

# Untangling syntactic and morphophonological effects on Brazilian Portuguese Wh-interrogatives<sup>1</sup>

Livia Oushiro  
livia.oushiro@usp.br  
Universidade de São Paulo

Ronald Beline Mendes  
rbeline@usp.br  
Universidade de São Paulo

## 1. Introduction

This paper analyzes variation among four different structures of Wh-Interrogatives in Brazilian Portuguese (BP), as shown in (1–4). Wh-Interrogatives are defined as sentences containing a wh-word (*o que* ‘what’, *quem* ‘who’, *qual (-is)* ‘which’, *quando* ‘when’, *quanto (-os, -a, -as)* ‘how much/many’, *como* ‘how’, *onde* ‘where’, *pra que* ‘what for’, *por que* ‘why’), both in main (1a–4a) and subordinate (1b–3b) clauses.

- (1) ‘Simple’:       a. *Onde você mora?*  
                      b. *Eu não sei onde você mora.*
- (2) Wh-que:       a. *Onde que você mora?*  
                      b. *Eu não sei onde que você mora.*
- (3) Cleft:         a. *Onde é que você mora?*  
                      b. *Eu não sei onde é que você mora.*
- (4) Wh-in-situ:   a. *Você mora onde?*  
                      b. *\*Eu não sei você mora onde.*  
                      ‘Where do you live?’ / ‘I don’t know where you live.’

In European Portuguese, (2) is ungrammatical (Lopes-Rossi, 1996; Kato & Miotto, 2005) and (4) is restricted to echo-questions (Ambar et al, 2001). In BP, we present token and type evidence that it is not the case. Both structures are relatively productive in spoken language (see Section 2), and can be equally employed in information, rhetorical, and semi-rhetorical questions (Oushiro, 2011). In this paper, we show that (i) these structures alternate as variants in BP, with semantic-pragmatic equivalence, and can be defined as two variables: the position of the wh-word in main clauses (in situ or not – 4a vs. 1a–3a), and the presence of the complementizer *que* (2a-b vs. 1a-b/3a-b); and (ii) we contrast different quantitative analyses correlating Syntactic Function and Wh-word with both variables, in order to untangle and better interpret the effects of syntactic and morphophonological factors.

## 2. Analyses

We analyzed these structures in a contemporary corpus of 53 sociolinguistic interviews (about half a million words) with native Paulistano speakers stratified according to their sex/gender, three age groups, and two levels of education. Quantitative analyses were performed in GoldVarb X.

Table 1 *General distribution*

	Variable position		Variable <i>que</i>	
	N	%	N	%
<i>‘Simple’</i>	289	28.9	721	50.8
<i>Wh-que</i>	408	40.8	579	40.7
<i>Cleft</i>	75	7.5	121	8.5
<i>Wh-in-situ</i>	227	22.7	N/A	N/A
	999	100.0	1421	100.0

<sup>1</sup> This research was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant number 2009/03190-0.

Table 1 shows the general distribution of the four structures in the two variables. Wh-in-situ accounts for 22.7% of Variable position from a total of 999 tokens, and wh-que interrogatives account for 40.7% of 1421 tokens of Variable *que*. These frequencies are the first evidence that wh-in-situ and wh-que alternate with the other two structures of Wh-interrogatives, i.e., that they are not marked structures with highly specific and categorical discourse-pragmatic constraints.

The extract in (5) is an example from our corpus that wh-in-situ is used alternatively to the structures with a preverbal wh-word (Variable position). The speaker Helena asks about her interlocutor's sister's age with a wh-in-situ (5a), and then asks about her parents' age with a 'simple' wh-interrogative (5b).<sup>2</sup>

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

Example (6) shows variation among the three structures of preverbal Wh-interrogatives (Variable *que*). The same speaker Gabriel employs 'simple' (6b), wh-que (6a, 6d) and cleft (6c) wh-interrogatives to ask about different accents.<sup>3</sup>

- (6) a. Gabriel: *como que é o sotaque do paulista?*  
(how that is the accent of Paulista?)  
Bruno: *ah... eu acho que é mais... é mais aberto pra você entender*
- b. Gabriel: *como carioca fala?*  
(how Carioca speaks?)  
Bruno: *é é a pronúncia deles né?*  
Gabriel: *a pronúncia deles é irritante?*  
Bruno: *a língua puxada né?*
- c. Gabriel: *mas como é que o carioca fala?*  
(how is-it that Carioca speaks?)  
Bruno: *ah não sei te falar não*
- d. Gabriel: *não... e como que o pessoal do interior fala?*  
(and how that the people in the country speak?)  
Bruno: *meio caipira né?*

Tokens were coded for 13 factor groups for Variable position<sup>4</sup> and 10 factor groups for Variable *que*.<sup>5</sup> For both variables, there were two similar factor groups: (i) Syntactic function (subject, direct object, indirect object, and different types of adverbial adjuncts), and (ii) Wh-word (*que* 'what', *qual* 'which', *quem* 'who', *quanto* 'how-much/how-many', *como* 'how', *onde* 'where', *quando* 'when', *pra que* 'what for', *por que* 'why'). Although these factor

<sup>2</sup> Translation of example (5): "H.: and how old is your sister? I.: uhn she's four years younger than me she's thirty-one H.: oh it's true... and your parents? how old are they? I.: my father is sixty and my mother is fifty-six..."

<sup>3</sup> Translation of example (6): "G.: what's Paulista speech like? B.: oh... I think it's more... it's easier to understand G.: how do Cariocas speak? B.: it's- it's their pronunciation you-know G.: their pronunciation is annoying? B.: they drawl you-know G.: but how do Cariocas speak? B.: oh I can't really say G.: and how do people in the country speak? B.: kinda like rednecks you-know?"

<sup>4</sup> Sex/gender, age, level of education, role of speaker, type of question, syntactic function, wh-word, type of verb, extension of sentence, elements before main clause, group of answers, givenness of the presupposition, type of question (2).

<sup>5</sup> Sex/gender, age, level of education, role of speaker, type of question, syntactic function, wh-word, type of verb, type of clause, following syllable stress.

groups are *not* the same, they are clearly not independent from each other – for instance, *por que* ‘why’ is always an adverbial adjunct of cause. These two factor groups were first analyzed separately in multivariate analyses for both variables. Table 2 thus shows the results of four different GoldVarb runs.

Table 2 Syntactic function and Wh-word in separate analyses for each variable

	Variable position				Variable <i>que</i>		
		%	N			%	N
<b>Wh-word<sup>a</sup></b>				<b>Wh-word<sup>c</sup></b>			
<i>quando</i>	.85	62.5	5	<i>que</i>	.90	87.8	122
<i>onde</i>	.80	47.4	36	<sup>+</sup> <i>porque/prá que</i>	.77	72.4	89
<i>o que</i>	.74	33.7	85	<i>o que</i>	.50	36.4	156
<i>quem</i>	.72	29.2	14	<sup>+</sup> <i>qual/qual+NP</i>	.47	43.5	50
<sup>+</sup> <i>quanto/quanto+NP</i>	.71	39.2	20	<i>que+NP</i>	.40	33.7	29
<i>que+NP</i>	.63	28.6	24	<i>como</i>	.39	30.6	85
<sup>+</sup> <i>por que/prá que</i>	.61	22.3	27	<i>quanto/quanto+NP</i>	.33	23.5	16
<i>que</i>	.27	7.3	9	<i>quem</i>	.31	21.2	21
<sup>+</sup> <i>qual/qual+NP</i>	.21	4.5	4	<i>onde/quando</i>	.17	12.9	11
<i>como</i>	.09	2.1	3				
<i>Range:</i>	<i>76</i>	<i>22.7</i>	<i>227</i>	<i>Range:</i>	<i>73</i>	<i>40.7</i>	<i>579</i>
<b>Syntactic Function<sup>b</sup></b>				<b>Syntactic Function<sup>d</sup></b>			
<sup>+</sup> other adv. adj.	.83	55.0	72	<sup>+</sup> adv.adj.cause/purp.	.75	70.9	90
indirect object	.79	48.1	26	adv.adj.time	.67	62.5	10
direct object	.58	24.4	81	direct object	.58	47.2	210
<sup>+</sup> adv.adj.cause/purp.	.53	22.0	27	subject	.49	35.4	46
adv. adj. manner	.19	4.5	5	indirect object	.36	25.3	25
subject	.16	3.6	2	<sup>+</sup> other adv.adj.	.35	26.7	70
<i>Range:</i>	<i>67</i>	<i>23.5</i>	<i>227</i>	<i>Range:</i>	<i>40</i>	<i>40.7</i>	<i>579</i>

<sup>a</sup>Input: 0.112,  $p < 0.05$ ; <sup>b</sup>Input: 0.115,  $p < 0.05$ ; <sup>c</sup>Input: 0.400,  $p < 0.02$ ; <sup>d</sup>Input: 0.398,  $p < 0.02$ . Mismatch between factor weights and percentages indicates interaction. <sup>+</sup> indicates amalgamation of factors.

Both Wh-word and Syntactic function are always selected first in their respective runs for both variables. Wh-word exhibits a greater range than Syntactic function (Cf. 76 vs. 67 for Variable position, 73 vs. 40 for Variable *que*), but this is probably due to its greater number of factors. These results are hard to interpret: there are problems in those analyses, interaction among factors (as indicated by the grey shaded cells), and too few tokens for some factors. From these results, it is not clear if these effects are mostly syntactic or morphological.

We then proceeded to analyses with both factor groups in the same run for each variable, containing only factors that are orthogonal to others (Table 3). The analysis of Variable position selects Syntactic function as the main factor group influencing variation, and Wh-word second. In Syntactic function, there is a coherent hierarchy of constraints: arguments with the function of subject strongly disfavor the use of wh-in-situ, followed by direct objects, whereas indirect objects and adverbials highly favor wh-in-situ. In this case, we can conclude that the greater the movement the wh-constituent would do, the greater the tendency for it to remain in situ. In Variable *que*, the opposite happens: Wh-word is the first group to be selected, whereas Syntactic function is the third, with a very small range. This means that for variation among the three structures of preverbal Wh-interrogatives, syntactic factors play a minor role, differently from Variable position.

To account for the fact that the wh-word *que* ‘what’ favors wh-*que* interrogatives so strongly, it is useful to go back to the analysis containing all factors in this group. Table 2

shows that the factors mostly favoring wh-*que* are *QUE*, *por QUE*, *pra QUE*, and *o QUE*. These words have in common the fact that they are all oxytones, whereas most of the disfavoring factors are paroxytones: *QUANdo*, *ONde*, *QUANto*, *COmo*, and the wh-words followed by an NP.<sup>6</sup> Since complementizer *que* is always unstressed in BP, it seems that Variable *que* follows the Principle of Rhythmic Alternation (Selkirk, 1984), as the use of the word *que* can avoid sequences of stressed syllables (Cf. e.g. *por QUE que VAle a PEEna?* vs. *por QUE VAle a PEEna?* ‘why is it worth it?’).

Table 3 *Syntactic function and Wh-word in joint analyses for each variable*

	Variable position <sup>a</sup>				Variable <i>que</i> <sup>b</sup>		
		%	N			%	N
<b>Syntactic function</b>				<b>Wh-word</b>			
adverbial adjunct	.87	55.0	72	<i>que</i>	.88	87.8	122
indirect object	.84	48.1	26	<sup>+</sup> <i>o que/qual(-is)</i>	.45	37.5	199
direct object	.51	24.4	81	<sup>+</sup> <i>que/qual/quanto+NP</i>	.35	31.0	44
subject	.20	6.4	16	<i>quem</i>	.32	21.2	21
	<i>Range:</i>	67	25.5	195	<i>Range:</i>	56	42.4
							386
<b>Wh-word</b>				<b>Syntactic function</b>			
<i>o que</i>	.76	33.7	85	direct object	.52	47.2	210
<i>quem</i>	.74	29.2	14	subject	.51	36.9	174
<i>que+NP</i>	.41	28.6	24	indirect object	.42	24.6	16
<i>que</i>	.21	7.3	9	adv. adjunct	.42	23.6	30
<i>qual(-is)</i>	.20	4.5	4				
	<i>Range:</i>	56	22.8	136	<i>Range:</i>	9	38.8
							430

<sup>a</sup>Input: 0.121,  $p < 0.05$ ; <sup>b</sup>Input: 0.416,  $p < 0.03$ . <sup>+</sup> indicates amalgamation of factors.

## Conclusion

Although both Syntactic function and Wh-word are selected as significant factor groups for both Variable position and Variable *que*, the contrastive analyses show that each dependent variable is mostly conditioned by constraints of different nature: Variable position by syntactic factors (the greater the movement, the greater the tendency for the wh-word to remain in situ), and Variable *que* by morphophonological factors, following the Principle of Rhythmic Alternation (Selkirk, 1984).

## References

- Ambar, Manuela, Kato, Mary A., Mioto, Carlos & Veloso, Rita. (2001) Padrões de interrogativas-Q no português europeu e no português brasileiro: uma análise inter e intra-linguística. *Boletim da Associação Brasileira de Linguística*, vol. 26, 400–404.
- Lopes-Rossi, Maria Aparecida Garcia. (1996) *A sintaxe diacrônica das Interrogativas-Q do português*. PhD Thesis. IEL/Unicamp.
- Mioto, Carlos & Kato, Mary. (2005) As interrogativas Q do português europeu e do português brasileiro atuais. *Revista da ABRALIN*, vol. 4, no. 1 e 2, 171–196.
- Oushiro, Livia. (2011) Wh-interrogatives in Brazilian Portuguese: the influence of common ground. *University of Pennsylvania Working Papers in Linguistics* 17(2), 145–154. Available at <<http://repository.upenn.edu/pwpl/vol17/iss2/17/>>.
- Selkirk, Elisabeth O. (1984) *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Massachusetts: MIT Press.

<sup>6</sup> Most nouns in Portuguese are paroxytones.