

Wh-interrogatives in Brazilian Portuguese: The Influence of Common Ground

Livia Oushiro*

1 Introduction

Wh-interrogatives have been extensively studied, in Portuguese and other languages, in terms of their syntactic structure (see Chomsky 1977, Kato 1987, Cheng 1991, Duarte 1992, Mioto and Figueiredo-Silva 1995, Mioto 1997, Cheng and Rooryck 2000, Ambar et al. 2001, Kato 2004, Kato and Mioto 2005, inter alia). However, studies that observe these structures in naturally occurring language are much scarcer (see Coveney 2002, Hall 2008, Oushiro 2011).

This paper analyzes discourse-pragmatic factors conditioning the variable use of wh-interrogatives in Brazilian Portuguese (BP), and more specifically the influence of common ground (Clark 1996, Stalnaker 2002) and other related aspects of face-to-face conversation. The main argument is that the alternation between different syntactic structures of wh-interrogatives in BP is highly sensitive to the *hic et nunc* of conversation, as speakers make their conversational contributions and common ground is updated.

Wh-interrogatives are broadly defined here as sentences containing an interrogative adjective, adverb or pronoun:¹ (*o que* ‘what’; *que* + NP ‘what + NP’; *qual(-is)* ‘which’; *qual(-is)* + NP ‘which + NP’; *quanto(-a, -os, -as)* ‘how much/how many’; *quanto(-a, -os, -as)* + NP ‘how much/how many + NP’; *quem* ‘who, whom’; *como* ‘how’; *quando* ‘when’; *onde* ‘where’; and *por que* ‘why’. In contemporary BP there are four different morphosyntactic structures of wh-interrogatives, as shown in (1):

- | | |
|-----------------------------------|--|
| (1) a. “Simple” wh-interrogative: | <i>Onde você mora?</i> (Where you live?) |
| b. Wh-que interrogative: | <i>Onde que você mora?</i> (Where that you live?) |
| c. Cleft wh-interrogative: | <i>Onde é que você mora?</i> (Where is-it that you live?) |
| d. Wh-in-situ: | <i>Você mora onde?</i> (You live where?)
‘Where do you live?’ |

When the wh-word is preverbal, there are three possibilities of realization: (i) wh-word + subject-verb, as in (1a); (ii) wh-word + complementizer *que* + subject-verb, as in (1b); and (iii) wh-word + *é que* + subject-verb, as in (1c).² The wh-word may also be realized in situ, in the position of its syntactic function, as in (1d). The focus of this paper is on the latter, wh-in-situ, in contrast to the three structures with a preverbal wh-word.

Typological studies describe four types of languages in regard to the movement of the wh-constituent (Ambar et al. 2001): (i) languages in which the wh-word is always in situ, such as Chinese; (ii) languages in which the wh-word always moves, such as Hungarian;³ (iii) languages that allow wh-in-situ when another wh-word has already moved, such as English (e.g., “Who bought what?”); and (iv) mixed languages that allow both structures, such as French and (European and Brazilian) Portuguese. However, several works (Lopes-Rossi 1993, Ambar et al. 2001, Kato and Mioto 2005, Pires and Taylor 2007, inter alia) notice that wh-in-situ is more productive in BP than

*This research was funded by a grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Grant 2009/03190-0), to whom I am grateful. I would also like to thank my dissertation advisor Ronald Beline Mendes, as well as Gregory Guy, James Walker, Anthony Naro, and Marta Scherre for their guidance and suggestions during the development of this research. They are all naturally exempted from any faults that may remain.

¹But see Section 2 for the definition of the envelope of variation.

²There are also two other structures of wh-interrogatives with preverbal wh-words in BP: “Wh-que é que,” e.g., *O que que é que você está querendo?* ‘What do you want?’ (Mioto 1997), and “É wh-que,” e.g., *É o que que ele quer?* ‘What does he want?’ (Lessa de Oliveira 2005). These two structures, however, are very infrequent in Paulistano Portuguese (i.e., from the city of São Paulo) and will not be discussed here.

³According to Ambar et al. (2001), the wh-word can be in situ in Hungarian only in “highly marked” contexts.

in European Portuguese (EP). This fact is generally explained by a pragmatic constraint to the use of *wh-in-situ* in EP, being restricted to echo-questions, with rising intonation and the function of requesting the repetition of something just said ((2) is extracted from Pires and Taylor 2007:2):

- (2) A: *A Maria comeu um gambá.* A: ‘Mary ate a skunk.’
 B: *A Maria comeu O QUÊ?* B: ‘Mary ate WHAT?’

In BP, such a discourse-pragmatic constraint does not seem to apply. In addition to echo questions, *wh-in-situ* can be employed with falling intonation, with semantic and functional equivalence to preverbal *wh-interrogatives*. In (3) below, both *wh-in-situ* (3a) and a “simple” *wh-interrogative* (3b) are employed by the same person, Helena, to make similar questions about her interlocutor’s sister’s and parents’ age.⁴

- (3) a. Helena: *e::... sua irmã tem **quantos anos**?* (your sister has how-many years?)
 Ingrid: *é ela é quatro anos mais nova que eu ela tem trinta e um*
 b. Helena: *ai é verdade... e seus pais?... **quantos anos** eles têm?* (how-many years they have?)
 Ingrid: *meu pai tem sessenta... e a minha mãe tem cinquenta e seis...*

This study describes and analyzes some of the discourse-pragmatic contexts in which the structure of *wh-in-situ* is favored in the Paulistano variety of BP. Section 2 presents the corpus, methodological issues, and the envelope of variation; Section 3 discusses two discourse-pragmatic factor groups, both concerning common ground among speakers, and presents the results of multivariate quantitative analyses in GoldVarb X; finally, Section 4 summarizes the main conclusions of this paper.

2 Corpus and Methodological Decisions

The corpus consists of spoken and written samples of contemporary Paulistano Portuguese. The written corpus comprises five issues of the weekly magazine *Veja* (about 170,000 words), Nov-Dec/2008; samples from the online version of the newspaper *Folha de São Paulo* (about 155,000 words), Dec/2008; and 1,470 school compositions by elementary and high school students attending private and public schools in São Paulo (about 230,000 words), collected in 2005. The spoken language corpus is made up of 53 sociolinguistic interviews collected by students of the Sociolinguistics course at University of São Paulo (USP) between 2003 and 2008. The interviewees were all born in the city and are stratified by gender, age (20-34; 35-49; 50 or more), and level of education (elementary school; college). The transcriptions of these interviews add up to about 500,000 words.

The choice for a robust corpus containing more than a million words is partially due to the challenge of obtaining a good amount of tokens of a syntactic variable. Obtaining tokens of interrogative structures is particularly challenging in sociolinguistic interviews, where speakers “tend to exhibit a more limited segment of their linguistic repertoire grammatically and stylistically” (Sankoff 1980:53): the role of making questions is the interviewer’s, not the interviewee’s. One way of dealing with this challenge was to analyze both the interviewees’ *and* interviewers’ speech, taking advantage of the fact that the spoken corpus had been collected by dozens of students. Out of the 53 student interviewers, 19 were from the city of São Paulo. Their tokens of *wh-interrogatives* were thus also included in the analyses.

Another issue to be addressed is the difference between *questions* and *interrogatives*, which are not taken as synonymous: in this study, the first term is used to refer to a *discourse function* (which does not always take the interrogative form), whereas the second is used to refer to a syntactic and prosodic structure (which does not always have the function of a question). As Milroy and Gordon (2003:170) put it, “(...) since there is no isomorphic relationship between function

⁴Translation of example (3): ‘H.: And how old is your sister? I.: uhn she’s four years younger than me, she’s thirty-one. H.: oh I see... and your parents? How old are they? I.: My father is sixty and my mother is fifty-six...’

and form, questions are not always realized syntactically as interrogatives and interrogative forms may realize many different functions.”

The envelope of variation of this analysis is defined by the intersection of form and function. The context in which the four structures in (1) can be alternatively employed as questions and can be considered “alternative ways of saying the same thing,” thus allowing for speaker’s optionality (Labov 1978), is defined as follows (Oushiro 2011:67):⁵

- (4) Envelope of variation of wh-interrogatives:
Full sentences containing only one wh-word in a finite main clause or in an embedded clause introduced by a complementizer, excluding the cases of crystallized/semilexicalized expressions and in which the wh-word is the subject of its clause.

Instances of sentences containing more than one wh-word are scarce and, when they do occur, each wh-word tends to be realized in situ (5). Sentences with non-finite verbs (6) or without verbs (7) do not allow the structures of wh-que and cleft wh-interrogatives. In embedded clauses introduced by a complementizer, wh-in-situ is also possible (8); but when the embedded clause is not introduced by a complementizer, wh-in-situ is ungrammatical in contemporary BP (9).

- (5) *Você deu o quê pra quem?*
‘You gave what to whom?’
- (6) a. *O que fazer?*
b. *Fazer o quê?*
c. **O que que fazer?*
d. **O que é que fazer?*
‘What to do?’
- (7) a. *Por que taxa do lixo?*
b. *Taxa do lixo por quê?*
c. **Por que que taxa do lixo?*
d. **Por que é que taxa do lixo?*
‘Why garbage tax?’
- (8) a. *Por que (que / é que) a senhora acha [que o público é tão diferente]?*
b. *A senhora acha [que o público é tão diferente por quê]?*
‘Why do you think [(that) the public is so different]?’
- (9) a. *Você já sabe [em quem (que / é que) vai votar nas próximas eleições]?*
b. **Você já sabe [vai votar em quem nas próximas eleições]?*
‘Do you already know [who you will vote for in the next election]?’

Tokens of certain expressions that seem to be crystallized or semilexicalized in a fixed formula were also excluded (10). This is a type/token question (Wolfram 1993, Tagliamonte 2006): similarly to studies of /-t, -d/ deletion in English, the inclusion of all instances of these very frequent tokens which do not tend to vary may skew the results of quantitative analyses.

- (10) a. *Onde já se viu?* (Where has it been seen? = ‘This is absurd!’)
b. *Quem sou eu?* (Who am I? = ‘I’m not important enough.’)
c. *Como chama? Como se diz?* (How calls? How is it said? = ‘What do you call it?’)

Finally, when the wh-word is the subject of its clause (11), there is a context of neutralization between the structures of “simple” wh-interrogatives and wh-in-situ, and it is not possible to determine, in principle, which structure was employed by the speaker.⁶

⁵See Oushiro 2011 for a more detailed discussion of the envelope of variation.

⁶Studies within a generative grammar framework (Cheng 1991, Cheng and Rooryck 2000, Kato 2004) argue that, in sentences such as *Quem disse isso?* ‘Who said this?’, the wh-word is only apparently in situ: the wh-word moves from Spec IP to Spec CP. In BP, evidence for that view is the possibility of producing wh-que interrogatives (*Quem que disse isso?*) and cleft wh-interrogatives (*Quem é que disse isso?*) (Kato 2004). For a variationist analysis, however, this is a false question; in any case, wh-in-situ would not be pos-

- (11) a. *Quem disse isso?* ('Who said this?')
 b. *O que aconteceu?* ('What happened?')

3 Wh-interrogatives and Common Ground

A token frequency distribution reveals that there is almost no variation in the written corpus (see Table 1). In the magazine, the online newspaper and the school compositions there was less than 4% of wh-in-situ and more than 93% of "simple" wh-interrogatives. On the other hand, the distribution of variants is much more balanced in sociolinguistic interviews: there were 22.7% of wh-in-situ as well as variation among preverbal wh-interrogatives.⁷

Wh-interrogatives	Magazine		Newspaper		School compositions		Sociolinguistic interviews	
	N	%	N	%	N	%	N	%
"Simple"	334	98.8	208	93.3	626	95.4	289	28.9
Wh-que	0	0	5	2.2	9	1.4	408	40.8
Cleft-wh	3	0.9	2	0.9	11	1.7	75	7.5
Wh-in-situ	1	0.3	8	3.6	10	1.5	227	22.7
Total	338	100.0	223	100.0	656	100.0	999	100.0

Table 1: Wh-interrogative frequency distribution in different corpora of Paulistano Portuguese.

The extremely unbalanced distribution in the written corpus does not allow for hypothesis testing in the quantitative analyses, since variables will certainly not be orthogonal to each other (Guy 1988). The results reported in Section 4 thus refer only to quantitative analyses of the spoken language sample. Nevertheless, the fact that there are so few wh-in-situ in written language, even in compositions by elementary and high school students, signals that orality may be an important factor in wh-interrogative variation. Out of the ten tokens of wh-in-situ in compositions, 9 represented dialogical situations in narratives (12);⁸ all 8 tokens of wh-in-situ in the online newspaper refer to comments of the readers, who engage in debates about controversial issues (13); and the only instance of wh-in-situ in the magazine was drawn from an interview, presumably spoken and later transcribed (14).⁹

- (12) – *Que mochila bonita!*
 – *Obrigado! Comprei aqui na escola!*
 – *Que legal! Você estuda já há quanto tempo aqui?!* (you study already how-much time here?)
 – *A 1 ano!*
- (13) *Mas se escolhendo homens com vida normal, se resume as chances, pq quem não fica com mulher nesta terra, fica com quem?* (stays with whom?) *Com homem, na falta o pior com crianças. A sociedade prefere o que?* (society prefers what?)
- (14) – *O senhor fazia o que antes de atuar na Abin?* (you did what before of to-work at Abin?)
 – *Fui do corpo de segurança do presidente Lula. (...)*

sible and variation in this context cannot be analyzed.

⁷In some varieties of BP, wh-in-situ is the most frequent structure of wh-interrogative. For instance, in Vitória da Conquista, in the state of Bahia, wh-in-situ gets up to 81.7% in the speech of one informant (Lessa de Oliveira 2005).

⁸The sample of school compositions also comprises essays and descriptive texts.

⁹Translation of examples (12–14): (12) "What a nice backpack!" "Thanks! I got it here at school!" "Cool! How long have you been studying here?!" "For a year!" (13) "But if choosing men with a regular life, opportunities are deemed, because the ones who are not with women on this earth, who are they with? Men, or in the lack of them, worse, children. What does society prefer?" (14) "What did you do before working at Abin?" "I was part of President Lula's security team."

This fact motivated the investigation of discourse-pragmatic factors on the variable use of wh-interrogatives, in face-to-face conversational interactions. In this paper, two factor groups are more closely analyzed: the Type of Question (“true” information questions, rhetorical questions, and semi-rhetorical questions) and the Givenness of the Presupposition (when last activated in the conversation, if at all).¹⁰

The hypotheses are based on the concepts of *common ground* among speakers (Clark 1996, Stalnaker 2002), *speaker’s presupposition* (Stalnaker 2002), and on the Conversation Analysis concepts of *turn-taking system* (Sacks, Schegloff, and Jefferson 1974) and *Question–Answer pair* (Schegloff 1972, Schegloff and Sacks 1973, Sacks, Schegloff, and Jefferson 1974). Two speakers’ *common ground* is “the sum of their mutual, common, or joint knowledge, beliefs, and suppositions” (Clark 1996:93). For Stalnaker (2002:701), this concept is closely related to that of *speaker’s presupposition*: “To presuppose something is to take it for granted, or at least act as if one takes it for granted, as background information—as *common ground* among participants in the conversation” (author’s emphasis). It is worth noting that, in this concept of presupposition, the speaker’s attitude is also a social attitude, since it takes into account what the speaker takes as common knowledge among the participants of the communicative interaction: one only presupposes something if one presupposes the others presuppose the same. For instance, when one asks “Why do you want to learn Greek?”, one not only presupposes the interlocutor “wants to learn Greek,” but also presupposes the interlocutor shares the same presupposition.¹¹

Common ground and speaker’s presuppositions ought to be viewed dynamically, since they are constantly updated in talk-in-interaction. This is to say that speakers (and the researcher) have to keep track of common ground; one way of checking if what is presupposed in a certain question was in fact shared among speakers is to examine the interlocutor’s answer or, at least, her reaction. Through her response, verbal or non-verbal, it is possible to verify if she sanctions the presupposition that had been made. Thus it is necessary to examine both question and answer when analyzing the flow of information in talk-in-interaction.

A Question–Answer pair (a type of the broader category “adjacency pairs;” see e.g., Schegloff 1972, Sacks 1987 [1973], Schegloff and Sacks 1973, Sacks, Schegloff, and Jefferson 1974) is characterized by: (i) two utterance length; (ii) adjacent positioning of component utterances; (iii) different speakers producing each utterance; (iv) relative ordering of parts (i.e., the first pair parts precede the second pair parts); and (v) discriminative relations (i.e., the pair type of which a first pair part is a member is relevant to the selection among second pair parts) (Schegloff and Sacks 1973:295–296). Upon the production of a first pair part, a basic conversational rule is that current speaker should stop talking and the interlocutor should start producing the second pair part. However, Schegloff (1972) also notices the possibility of certain deviations from this rule.¹² Schegloff (1972:77) defines the “basic organization” of the Question–Answer pair as the following: “(...) by basic I intend that although other actual sequences may be found empirically, their analysis will be accomplished best by seeing them as modifications of this ‘basic organization’ (...).”

Oushiro and Nasser (2010) propose a typology of questions based on the turn-taking system (Sacks, Schegloff, and Jefferson 1974) and the update of common ground among interlocutors (Clark 1996, Stalnaker 2002). This typology consists of three broad types of questions: “true” in-

¹⁰These hypotheses were tested in multivariate analyses in GoldVarb X which also included eight other factor groups. Among other results, the analyses indicate that variation in the position of the wh-word is a case of stable variation in the community (wh-in-situ being favored by younger (.53) and older speakers (.54) and disfavored by middle-aged speakers (.37)), not correlated with speakers’ sex/gender, level of education nor role in the sociolinguistic interview (interviewer or interviewee). Wh-in-situ is favored in shorter sentences, with less than six words (.54), and when the wh-word is an adverbial adjunct (.87). Other discourse-pragmatic factor groups correlated with the variable use of wh-interrogatives include the presence of a topic phrase in the sentence (.66) and the degree of predictability of the answer (.84 for more predictable answers). See Oushiro (2011) for a detailed account.

¹¹It is also important to stress that this concept of presupposition is different from that in formal syntax and semantics (cf. e.g., Zubizarreta 1997), which analyzes presuppositions in isolated propositions with minimal contribution from the context of use.

¹²For instance, it is possible to insert a Question–Answer pair (Q1 and A1) within another Question–Answer pair (Qb and Ab): “A: (Qb) Are you coming tonight? B: (Q1) Can I bring a friend? A: (A1) Sure. B: (Ab) I’ll be there.” (Schegloff 1972:78).

formation questions, rhetorical questions, and semi-rhetorical questions. The last two types are normally classified equally as rhetorical questions, that is, that do not require an answer from the interlocutor, but Oushiro and Nasser's (2010) proposal highlights their different mechanisms in talk-in-interaction. This typology was applied to the factor group Type of Question.

"True" information questions (15) are the ones closest to the prototype of Question–Answer pair, following the five characteristics defined above for an adjacency pair.¹³ Not only do they pass the turn, but also require an answer from the interlocutor and thus seek to update common ground. Rhetorical questions (16) perform the same function of assertions, as they are characterized by the obviousness of the presupposition (Rohde 2006); they cannot be characterized solely by their syntactic form, as the obviousness of the presupposition can only be determined from the current common ground among the participants of the conversational interaction. Notice that both questions in (16), in a different context, could be considered "true" information questions. As the presupposition is obvious, rhetorical questions can be replied to with "It's true" and do not update common ground among speakers. Semi-rhetorical questions (17) do not pass the turn. Unlike rhetorical questions, they do require an answer, as the presupposition is not obvious from current common ground; unlike "true" information questions, however, the answer is required from the current speaker and not from the interlocutor. As the presupposition is not obvious, they also seek to update common ground.

(15) "True" information questions

- a. Carla: *há quanto tempo ele tá lá?* (how-much time he is there?)
Pedro: *faz... há quatro anos...*
- b. Fabiana: *sua fita tem quantos minutos?* (your tape has how-many minutes?)
Suzana: *tem quarenta e cinco- não! noventa*

(16) Rhetorical questions

- a. José: *então a gente tá gastando... com duas eleições... quanto é que você gasta aí nesses dois turnos?* (how-much is-it that you spend there in-those two turns?) *quanto que vai em dinheiro?* (how-much that goes in money?)
Jorge: *é verdade...*
- b. Clara: *eu tenho que ter um carro... dá alguma coisa nela de noite eu vou aonde?* (I go where?)

(17) Semi-rhetorical questions

- a. Cecília: *vale a pena... é:: por que que vale a pena?* (why that worth the pain?) *porque nós raramente vamos ao cinema...*
- b. Marco: *e a televisão te traz o quê?...* (and TV brings what?) *só te traz coisa errada... é novela... ensina o quê?...* (soap opera teaches what?) *ensina o camarada a beber ensina a prostituição...*

The quantitative results for Type of Question (Table 2) show that semi-rhetorical questions, that is, those that require the current speaker to give an answer to his own question, favor the use of wh-in-situ (.68), whereas "true" information questions and rhetorical questions slightly disfavor its use (.47 and .45 respectively). This effect may be the manifestation of a strategy of turn-keeping. According to Sacks, Schegloff, and Jefferson (1974), speakers and listeners project the end of a conversational turn through the identification of linguistic units, which range from a single word to a whole sentence. The use of a wh-word in the beginning of a sentence may create the expectation that the interlocutor should take the next turn and answer the question; as semi-rhetorical questions do not aim at obtaining information from the interlocutor (on the contrary,

¹³In (15–23), (a) is an instance of wh-interrogative with a preverbal wh-word ("simple" wh-interrogative, wh-que interrogative, or cleft wh-interrogative) and (b) is an instance of wh-in-situ. Translation of examples (15–17): (15a) "C.: How long has he been there? P.: It's been... it's been four years." (15b) "F.: How long is your tape? S.: Forty-five- no! ninety." (16a) "J1.: So we are spending money on two elections... How much is spent on those two voting turns? How much money is spent? J2.: True..." (16b) "C.: I need to have a car... If something should happen to her at night, where will I go?" (17a) "C.: It's worth it... uh why is it worth it? Because we rarely go to the movies..." (17b) "M.: And what does TV bring you?... only wrong things... what do soap operas teach?... they teach drinking, prostitution..."

they require that the current speaker answer the question), wh-in-situ may better serve the purpose of keeping the conversational turn.¹⁴ In addition, notice that, in (17b), the speaker Marco could have said *e a televisão... só te traz coisa errada... novela... ensina o camarada a beber prostituição* ('TV only brings only wrong things... soap operas teach drinking prostitution'), without an interrogative pronoun. In employing a wh-word, the speaker is able to introduce new information in smaller chunks, which may facilitate cognitive processing.

Type of question	Weight	%	N <i>in-situ</i> /N total
Semi-rhetorical questions	.68	34.7	51/147
"True" information questions	.47	21.6	129/597
Rhetorical questions	.45	18.4	47/255
<i>Range: 23</i>		22.7	227/999

Table 2: Wh-in-situ and Type of Question. Input: 0.165. $p < 0.02$.

The second factor group, Givenness of the Presupposition, examines if presuppositions and referents recently activated in discourse have an influence on the choice of different syntactic structures of wh-interrogatives. The factors exemplified below are organized in different degrees of activation, from previous sentence (presupposition maximally activated) (18) to not at all so far in the recorded conversation (presupposition minimally activated) (23).¹⁵ If a referent (but not the full presupposition) had been activated in one of the five previous sentences, the instance was coded in a separate factor (22). In this factor group, rhetorical questions are excluded, as their presupposition is always activated; this hypothesis applies to "true" information questions and semi-rhetorical questions.¹⁶ In the following examples, the numbers in parentheses indicate the count of previous sentences.

- (18) Presupposition activated in the last sentence
- Pedro: *eu acho horrível... (1) acho horrível*
Carla: **por que** *você acha horrível?* (why you think horrible?)
 - Aline: *assim... (1) eles tão falando muito da saúde...*
Marcio: *é né?*
Aline: *mas tão falando da saúde por quê?* (but are talking of health why?)
- (19) Presupposition activated in second to last sentence
- Mariana: (2) *tem o caso das pessoas que têm dinheiro demais... e aí... (1) que que você acha? como- como que elas vivem com tanto dinheiro né?* (how that they live with so-much money?)
 - Gabriela: (2) *e você sabia que hoje em dia "a gente" é considerado um pronome pessoal do mesmo jeito que "eu você tu eles"?*
Lucia: (1) *não sabia*
Gabriela: *quando você...*

¹⁴This is not to suggest that speakers consciously make this decision.

¹⁵Translation of examples (18–23): (18a) "P.: I find it horrible... I find it horrible C.: Why do you find it horrible?" (18b) "A.: Like... they're talking a lot about health... M.: Yeah aren't they? A.: But why are they talking about health?" (19a) "M.: There are cases of people with too much money... and then... what do you think? How- how do they live with so much money right?" (19b) "G.: And did you know that, nowadays, "a gente" is considered a personal pronoun just like "eu você tu eles?" L.: I didn't. G.: When you... L.: But where is it so?" (20a) "B.: "I'm an outlaw sentenced for life and today I decided to wear my ball and chain." Then I said "No, I found it interesting! I didn't come to criticize... Why are you wearing them?" (20b) "V.: He died inside of- this brother of mine who died, he died in his house in the bathroom... He used to say he never had any health problems... and he wouldn't go to the doctor... R.: How old was he when he died?" (21a) "C.: Where again is this school you work at?" (21b) "R.: How old is he again?" (22a) "C.: I'd forgotten your sister's name... how old is she?" (22b) "A.: He doesn't know yet but he likes computers... J.: How old is he? Is he eighteen or -...?" (23a) "C.: What do you do for leisure?" (23b) "M.: Where do you come from?"

¹⁶The quantitative results reported in Table 3 refer to different runs in GoldVarb X which did not include the factor group Type of Question. All other factor groups were kept the same in the multivariate analyses.

- Lucia: *mas é considerado aonde?* (but is considered where?)
- (20) Presupposition activated in the third, fourth or fifth previous sentence
- a. Beatriz: “*eu sou bandido tô cumprindo prisão perpétua e (4) hoje eu resolvi por minha bola e minha corrente no pé*” aí (3) *eu falei assim* (2) “*não mas eu achei interessante... (1) eu não vim criticar... por que que o senhor tá usando isso?*” (why that you are wearing it?)
- b. Valter: *ele morreu dentro d/ (3) esse meu irmão que morreu morreu dentro da da casa dele dentro do banheiro... (2) ele falava que nunca tinha problema de saúde nenhuma... e (1) ele não ia no médico*
Rafael: *morreu com quantos anos?* (died with how-many years?)
- (21) Presupposition mentioned in conversation, but not in the previous five sentences
- a. Carolina: *aonde que é o colégio que cê dá aula mesmo?* (where that is the school that you give classes again?)
- b. Rafael: *agora tem que idade mesmo?* (now has which age again?)
- (22) Only one referent activated in one of the five previous sentences, but not full presupposition
- a. Carla: *eu já tinha esquecido o nome da sua irmã... quantos anos ela tem?* (how-many years she has?)
- b. Amanda: *não sabe ainda mas ele gosta de computador...*
Joaquim: *ele tá com quantos anos? tá com dezoito ou dez-...?* (he is with how-many years?)
- (23) Presupposition not previously activated
- a. Carlos: *que que cê faz como lazer assim?* (what that you do as leisure like?)
- b. Marina: *você vem de que canto pra cá?* (you come from which corner to here?)

Givenness of the presupposition	Weight	%	N <i>in-situ</i> /N total
Run 1: ^a			
+Presupp. 1 st or 2 nd preceding sentence	.66	35.2	56/159
Referent activated in 5 previous sent.	.59	27.8	42/151
+Presupp. 3 rd sentence or before	.49	22.5	20/89
Presupp. not activated	.39	19.0	62/345
<i>Range: 27</i>		24.2	180/744
Run 2: ^b			
+presupp./ref. 1 st or 2 nd preceding sentence	.63	34.0	89/202
+presupp./ref. 3 rd sentence or before	.51	21.2	29/137
Presupp. not activated	.38	18.0	62/345
<i>Range: 25</i>		24.2	180/744

Table 3: Wh-in-situ and Givenness of the Presupposition. ^aInput: 0.115, $p < 0.05$. Log-likelihood: -388.695. ^bInput: 0.115, $p < 0.05$. Log-likelihood: -389.425. + amalgamation of factors.

Table 3 shows the results of two different runs in GoldVarb X. In the first one, the two maximally activated degrees (last and second to last sentence) were joined in a single factor, as well as the two intermediate degrees (presupposition activated in the third, fourth or fifth previous sentences, and mentioned in the conversation but not in the previous five sentences) ($\chi^2 = 0.77(2)$, $p > 0.50$). In this run, the factor for referents (only one referent activated in one of the five previous sentences, but not full presupposition) is the second factor favoring wh-in-situ (.59), right between the factors referring to presuppositions activated in the last or second to last sentence (.66), and presuppositions activated in the third previous sentence or before (.49). In the second run, the factor for referents was then separated into two factors parallel to those of presuppositions (referent activated in last or second to last previous sentence; referent activated in third, fourth, or fifth previous sentence) ($\chi^2 = 1.46(1)$, $p > 0.20$). The first observation to be made over these results is that there does not seem to be a significant difference whether a referent or a full presupposition was activated in discourse.

Results also show that the more activated the presupposition or one referent, the greater the tendency to employ wh-in-situ (.63), and that this tendency drops if presuppositions or referents

had been activated earlier in discourse. In addition, notice that there are no significant differences from the third previous sentence or before; in other words, the significant difference is found between presuppositions activated in the last/second to last sentence, which is indicative of the importance of the *hic et nunc* in conversation: it suggests that speakers are highly sensitive to the flow of information and follow the update of common ground turn by turn.

A similar hypothesis has been tested by Weiner and Labov (1983 [1977]), in their study of the alternation between active and agentless passive voice in English, regarding the “given” status of referents: they verified that recently activated referents have an influence on the use of the agentless passive voice, but also that there is a strong mechanical tendency to preserve parallel structures in discourse. The fact that speakers tend to employ wh-in-situ in instances such as the ones exemplified in (18) and (19) above can also be compared to Weiner and Labov’s (1983 [1977]) results: here too there is a tendency to preserve parallelism. The parallel structures of (18b) and (19b) are underlined in (24–25):

- (24) *eles tão falando muito da saúde... (...) mas tão falando da saúde por quê?* (they are talking a-lot about health... (...) but they’re talking about health why?)
- (25) (...) *hoje em dia “a gente” é considerado um pronome pessoal do mesmo jeito que “eu você tu eles”?* (...) *mas é considerado aonde?* (nowadays “a gente” is considered a personal pronoun just like “eu você tu eles?” (...) but is considered where?)

Results for both factor groups can also be the consequence of a more general tendency of discourse organization, namely to postpone the introduction of new information (Chafe 1974). As semi-rhetorical questions introduce new information, the structure of wh-in-situ better aligns with the sequence given–new. Similarly, the more activated a presupposition or a referent, the greater the tendency to organize discourse from these (maximally) given pieces of information towards new ones.

4 Conclusions

In BP, and more specifically in the Paulistano variety of spoken language, wh-in-situ can be employed alternatively to the structures with a moved wh-word, as it does not suffer categorical discourse-pragmatic restrictions, as do European Portuguese and other Romance languages. Its productivity in contemporary spoken language (22.7% in this corpus) is an evidence of its variable use, performing the functions of “true” information questions as well as rhetorical and semi-rhetorical questions.

The analyses of the factor groups Type of Question and Givenness of the Presupposition show that speakers are highly sensitive to the flow of information in talk-in-interaction, which has consequences for the alternation between different syntactic structures of wh-interrogatives. Wh-in-situ is favored when the presupposition had been most recently activated and in semi-rhetorical questions, which can be a strategy for turn-keeping, cognitive processing and general discourse organization. Both factor groups suggest that common ground among interlocutors plays an important role in the variable use of wh-interrogatives and that it must be examined in the dynamics of talk-in-interaction, since it is constantly updated throughout conversation.

References

- Ambar, Manuela, Mary A. Kato, Carlos Mioto, and Rita Veloso. 2001. Padrões de interrogativas-Q no português europeu e no português brasileiro: uma análise inter e intra-linguística. *Boletim da Associação Brasileira de Linguística* 26:400–404.
- Chafe, Wallace. 1974. Language and consciousness. *Language* 50:111–133.
- Cheng, Lisa Lai-Shen. 1991. On the Typology of Wh-questions. Doctoral dissertation, MIT.
- Cheng, Lisa Lai-Shen, and Johan Rooryck. 2000. Licensing wh-in-situ. *Syntax* 31:1–19.
- Chomsky, Noam. 1977. On wh-movement. In *Formal Syntax*, ed. P. Culicover, T. Wasow, and A. Akmajian, 71–132. New York: Academic Press.
- Clark, Herbert. 1996. *Using Language*. Cambridge: Cambridge University Press.

- Coveney, Aidan. 2002 [1996]. *Variability in Spoken French: a Sociolinguistic Study of Interrogation and Negation*. Bristol, UK: Elm Bank, 2nd ed.
- Duarte, Maria Eugênia Lamoglia. 1992. A perda da ordem V(erbo) S(ujeito) em interrogativas qu- no português do Brasil. *DELTA* 8:37–52.
- Guy, Gregory R. 1988. Advanced Varbrul analysis. In *Linguistic Change and Contact*, ed. K. Ferrara et al., 124–136. Austin: University of Texas.
- Hall, Damien John. 2008. A Sociolinguistic Study of the Regional French of Normandy. Doctoral dissertation, University of Pennsylvania.
- Kato, Mary A. 1987. A inversão da ordem SV em interrogativas no português: uma questão sintática ou estilística? *DELTA* 3:243–252.
- Kato, Mary A. 2004. Two types of wh-in-situ in Brazilian Portuguese. Paper presented at Georgetown Round Table, Washington D.C.
- Kato, Mary A., and Carlos Mioto. 2005. A multi-evidence study of European and Brazilian Portuguese wh-questions. In *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, ed. S. Kepsner and M. Reis, 307–328. Berlin, New York: Mouton de Gruyter.
- Labov, William. 1978. Where does the linguistic variable stop? A response to Beatriz Lavandera. *Working Papers in Sociolinguistics* 44.
- Lessa de Oliveira, Adriana Stella Cardoso. 2005. Aquisição da linguagem e variação linguística. *Estudos Linguísticos* 34:409–414.
- Lopes-Rossi, Maria Aparecida Garcia. 1993. Estudo diacrônico sobre as interrogativas no português do Brasil. In *Português Brasileiro: uma Viagem Diacrônica*, ed. I. Roberts and M. A. Kato, 307–342. Campinas: Unicamp.
- Milroy, Leslie, and Matthew Gordon. 2003. *Sociolinguistics: Method and Interpretation*. Malden, Mass.: Blackwell Pub.
- Mioto, Carlos. 1997. Wh é que ≠ wh que. *Estudos Linguísticos* 26:648–654.
- Mioto, Carlos, and Maria Cristina Figueiredo-Silva. 1995. Wh é que = wh que? *DELTA* 11:301–311.
- Oushiro, Livia. 2011. Uma Análise Variacionista para as Interrogativas-Q. Master's dissertation, Universidade de São Paulo.
- Oushiro, Livia, and Juliana Antunes Nasser. 2010. O par pergunta-resposta em entrevistas sociolinguísticas. In *Anais do Congresso Internacional Linguagem e Interação II*, ed. M. Sell and A.M.M. Guimarães, 768. São Leopoldo: Casa Leiria.
- Pires, Acrísio, and Heather Lee Taylor. 2007. The syntax of wh-in-situ and common ground. Paper submitted to *Romance Languages: Structure, Interfaces, and Microparametric Variation*, ed. P. Masullo. Amsterdam: John Benjamins. URL <http://www.ling.umd.edu/~htaylor/papers/PiresTaylor07LSRL.pdf>.
- Rohde, Hannah. 2006. Rhetorical questions as redundant interrogatives. *San Diego Linguistics Papers* 2:134–168.
- Sacks, Harvey. 1987 [1973]. On the preferences for agreement and contiguity in sequences in conversation. In *Talk and Social Organization*, ed. G. Button and J.R.E Lee, 54–69. Clevedon, UK: Multilingual Matters.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735.
- Sankoff, Gillian. 1980. A quantitative paradigm for the study of communicative competence. In *The Social Life of Language*, ed. G. Sankoff, 47–79. Philadelphia: University of Pennsylvania Press.
- Schegloff, Emanuel A. 1972. Notes on a conversational practice: formulating place. In *Studies in Social Interaction*, ed. D. Sudnow, 75–119. New York: The Free Press.
- Schegloff, Emanuel A., and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8:289–327.
- Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25:701–721.
- Tagliamonte, Sali. 2006. *Analysing Sociolinguistic Variation*. São Paulo: Cambridge University Press.
- Weiner, Judith, and William Labov. 1983 [1977]. Constraints on the agentless passive. *Journal of Linguistics* 19:29–58.
- Wolfram, Walt. 1993. Identifying and interpreting variables. In *American Dialect Research*, ed. D. R. Preston, 193–221. Amsterdam/Philadelphia: John Benjamins.
- Zubizarreta, Maria Luiza. 1997. Focus, Prosody, and Word Order. Doctoral dissertation, University of Southern California in Los Angeles.

Departamento de Linguística
 Universidade de São Paulo
 Av. Prof. Luciano Gualberto, 403
 Cidade Universitária – São Paulo – SP 05508–900
 livia.oushiro@usp.br