. 307–328.

Kato, Mary A., & Ribeiro, Ilza. (2009). Cleft sentences from Old Portuguese to Modern Portuguese. In A. Dufter, & D. Jacob (Eds.), *Focus and Background in Romance Languages*. Amsterdam: Benjamins. 123–154.

Koenker, Roger W. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.

Labov, William. (1994). *Principles of Linguistic Change. Volume I: Internal Factors*. Oxford: Blackwell.

Levshina, Natalya. (2015). *How To Do Linguistics With R*. Amsterdam: John Benjamins.

Lopes, Célia R. dos Santos. (2015). Tópicos de história do português pelo viés da gramaticalização. *LaborHistórico* 1(2):197–209.

Lopes, Célia R. dos Santos, & Márcia, C. de Brito Rumeu. (2015). A difusão do você pelas estruturas sociais carioca e mineira dos séculos XIX e XX. *LaborHistórico* 1(1):12–25.

Lopes Rossi, Maria A. (1996). *A sintaxe diacrônica das interrogativas-Q do português*. Doctoral dissertation, UNICAMP.

Mathieu, Eric. (2004). The mapping of form and interpretation: the case of optional wh-movement in French. *Lingua* 114(9–10):1090–1132.

Oushiro, Livia. (2011). *Um análise variacionista para as Interrogativas-Q*. Doctoral dissertation, Universidade de São Paulo.

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

Rosemeyer, Malte. (2018a). The pragmatics of Spanish postposed-*wh*-interrogatives. *Folia Linguistica* 52(2):283–317.

______. (2018b). PorThea. A historical corpus of Portuguese theater plays. Available online at http://www.romanistik.uni-freiburg.de/rosemeyer/05corpus.html. Last access February 8, 2019.

______. (Forthcoming). Brazilian Portuguese *in-situ-wh*-interrogatives between rhetoric and change. *Glossa*.

Schmid, Hans-Jörg. (2015). A blueprint of the Entrenchment-and- Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association* 3:3–25.

Szmrecsanyi, Benedikt. (2016). About text frequencies in historical linguistics: Disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* 12(1):153–171.

Van de Velde, Freek, & Petré, Peter. (Forthcoming). Historical linguistics. In D. Knight, & S. Adolphs (Eds.), *The Routledge Handbook of English Language and Digital Humanities*. London: Routledge.

Waltereit, Richard. (2018). Inferencing, reanalysis, and the history of the French *est-ce* que question. *Open Linguistics* 4:35–48.

Weinreich, Uriel, Labov, William, & Herzog, Marvin. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann, & Y. Malkiel (Eds.), *Directions for Historical Linguistics*. Austin: University of Texas Press. 95–188.

https://doi.org/10.1017/S0954394519000097 Published online by Cambridge University Press