UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM SEMIÓTICA E LINGUÍSTICA GERAL

Uma análise variacionista para as Interrogativas-Q

LIVIA OUSHIRO

Dissertação apresentada ao Programa de Pós-Graduação em Semiótica e Linguística Geral do Departamento de Linguística da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de mestre em Linguística.

Orientador: Prof. Dr. Ronald Beline Mendes

São Paulo
2011

*Para o Pedro,
olhos de girassol,
que torna minha vida mais perrerez.*

Agradecimentos

Essa dissertação é fruto de uma trajetória acadêmica de mais de 10 anos. Nesse período, muitas pessoas contribuíram para meu desenvolvimento acadêmico e pessoal.

Agradeço, antes de tudo, a minha família e a meus amigos. A meus pais, Kengo e Iuri, por apoiarem desde sempre meus estudos e por me ensinarem – não só através de palavras, mas principalmente através de exemplos cotidianos – a ser uma pessoa responsável e feliz; a minha irmã, Luciana, por estar sempre ao meu lado em São Paulo; ao Pedro, pelos mais de 10 anos de convivência que ele preencheu diariamente com amor e com cuidados; à Júlia, pela doçura com que trata a todos; à Ana Maria e ao Rubens, por terem me acolhido em sua família e por me fazerem sentir em casa sempre que vou a Cotia.

Agradeço à Camila Zanon, ao Tiago Penna, ao Uiran Gebara e ao Luciano Ferreira por todas as noites de sábado compartilhadas; por todas as conversas produtivas e improdutivas; por me alertarem de todos os percalços da vida na pós-graduação; ou, mais simplesmente, por sua amizade sincera.

Agradeço ao professor Ronald Beline Mendes pela orientação firme com que ele tem guiado meu desenvolvimento acadêmico nos últimos quatro anos; por saber a medida e a hora exata de quando criticar meu trabalho e de quando tecer elogios; por me mostrar que sempre é possível melhorar.

Agradeço aos professores Evani Viotti e Leland McCleary pelo acolhimento em seu grupo de pesquisa, o LLIC, e por serem, para mim, dois grandes modelos de pesquisadores.

Agradeço aos professores da Letras da USP, a quem devo minha sólida formação. Em especial, agradeço ao professor Marcos Lopes, por todo o incentivo desde a graduação e por sua capacidade de abrir horizontes; ao

professor Mário Viaro, pelas mais inspiradoras aulas de Fonética, Fonologia e Morfologia; à professora Esmeralda Negrão, que instigou meu gosto pela Sintaxe, por ser um exemplo de retidão acadêmica; aos professores Ana Paula Scher e Waldir Beividas, enquanto chefe de departamento e coordenador da pós-graduação, pelo apoio constante aos alunos de Linguística; à professora Elaine Grolla, por compartilhar o interesse no tema desta pesquisa e por me fornecer tantas referências sobre Interrogativas-Q.

Agradeço aos professores Gregory Guy e Paulo Chagas de Souza pelas sugestões oferecidas a esta pesquisa por ocasião do exame de qualificação; aos professores Anthony Naro, Marta Scherre, James Walker, Scott Schwenter e Sebastião Carlos Gonçalves, agradeço igualmente por todos os comentários. Todos esses professores contribuíram para o enriquecimento desta dissertação.

Agradeço a todos os colegas do Grupo de Estudos e Pesquisa em Sociolinguística, pelas tardes de discussão na sala 236 e pelos fins de tarde no Duetos. Sinto imenso prazer e orgulho em ver nosso grupo crescer, fato que me incentiva ainda mais a desenvolver pesquisas em nosso campo. Aos colegas do Grupo de Estudos do R, agradeço pela ajuda na resolução de quebra-cabeças de programação e pelas divertidas reuniões que tivemos; a todos os colegas do LLIC, pelas estimulantes discussões e pelo companheirismo.

Agradeço aos amigos linguistas que conheci nos últimos quatro anos e que já fazem parte do rol de pessoas queridas: Eneida Leal, Fernanda Canever, Juliana Nasser, Fátima de Almeida Baia, Danilo Carlos, Ivan Zanni, Marcus Lunguinho, Melissa Cruz e Ivanete Belém do Nascimento. A experiência da pós-graduação é ainda mais gratificante por causa de vocês.

Agradeço à Érica, ao Robson e ao Ben Hur, por me auxiliarem com todos os trâmites da universidade. Pelo financiamento desta pesquisa, agradeço ao CNPq (pelos seis primeiros meses) e à FAPESP (pelos 18 meses restantes).

Resumo

A investigação de variáveis sintáticas e morfossintáticas dentro do quadro da Sociolinguística Variacionista tem recebido, em geral, menor atenção do que o estudo de variáveis fonológicas.

Este trabalho analisa a variação entre quatro estruturas de Interrogativas-Q em amostras de fala e de escrita do português paulistano contemporâneo: (i) *interrogativas-qu* (“Onde você mora?”); (ii) *interrogativas qu-que* (“Onde que você mora?”); (iii) *interrogativas é-que* (“Onde é que você mora?”); e (iv) *interrogativas qu-in-situ* (“Você mora onde?”). A equivalência semântica entre as formas interrogativas se estabelece através do conceito de *pressuposição do falante* (Stalnaker, 2002). Além disso, este trabalho propõe o conceito de *competência comunicativa* (Hymes, 1991 [1979]) como critério para determinar o envelope de variação: diferentes estruturas são consideradas variantes se forem factualmente possíveis, factíveis, adequadas e empregadas nos mesmos contextos. A análise qualitativa com base nesses conceitos define dois envelopes de variação – e, portanto, duas variáveis: uma que envolve a alternância na posição do constituinte interrogativo (*in situ* ou não), e outra que encerra as três estruturas com constituinte interrogativo pré-verbal (*-qu*, *qu-que* e *é-que*).

Os resultados das análises quantitativas mostram que *interrogativas qu-in-situ* são favorecidas principalmente por fatores morfossintáticos e discursivo-pragmáticos; fatores extralinguísticos, como o sexo/gênero e a faixa etária

do falante, também se correlacionam indiretamente através do emprego de diferentes estratégias discursivas. O uso de *interrogativas qu-que*, por sua vez, demonstra uma provável mudança linguística em progresso, uma vez que a análise em tempo aparente revela o favorecimento da estrutura por falantes mais jovens. Neste caso, a variação é influenciada principalmente por fatores sintáticos e prosódicos.

Além de propor critérios para o estudo de variáveis morfossintáticas, este trabalho discute os resultados das análises quantitativas em perspectiva com outros níveis de variação linguística, com vistas a integrá-los em um quadro mais amplo da Teoria da Variação.

Palavras-chave: Interrogativas-Q; interrogativas qu-*in-situ*; interrogativas qu-que; variação morfossintática; português paulistano

Abstract

The investigation of syntactic and morphosyntactic variables within the framework of Variationist Sociolinguistics has received, in general, less attention than the study of phonological variables.

This study analyzes variation among four different structures of Wh-Interrogatives in the speech and writing of native paulistanos: (i) ‘simple’ wh-interrogatives (as in *Onde você mora?* ‘Where you live?’); (ii) wh-que interrogatives (as in *Onde que você mora?* ‘Where-that you live?’); (iii) cleft wh-interrogatives (as in *Onde é que você mora?* ‘Where is-it that you live?’); and (iv) wh-in-situ (as in *Você mora onde?* ‘You live where?’). Semantic equivalence among the four structures is established through the concept of *speaker’s presupposition* (Stalnaker, 2002). Further, this study proposes the concept of *communicative competence* (Hymes, 1991 [1979]) as a criterium for determining the envelope of variation: different structures are considered variants if they are factually possible, feasible, appropriate, and performed in the same contexts. The qualitative analysis based on these concepts defines two different envelopes of variation – hence, two variables: one which involves the alternation in the position of the wh-word (in situ *vs.* moved), and another which comprises the three structures with a moved wh-word (‘simple’, wh-que and cleft-wh).

Results of the quantitative analyses show that wh-in-situ is mostly favored by morphosyntactic and discourse-pragmatic factors, and that non-linguistic

factor groups such as gender and age are also indirectly correlated with the variation through the use of different discourse strategies. The use of *wh*-*que*, on the other hand, is probably undergoing change, as apparent time analyses show a favoring effect by younger speakers. Variation in this case is mostly influenced by syntactic and prosodic factors.

In addition to proposing criteria to the study of morphosyntactic variables, this dissertation discusses the results of the quantitative analyses in relation to other levels of linguistic variation, aiming at their integration into a larger framework of Variation Theory.

Keywords: Wh-Interrogatives; *wh*-in-situ; *wh*-*que* interrogatives; morphosyntactic variation; Paulistano Portuguese

Sumário

Resumo	v
Abstract	vii
Lista de Tabelas	xiii
Lista de Figuras	xiv
Introdução	1
1 Pressupostos teóricos e metodológicos	10
1.1 A análise da língua em uso e da variação	10
1.2 O paradigma quantitativo	13
1.3 O estudo sociolinguístico de variáveis morfossintáticas	17
2 Dados	28
2.1 Corpus	29
2.2 Dados de documentadores: o Paradoxo do Observador revisitado	37
2.3 Interrogativas e perguntas: forma e função	43
3 Análise qualitativa	52
3.1 Equivalência semântica	53
3.2 Os envelopes de variação	56
3.2.1 Formas possíveis: análise de gramaticalidade	57
3.2.2 Formas factíveis: análise de viabilidade	61
3.2.3 Formas adequadas: análise pragmática	62

3.2.4	Formas empregadas: análise de produção	64
3.3	Grupos de fatores	68
3.3.1	Grupos de fatores extralinguísticos	69
3.3.2	Grupos de fatores linguísticos	72
3.3.3	Interações entre grupos de fatores	86
4	A variável (Posição do constituinte interrogativo)	89
4.1	Envelope de variação e Distribuição geral dos dados	90
4.2	Grupos de fatores extralinguísticos	91
4.3	Grupos de fatores linguísticos	100
5	A variável (Constituinte interrogativo pré-verbal)	117
5.1	Envelope de variação e Distribuição geral dos dados	118
5.2	Grupos de fatores extralinguísticos	119
5.3	Grupos de fatores linguísticos	124
	Conclusão	133
	Referências	139
	Anexos	148
	A – Questionário enviado aos documentadores	148
	B – <i>Script</i> para extração de dados de interrogativas de constituinte com o programa R	150
	C – Roteiro da entrevista sociolinguística	153

Lista de Tabelas

1	Comparação da frequência de Interrogativas-Q em diferentes <i>corpora</i> do português paulistano	7
1.1	Número de ocorrências de diferentes variáveis em 32 entrevistas da Amostra Censo 00	23
2.1	Número de informantes por célula no <i>corpus</i> de entrevistas sociolinguísticas	31
2.2	Perfil sociolinguístico dos documentadores	32
2.3	Distribuição geral de Interrogativas-Q no <i>corpus</i> de entrevistas sociolinguísticas	33
2.4	Distribuição geral de Interrogativas-Q no <i>corpus</i> de língua escrita	35
3.1	Interrogativas-Q	53
3.2	Possibilidades de estruturação sintática das Interrogativas-Q	60
3.3	Distribuição geral de ocorrências de Interrogativas-Q no <i>corpus</i> de entrevistas sociolinguísticas de acordo com os contextos sintáticos relevantes	67
3.4	Grupos de fatores extralinguísticos (Envelopes de Variação I e II)	69
3.5	Grupos de fatores linguísticos (Envelopes de Variação I e II)	73
3.6	Grupos de fatores linguísticos (Envelope de Variação I)	77
3.7	Grupos de fatores linguísticos (Envelope de Variação II)	84

4.1	Distribuição de Interrogativas-Q no <i>corpus</i> de entrevistas sociolinguísticas no Envelope de Variação I e no Quadro Geral	90
4.2	Distribuição de ocorrências de <i>interrogativas qu-in-situ</i> em relação a Quem Fala, Sexo/Gênero e Escolaridade	91
4.3	Distribuição de ocorrências de <i>interrogativas qu-in-situ</i> em relação ao cruzamento de Quem fala e Sexo/gênero	93
4.4	Correlação entre <i>interrogativas qu-in-situ</i> e Faixa etária	94
4.5	Correlação entre <i>interrogativas qu-in-situ</i> e o cruzamento de Quem fala e Faixa etária	95
4.6	Correlação entre <i>interrogativas qu-in-situ</i> e grupos de fatores extralinguísticos	97
4.7	Correlação entre <i>interrogativas qu-in-situ</i> e a ocorrência de outro elemento antes da oração principal	101
4.8	Correlação entre <i>interrogativas qu-in-situ</i> e Sinceridade Pragmática da Pergunta	102
4.9	Correlação entre <i>interrogativas qu-in-situ</i> e o cruzamento de Sinceridade pragmática da pergunta e Sexo/gênero	104
4.10	Correlação entre <i>interrogativas qu-in-situ</i> e grupos de fatores de subconjuntos de dados	106
4.11	Correlação entre <i>interrogativas qu-in-situ</i> e Grau de ativação do fundo comum recodificado	108
4.12	Correlação entre <i>interrogativas qu-in-situ</i> e grupos de fatores morfossintáticos	109
4.13	Correlação entre <i>interrogativas qu-in-situ</i> e grupos de fatores Função sintática e Constituinte interrogativo recodificados . . .	112
4.14	Correlação entre <i>interrogativas qu-in-situ</i> e Número de palavras na oração	114
4.15	Correlação entre Número de palavras na oração e Posição do constituinte interrogativo	114
4.16	Grupos de fatores correlacionados ao emprego de <i>interrogativas qu-in-situ</i>	115

5.1	Distribuição de Interrogativas-Q no <i>corpus</i> de entrevistas sociolinguísticas no Envelope de Variação II e no Quadro Geral	118
5.2	Distribuição de ocorrências de <i>interrogativas qu-que</i> em relação a Quem fala, Escolaridade, Sexo/gênero e Espontaneidade da pergunta	119
5.3	Correlação entre <i>interrogativas qu-que</i> e grupos de fatores extralinguísticos	120
5.4	Correlação entre <i>interrogativas qu-que</i> e Faixa etária recodificado	121
5.5	Correlação entre <i>interrogativas qu-que</i> e Faixa etária em duas análises contrastivas	122
5.6	Correlação entre <i>interrogativas qu-que</i> e os cruzamentos de Quem fala-Faixa etária e Quem fala-Sexo/gênero	124
5.7	Distribuição de ocorrências de <i>interrogativas qu-que</i> em relação a Tipo de verbo	125
5.8	Correlação entre <i>interrogativas qu-que</i> e Tipo de oração (2) . .	126
5.9	Correlação entre <i>interrogativas qu-que</i> e Função sintática . . .	127
5.10	Correlação entre <i>interrogativas qu-que</i> e grupos de fatores linguísticos	128
5.11	Correlação entre <i>interrogativas qu-que</i> e o cruzamento de Constituinte Interrogativo e Tonicidade da Sílabas Seguinte . .	130
5.12	Grupos de fatores correlacionados ao emprego de <i>interrogativas qu-que</i>	131
5.13	Resumo dos grupos de fatores analisados	137

Lista de Figuras

1.1	Cálculo de proporção de variantes linguísticas em relação à variável	15
3.1	Definição do envelope de variação a partir do conceito de competência comunicativa	57
4.1	Pesos relativos para <i>interrogativas qu-in-situ</i> por Faixa etária e Quem fala	96
4.2	Pesos relativos para <i>interrogativas qu-in-situ</i> em relação ao Sexo/gênero e Sinceridade pragmática da pergunta	104
1	Planilha de codificação com dados extraídos através do programa R	153

Introdução

*I keep six honest serving-men:
(They taught me all I knew)
Their names are What and Where and When
And How and Why and Who.*

R. Kipling. *The Elephant's Child*.

Este trabalho é uma investigação a respeito do emprego de diferentes formas interrogativas no português paulistano atual, a partir da perspectiva teórica e metodológica da Sociolinguística Variacionista (Labov, 2006 [1966], 1972). Mais especificamente, investiga-se o emprego de interrogativas de constituinte – também chamadas de Interrogativas-Q (Miotto & Kato, 2005) – ou seja, sentenças que contêm um pronome, advérbio ou adjetivo interrogativo: (*o*) *que*, *que* + NP, *qual(-is)*, *qual(-is)* + NP, *quanto(-a, -os, -as)*, *quanto(-a, -os, -as)* + NP, *quem*, *como*, *quando*, *onde* e *por que*.

No português brasileiro (PB), existem pelo menos quatro possibilidades de estruturação sintática dessas interrogativas: as *interrogativas-qu*, as *interrogativas qu-que*, as *interrogativas é-que* e as *interrogativas qu-in-situ*. As sentenças em (1) apresentam um exemplo dessas estruturas com o advérbio “onde”.

	Tipo de Interrogativa-Q	Exemplo	Estrutura
(1)	a. <i>Interrogativa-qu</i> :	Onde você mora?	Q +
	b. <i>Interrogativa qu-que</i> :	Onde que você mora?	Q que +
	c. <i>Interrogativa é-que</i> :	Onde é que você mora?	Q é que +
	d. <i>Interrogativa qu-in-situ</i> :	Você mora onde?	+ Q

Quando o constituinte interrogativo encontra-se no início da oração, em posição pré-verbal, há três possibilidades de realização, ilustradas em (1a)-(1c): *Q*, *Q que* e *Q é que*. Além dessas estruturas, o PB também permite que o constituinte seja realizado *in situ*, na posição de sua função sintática, como em (1d).¹

Tais estruturas são empregadas de forma alternativa em determinados contextos, com equivalência semântica e funcional. As ocorrências (2) e (3) ilustram o emprego das quatro formas na língua em uso.² No exemplo (2), o documentador Gabriel emprega uma *interrogativa qu-que* (2-a) e uma *interrogativa é-que* (2-e) para perguntar sobre sotaques, e uma *interrogativa-qu* (2-b), uma *interrogativa é-que* (2-c) e uma *interrogativa qu-que* (2-d) para perguntar sobre o modo como cariocas e interioranos falam:

- (2) a. Gabriel: **como que** é o sotaque de paulista? (M1S-DOC)
 Bruno: ah... eu acho que é mais... como se diz... é o mais aberto pr'ocê entender que nem tem outras pronúncias aí que o pessoal fala de- do jeito deles e cê não entende entendeu?
 Gabriel: qual o sotaque que- cê falou que o que você mais gosta é o de paulista... e o que você menos gosta?
 Bruno: é o carioca
 Gabriel: por quê?
 Bruno: ah eu acho um estilo meio chato né?
 b. Gabriel: **como** carioca fala?

¹ É possível ainda citar duas outras estruturas de interrogativas de constituinte atestadas no PB: *Q que é que* – p.ex., “O que que é que você está querendo?” (Miotto, 1997); e *É Q que* – p.ex., “É o que que ele quer?” (Lessa de Oliveira, 2005). No entanto, pelo número bastante reduzido de ocorrência dessas estruturas no *corpus* analisado neste estudo (três de *Q que é que* e apenas uma de *É Q que*), elas não serão tratadas aqui.

² Todas as citações do *corpus* de entrevistas sociolinguísticas são acompanhadas da identificação do falante (por pseudônimo), de seu perfil sociolinguístico (M – homem, F – mulher; 1 – de 20 a 30 anos, 2 – de 35 a 45 anos, 3 – mais de 50 anos; G – até Ensino Fundamental II, S – Ensino Superior) e de seu papel na entrevista (DOC-documentador; INF-informante).

- Bruno: é é a pronúncia deles né?
 Gabriel: a pronúncia deles é irritante?
 Bruno: é a língua puxada né?
 c. Gabriel: mas **como é que** o carioca fala?
 Bruno: ah... não sei te falar não
 Gabriel: cê gosta de imitar o pessoal falando ou não?
 Bruno: não não gosto
 d. Gabriel: não... e **como que** o pessoal do interior fala?
 Bruno: meio caipira né?
 e. Gabriel: **como é que** é o sotaque de caipira?
 Bruno: “uai” assim né?

Em (3), a documentadora Helena emprega uma *interrogativa qu-in-situ* (3-a) e uma *interrogativa-qu* (3-b) para perguntar a idade da irmã e dos pais da informante:

- (3) a. Helena: e::... sua irmã tem **quantos anos**? (F1S-DOC)
 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?
 Ingrid: meu pai tem sessenta... e a minha mãe tem cinquenta e seis...

A existência de múltiplas formas, possivelmente para a mesma função, chama a atenção de um olhar variacionista sobre o uso linguístico, que entende a variação como um fenômeno intrínseco à gramática dos falantes: parte-se da premissa de que a alternância entre diferentes formas, verificável em todas as línguas e em diversos níveis de análise, faz parte da *competência comunicativa* dos falantes (Hymes, 1991 [1979]), e que a língua falada por uma comunidade, longe de ser assistemática, exhibe *heterogeneidade ordenada* (Weinreich *et al.*, 1968).

Nesse sentido, o objetivo geral desta pesquisa é descrever e buscar explicações para os padrões de uso das formas em alternância. De maneira mais específica, este estudo procura encaminhar respostas para as seguintes questões:

- (i) Que critérios podem ser aplicados para determinar a factual opcionalidade de emprego entre diferentes estruturas morfossintáticas?;
 (ii) Nos contextos em que as estruturas de Interrogativas-Q são, de fato,

variantes de uma variável, quais fatores linguísticos e não-linguísticos influenciam a variação entre essas estruturas?;

- (iii) Que explicações podem ser fornecidas para a alternância?;
- (iv) Como a variação morfossintática se encaixa no quadro mais amplo da Teoria da Variação?

Duas das quatro estruturas apresentadas em (1) despertam um interesse especial: as *interrogativas qu-que* (1b) e as *interrogativas qu-in-situ* (1d). Nas gramáticas tradicionais do PB, tais formas raramente são mencionadas.³ Bechara (2005:597) é o único a mencionar a possibilidade de estruturação sintática das *interrogativas qu-que* ao citar como exemplo a sentença “Quanto que é a conta?”; o autor classifica a palavra *que* após o constituinte interrogativo como uma “expressão expletiva ou de realce”, que “não exerce função gramatical”. Almeida (1962:407) menciona a estrutura das *interrogativas qu-in-situ* ao citar a possibilidade de “transposição” da palavra interrogativa ao verbo em seu capítulo sobre colocação dos termos da oração, fornecendo o exemplo da sentença “Receias o quê?”. Além dessas breves menções, pouco se discorre sobre suas “regras” de emprego ou sobre seus efeitos discursivos.

Em contraste com as lacunas das gramáticas normativas, diversas pesquisas linguísticas têm sido desenvolvidas desde meados da década de 1980 a respeito da estrutura das Interrogativas-Q. No PB, tais estudos se concentram no arcabouço da Teoria da Regência e Ligação (Cf. p.ex. Kato 1987, Miotto 1989), Teoria de Princípios e Parâmetros (Cf. p.ex. Lopes-Rossi 1993; Miotto & Figueiredo-Silva 1995; Miotto 1997; Ambar *et al.* 2001) e do Programa Minimalista (Cf. p.ex. Lopes-Rossi 1996, Pires & Taylor 2007), e se preocupam em descrever as propriedades formais dos diferentes tipos de Interrogativas-Q em termos de sua representação sintática de acordo com as propostas do Programa Gerativista.

Segundo esses estudos, as *interrogativas qu-que*, como no exemplo (4) abaixo, seriam consideradas agramaticais no português europeu (PE) (Lopes-Rossi, 1996; Ambar *et al.*, 2001; Miotto & Kato, 2005). A explicação seria a

³ Foram consultadas as gramáticas de Almeida (1962), Bechara (2005), Cunha & Cintra (2007) e Luft (s/d).

perda da ordem VS em construções interrogativas no PB que, diferentemente do PE, permite sentenças como (5), com a ordem SV; como no PB o verbo não mais se move para V2 no núcleo do CP, essa posição estaria disponível para ser preenchida por um complementizador *que*.

- (4) Onde que você estava em 82, Artur Jorge? *PE / PB
(5) Como o Adriano Pinto reagiu? *PE / PB

(Exemplos extraídos de [Miotto & Kato 2005](#):172.)

Quanto às *interrogativas qu-in-situ*, embora não sejam agramaticais no PE, diversos trabalhos ([Lopes-Rossi, 1993, 1996](#); [Ambar et al., 2001](#); [Miotto & Kato, 2005](#)) notam sua baixa produtividade no PE comparativamente ao PB. A literatura explica esse fato por uma restrição pragmática ao emprego dessa forma interrogativa na variedade europeia: nela, as *interrogativas qu-in-situ* estariam restritas ao contexto de perguntas-eco, com entonação ascendente, em que se busca a repetição de algo já mencionado (6), enquanto o PB também permite o emprego de *interrogativas qu-in-situ* com entonação descendente, interpretadas como interrogativas comuns (7).⁴

- (6) O João viu QUEM? (entonação ascendente) PE / PB
(7) O João viu quem? (entonação descendente) *PE / PB

(Exemplos extraídos de [Kato 2004](#):1.)

Dessa forma, os esforços têm se concentrado na comparação entre o PB e o PE, enfatizando diferenças entre as duas variedades. Não se tem notícia, até o momento, de estudos sobre interrogativas de constituinte do PB que tenham se desenvolvido a partir da perspectiva sociolinguística. No francês europeu, pelo menos dois estudos variacionistas reportam a existência da

⁴ Certas análises dentro do arcabouço da Teoria Gerativa (p.ex. [Cheng 1991](#); [Kato 2004](#); [Cheng & Rooryck 2000](#)) argumentam que, em perguntas regulares como em (7) (i.e., não-eco), o constituinte interrogativo encontra-se apenas aparentemente *in situ*, tendo se movido para uma projeção em CP, assim como os demais constituintes. No entanto, a análise aqui desenvolvida, que parte de uma premissa diferente daquela da Teoria Gerativa, permanece cética quanto ao movimento ou não de constituintes interrogativos superficialmente *in situ*. Interessa examinar a variação entre estruturas que, para o falante na língua em uso, apresentam-se como “formas alternativas de dizer a mesma coisa” ([Labov, 1972, 1978](#)).

estrutura *qu-que* (Coveney, 2002 [1996]; Hall, 2008), o que dá indícios de que a variação em interrogativas não é exclusividade nem idiossincrasia do PB. A ideia central que percorre todo o presente estudo é a premissa de que uma teoria que abarque os fatos variáveis da língua leva a descrições mais adequadas da competência linguística dos falantes (Weinreich *et al.*, 1968).

Ao mesmo tempo, é importante notar que existem diferenças não apenas entre variedades transcontinentais da língua, mas também entre dialetos do PB. Lessa de Oliveira (2005: 412), por exemplo, reporta que as *interrogativas qu-in-situ* são as formas mais frequentes no dialeto da cidade de Vitória da Conquista (sudoeste do estado da Bahia), chegando a 81,7% na fala de uma informante.⁵ No português paulistano, a taxa de emprego dessa forma interrogativa em geral fica em torno de 10% em um quadro de distribuição geral (ver Tabela 1 adiante). Aqui, chama-se atenção para o fato de que generalizações como “português brasileiro” e “português europeu” muitas vezes escondem particularidades da língua em uso em diferentes comunidades de fala, cada qual com suas variações internas.

De fato, diferentes estudos reportam distribuições bastante díspares das formas interrogativas de constituinte, mesmo quando se trata, em princípio, de uma mesma comunidade de fala. Na Tabela 1, comparam-se os resultados de frequência das Interrogativas-Q em *corpora* de língua falada e escrita do português paulistano: uma amostra da década de 1970 do *corpus* compartilhado do Projeto de Estudo da Norma Linguística Urbana Culta de São Paulo (NURC-SP) (Oushiro, 2009); uma amostra sincrônica (2003-2006) de língua oral (*Corpus* piloto); e uma amostra sincrônica de língua escrita (Kato & Miotto, 2005).⁶ Ainda que se possam fazer ressalvas devido ao número diferente de ocorrências em cada uma das amostras, tais dados são comparáveis: tanto o *corpus* do NURC quanto o piloto foram compostos de 12 gravações com informantes cultos, estratificados de acordo com o seu sexo/gênero, três

⁵ Lessa de Oliveira (2005) apresenta a distribuição de diferentes formas de interrogativas de constituinte na fala de duas informantes, além das taxas de emprego de suas respectivas filhas, a fim de dar conta do fato de que as *interrogativas qu-in-situ* foram a primeira forma a ser adquirida pelas duas crianças, diferentemente do que ocorre no português paulistano (Cf. Grolla 2005).

⁶ Kato & Miotto (2005) realizaram o levantamento de dados em dois jornais, um do PB (jornal *Folha de São Paulo*), e um do PE (jornal *O Público*), com o intuito de traçar comparações interlinguísticas. No PE, não houve ocorrências de *interrogativas qu-que*; houve apenas 4 (1,2%) de *interrogativas qu-in-situ*, 72 de *interrogativas é-que* (22,0%) e 252 de *interrogativas-qu* (76,8%) (Kato & Miotto, 2005: 314).

faixas etárias e dois tipos de elocução (aula ou entrevista sociolinguística); a amostra de língua escrita refere-se a dados do jornal *Folha de São Paulo* e, portanto, também pode-se assumir que são textos escritos por pessoas mais escolarizadas.

Tabela 1 Comparação da frequência de Interrogativas-Q em diferentes *corpora* do português paulistano

	NURC-SP		<i>Corpus</i> piloto		Jornal	
	década de 1970		década de 2000		década de 2000	
	N	%	N	%	N	%
<i>Interrogativas-qu</i>	144	56,0	191	40,5	238	65,7
<i>Interrogativas qu-que</i>	41	16,0	205	43,4	4	1,1
<i>Interrogativas é-que</i>	56	21,8	38	8,1	88	24,3
<i>Interrogativa qu-in-situ</i>	16	6,2	38	8,1	32	8,8
Total	257	100,0	472	100,0	362	100,0

No estudo piloto, verificou-se que as *interrogativas qu-que* foram as mais frequentes (43,4%), juntamente com as *interrogativas-qu* (40,5%). Em comparação com os dados da década de 1970 (NURC-SP), as *interrogativas qu-que* tiveram um aumento considerável em sua taxa de emprego, dado que sua frequência mais do que dobrou em um período de 30 anos (de 16,0% a 43,4%). Por outro lado, apesar de as *interrogativas qu-que* serem atualmente bastante frequentes na língua oral, elas são praticamente ausentes do registro escrito (1,1%), como se vê nos dados de jornal (Kato & Miotto, 2005). Essas observações permitem supor que o emprego de *interrogativas qu-que* esteja correlacionado a fatores prosódicos e ao meio discursivo, além de levantar uma hipótese de mudança em progresso. Com base nesses dados, no entanto, não é evidente qual outra forma de interrogativa-Q estaria sendo desfavorecida em favor das *interrogativas qu-que*: tanto as *interrogativas-qu* quanto as *interrogativas é-que* parecem boas candidatas a tal.

Por outro lado, a distribuição de *interrogativas qu-in-situ* na Tabela 1 revela que a estrutura é relativamente produtiva no português paulistano atual (cerca de 8% em um quadro de distribuição geral). Como se verá no

capítulo 3, o emprego dessa estrutura é na verdade muito mais produtivo no português paulistano (cerca de 22% na língua oral) quando se consideram os contextos em que de fato é opcional. Ainda que sua taxa global pareça ter permanecido a mesma desde a década de 1970 (6,2% *vs.* 8,1%) – o que pode ser um indício de variação estável ou, colocado de outro modo, de que não há mudança em progresso no emprego de *interrogativas qu-in-situ* – interessa investigar os fatores que influenciam o seu emprego na comunidade de fala paulistana.

Uma distribuição geral de frequências não permite uma análise detalhada da variação e da língua em uso; trata-se apenas de um primeiro passo no estudo da heterogeneidade ordenada. Como se verá no capítulo 1, a definição do envelope de variação de variáveis (morfos)sintáticas não é sem controvérsia. Desde o debate entre Lavandera (1978) e Labov (1978), o estudo de variáveis além do nível fonológico levanta questões a respeito da equivalência semântico-pragmática das possíveis variantes e, portanto, sobre a factual opcionalidade do falante ao empregar diferentes formas. Desse modo, este estudo se enquadra dentro de um debate mais amplo acerca dos critérios para o estudo de variáveis além do nível fonológico. Em especial, discutem-se problemáticas que, de uma maneira geral, têm deixado o estudo da variação morfossintática em situação menos privilegiada do que o da variação fonológica na formulação de amplas generalizações (Cf. p.ex. Labov 1994, 2001; Labov *et al.* 2006).

O capítulo 2 relaciona as questões teórico-metodológicas levantadas no capítulo anterior com o exame das Interrogativas-Q. Além de apresentar o material de análise, que compreende amostras de língua oral e escrita (53 entrevistas sociolinguísticas com informantes paulistanos, 1470 redações escolares, amostras do jornal Folha de São Paulo *online* e 5 edições da revista Veja), o capítulo expõe as decisões metodológicas tomadas neste estudo: a análise de dados tanto de informantes quanto de documentadores e a definição do objeto de estudo como formas *interrogativas*, que diferem conceitualmente de *perguntas*.

O capítulo 3 define o conceito de *pressuposição do falante* (Stalnaker, 2002) como critério de equivalência semântica entre as interrogativas de constituinte

e o conceito de *competência comunicativa* (Hymes, 1991 [1979]) como critério de verificação dos contextos em que as formas se alternam. Nesse capítulo se desenvolve uma análise qualitativa minuciosa dos contextos de uso das Interrogativas-Q a partir da qual são definidos dois envelopes de variação, um que envolve a alternância na posição do constituinte interrogativo (Q *in situ* vs. Q movido) e outro que envolve as três estruturas com constituinte interrogativo pré-verbal (Q que vs. Q + Q é-que). O capítulo se encerra com a apresentação de hipóteses, operacionalizadas na forma de grupos de fatores, que têm o objetivo de buscar explicações para a variação observada.

Os capítulos 4 e 5 trazem análises quantitativas, realizadas com o programa GoldVarb X, que enfocam respectivamente as *interrogativas qu-in-situ* e as *interrogativas qu-que*; no capítulo 4, dá-se especial atenção a questões discursivo-pragmáticas e, no capítulo 5, a questões de natureza sintática e prosódica.

A Conclusão faz um resumo dos resultados e das contribuições deste estudo dentro do quadro mais amplo da Teoria da Variação.

1

Pressupostos teóricos e metodológicos

Este capítulo apresenta os preceitos teóricos e metodológicos gerais que norteiam esta pesquisa e que formam a base das discussões específicas, desenvolvidas no próximo capítulo, sobre o estudo de Interrogativas-Q. Após descrever a visão de língua adotada na Sociolinguística e os fundamentos teóricos do paradigma quantitativo, discutem-se quatro problemáticas centrais do estudo da variação morfossintática: a questão da equivalência semântica, a definição do envelope de variação, a relativa escassez de dados e a falta de correlação com grupos de fatores sociais.

1.1 A análise da língua em uso e da variação

A variação linguística, ainda que constatada desde tempos antigos, adquiriu estatuto de uma disciplina com metodologia e orientação próprias somente a partir da década de 1960. Segundo [Chambers \(2002\)](#), pode-se atribuir esse fato, por um lado, à herança dos estudos clássicos, cujas investigações linguísticas se concentraram fundamentalmente na categorização das formas (em detrimento do uso); e, por outro, ao advento de tecnologias que possibilitaram a coleta e manipulação de grandes quantidades de dados linguísticos, viabilizando um estudo sistemático da variação – tarefa até então hercúlea, senão impossível.

A maior parte dos estudos linguísticos a partir do século XIX, de fato, ocupou-se da busca pela invariância (Guy, Ms.). Tomem-se como exemplos as regras de mudanças fonéticas sem exceção, postuladas pelos neogramáticos; a definição de fonemas e morfemas como unidades abstratas invariantes e formadoras da *langue* estruturalista; a noção gerativo-transformacional de que diferentes realizações “superficiais” correspondem a uma única “estrutura profunda”; e, mais recentemente, a busca por universais linguísticos.

Tal comprometimento teórico parece se basear em uma premissa que equaciona a variação com a aleatoriedade e a sistematicidade com o categórico (Guy, Ms.). Tanto o estruturalismo desenvolvido com base no Curso de Linguística Geral de Saussure quanto o Programa Gerativista de Chomsky operam sobre uma distinção conceitual entre sistema/conhecimento, por um lado, e os usos linguísticos, por outro. Nessas delimitações, a língua em uso está à parte do programa de estudo devido à perspectiva de que o sistema linguístico não coincide com aquilo que é falado pelas pessoas em seu cotidiano; estuda-se, portanto, o conhecido “falante-ouvinte ideal” que vive em uma “comunidade de fala homogênea” (Chomsky 1965: 3-4, apud Weinreich *et al.* 1968).

Hymes (1991 [1979]), em um artigo já clássico, apresenta uma crítica à divisão chomskyana entre *competência* e *performance*. Segundo o autor, o conceito de Chomsky sobre a competência dos falantes, que se restringe ao conhecimento das regras do sistema linguístico, deve ser estendido para abarcar o conhecimento que os falantes têm a respeito dos modos de uso das formas linguísticas. Tal conhecimento é denominado de *competência comunicativa*.

Segundo esse conceito, os falantes de uma língua possuem conhecimento não apenas de suas regras gramaticais, mas também dos modos apropriados de aplicá-las, de forma semelhante ao conhecimento de outras regras culturais que os falantes empregam em outros aspectos da vida social. Em situações conversacionais, os falantes sabem, de acordo com a competência adquirida no convívio social, quando e o que falar, e para quem, onde, de que modo. Curiosamente, Hymes (1991 [1979]) fornece exemplos a respeito de comportamentos

apropriados, em diferentes grupos, do emprego de perguntas:¹

(...) entre os araucanos do Chile, repetir uma pergunta é insultar; entre os tseltalanos de Chiapas, México, não se deve fazer uma pergunta direta (e esta não deve ser respondida); entre os caxinauás do Brasil, uma resposta direta para uma primeira pergunta implica que aquele que responde não tem tempo para conversar, enquanto uma resposta vaga implica que a pergunta será respondida diretamente da segunda vez, e que a conversa pode continuar. (Hymes 1991 [1979]: 16. Tradução própria.²)

Desse modo, o conhecimento do falante compreende não apenas julgamentos de gramaticalidade e aceitabilidade de sentenças, mas quatro tipos de distinções:

- (1) se (e até que ponto) algo é formalmente *possível*;
- (2) se (e até que ponto) algo é *factível* em virtude dos meios disponíveis de implementação;
- (3) se (e até que ponto) algo é *apropriado* (adequado, feliz, bem-sucedido) em relação ao contexto em que é usado e avaliado;
- (4) se (e até que ponto) algo de fato é feito, de fato é *empregado*, e o que a sua realização acarreta. (Hymes 1991 [1979]: 19. Grifos do autor. Tradução própria.³)

Uma sentença pode ser gramatical, mas estranha, cuidadosa e pouco frequente. Nesta pesquisa, o conceito de competência comunicativa é empregado na definição do envelope de variação (ver seção 3.2); uma forma é considerada variante se ela é possível, factível, apropriada e empregada.

Por ora, importa destacar o objeto de estudo da Sociolinguística, redefinido a partir de outra visão de língua: estuda-se a *competência comunicativa* dos falantes, que deve ser examinada na língua em uso.

¹ Adiante, na seção 2.3, será feita uma distinção entre os termos “pergunta” e “interrogativa”.

² Texto original: “(...) among the Araucanians of Chile, to repeat a question is to insult; among the Tzeltal of Chiapas, Mexico, a direct question is not properly asked (and to be answered “nothing”); among the Cahinahua of Brazil, a direct answer to a first question implies that the answerer has no time to talk, a vague answer that the question will be answered directly the second time, and that talk can continue.”

³ Texto original: “(1) whether (and to what degree) something is formally *possible*;
(2) whether (and to what degree) something is *feasible* in virtue of the means of implementation available;
(3) whether (and to what degree) something is *appropriate* (adequate, happy, successful) in relation to a context in which it is used and evaluated;
(3) whether (and to what degree) something is in fact done, actually *performed*, and what its doing entails.”

Essa perspectiva teórica não só permite como também exige que se adotem os aspectos até então “marginais” nos estudos linguísticos – a saber, *parole* e *performance*. Na concepção da Sociolinguística Variacionista, reconhece-se a variabilidade (não apenas entre gramáticas de diferentes línguas, mas na gramática de uma mesma comunidade e, em última instância, na fala de um mesmo indivíduo) como parte do sistema linguístico. A “língua-uso” é a um tempo heterogênea e sistemática: é a ausência de heterogeneidade ordenada que seria disfuncional em uma língua usada por uma comunidade complexa e real, naturalmente heterogênea (Weinreich *et al.*, 1968).

O estudo da variação parte da premissa de que, em uma comunidade de fala, a gramática é inerentemente variável e que formas alternativas de dizer “a mesma coisa” estão disponíveis para os membros dessa comunidade (Weinreich *et al.*, 1968; Labov, 1972). Assim, Weinreich *et al.* (1968) propõem o conceito de *variável sociolinguística*, cujas variantes são empregadas de acordo com a influência de fatores linguísticos e sociais.

O olhar simultâneo tanto para a estrutura linguística quanto para a estrutura social requer o exame de uma grande quantidade de dados (Guy, 1993): raramente se pode falar em termos categóricos acerca de características dialetais, pois os dados observados nunca se organizam perfeitamente em regras isentas de variação interna; ao contrário, deve-se falar em variabilidade, tendências e relações quantitativas. A variável sociolinguística deve obedecer a condições claramente delineadas a fim de fazer parte de uma estrutura linguística pois, caso contrário, permitir-se-iam regras em que “frequentemente”, “ocasionalmente” ou “às vezes” se aplicam (Weinreich *et al.*, 1968) – a chamada “variação livre”. Para evitar tais regras impressionísticas, faz-se necessária a apresentação de evidência quantitativa de covariação entre a variável em questão e outros elementos intra e extralinguísticos.

1.2 O paradigma quantitativo

O paradigma quantitativo se baseia no conceito de competência comunicativa: os membros de uma comunidade têm internalizadas não apenas as regras da gramática mas também as regras de seus usos apropriados, de forma

semelhante ao domínio de outras regras culturais (Sankoff, 1980). A partir da premissa de que as flutuações observadas na língua em uso são ordenadas e formam padrões, o paradigma quantitativo busca modelar a competência comunicativa do falante através de um conjunto de princípios teóricos.

O mais central deles é o Princípio da Modelagem Quantitativa (Bayley, 2002), que afirma (a) a necessidade de observar formas linguísticas variáveis em seu contexto de uso, seja o contexto linguístico ou não-linguístico; e (b) contanto que haja dados suficientes, a possibilidade de derivar afirmações acerca da probabilidade de coocorrência de uma forma linguística variável e as características contextuais. Tais afirmações expressam, em termos quantitativos, a força da associação entre a característica contextual e a variável linguística (Bayley, 2002:118).

O Princípio de Múltiplas Causas (Bayley, 2002) afirma a improbabilidade de que um único fator contextual possa dar conta da variabilidade que se observa nos dados. Uma vez que cada ocorrência de uma determinada variável se dá necessariamente dentro de um contexto enunciativo, social e linguístico, há uma gama de fatores que pode influenciar o emprego de formas linguísticas em alternância; desse modo, a análise multivariada é uma tentativa de modelar as ocorrências de variáveis como “uma função de várias forças simultâneas, interseccionadas e independentes, que podem agir em diferentes direções” (Guy & Zilles, 2007:50). Nesse sentido, torna-se tanto mais necessária uma medida numérica não só da força de associação entre características contextuais e a variável linguística, mas também de sua direção (favorecimento ou desfavorecimento).

Para que se possa modelar a força relativa de diversos fatores no emprego de variáveis linguísticas, é necessário examinar tanto as ocorrências em que determinada forma linguística foi empregada quanto as instâncias de não-ocorrência. O Princípio de Contabilidade, tal como definido por Labov (1969), afirma que:

(...) qualquer forma variante (um membro de um conjunto de modos alternativos de ‘dizer a mesma coisa’) deve ser reportada com a proporção de casos nos quais a forma de fato ocorreu no contexto relevante, comparada com o número total de casos em que ela poderia

ter ocorrido. (Labov 1969:738. Tradução própria.⁴)

Esse princípio segue da observação de que a língua é normalmente percebida como categórica, tanto pelas pessoas em geral (observem-se comentários comuns na forma de afirmações como “paulistanos falam po[r]ta, pessoas do interior falam po[ɹ]ta”) quanto por linguistas em busca da invariância. Sua motivação é a perspectiva de que a análise linguística deve objetivar a descrição de padrões na comunidade de fala em vez de idiossincrasias, e que essas podem ser assim categorizadas de modo mais apropriado através de uma análise que não se baseie em instâncias isoladas.

Um modo alternativo de definir o Princípio de Contabilidade é apresentado, de forma bastante didática, através daquilo que Buchstaller (2009) chama de “construção e quantificação do denominador” (ver Figura 1.1).⁵

Figura 1.1 Cálculo de proporção de variantes linguísticas em relação à variável

$$\frac{(\sum \text{ocorrências da variante A})}{(\sum \text{ocorrências da variante A} + \sum \text{ocorrências das demais variantes})}$$

(Figura adaptada de Buchstaller 2009.)

As ocorrências de uma variante linguística em estudo devem ser apresentadas em proporção ao total de ocorrências de todas as variantes possíveis no contexto variável; em outras palavras, trata-se de uma razão matemática em que o numerador é o número total de ocorrências da variante em análise e o denominador é o total de ocorrências da variável, incluindo as ocorrências da variante sob estudo e de todas as demais variantes.

Entretanto, a decisão sobre as formas que devem ser somadas no denominador nem sempre é tão direta, sobretudo em casos de variação morfossintática,

⁴ Texto original: “(...) any variable form (a member of a set of alternative ways of ‘saying the same thing’) should be reported with the proportion of cases in which the form did occur in the relevant environment, compared to the total number of cases in which it might have occurred.”

⁵ Buchstaller (2009) emprega o termo “denominador” de forma mais ampla do que a definição da variável sociolinguística com suas possíveis variantes, visto que também discute outros métodos de normalização de frequência (ocorrências por mil palavras, por duração de gravação, por linhas de transcrição etc.) que podem servir como denominadores estáveis. Esses outros métodos são úteis em casos de usos linguísticos para os quais a tarefa de definição da variável sociolinguística pode não ser viável (p.ex., o emprego de *like* como um marcador discursivo. Cheshire *et al.* 2005 apud Buchstaller 2009).

e certamente tem consequências para os resultados quantitativos que serão obtidos. Uma discussão mais detalhada dessa questão é desenvolvida adiante na seção 1.3.

Dois outros princípios baseados no conceito de heterogeneidade ordenada concernem à relação entre a variação idioletal (do indivíduo) e dialetal (de grupos sociais). Guy (1991 apud Bayley 2002) postula que (a) os falantes individuais podem diferir em sua taxa de emprego de uma regra variável, i.e. em sua probabilidade de *input* para a regra; e que (b) os indivíduos de uma mesma comunidade de fala devem ser similares ou idênticos nos valores atribuídos a restrições linguísticas sobre a regra. Assim, as taxas de emprego de uma variante podem ser tão díspares quanto 10% e 70% para falantes distintos de uma mesma comunidade, mas suas tendências de emprego em relação a diferentes contextos linguísticos devem ser similares. Tal, de fato, é o critério de Labov (2006 [1966]) para a definição de uma comunidade de fala, que se define pelo compartilhamento de normas nos usos linguísticos.⁶

Todos esses princípios – de Modelagem Quantitativa, de Múltiplas Causas, de Contabilidade e aqueles que dizem respeito à relação entre variação idioletal e dialetal – também esclarecem a importância de se desenvolverem análises quantitativas de *tendências*, e não apenas de frequências, visto que falantes individuais podem apresentar taxas de uso bastante díspares entre si. Uma análise de tendências pode identificar padrões gerais relevantes para um grupo maior de indivíduos além daqueles cuja fala foi analisada.

Dentro de uma visão de língua que entende a heterogeneidade ordenada como parte do sistema linguístico, os princípios do paradigma quantitativo devem ser aplicados a variáveis de todos os níveis de análise, dado que a variação linguística não se restringe apenas à fonologia, à morfologia, à sintaxe ou à pragmática. Contudo, estudos sobre variáveis além do nível fonológico

⁶ É possível, no entanto, que esses dois últimos princípios se apliquem mais propriamente à variação fonológica. Guy (1980, apud Bayley 2002) fundamenta esses princípios em uma análise comparativa do apagamento de /-t, -d/ em amostras de falantes de Nova Iorque e da Filadélfia, mas seus resultados são consistentes contanto que haja um número suficiente de ocorrências (cerca de 20 dados por célula). Em um estudo de variação sintática, essa quantidade de dados por falante nem sempre é obtida. Ademais, Labov (2006 [1966]:5) cita o estudo de Zwicky (2002) como uma “forte demonstração da existência de gramáticas individuais para fenômenos sintáticos menos frequentes”. No presente estudo, diante do número reduzido de dados por falante, não é possível verificar a relação entre variação idioletal e dialetal postulada por Guy (1991).

na tradição sociolinguística geralmente têm se cercado de um debate teórico e metodológico, cujas principais questões são explicitadas a seguir.

1.3 O estudo sociolinguístico de variáveis morfossintáticas

Os estudos quantitativos inaugurais da variação sociolinguística trataram de variáveis fonéticas/fonológicas: a alternância na altura da primeira vogal dos ditongos /ay/ (como em *pride* ‘orgulho’) e /aw/ (como em *house* ‘casa’) na comunidade da ilha de Martha’s Vineyard (Labov, 1963, 1972); a pronúncia do /r/ pós-vocálico (como em *card* ‘cartão’) no inglês nova-iorquino (Labov, 2006 [1966], 1972). Labov (2006 [1966]) justifica a escolha dessas variáveis pela alta frequência de ocorrência de suas variantes, por terem certa imunidade à supressão consciente, pelo fato de serem unidades integrantes de estruturas maiores e pela facilidade de quantificá-las em uma escala linear.

Após esses primeiros estudos, não se passou muito tempo até que o conceito de variável sociolinguística e o instrumental quantitativo de análise fossem empregados em estudos “para além” do nível fonológico. Alguns exemplos são Labov 1969, sobre a regra variável de apagamento da cópula no inglês de negros norte-americanos; Weiner & Labov 1983 [1977], sobre a alternância entre voz ativa e voz passiva no inglês; Sankoff & Thibault 1977, sobre o emprego variável de *être* ‘ser, estar’ e *avoir* ‘ter’ em alguns contextos em que ocorrem como verbos auxiliares no francês de Montreal; Labov & Labov 1977, sobre a aquisição da regra de inversão em Interrogativas-Q por uma criança (Jessie) falante de inglês.

Essa extensão conceitual, no entanto, não foi sem controvérsias. Abaixo se discutem quatro das principais problemáticas pertinentes ao estudo de variáveis além do nível fonológico, intimamente ligadas entre si: a questão da equivalência semântica entre as variantes; a definição do envelope de variação; a relativa escassez de dados; e a (frequente) falta de correlação entre variáveis morfossintáticas e variáveis sociais.

A questão da equivalência semântica

O estudo da variação pressupõe a possibilidade de empregar “formas alternativas de dizer a mesma coisa”: se não há opcionalidade, observa-se o uso categórico de determinada forma; e se duas formas não possuem equivalência semântica, é possível questionar essa opcionalidade.

De acordo com a formulação de Labov (1972, 1978), certas variantes são semanticamente equivalentes quando sua referência ou valor de verdade são idênticos – ou, como o autor prefere definir, quando se referem a um mesmo “estado de coisas”.

Em artigo de 1978, Lavandera questiona se é possível manter o requisito de equivalência semântica para unidades linguísticas além do nível fonológico; segundo a autora, unidades como morfemas, itens lexicais ou construções sintáticas possuem, por definição, um significado, ao contrário de fonemas que, igualmente por definição, não possuem uma “constância de referência”. Assim, é muito menos problemático argumentar que “po[ɹ]ta” e “po[r]ta” possuem a mesma referência do que um par de construções sintáticas como “O armário de bebidas foi arrombado” e “Arrombaram o armário de bebidas”.⁷ O critério de equivalência semântica, segundo Lavandera (1978: 175), conduz o pesquisador à “exoneração engenhosa de possíveis diferenças no significado referencial”. Desse modo, em lugar do critério do valor de verdade, a autora propõe um critério de comparabilidade funcional.

Em resposta ao artigo de Lavandera, Labov (1978) volta a defender a definição do conjunto de variantes de modo restrito e reafirma a importância de isolar o contexto em que a variação relevante ocorre, eliminando casos em que os itens desempenham diferentes funções linguísticas, em que a variação é neutralizada ou em que a regra é categórica. Ignorar essas condições

⁷ Em inglês, “The liquor closet was broken into” e “They broke into the liquor closet”, respectivamente. O exemplo fornecido por Lavandera (1978) para a voz ativa difere de certa forma do exemplo aqui apresentado para o português. O inglês, enquanto língua de sujeito obrigatório, requer a presença de um pronome, “they”, o que causa uma ambiguidade: “they” pode ter uma leitura de sujeito indeterminado, caso em que não se sabe quem arrombou o armário de bebidas, ou pode ter uma referência específica, caso em que o pronome é usado para retomar um referente já mencionado. No português, esta ambiguidade da voz ativa possivelmente pode ser desfeita através de duas construções distintas: “Arrombaram o armário de bebidas” vs. “Eles arrombaram o armário de bebidas”. No entanto, o problema da referência permanece: podemos nos perguntar qual dessas construções na voz ativa seria a equivalente da voz passiva no português.

significaria perder a precisão da análise.

Para as Interrogativas de constituinte, o critério de equivalência semântica com base na referência parece impor um desafio: como atribuir um valor de verdade a sentenças como “Onde (que/ é que) você mora?” e “Você mora onde?”? Este estudo adotará o conceito de *pressuposição do falante* (Stalnaker, 2002) (ver Capítulo 3), que estabelece a equivalência semântica entre as variantes diante de um mesmo “estado de coisas” (Labov, 1978).

A definição do envelope de variação

A definição do envelope de variação – ou a tarefa de “fechar o conjunto” de variantes (Labov, 1996) – segue diretamente do Princípio de Contabilidade (Labov, 1969) e se liga à questão da equivalência semântica. Se duas ou mais formas não podem ser consideradas opcionais (seja por critérios funcionais ou formais), o Princípio de Contabilidade se vê em risco.

Entretanto, ainda que se possa afirmar a equivalência semântica de certas formas linguísticas, é possível que o pesquisador não inclua uma variante em potencial no envelope de variação. Um exemplo dessa situação é o estudo de Labov (1969) a respeito da realização variável da cópula no inglês de negros norte-americanos.

Neste artigo, Labov identifica três formas em alternância do verbo *be* ‘ser, estar’ finito: uma forma plena (1-a), uma forma contraída (1-b) e uma forma apagada (1-c).

- (1) a. He is better than the girls is, now.
- b. Tha’s my daily routine: women.
- c. She the first one started us off.

(Exemplos extraídos de Labov 1969: *passim*.)

O critério de equivalência semântica com base no valor de verdade provavelmente não traria restrições à comparação entre “She is the first one...”, “She’s the first one...” e “She the first one...”: pode-se considerar que as três formas se referem a um mesmo estado de coisas.

O autor, no entanto, nota que o apagamento da cópula sofre as mesmas

restrições dos contextos em que também não é possível contraí-la no inglês padrão:

(2)	Inglês padrão	Inglês de afro-americanos
a.	*He's as nice as he says he's.	a'. *He's as nice as he says he.
b.	*How beautiful you're!	b'. *How beautiful you!
c.	*Here I'm.	c'. *Here I.
d.	Are you going? *I'm.	d'. Are you going? *I.
e.	*He's unfortunately.	e'. *He unfortunately.

(Exemplos extraídos de Labov 1969 : 722.)

Através dos exemplos em (2), verifica-se que a contração não é possível em orações que não têm o predicativo posposto ao verbo *be*, seja em final de oração (2-a)-(2-c), em respostas curtas (2-d) ou no meio da sentença (2-e). Como o apagamento da cópula sofre as mesmas restrições gramaticais (exemplos (2)'), o autor conclui que o envelope de variação é composto apenas pelas formas contraídas e apagadas do verbo *be* finito, e descarta as ocorrências de pronúncia plena; a regra variável formulada diz respeito ao apagamento da cópula contraída.

Outra decisão metodológica tomada por Labov (1969) foi considerar inicialmente apenas ocorrências de *is* 'é, está'; as ocorrências de *are* 'são, somos, estão, estamos' foram analisadas separadamente, pois sua taxa de apagamento é maior, e as ocorrências de *am* 'sou, estou' foram descartadas por ocorrer quase que categoricamente (99%) com a forma contraída *I'm*. Essa decisão, no entanto, levanta a seguinte questão: trata-se de duas variáveis distintas, uma que diz respeito ao apagamento de *is* e outra ao apagamento de *are*, ou ambas poderiam ser consideradas uma mesma variável, "o apagamento da cópula"?

Rickford *et al.* (1991) analisam ambas decisões teórico-metodológicas sobre quais formas devem integrar o envelope de variação, em um artigo que discute diferentes análises realizadas por diversos autores.⁸ Para a primeira dessas questões – quais os efeitos de se incluírem as formas plenas, contraídas

⁸ Dentre outros, Labov 1969, Wolfram 1974, Baugh 1979, Romaine 1982, Poplack & Sankoff 1987. Cf. referências em Rickford *et al.* (1991).

e apagadas, ou apenas um subconjunto delas –, os autores constataam grandes diferenças de resultados quantitativos a depender de quais variantes são incluídas na análise. Para a segunda dessas questões – se *is* e *are* podem ser consideradas uma única variável ou se devem ser consideradas separadamente –, os efeitos parecem não diferir significativamente.

Buchstaller (2009) oferece outro exemplo contrastivo sobre a possibilidade de diferentes decisões acerca do envelope de variação. A autora compara os estudos de Tagliamonte & Roberts (2005) e Rickford *et al.* (2007) (apud Buchstaller 2009) a respeito do emprego de intensificadores (*so* ‘tão’, *really* ‘realmente’, *totally* ‘totalmente’, *very* ‘muito’, *pretty* ‘consideravelmente’, *all* ‘completamente’ etc.).

Tagliamonte & Roberts (2005), com base em um *corpus* composto pelas oito primeiras temporadas da série de TV *Friends*, optam por um ponto de partida formal, evidenciando a dificuldade de definição do envelope de variação no caso dessa variável:

(...) é uma tarefa fácil encontrar os intensificadores em si, mas é difícil, se não impossível, encontrar onde eles poderiam ter ocorrido, mas não ocorreram. Pelo fato de os adjetivos serem a forma mais comumente intensificada na língua (...), selecionamos esse contexto para nosso estudo. (Tagliamonte & Roberts (2005, 285) apud Buchstaller 2009. Tradução própria.⁹)

O revés a que Tagliamonte & Roberts (2005) se referem concerne ao Princípio de Contabilidade. A solução dos autores é definir um denominador estável (frases adjetivais) através do qual é possível quantificar a ocorrência relativa das variantes. Em ordem de frequência, os intensificadores mais empregados são *so* (44,1%), *really* (24,6%), *very* (14,2%), *pretty* (6,1%) e *totally* (2,8%).

Por outro lado, em um estudo sobre o emprego do intensificador *all* em conversas entre jovens da Califórnia, Rickford *et al.* (2007) seguem um critério funcional para delimitar o envelope de variação. O primeiro passo de sua análise foi identificar todas as formas com as quais a variante compete e que

⁹ Texto original: “(...) it is an easy task to find the intensifiers themselves but difficult, if not impossible, to find where they could have occurred, but did not. Because adjectives are the most commonly intensified form in language (...), we selected this context for our study.”

têm a função de reforçar a propriedade denotada pela cabeça que modifica, seja ela adjetival, nominal ou verbal. Para intensificadores de adjetivos, Rickford *et al.* (2007) identificam 32 variantes, cinco das quais têm uma frequência relativa acima de 2%: *really* (52,2%), *so* (18,9%), *very* (9,4%), *all* (7,4%) e *totally* (2,7%).

Apesar das semelhanças entre ambas as listas de frequência (*so*, *really*, *very* e *totally* figuram em ambas as listas entre as cinco formas mais frequentes, com pequenas diferenças de ordenamento), os resultados desses dois estudos são comparáveis apenas até certo ponto, visto que se baseiam em diferentes envelopes de variação.

Através desses casos, ressalta-se a importância de reportar as decisões teórico-metodológicas tomadas nas análises desenvolvidas, a fim de sinalizar que comparações podem ser traçadas com outros estudos. Para as Interrogativas-Q, que apresentam quatro possibilidades de estruturação morfo-sintática, é possível delimitar mais de um envelope de variação. As decisões tomadas neste estudo serão discutidas nos capítulos 2 e 3.

A relativa escassez de dados

Comparativamente a estudos de variação fonético-fonológica, pesquisas sobre variáveis morfo-sintáticas podem muitas vezes não contar com uma abundância de dados.

Para ilustrar a diferença entre o número de dados que em geral se obtém, é útil comparar estudos feitos com base em um mesmo *corpus* – ou seja, estudos que se valeram de informantes estratificados a partir dos mesmos fatores extralinguísticos e de um material linguístico de mesma extensão em número de palavras ou horas de gravação. Na Tabela 1.1, compara-se o número de ocorrências de três variáveis na Amostra Censo 00, coletada pelos pesquisadores do Programa de Estudos sobre o Uso da Língua (PEUL) em 1999-2000 (Paiva & Duarte, 2003). Essa amostra é composta de 32 gravações sociolinguísticas com informantes cariocas estratificados de acordo com o seu sexo/gênero, quatro faixas etárias e três níveis de escolaridade.

Para a variável fonética – a pronúncia de (ey) como ditongo (*peixe*) ou

Tabela 1.1 Número de ocorrências de diferentes variáveis em 32 entrevistas da Amostra Censo 00

Variável	Número de ocorrências		
	N total	N valor de aplicação ^a	%
Pronúncia de (ey)	1445	1108	76,7
Pronome 1PP	968	768	79,3
Orações relativas	583	32	5,5

^a Pronúncia de (ey): monotongo [e]; Pronome de 1PP: *a gente*; Orações relativas: relativas com anáfora.

monotongo (*pexe*) (Paiva, 2003) –, o total de ocorrências foi de 1445; para a variável morfológica – a alternância entre os pronomes *nós* e *a gente* (Omena, 2003) –, o número de ocorrências cai em um terço, para 968; no caso da variável sintática – a variação em orações relativas, com foco na variante com pronome anafórico (resumptivo) (Mollica, 2003) –, o número de ocorrências cai ainda mais, para 583, menos da metade daqueles da variável fonética. Aqui não se objetiva generalizar essas proporções, visto que o número de ocorrências de uma determinada variável também depende fundamentalmente das decisões do pesquisador a respeito do envelope de variação.¹⁰ Entretanto, esses dados ilustram aquilo que em geral se observa nos estudos de variáveis de diferentes níveis de análise linguística.

Mais do que assinalar o simples número de ocorrências de determinada variável – já que, em princípio, sempre se pode aumentar o tamanho do *corpus* de análise e, assim, obter mais dados – é importante verificar a proporção da variante em estudo em relação ao total de ocorrências da variável. Note-se que, no estudo de Mollica (2003) sobre orações relativas, a taxa de emprego da variante sob análise – a construção com anáfora – representa apenas 5,5% de seus dados. Em alguns casos, uma proporção muito baixa ou muita alta da variante de interesse pode inviabilizar o teste de hipóteses quantitativamente,

¹⁰ Por exemplo, em um estudo sobre o emprego variável do pronome de primeira pessoa do plural, seria possível examinar apenas os dados em caso nominativo, ou apenas os casos em que houve concordância com o verbo etc.

devido à distribuição desequilibrada dos dados.¹¹

Hall (2008), por exemplo, em seu estudo sobre o francês regional da Normandia, dedica um capítulo à análise da estrutura com complementizador duplamente preenchido, similar às *interrogativas qu-que* do presente estudo. No entanto, as análises quantitativas sobre a variável referem-se a dados obtidos através de testes de percepção e julgamento, que o autor chama de “Métodos formais”; os dados colhidos das entrevistas sociolinguísticas propriamente não foram frequentes o suficiente para que se pudessem testar hipóteses (grupos de fatores) com um nível confiável de relevância estatística.

Coveney (2002 [1996]), de modo semelhante, também identificou a ocorrência da estrutura *qu-que* no francês falado, mas não obteve nenhuma em suas entrevistas:

(...) não houve ocorrências de [QkSV] produzidas pelos informantes — mas observações fora das entrevistas gravadas sugeriram que [QkSV] é de fato amplamente empregado por crianças e por alguns adultos. (Coveney 2002 [1996]: 103. Tradução própria.¹²)

Tanto Hall (2008) quanto Coveney (2002 [1996]) afirmam que a estrutura com complementizador duplamente preenchido ou [QkSV]¹³ é estigmatizada no francês e associada a populações rurais sem letramento, fato que pode explicar a sua baixa produtividade em situação de entrevista sociolinguística.

A questão da obtenção de dados e de sua distribuição também é relevante nesta pesquisa. Trabalhou-se com um *corpus* robusto, com cerca de um milhão de palavras, a fim de que o número total de ocorrências permitisse uma análise multivariada. A questão da distribuição dos dados virá à tona na comparação entre as amostras de língua oral e língua escrita; enquanto a primeira exhibe ampla variação entre estruturas, a segunda apresenta o uso

¹¹ De fato, a proporção de 5,5%-94,5% das variantes da variável “orações relativas” na Tabela 1.1 está bastante próxima do limite sugerido por Guy & Zilles (2007: 60) de 5%-95% para uma análise confiável, já que distribuições piores fatalmente resultarão em interação entre os grupos de fatores, inviabilizando uma análise multivariada.

¹² Texto original: “(...) no occurrences of [QkSV] were produced by the informants — but observations outside of the recorded interviews suggested that [QkSV] is in fact quite widely used by children and some adults.”

¹³ Na notação de Coveney (2002 [1996]), *Q* se refere ao constituinte interrogativo, *k* ao complementizador e *SV* a “sujeito-verbo”.

quase categórico de uma das variantes, as *interrogativas-qu*. Essa discussão será desenvolvida no capítulo 2.

A falta de correlação com grupos de fatores sociais

Não raro em estudos sociolinguísticos de variáveis (morfo)sintáticas é o fato de não se observar correlação entre o fenômeno linguístico variável e grupos de fatores sociais (Cheshire, 2005), como o sexo/gênero do falante, sua condição socioeconômica, a situação de fala etc. – exceção feita, em certos estudos, ao grupo de fatores Faixa Etária, possível indicador de mudanças em progresso. Esse foi o caso, por exemplo, do estudo de Mendes (2005) a respeito da variação entre as perífrases de “estar + gerúndio” (p.ex. “Está chovendo muito”) e “ter + particípio” (p.ex. “Tem chovido muito”) em contextos de aspecto durativo e iterativo. O autor verificou uma correlação entre o emprego da forma “estar + gerúndio” e a faixa etária mais jovem dos falantes, mas não observou uma influência significativa do sexo/gênero do informante tampouco do tipo de entrevista (Diálogo entre informante e documentador ou Diálogo entre dois informantes).

A falta de correlação com grupos de fatores não-linguísticos já foi alvo de críticas de pesquisadores mais preocupados com o significado social da variação. A esse respeito, afirma Lavandera (1978):

A noção de variável sociolinguística foi originalmente apresentada para explicar os casos de variação que demonstravelmente carregam *significado social e estilístico* e, além disso, em que o significado social e estilístico se manifesta através de consistentes diferenças nas frequências que covariam com outros fatores linguísticos e extralinguísticos. (Lavandera 1978:177-178. Grifo acrescido. Tradução própria.¹⁴)

De fato, a variável sociolinguística, tal como definida originalmente por Labov (2006 [1966], 1972), tem o objetivo de investigar a variação entre formas “idênticas em referência ou valor de verdade, mas opostas em seu significado social e/ou estilístico”. Nesse aspecto, Lavandera se vê justificada

¹⁴ Texto original: “The notion of the sociolinguistic variable was originally introduced to account for those cases of variation which could be shown to carry *social and stylistic significance*, and furthermore, where social and stylistic significance was manifested by consistent differences in frequencies which covaried with other linguistic and extralinguistic factors.”

em apontar que estudos como o de [Weiner & Labov \(1983 \[1977\]\)](#), sobre a alternância entre voz ativa sem agente e voz passiva no inglês – que não revela qualquer correlação entre a variável linguística e fatores sociais –, não têm o propósito de estudar a comunidade de fala. Para Lavandera, algo se perde na extensão conceitual da variável sociolinguística a fenômenos linguísticos que não demonstram estratificação social.

Entretanto, observar que não há correlação entre determinada variável linguística e grupos de fatores sociais configura antes uma constatação empírica do que um impedimento teórico-metodológico para o desenvolvimento de tais análises. Se é o caso que a maior parte das variáveis além do nível fonológico tende a não revelar correlações sociais, impõe-se uma questão interpretativa que cabe à Teoria da Variação: por que variáveis morfosintáticas tendem a não se distribuir socialmente do mesmo modo que variáveis fonético-fonológicas?

Bell (1984, apud [Cheshire 2005](#)) sugere que a relativa infrequência de formas sintáticas em relação a variáveis fonético-fonológicas faz com que elas tenham uma menor probabilidade de ser associadas a um grupo social específico e ser socialmente avaliadas como marcadores de identidade social.

[Cheshire \(2005\)](#) aventa uma explicação com base na relação forma-função do signo linguístico: por ser arbitrária para variáveis fonológicas, os falantes poderiam usá-las para marcar seu pertencimento a diferentes grupos sociais; para estruturas (morfo)sintáticas, usadas na construção de sentidos proposicionais no discurso, o encaixamento e a estratificação social da variação seriam mais complexos e menos diretos, mas de modo algum inexistentes.

Labov (1993, apud [Meyerhoff 1999](#)) argumenta que a estrutura sintática não está diretamente disponível à percepção dos falantes do mesmo modo que variantes fonéticas e requer uma análise mais abstrata da variação. Esse nível de abstração removeria as variáveis sintáticas daquilo que Labov chama de “monitor sociolinguístico”, usado para associar significados não linguísticos a formas linguísticas. Essa hipótese é resumida no Princípio da Interface:

Membros da comunidade de fala avaliam [socialmente] as formas superficiais da língua, mas não avaliam características estruturais mais

abstratas. (Labov, 1993, 4, apud Meyerhoff 1999. Tradução própria.¹⁵)

Embora as explicações acima não se excluam mutuamente, elas apontam para direções bastante diferentes, fato que revela que a relação entre condicionamentos linguísticos e não linguísticos para variáveis de diferentes níveis de análise ainda não está totalmente clara. Após mais de 40 anos de pesquisas sociolinguísticas, já é possível traçar amplas generalizações sobre o encaixamento social de variáveis fonético-fonológicas (Cf. p.ex. Labov 1994, 2001; Labov *et al.* 2006). O mesmo não se pode afirmar sobre a variação (morfo)sintática, discursiva ou pragmática. Nesse sentido, faz-se ainda mais necessário o desenvolvimento de estudos nesses âmbitos, com vistas à ampliação do entendimento que se tem da variação linguística como parte da competência comunicativa dos falantes.

¹⁵ Texto original: “Members of the speech community [socially] evaluate the surface forms of language but not more abstract structural features.”

2

Dados

Ao tratar de algumas dificuldades metodológicas que o pesquisador enfrenta na coleta de amostras de fala representativas de uma comunidade, adequadas aos objetivos de sua pesquisa, Sankoff (1980:53) menciona o problema de obter certos tipos de dados e cita como exemplo justamente a produção de formas interrogativas: ora, o sociolinguista, com o objetivo de chegar a generalizações a respeito da fala de um grupo maior de indivíduos além daqueles diretamente consultados, volta-se a métodos de amostragem estatística que o coloca em contato com pessoas que não fazem parte de seu convívio diário; esse mesmo fato torna mais difícil a obtenção de gravações em situações diversas de fala e, em geral, limita-o à situação de entrevista sociolinguística; nela, os entrevistados tendem a exibir um repertório linguístico mais restrito, o que resulta em amostras de dados não tão diversificados quanto à gramática ou ao estilo; assim, pessoas sendo entrevistadas têm pouca probabilidade de empregar interrogativas.

A questão colocada por Sankoff (1980) é inquestionavelmente relevante para o estudo da variação. Este capítulo trata de algumas decisões metodológicas tomadas neste estudo a respeito do processo de obtenção e seleção de dados.

2.1 Corpus

O material analisado neste trabalho consiste em diversas amostras do português paulistano contemporâneo:

- (i) 53 entrevistas sociolinguísticas (2003-2008) – cerca de 500 mil palavras;
- (ii) 5 edições da revista *Veja* (nov-dez/2008) – cerca de 170 mil palavras;
- (iii) jornal *Folha de São Paulo online* (dez/2008-ago/2010) – cerca de 155 mil palavras;
- (iv) 1470 redações escolares (2005) – cerca de 230 mil palavras.

As amostras de língua oral (i) e de língua escrita (ii-iv) se distribuem equilibradamente, cada qual com cerca de meio milhão de palavras. A escolha de um *corpus* robusto deveu-se à frequência relativamente reduzida das formas linguísticas investigadas. Uma análise preliminar, em uma amostra de 12 entrevistas sociolinguísticas, forneceu um total de apenas 472 dados das quatro estruturas de interrogativas de constituinte (ver Tabela 1, p. 7, “Corpus piloto”). Esse número de dados, ainda que permita o teste de hipóteses quantitativamente, pode prejudicar a análise de certas variáveis independentes que possuem muitos fatores e lançar dúvidas sobre a relevância estatística dos resultados obtidos.

A seleção de amostras de língua escrita também se norteou pela possibilidade de obtenção de um maior número de dados. O levantamento conduzido por Kato & Miotto (2005) já revelava que a distribuição das formas interrogativas de constituinte parece ser bastante diferente na língua escrita (ver Tabela 1, 7, “Jornal”). Em especial, nota-se que as *interrogativas qu-que* são muito pouco produtivas nesse meio (apenas 1,1%), sobretudo se comparadas ao seu emprego na língua oral (43,4%). Por outro lado, a proporção de ocorrências de *interrogativas qu-in-situ* (8,8%) parece permitir o teste de hipóteses quantitativamente.

Como se comentou na Introdução, Kato & Miotto (2005) haviam feito esse levantamento de ocorrências com o intuito de comparar a produtividade das

quatro estruturas no PB e no PE, mas fizeram-no apenas através de uma análise de frequências. Procedeu-se então à análise de amostras semelhantes (jornal *online* e revista) com o intuito de identificar fatores que favorecem ou desfavorecem o emprego das *interrogativas qu-que* e *qu-in-situ*.

Amostra de língua oral

As 53 entrevistas sociolinguísticas foram coletadas pelos alunos do curso de Sociolinguística da USP entre os anos de 2003 e 2008. Os informantes dessa amostra foram estratificados de acordo com três variáveis sociais:

- (i) sexo/gênero (masculino, feminino);
- (ii) faixa etária (de 20 a 30 anos, de 35 a 45 anos, mais de 50 anos);
- (iii) escolaridade (até Ensino Fundamental II, curso superior)

O grupo de fatores Sexo/gênero, assim como a Faixa etária, é um dos fatores externos mais analisados em estudos sociolinguísticos e, atualmente, é considerado o principal motivador da variação e da mudança linguística (Labov, 2001). A constatação de que esse grupo de fatores não se correlaciona a variáveis morfossintáticas é relativamente comum (Cheshire, 2005); no entanto, como apontam Milroy & Milroy (1997, apud Cheshire 2005), trata-se de uma categoria social ampla que pode ser facilmente levada em consideração na montagem de um *corpus*, o que facilita replicar e comparar estudos em diferentes comunidades.

A Faixa etária pode ser um indicador de movimentos no sistema da língua de acordo com o construto de *mudança linguística em tempo aparente* (Labov, 2001): se se assume que o estado atual da língua de um falante adulto reflete o estado da língua adquirida até o período de sua puberdade, pode-se projetar uma gradação etária em que falantes de diferentes idades em dado momento refletiriam o estado da língua em diferentes momentos passados. Essa hipótese é investigada em relação a possíveis mudanças em curso no emprego de *interrogativas qu-in-situ* e *qu-que*.

A Escolaridade foi escolhida devido à constatação de que as *interrogativas qu-que* são praticamente ausentes na escrita, conforme se viu na Tabela 1 (p. 7): ainda que não pareça ser estigmatizada na língua oral, é possível que pressões normativas desfavoreçam o emprego desse tipo de interrogativas entre informantes mais escolarizados (aqueles com nível superior em curso ou concluído) em relação aos menos escolarizados (aqueles que estudaram até o Ensino Fundamental II).

A amostra de entrevistas sociolinguísticas conta com 4-5 informantes por célula distribuídos conforme a Tabela 2.1.

Tabela 2.1 Número de informantes por célula
no *corpus* de entrevistas sociolinguísticas

Total:	Homens (26)		Mulheres (27)	
53 informantes	E.Fund. (12)	E.Sup. (14)	E.Fund. (12)	E.Sup. (15)
20-30 anos (18)	4	5	4	5
35-45 anos (18)	4	5	4	5
+50 anos (17)	4	4	4	5

Além da fala dos informantes, decidiu-se analisar a fala dos documentadores. Pode-se imaginar que, em virtude do próprio papel que desempenham na entrevista, muitos dados de interrogativas de constituinte se encontram em sua fala. Das 53 gravações selecionadas, foi possível obter o endereço eletrônico de 46 alunos-documentadores através dos registros mantidos pelo professor do curso; a eles foi enviado um questionário (ver Anexo A, p. 148) que teve o duplo objetivo de obter a autorização para o uso/análise dos dados de sua fala, assim como maiores informações a respeito de seu perfil sociolinguístico (idade, cidade de nascimento, bairros/locais em que já morou, ocupação, escolaridade e profissão dos pais e irmãos) e da situação da entrevista gravada (grau de relação com o informante, local da gravação). Dos 46 alunos contactados, 31 responderam ao questionário, dentre os quais 19 são paulistanos. Os dados destes últimos foram incluídos na análise quantitativa.

Tabela 2.2 Perfil sociolinguístico dos documentadores

Total: 19 documentadores	Homens (6)	Mulheres (13)
20-30 anos (16)	4	12
35-45 anos (2)	1	1
+50 anos (1)	1	0

A Tabela 2.2 indica que a distribuição desses falantes é desbalanceada, sobretudo quanto à faixa etária: a grande maioria dos documentadores, alunos universitários, tem menos de 30 anos, e cerca de dois terços são mulheres. Desse modo, realizaram-se análises quantitativas em que foram controlados os grupos de fatores Faixa etária e Sexo/gênero em relação ao papel do falante na entrevista sociolinguística, a fim de verificar se o número desigual de falantes por perfil afeta os resultados. Além disso, a análise da fala dos documentadores deve passar por uma revisão do Paradoxo do Observador (ver seção 2.2).

Ainda que tecnicamente os alunos-documentadores se transformem em “informantes” neste estudo, será mantido o termo “documentador” para se referir a esses falantes, pois tal diferenciação é relevante para alguns grupos de fatores.

A Tabela 2.3 apresenta a distribuição geral de ocorrências das quatro estruturas de interrogativas de constituinte no *corpus* de 53 entrevistas sociolinguísticas. Nela são computados os dados que permitem a variação com pelo menos duas outras estruturas, mas a análise qualitativa do capítulo 3 redefinirá essa distribuição global nos contextos em que as estruturas são intercambiáveis.

Tabela 2.3 Distribuição geral de Interrogativas-Q
no *corpus* de entrevistas sociolinguísticas

	N	%
<i>Interrogativas-qu</i>	721	43,8
<i>Interrogativas qu-que</i>	579	35,1
<i>Interrogativas é-que</i>	121	7,3
<i>Interrogativas qu-in-situ</i>	227	13,8
Total	1648	100,0

Amostras de língua escrita

No caso das amostras de língua escrita, é provável que haja maior planejamento do texto, algo que pode ter consequências no emprego de diferentes estruturas morfossintáticas.

A amostra da revista *Veja* pode ser considerada aquela cuja linguagem é mais cuidada, visto que se trata de uma revista semanal de ampla tiragem (mais de 1,2 milhões de exemplares), com um tempo relativamente maior de planejamento do texto. Apesar de ser uma revista de circulação nacional, considera-se que é representativa do português paulistano, pois sua sede editorial se encontra na cidade de São Paulo. Foram consideradas todas as seções de texto (Editorial, Entrevista, cartas de leitores, colunas de articulistas, notícias etc.), exceto as propagandas comerciais, que, além de possivelmente repetidas em mais de uma edição, são de responsabilidade de agências publicitárias – em princípio – mais do que do corpo editorial da revista.

A amostra do jornal *Folha de São Paulo online*, por sua vez, possui a característica da dinamicidade da informação veiculada na Internet. O site ostenta em sua página inicial o fato de ser “o primeiro jornal *em tempo real* em língua portuguesa”, o que demonstra uma preocupação em publicar notícias tão logo elas ocorram. Isso pode ter consequências na forma do texto escrito, que em geral é mais curto e direto: há notícias que chegam a ter apenas um parágrafo e menos de 100 palavras. De forma semelhante à amostra da revista *Veja*, foram desconsideradas propagandas publicitárias e procurou-se

manter um equilíbrio entre diferentes tipos de notícias (política, esportes, ciências, economia, informática etc.). Também foram incluídos comentários de internautas (cerca de 35 mil palavras), que podem ser representativos de um texto mais espontâneo.

A amostra de redações, por fim, poderia revelar maior variação. Não é raro encontrar exemplos de fenômenos linguísticos em alternância refletidos na escrita de alunos que estão em fase de aprendizagem da norma culta: cliticização de verbos auxiliares (“*tafazendo*”), ditongação de vogais (“*mais*” por “*mas*”; “*feiz*” por “*fez*”), emprego de pronome reto em lugar do oblíquo (“eles vão destruir *ela*”), ausência de concordância nominal (“os carro”) e verbal (“eles foi”). A amostra de 1470 redações foi coletada pelos alunos do curso de Sociolinguística da USP em 2005 e é estratificada de acordo com o sexo/gênero do aluno (masculino, feminino), cinco séries escolares (5^a, 6^a, 7^a, 8^a séries do Ensino Fundamental e 3^o ano do Ensino Médio); e três tipos de escolas (Particular-Classe A, Particular-Classe B, Pública).

Os dados de interrogativas de constituinte da amostra de língua escrita foram extraídos automaticamente com o auxílio do programa de computador R¹ (ver *script* no Anexo B, p. 151), que gerou uma planilha de ocorrências pronta para ser codificada.

No entanto, uma primeira análise dos dados já revela que, nesse *corpus*, o teste quantitativo de hipóteses é inviabilizado pela distribuição extremamente desbalanceada das quatro estruturas de interrogativas de constituinte, fato que leva à não-ortogonalidade entre os fatores de cada grupo (Guy & Zilles, 2007: 171). A Tabela 2.4 mostra que, para todas as amostras de língua escrita, as *interrogativas-qu* são quase categóricas, com taxas de 93,3% a 98,8% dos dados.² O fato de haver pouquíssimas ocorrências de *interrogativas qu-que*, *é-que* e *qu-in-situ* em todas as amostras de língua escrita, mesmo em redações de alunos em fase de aprendizagem da norma culta, sinaliza que a oralidade

¹ Ver página do Projeto R na Internet: <<http://www.r-project.org/>>.

² Há diferenças consideráveis nas taxas de emprego das quatro estruturas de Interrogativas-Q entre o levantamento de Kato & Miotto (2005) (jornal Folha de São Paulo *online*) e o deste estudo (Cf. Tabela 1, p. 7, e Tabela 2.4 adiante). Essas diferenças provavelmente se devem a diferentes critérios de seleção de dados. Kato & Miotto (2005) apenas apresentam as porcentagens, sem especificar os critérios. Os dados da Tabela 2.4 se referem a Interrogativas-Q que permitem a variação com pelo menos duas outras estruturas. Ver definição dos envelopes de variação, p. 67.

pode ser um fator importante na variação.

Tabela 2.4 Distribuição geral de Interrogativas-Q
no *corpus* de língua escrita

	Redações		Folha de SP		Veja		Total	
	N	%	N	%	N	%	N	%
<i>Interrogativas-qu</i>	626	95,4	208	93,3	334	98,8	1168	96,0
<i>Interrogativas qu-que</i>	9	1,4	5	2,2	0	0,0	14	1,1
<i>Interrogativas é-que</i>	11	1,7	2	0,9	3	0,9	16	1,3
<i>Interrogativas qu-in-situ</i>	10	1,5	8	3,6	1	0,3	19	1,6
Total	656	100,0	223	100,0	338	100,0	1217	100,0

Na amostra de redações, nove das 10 ocorrências de *interrogativas qu-in-situ* representam situações de diálogo, como mostram os exemplos em (1).

- (1) *Interrogativas qu-in-situ* em redações
- a. Nesta semana entrevistei a cantora Pitty.
 - Pitty **você fez franja e cortou o cabelo por quê?**
 - Bom, cortei o cabelo, para inovar, mudar de estilo. (F6B)
 - b.
 - Que mochila bonita!
 - Obrigado! Comprei aqui na escola!
 - Que legal! **Você estuda já há quanto tempo aqui?!**
 - A 1 ano! (F8B)

De modo semelhante, 4 de 9 ocorrências de *interrogativas qu-que* são de situações de diálogo representadas em narrativas (2).³

- (2) *Interrogativas qu-que* em redações
- a.
 - Mas, já que falou do show, **pode dizer onde que vai apresentar, os dias e os preços?**
 - Claro! Bom, para o pessoal de São Paulo vai ser dia 25 de Setembro (...). (F6B)
 - b.
 - **O que que foi papai?**
 - Nada! Nada! (M5A)

³ O *corpus* de redações escolares também compreende textos descritivos e dissertativos, embora esta não tenha sido uma variável controlada na amostragem.

Na amostra do jornal Folha de São Paulo, todas as oito ocorrências de *interrogativas qu-in-situ* são provenientes de Comentários de Leitores (3). Dentre as cinco ocorrências de *interrogativas qu-que*, três são igualmente de leitores (4-a), uma é do Blog do Folhateen (4-b) e uma é da parte de Informática (4-c):

- (3) *Interrogativas qu-in-situ* no jornal Folha de São Paulo
- a. Tem de mandar para o Papa, sim! Afinal, quando simples Cardeal, **ele acobertou quantos casos?** Que se dê a desculpa de que, como Cardeal, nada podia; porém, agora é Papa (...)
 - b. Mas se escolhendo homens com vida normal, se resume as chances, pq quem não fica com mulher nesta terra, **fica com quem?** Com homem, na falta o pior com crianças. **A sociedade prefere o que?**
- (4) *Interrogativas qu-que* no jornal Folha de São Paulo
- a. Desculpa por me entrometer, mas não é que exista falta de interesse do governo em resolver o problema da segurança.... o que realmente acontece é que os verdadeiros “cabeças” do crime organizado estão “governando” o país, então se eles resolverem o problema da segurança eles estarão perdendo dinheiro com isso, pensa comigo, **como que com o exército nas fronteiras a droga entra no país?** é pq tem algum político influente que manda o exército liberar!
 - b. Quem depois da vitória sobre Portugal e de um bom primeiro tempo imaginaria essa despedida da copa do mundo? **Quem que não ficou com pena do Julio Cezar?**
 - c. No Brasil, a produtora O2 Digital criou em junho a websérie “**O que que é isso?**”, para um portal de serviços de internet.

A parte de Comentários de Leitores consiste em um espaço, ao fim das notícias, em que os internautas podem emitir suas opiniões ou discutir a notícia veiculada. Trata-se de um modo textual que é mais dialógico do que o restante do jornal, pois os autores dos comentários têm a oportunidade de responder outros internautas como se estivessem em um debate. Nos exemplos (3-a), (3-b) e (4-a) acima, é possível observar que o texto adquire características de oralidade: além de abreviações típicas de salas de bate-papo (“pq”), vários trechos remetem o leitor a um texto falado (“Tem de mandar para o Papa, sim! Afinal...”, “Desculpa por me entrometer...”, “pensa comigo...”). O Blog do Folhateen (4-b), assim como os Comentários

de Leitores, é composto por textos enviados pelos internautas. Apesar de claramente haver uma seleção de textos a serem publicados, trata-se também de um texto mais “informal”. A ocorrência da seção de Informática (4-c), por sua vez, refere-se a um serviço de um portal da Internet que se nomeia a partir de uma expressão bastante comum na língua oral, “O que que é isso?”.

Na amostra da revista *Veja*, não houve qualquer ocorrência de *interrogativa qu-que*, e o único dado de *interrogativa qu-in-situ* foi retirado de uma entrevista, presumivelmente falada e depois transcrita:

- (5) *Interrogativa qu-in-situ* na revista *Veja*
– **O senhor fazia o que** antes de atuar na Abin?
– Fui do corpo de segurança do presidente Lula. (...)

Desse modo, ainda que haja poucas ocorrências de *interrogativas qu-que* e *qu-in-situ* nas amostras de língua escrita, grande parte delas aponta para a importância da oralidade no emprego dessas formas. Entretanto, o termo “oralidade” é demasiado amplo: que aspectos da fala poderiam estar correlacionados à variação? Neste trabalho (ver Capítulo 3), na amostra de entrevistas sociolinguísticas, foram investigados grupos de fatores discursivo-pragmáticos, relacionados com a fala-em-interação, bem como grupos de fatores prosódicos, a fim de descrever mais detalhadamente as características que podem explicar a ampla diferença distribucional de Interrogativas de constituinte entre fala e escrita.

2.2 Dados de documentadores: o Paradoxo do Observador revisitado

A dificuldade de obter amostras de fala espontânea para a análise sociolinguística já é bem conhecida: trata-se daquilo que Labov (2006 [1966]) chamou de *paradoxo do observador*:

Alguns dos exemplos mais convincentes do significado social de uma variável ocorre quando o linguista é simplesmente um observador anônimo. Nessas situações, podemos observar o comportamento linguístico

sem o *efeito enviesador de atenção consciente à fala*, que é *característico da entrevista linguística*. (...) Contudo, a interação conversacional anônima e casual é normalmente o tipo de observação menos passível de ser controlado: não se pode esperar que se aprenda muito a partir de tais anotações aleatórias a menos que a variação ao longo do eixo social e estilístico seja rigidamente controlada. (Labov 2006 [1966]:38. Grifos acrescidos. Tradução própria.⁴)

Nesse trecho, Labov discute a posição do linguista que busca estudar a língua em uso sistematicamente. Um modo de fazê-lo é através de gravações de entrevistas sociolinguísticas, em que as variáveis sociais e estilísticas podem ser controladas; no entanto, nessa situação, é provável que as pessoas, sabendo-se observadas, distanciem-se de seu modo mais espontâneo de fala.

A gravação da fala apenas de pessoas próximas, com quem o pesquisador tem contato diariamente, poderia providenciar amostras de fala mais naturais. Neste caso, contudo, a tendência seria a de registrar a fala de pessoas com um perfil social próximo ao do documentador, geralmente um professor ou aluno universitário, e o estudo ficaria restrito a uma descrição de sua própria variedade de fala, perdendo-se grande parte do interesse sociolinguístico. Ainda que se registrassem apenas pessoas de convívio próximo, a simples presença do gravador pode criar o efeito de monitoramento de fala e levar a um aumento da formalidade. Labov nota que esse efeito tende a diminuir ao longo da entrevista, mas que ele deve ser interpretado como uma “interferência constante” na espontaneidade do procedimento (Labov, 2006 [1966]:90). O roteiro de perguntas também restringe os possíveis tópicos sobre os quais ambos os participantes poderiam preferir falar naquele momento, naquele local, naquela situação, mas é necessário a fim de garantir a comparabilidade dos dados.

Assim, a presença do gravador, o *frame* de entrevista e a existência de um roteiro de perguntas contribuem, todos, para o distanciamento da fala espontânea e cotidiana. Em poucas palavras: o maior desafio do pesquisador é

⁴ Texto original: “Some of the most convincing illustrations of the social significance of a variable occur when the linguist is simply an anonymous observer. In such situations we can observe linguistic behavior without the biasing effect of conscious attention to speech, which is characteristic of the linguistic interview. (...) Yet the anonymous and casual speech exchange is usually the most uncontrolled type of observation: we cannot hope to learn very much from such random jottings unless the variation along the social axis, and the stylistic axis, is tightly controlled.”

tentar superar o viés do monitoramento de fala dentro da situação controlada da entrevista sociolinguística, ela mesma causadora do viés.

Tal questão normalmente é tratada como um problema metodológico que deve ser remediado através do desenvolvimento de técnicas melhores e mais eficientes; sem dúvida, todos os esforços do investigador nesse sentido se justificam. Entretanto, é necessário ter em mente que sempre haverá resquícios de problemas metodológicos. Garfinkel (1967:189-90) avalia que toda pesquisa empírica que se utiliza de dados coletados apresenta dois tipos de problemas: um primeiro, chamado de “problemas metodológicos gerais”, e um segundo, que o autor denomina “problemas normais e naturais”. Estes se referem a efeitos que decorrem naturalmente das regras inerentes às práticas sociais. Sobre eles, assumem-se as palavras do autor:

Não estamos dizendo isso de forma irônica. *Não* estamos dizendo: “O que mais você pode esperar?!” Em vez disso, o termo “normal, natural” é usado em um sentido sociológico convencional e quer dizer “de acordo com as regras práticas vigentes.” (Garfinkel 1967:191. Grifos do autor. Tradução própria.⁵)

Os problemas “normais e naturais” ocorrem porque os membros da sociedade buscam agir de acordo com as regras das práticas cotidianas que, de seu ponto de vista, representam a forma como tais atividades devem ser conduzidas.

Assim, a respeito da situação de entrevistas sociolinguísticas, a possibilidade de *total supressão* do efeito de monitoramento deve ser descartada. Uma propriedade fundamental da conversa espontânea e cotidiana é seu caráter privado – um acordo tácito entre interlocutores, parte da ética prática dos membros da sociedade. Garfinkel (1967:75) ilustra a operação dessas regras sociais em seus experimentos de ruptura, conduzidos na década de 1960, que tinham o objetivo de quebrar regras sociais tácitas a fim de estudar o seu funcionamento. Entre os experimentos, muitos dos quais conduzidos por seus alunos, estavam a tarefa de se comportar como um hóspede dentro de casa, agir como se o interlocutor estivesse escondendo motivos escusos,

⁵ Texto original: “We do *not* mean this ironically. We are *not* saying, ‘What more can you expect?!’ Rather, the term ‘normal, natural’ is used in a conventional sociological sense to mean ‘in accord with prevailing rules of practice.’”

aproximar-se do rosto de outra pessoa até que quase se tocassem... (Cf. Garfinkel 1967: cap.2).

Em um desses experimentos, o pesquisador iniciava uma conversa enquanto portava um gravador escondido sob a roupa; durante o colóquio, o pesquisador abria o casaco, revelando o aparelho, e dizia: “Está vendo o que eu tenho?” A reação dos sujeitos era de uma pausa, seguida da pergunta: “O que você vai fazer com isso?” A justificativa para o desconforto, formulada pelos próprios sujeitos, era a expectativa de que a conversa era privada; a quebra de tal acordo implícito conduzia a novas possibilidades, e o pesquisador e o sujeito, então, viam-se diante da necessidade de negociar um novo acordo.

Tendo isso em mente, é necessário considerar que a situação de entrevista sociolinguística produz, inevitavelmente, uma conduta que pode divergir dos modos como as pessoas se comunicam cotidianamente em outras situações. Se se considera que a fala dos indivíduos entrevistados afasta-se da naturalidade da conversa cotidiana devido ao monitoramento da fala, isto é tanto mais verdadeiro para o documentador que coleta a entrevista: ele, mais do que o informante, tem plena consciência do roteiro a ser seguido e dos propósitos da gravação.

Nos estudos variacionistas, não é usual a análise da fala das pessoas que coletaram as entrevistas sociolinguísticas; em muitos casos, é o próprio pesquisador que faz a coleta, de modo que não seria possível projetar tendências na comunidade a partir de sua própria fala. Assim, o papel do documentador em geral é tão somente o de proporcionar as condições para que os informantes de determinados perfis sociolinguísticos se expressem verbalmente.

Contudo, este estudo dispõe de um *corpus* de entrevistas coletadas por dezenas de alunos do curso de Sociolinguística da USP entre os anos de 2003 e 2008. Há vantagens e desvantagens na utilização dos dados da fala desses alunos que devem ser explicitadas.

Por um lado, é certo que em alguns aspectos seus perfis sociolinguísticos são semelhantes (alunos universitários entre 20 e 25 anos), embora haja uma variabilidade razoável em relação ao sexo/gênero dos alunos. A espontaneidade da fala dos documentadores pode ser questionada devido à existência

de um roteiro previamente estabelecido (ver Anexo C, p. 154) que todos os alunos deveriam seguir; é razoável supor que a entrevista tenha sido precedida de uma preparação através da memorização de temas e de perguntas a serem feitas. Assim, ocorre um monitoramento da fala também pelo fato de que parte da avaliação na disciplina seria feita a partir do cumprimento ou não desse roteiro. Há então a figura de dois observadores (além do gravador): para o informante, é o aluno-documentador que coleta a entrevista; para o aluno, é o professor da disciplina que o aprovará ou não ao fim do semestre.

Por outro lado, na época em que coletaram as gravações, os alunos não tinham consciência de que a sua fala seria analisada futuramente (de fato, tampouco esta autora havia definido o seu objeto de estudo). Quanto ao fenômeno linguístico específico desta análise, é bastante provável que os alunos não estivessem monitorando a sua fala; nenhuma das formas de interrogativas de constituinte é estigmatizada na língua oral, ao menos em São Paulo, e todas parecem ser empregadas de forma semelhante por falantes com maior ou menor nível de escolarização.

Ao mesmo tempo, as mesmas características que exacerbam o paradoxo do observador podem ser revertidas em benefício desta análise: se o intuito da entrevista sociolinguística é o de ter uma base de comparação dos dados dos falantes, o próprio roteiro estabelece uma série de pontos que se repetem em todas elas sistematicamente, tornando os dados de documentadores tanto mais comparáveis.⁶ Mais importante é notar que há variabilidade nesses dados, não obstante o roteiro: nele, todas as interrogativas de constituinte foram formuladas como *interrogativas-qu* (exs. (6-a) e (7-a)), mas é frequente a ocorrência de realizações alternativas (exs. (6-b)-(6-c) e (7-b)-(7-c)).

- (6) a. ROTEIRO: Como você acha que as pessoas pronunciam o ERRE em São Paulo, em palavras como “porta” e “caderno”?
- b. Jorge: e na hora de falar assim... você acha que em São Paulo **como que** o pessoal pronuncia o “erre”? (M1S-DOC)
- c. Renata: **como é que** você acha que as pessoas pronunciam o “erre” em São Paulo?... em palavras como “porta” “caderno”? (F2S-DOC)

⁶ Entre os anos de 2003 e 2008, houve algumas modificações no roteiro de entrevista, mas o seu desenho geral permaneceu o mesmo.

- (7) a. ROTEIRO: Quem você acha que fala mais assim? Empregadas domésticas, caixas de banco, atendentes de lojas, operadoras de telemarketing ou seus amigos?
 b. Helena: e pra você **quem que** fala mais assim? o telemarketing ou então empregada doméstica... caixa de banco...? (F1S-DOC)
 c. Carlos: cê acha que... você ouve mais isso... **em que tipo de pessoa?** de repente empregada... (cara) que trabalha em banco... atendente de loja... operadora de telemarketing...? (M3S-DOC)

Além disso, pode-se afirmar que o monitoramento da fala deve-se a um monitoramento dos papéis atribuídos a cada um dos falantes em uma situação de entrevista. Essa consciência nada mais é do que manifestação da competência comunicativa dos falantes, que sabem, a partir de seu conhecimento das regras culturais, que nesta situação uma pessoa deve falar mais e outra deve falar menos; uma deve fazer perguntas, outra deve respondê-las. Uma evidência de que os falantes seguem essas normas na condução de suas ações na entrevista é o fato de que “desvios” da regra normalmente são marcados explicitamente:

- (8) João: eu não vou resistir... **eu sei que você é a entrevistada (mas) eu não resisto...** ((risos)) (M2S-DOC)
 Cecília: ((risos))
 João: mas eu vou em médico... eu também fic-... esse negócio que você falou da sua vizinha... meu... eu vou em médico eu fico assim meu... eu pego assim- eu- eu- eu não fico à vontade com médico...
 (9) Jefferson: é **engraçado... eu sou o entrevistado mas eu acabo fazendo perguntas...** ahn... do que que você tem mais medo... da polícia ou do ladrão? (M2G-INF)

Em (8), o documentador João introduz seu comentário pessoal sobre como ele se sente quando vai ao médico com uma lembrança do papel que cada um dos falantes tem naquela interação, sinalizando sua consciência de que seu comportamento não é o mais adequado naquele momento. De forma muito semelhante, em (9) o informante Jefferson anuncia verbalmente a quebra da norma tácita antes de fazer sua pergunta.

Quando não há marcação explícita para a reversão de papéis, a reação do interlocutor chega a ser semelhante à dos sujeitos dos experimentos de ruptura de Garfinkel (1967), estupefatos por se verem em uma posição não

esperada:

- (10) Pedro: (...) o que você acha da religião? (M1S-INF)
 Carla: h-h-... não- EU tô fazendo a entrevista cê dá licença? (F1S-DOC)
- (11) Ulisses: hoje a gente tem dicionário até de... língua portuguesa com gíria... não devia ter... o que é é... o que não é não é... é isso que eu penso assim... e você pensa o quê? (M3G-INF)
 Francisco: ((risos)) (M1S-DOC)
 Ulisses: não não... o que você pensa?
 Francisco: não a entrevista aqui... é o que o senhor acha... mas ahh...

Nesta pesquisa, assume-se, em consonância com a literatura sociolinguística tradicional, que a situação de entrevista diverge daquela de fala espontânea e cotidiana que o sociolinguista desejaria em princípio investigar, mas fornece uma sólida base de comparação através da qual variáveis sociais e estilísticas podem ser controladas. Assume-se também que a fala do documentador, mais do que a fala do informante, é duplamente marcada pelo paradoxo do observador. É certo que o papel atribuído ao documentador na entrevista sociolinguística tem consequências em sua fala, mas isso também é verdadeiro para a fala do informante. Tal conduta não é senão reflexo da consciência dos falantes a respeito dos modos apropriados de agir segundo normas – tácitas, mas reconhecíveis – de convívio social (Garfinkel, 1967). O conhecimento dessas normas faz parte de sua competência comunicativa (Hymes, 1991 [1979]).

2.3 Interrogativas e perguntas: forma e função

Até o momento, os termos “interrogativas” e “perguntas” vêm sendo empregados de forma relativamente indistinta. No entanto, ambos os termos, muitas vezes tomados como sinônimos, precisam ser diferenciados. Aqui, entende-se “pergunta” como expressões que desempenham *funções discursivas*, enquanto o termo “interrogativa” é usado para se referir a *estruturas sintáticas*.

Embora muitas vezes uma expressão possa ser, ao mesmo tempo, uma

interrogativa e uma pergunta, esta relação forma-função nem sempre coincide: uma mesma forma pode exercer mais de uma função, e uma mesma função pode ser expressa de diversas formas. De fato, é exatamente tal relação assimétrica que caracteriza a variação linguística.

Conforme lembram Milroy & Gordon (2003),

(...) uma vez que não há relação isomórfica entre função e forma, perguntas nem sempre são realizadas sintaticamente como interrogativas e formas interrogativas podem realizar muitas funções diferentes. (Milroy & Gordon 2003: 170. Tradução própria.⁷)

Na seção 1.3, apresentaram-se diferentes critérios possíveis para a definição do envelope de variação. Labov (1969) e Tagliamonte & Roberts (2005, apud Buchstaller 2009) partem de critérios formais em seus estudos sobre a realização variável da cópula e intensificadores de adjetivos, respectivamente; Rickford *et al.* (2007, apud Buchstaller 2009) partem de um critério funcional de equivalência em seu estudo sobre intensificadores.

Coveney (2002 [1996]) também adota um critério funcional em seu estudo sobre variação de *interrogations* em francês. Para o autor, primeiro é necessário identificar a função comunicativa de todos os enunciados a fim de estabelecer o conjunto de variantes possíveis e disponíveis à seleção do falante – diferentes grupos de variantes interrogativas podem servir a diferentes funções comunicativas (pedidos de informação, pedidos de ação, ofertas, perguntas-eco etc.).

Entretanto, definir um conjunto de funções possíveis para as formas interrogativas está longe de constituir uma tarefa objetiva. Para citar apenas alguns exemplos, comparem-se as tipologias de Martín (1995), Athanasiadou (1991) e Lima (2005). Essas três obras são citadas por Silva (2006).

Martín (1995, apud Silva 2006: 267) identifica três tipos de perguntas, cada qual subdividida em dois subtipos: (a) as *perguntas inquisitivas* são perguntas “autênticas” que buscam uma informação, e podem ser totais (“sim” ou “não”) ou parciais (algum “elemento concreto”); (b) as *perguntas inquisitivo-expressivas* retêm o caráter inquisitivo, mas são acompanhadas de valores

⁷ Texto original: “(...) since there is no isomorphic relationship between function and form, questions are not always realized syntactically as interrogatives and interrogative forms may realize many different functions.”

expressivos (irritação, impaciência, ironia etc.), podendo ser subdivididas em negativas e afirmativas; e (c) as *perguntas retóricas* têm uma resposta conhecida que o próprio locutor se encarrega de fornecer, podendo ser auto-interrogativas (em que o próprio falante anuncia a resposta) ou interrogativo-exclamativas.

No estudo conduzido por Athanasiadou (1991, apud [Silva 2006](#):266) a respeito da situação específica de interação conversacional em sala de aula, o autor delimita quatro tipos de perguntas que, segundo [Silva \(2006\)](#), podem ser empregadas tanto por professores quanto por alunos com frequência variável: (a) as *perguntas de informação ou referenciais* são perguntas autênticas, que buscam uma informação desconhecida; (b) as *perguntas expositivas* têm o objetivo de checar se o interlocutor sabe a resposta correta; (c) as *perguntas retóricas* intencionam dar informação ao interlocutor sem que este responda de fato; e (d) as *perguntas indiretas* exigem alguma ação por parte do interlocutor.

Lima (2005, apud [Silva 2006](#):267-271), também em um estudo sobre interação conversacional em sala de aula, começa com uma divisão das perguntas em duas categorias, do professor ao aluno e do aluno ao professor, cada qual subdividida em sete categorias de caráter fortemente funcional e tópico-discursivo. A autora classifica as perguntas do professor ao aluno em: (a) *perguntas de verificação de leitura explícita*; (b) *perguntas de compreensão*; (c) *perguntas gatilho ou “prefácio” para a explicação/revisão*; (d) *perguntas retóricas*; (e) *perguntas de organização da classe*; (f) *perguntas dirigidas*; e (g) *perguntas esvaziadas*. As perguntas do aluno ao professor classificam-se em: (a) *perguntas de confirmação*; (b) *perguntas metalinguísticas*; (c) *perguntas de esclarecimento*; (d) *perguntas de incompreensão explícita*; (e) *dúvidas sobre tópicos exteriores ao tema em desenvolvimento*; (f) *perguntas como respostas*; e (g) *perguntas de organização*.

Esses exemplos deixam claro que o conjunto de possíveis funções desempenhadas por perguntas pode se desdobrar em listas cada vez mais extensas ([Oushiro & Nasser, 2010](#)). Neste estudo, em lugar de tentar definir uma lista de funções, apresenta-se uma visão estrutural com base em conceitos da Análise da Conversa Etnometodológica ([Sacks 1992](#); [Sacks et al. 1974](#),

entre outros), ao mesmo tempo em que se apontam os casos em que sua distribuição coincide com a de formas interrogativas. Como se verá, mesmo o conjunto de sobreposição entre forma e função não esclarece os limites das formas variáveis. Desse modo, opta-se por um ponto de partida formal – à semelhança de Labov (1969) e Tagliamonte & Roberts (2005) – para a definição do envelope de variação.

A definição de *perguntas*, à primeira vista, parece simples, dado que todos os falantes têm uma intuição sobre o que elas são. Em sua versão prototípica, de fato, forma e função se alinham:

- (12) Rosa: faz tempo que você mora aqui? (F1S-DOC)
Edith: não... faz um ano que eu estou morando aqui
- (13) Rosa: **como** foi a mudança pra cá? (F1S-DOC)
Edith: ah foi muita expectativa né

O exemplo (12), uma pergunta “sim ou não” (ou “pergunta total”) (Silva, 2006), é dado para lembrar que perguntas não se limitam àquelas de constituinte, embora os casos que interessam primariamente neste estudo sejam aqueles como em (13), com um constituinte interrogativo. Ambos os exemplos acima têm, além da função prototípica de perguntas (i.e. obter uma informação), uma forma interrogativa: no caso de (12), a forma é marcada principalmente pela entonação; em (13), pela entonação e pelo constituinte interrogativo.

No entanto, sabe-se que nem todas as perguntas têm a função de obter uma informação do interlocutor. Tal é o caso de perguntas como (14) e (15) abaixo, tradicionalmente chamadas de retóricas:

- (14) Cecília: os caras todos armados ali e a gente ia fazer **o quê?** ((risos))
(F2S-INF)
- (15) Fábio: **por que que** o médico não consegue ser um bom médico lá na periferia... lá no rincão lá do Nordeste?... porque às vezes lá ele não tem... não é reconhecido... (M1G-INF)

Em (14), tanto a falante quanto o seu interlocutor sabem que não se espera uma resposta para a pergunta (ainda que ela possa ser enunciada), dada

a sua obviedade (Rohde, 2006); em (15), é o próprio falante que fornece a resposta, de forma que a pergunta feita, assim como em (14), não busca obter uma informação.⁸

A característica que parece reunir os exemplos acima sob um mesmo rótulo de pergunta é a relação dessas expressões com uma resposta, ainda que esta nem sempre seja enunciada explicitamente. A noção de um *par* parece fundamental na definição de perguntas.

Em uma série de artigos, Sacks, Schegloff e Jefferson (Sacks 1992, 1987 [1973]; Schegloff 1972*b,a*; Schegloff & Sacks 1973; Sacks *et al.* 1974) desenvolveram o conceito de *par adjacente* ao analisar diversas sequências conversacionais recorrentes na interação. Dentre os exemplos de sequências, além do par pergunta-resposta, estão saudação-saudação, interpelação-resposta, convite-aceitação/recusa, desafio-aceitação/recusa etc. Tais sequências compõem-se de duas partes (denominadas *primeira parte do par* e *segunda parte do par*) e se caracterizam pelos seguintes atributos: (a) extensão de dois enunciados; (b) posição adjacente dos enunciados que as compõem; (c) produção de cada enunciado por diferentes falantes; (d) ordem relativa das partes; (e) relações discriminativas, i.e. o tipo de par do primeiro enunciado é relevante para a seleção do segundo (Schegloff & Sacks, 1973 : 295-6).

Ao se produzir a primeira parte de um par, uma regra básica de seu funcionamento é que o falante deve parar e o próximo falante deve começar a produzir a segunda parte do par correspondente. Entretanto, a ocorrência de desvios dessa regra também é possível; Schegloff (1972*a*) nota que é possível a inserção de uma sequência pergunta-resposta dentro de outro par pergunta-resposta:

- (16) A: (Pb) Você vem hoje à noite?
 B: (P1) Posso trazer outra pessoa?
 A: (R1) Claro.
 B: (Rb) Eu vou lá.

(Exemplo extraído de Schegloff 1972*a* : 78. Tradução própria.⁹)

⁸ Na seção 3.3, p. 68, será feita uma distinção entre as perguntas dos exemplos (14) e (15) quanto à sua organização no mecanismo de troca de turnos e à atualização do fundo comum entre os interlocutores.

⁹ Texto original: “A: (Qb) Are you coming tonight?
 B: (Q1) Can I bring a guest?”

Em (16), B responde a pergunta de A (Pb) inicialmente com outra pergunta (P1); quando A fornece a resposta de P1 (R1), B fornece a resposta de Pb (Rb). A adjacência, portanto, é melhor entendida como uma tendência à contiguidade em vez de uma regra indisputável. Schegloff (1972a) também aponta outros desvios da regra, como perguntas que não passam o turno ao interlocutor e perguntas em que o interlocutor tipicamente replica com um pedido de resposta (p.ex. charadas). Assim o autor esclarece o sentido de “organização básica” do par pergunta-resposta:

(...) por “básica” quero dizer que, embora outras sequências possam ser encontradas empiricamente, elas serão melhor analisadas se vistas como modificações dessa “organização básica” (...) (Schegloff 1972a:77. Tradução própria.¹⁰)

Este trabalho não propõe uma definição acabada de perguntas, mas apresenta uma noção que parece estar de acordo com o conhecimento que os falantes têm de seu funcionamento: perguntas são sequências que se ligam a uma contraparte – uma resposta –, com a qual formam um par. Tal contraparte pode ser uma informação requerida do interlocutor (ex. (13)), seu caso prototípico, mas diversas modificações podem ser feitas a partir de sua configuração básica: a resposta pode não estar adjacente (ex. (16)); pode já ser óbvia e, portanto, não ser requerida (ex. (14)); pode ser fornecida pelo próprio falante que fez a pergunta (ex. (15)).

Do mesmo modo que se pode considerar que existe uma função prototípica de perguntas de constituinte a partir da qual pode haver “desvios” de sua configuração básica, é possível conceber uma forma prototípica de interrogativa de constituinte que apresenta variações. Tal é o caso do primeiro exemplo apresentado na Introdução, reproduzido abaixo:

- (17) a. Onde você mora?
 b. Onde que você mora?
 c. Onde é que você mora?

A: (A1) Sure.

B: (Ab) I'll be there.”

¹⁰ Texto original: “(...) 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’ (...)”

- d. Você mora onde?

Pode-se considerar que a forma em (17-a) é a mais prototípica de interrogativa de constituinte: essa é a forma mais produtiva na língua escrita, aquela apresentada nas gramáticas normativas e ainda bastante empregada na língua oral. Além das quatro formas apresentadas em (17), há outras possibilidades *formais*:

- (18) a. Edilson: bom desculpa não entendi **o que que** você quer (M2S-INF)
 b. Fernanda: então eu gostaria de saber **o que** você acha sobre a saúde... de São Paulo (F1S-DOC)

Em (18), as orações com constituinte interrogativo encontram-se subordinadas aos respectivos verbos das orações principais (“entendi” e “saber”). Em geral, considera-se que tais construções também são perguntas, mas menos prototípicas do que aquelas em (17): são chamadas tradicionalmente de “interrogativas indiretas” (Bechara, 2005; Cunha & Cintra, 2007). Para casos como (18), ainda é possível considerar que os enunciados têm uma contraparte – uma resposta – com a qual formam um par adjacente.

Exemplos de interrogativas indiretas em manuais de gramática são tipicamente apresentados em orações subordinadas a *verba cognoscendi*, que são mais propícios à inferência de que o enunciado busca uma informação e, portanto, aproximam essas estruturas sintáticas do sentido de pergunta. No entanto, não é possível fazer uma relação direta entre esse tipo de verbo e uma função discursiva específica.

- (19) Fabrício: (...) só vou entender um quadro se tiver um sentido mais explícito... num... num quadro... não sei **quem** fez aquele dos retirantes... você vê que é uma coisa que está impregnada a pobreza... mas não... fora isso ahn fica meio difícil... (M1S-INF)

No exemplo acima, diferentemente de (18-b), a presença do verbo “saber” não garante que o informante Fabrício esteja dirigindo uma pergunta a seu interlocutor sobre quem é o pintor do quadro – com efeito, nesse trecho, o informante continua seu turno sem ceder a palavra ao documentador.

Com outros termos que não *verba cognoscendi*, a relação com uma segunda

parte do par torna-se ainda menos óbvia: ocorrências como (20) podem ser consideradas simples asserções.

- (20) a. Eduardo: isso vai depender... **com quem** estou conversando... (M2S-INF)
 b. Claudia: ela nunca falou pra gente **por que que** ela veio pra cá... (F1G-INF)

Ainda que não sejam perguntas, tais construções permitem a variação com outras formas de interrogativas de constituinte. Mais especificamente, são as três formas com constituinte pré-verbal (*interrogativas-qu*, *qu-que* e *é-que*) que se alternam nesses contextos de encaixamento. O exemplo (20-a), uma *interrogativa-qu*, também permite a construção *qu-que* (21-a) e *é-que* (21-b); por sua vez, a estrutura (20-b), uma *interrogativa qu-que*, permite sintaticamente as construções alternativas *-qu* (22-a) e *é-que* (22-b):

- (21) a. isso vai depender **com quem que** estou conversando
 b. isso vai depender **com quem é que** estou conversando
- (22) a. ela nunca falou (...) **por que** ela veio pra cá
 b. ela nunca falou (...) **por que é que** ela veio pra cá

Os exemplos (18) a (20) mostram que o limite no qual formas interrogativas deixam de ter a função de pergunta (i.e., ter uma relação com uma segunda parte do par) não é tão nítido. Observem-se ainda os seguintes casos:

- (23) Carlos: e você tá morando...? (M3S-DOC)
 Leandro: na Bela Vista
- (24) Jean: nome completo então? (M1S-DOC)
 Marco: é Marco F.N.
 (...)
 Jean: endereço?
 Marco: é... residencial ou comercial?
 Jean: residencial
 Marco: residencial é rua...
- (25) Mônica: você tem... você torce pra algum time? (F1S-DOC)
 Márcia: é... eu torço pro Corinthians mas eu não sou fanática de bater em ninguém

Há inúmeras formas de fazer uma pergunta parcial, que busca informações específicas referentes a constituintes interrogativos (*o que, quem, onde, quando* etc.), sem a necessidade de enunciar esse constituinte. Em (23), não foi necessário que Carlos dissesse “onde”, pois uma entonação ascendente foi o suficiente para que Leandro fornecesse a resposta pretendida. Em (24), o documentador Marco só enuncia sintagmas nominais (“nome completo”, “endereço”) em vez de todo o sintagma verbal (“Qual é o seu nome completo?”, “Qual é o seu endereço?”) para obter as informações necessárias. Em (25), a pergunta “sim ou não” (“Você torce para algum time?”) teve o mesmo efeito de uma pergunta como “Pra que time (que/é que) você torce?”/ “Você torce pra que time?”, pois Márcia já fornece tal informação (“Corinthians”); de fato, os falantes sabem que uma possível resposta para aquele tipo de pergunta não se restringe à simples afirmação ou negação da proposição.¹¹

De um ponto de vista funcional, seria possível considerar que as formas em (23)-(25) exercem a mesma função que suas contrapartes com constituinte interrogativo. No entanto, incluir todas essas formas no envelope de variação – e outras tantas que a imaginação permitisse – não parece viável em uma análise quantitativa de covariação. Um ponto de partida estritamente funcional tornaria a tarefa de “fechar o conjunto” de variantes (Labov, 1996) quase impossível. Este trabalho adota um critério formal para que as variantes não se estendam em uma listagem sem fim; tal critério é a definição do objeto de estudo como *interrogativas de constituinte* – que podem ou não ser perguntas.

Assim, os dados que se analisam referem-se a enunciados que contêm um constituinte interrogativo: (*o*) *que, que + NP, qual(-is), qual(-is) + NP, quanto(-a, -os, -as), quanto(-a, -os, -as) + NP, quem, como, quando, onde e por que*. Esses enunciados se distribuem entre as quatro estruturas morfossintáticas apresentadas na Introdução, que são discutidas mais extensamente a seguir.

¹¹ Clark (1996:217-8) reporta uma situação semelhante para o inglês: em um de seus estudos, 50 restaurantes na Califórnia foram contactados por telefone para perguntar: “Vocês aceitam cartão de crédito?”. Além das respostas “sim” e “não”, outras duas formas de respostas foram “Sim, aceitamos Mastercard e Visa” ou simplesmente “Aceitamos Mastercard e Visa” (Clark, 1979, apud Clark 1996). É possível, entretanto, que haja diferenças culturais em comunidades específicas e que, em algumas, uma pergunta como “Você torce para algum time?” de fato estimule apenas uma resposta “sim” ou “não”.

3

Análise qualitativa

No capítulo anterior, foi estabelecido um ponto de partida formal para a definição do envelope de variação: o conjunto de dados analisados se define inicialmente pelas formas interrogativas de constituinte, na fala tanto de informantes quanto de documentadores paulistanos.

Este capítulo apresenta a análise qualitativa dos dados, através da qual se definem os contextos em que as formas de Interrogativas-Q podem ser consideradas opcionais. Primeiramente, discute-se a equivalência semântica entre as estruturas interrogativas em foco e propõe-se o conceito de *pressuposição do falante* (Stalnaker, 2002) como aquele que define tais estruturas como formas variantes. Em seguida, a partir do conceito de *competência comunicativa* (Hymes, 1991 [1979]), verifica-se em que contextos as formas são (a) possíveis; (b) factíveis; (c) apropriadas; e (d) empregadas. Com base nesses parâmetros, definem-se dois envelopes de variação para o emprego variável de Interrogativas-Q: em um deles, a variável é definida pela posição do constituinte interrogativo, de maneira que a forma *in situ* é vista em oposição às outras três formas; no outro, analisa-se a variação entre as estruturas cujo constituinte interrogativo está em posição pré-verbal. Por fim, apresentam-se os grupos de fatores estabelecidos com base nas hipóteses aventadas para explicar os empregos das estruturas variantes dessas duas variáveis.

3.1 Equivalência semântica

Há quatro estruturas produtivas de Interrogativas-Q no português paulistano atual, apresentadas na Introdução e novamente na Tabela 3.1 abaixo:

Tabela 3.1 Interrogativas-Q

Tipo de Interrogativa-Q	Exemplo	Estrutura
a. <i>Interrogativa-qu</i> :	Onde você mora?	Q +
b. <i>Interrogativa qu-que</i> :	Onde que você mora?	Q que +
c. <i>Interrogativa é-que</i> :	Onde é que você mora?	Q é que +
d. <i>Interrogativa qu-in-situ</i> :	Você mora onde?	+ Q

Em primeiro lugar, é necessário determinar se todas as formas de Interrogativas-Q são semanticamente equivalentes, de modo que possam ser consideradas opcionais. Labov (1972, 1978) propõe que duas ou mais formas sejam consideradas equivalentes quando têm um mesmo valor de verdade ou se referem a um mesmo estado de coisas. Para algumas formas interrogativas em orações encaixadas, que integram asserções, esse critério se aplicaria:

- (1) a. Marco: aí ele pegava um livro... e via **qual** era os problemas que aquele antibiótico ia dar... (M3G-INF)
 b. ele (...) via **qual que** era os problemas que aquele antibiótico ia dar
 c. ele (...) via **qual é que** era os problemas que aquele antibiótico ia dar

Pode-se assumir que, se é verdade que “ele via qual era os problemas (...)” (1-a), também é verdade que “ele via qual que era os problemas (...)” (1-b) e “ele via qual é que era os problemas (...)” (1-c).

No entanto, as Interrogativas-Q não englobam apenas asserções, mas também formas interrogativas que exercem a função de perguntas.¹ Como atribuir um valor de verdade a sentenças que não são asserções? Parece não ser possível pautar-se em uma semântica verifuncional à la Frege.

¹ Esse será, de fato, o caso de todas as ocorrências definidas no Envelope de Variação I (ver seção 3.2, p. 56): pelas restrições do envelope de variação, todos os dados da variável (Posição do constituinte interrogativo) são perguntas.

Propõe-se alternativamente um critério de equivalência semântica com base na pressuposição. Ressalta-se, no entanto, que se adota o conceito de *pressuposição do falante* (Stalnaker, 2002), em vez de um conceito de pressuposição da sintaxe formal.

Na perspectiva da sintaxe formal, todas as perguntas de constituinte podem ser parafraseadas substituindo-se o constituinte interrogativo por um indefinido (Zubizarreta, 1997):

(2)	a.	O que aconteceu?	a'. <i>Algo</i> aconteceu.
	b.	O que o João fez?	b'. O João fez <i>algo</i> .
	c.	Quando ele nasceu?	c'. Ele nasceu <i>em algum momento</i> .
	d.	Por que ele saiu?	d'. Ele saiu <i>por algum motivo</i> .

É certo que se pode assumir que as pressuposições das sentenças em (2) seriam as mesmas, caso fossem enunciadas como *interrogativas qu-que, é-que* ou *qu-in-situ*.

(3)	a.	Quando ele nasceu?
	b.	Quando que ele nasceu?
	c.	Quando é que ele nasceu?
	d.	Ele nasceu quando?
		Pressuposição: Ele nasceu <i>em algum momento</i> .

Entretanto, observe-se a ocorrência abaixo:

- (4) Marina: sabe aquela adolescente que não sabe o que fala... depois se arrepende? falei... “**onde** cê vai arrumar um namorado?” (F2G-INF)

Seguindo o raciocínio estabelecido acima, uma sentença como “onde cê vai arrumar um namorado?” deve pressupor que “ela vai arrumar um namorado *em algum lugar*”. No entanto, ao se deparar com esse trecho, um falante do PB facilmente pode inferir que, na verdade, a informante não quis dizer que “a adolescente iria arrumar um namorado em algum lugar” mas, justamente ao contrário, que ela não o arranjará “em lugar nenhum”. Esse exemplo deixa claro que uma noção formal de pressuposição, que se restringe a proposições isoladas, não dá conta de pressuposições feitas por falantes e ouvintes na

língua em uso.

Contra uma noção estrita de pressuposição, Stalnaker (2002) apresenta o conceito de *pressuposição do falante*:

Pressupor algo é assumi-lo como dado, ou ao menos agir como se fosse dado, como informações de base – como *fundo comum* entre os participantes da conversação. (Stalnaker 2002:701. Grifo do autor. Tradução própria.²)

Esse conceito está intimamente ligado com aquele de *fundo comum*. Clark (1996:93), com base em Stalnaker, define o fundo comum entre duas pessoas como “a soma de seu conhecimento, crenças e suposições mútuas, comuns e conjuntas”,³ dentro de uma perspectiva que entende a língua e os atos comunicativos como ações conjuntas entre falantes.

É interessante notar que, dentro desse conceito de pressuposição – e, intrinsicamente, de fundo comum –, a atitude do falante é também uma atitude social, pois sua definição leva em conta aquilo que o falante supõe ser parte do conhecimento compartilhado entre os participantes da interação comunicativa: só se pressupõe algo se se pressupõe que os demais também pressupõem o mesmo, ou seja, se tal conhecimento faz parte do fundo comum entre os interlocutores. Tal abordagem parece mais compatível com os objetivos da Sociolinguística Variacionista, dado que apresenta uma noção de pressuposição que leva em conta o ato comunicativo e seus participantes e, portanto, a língua em uso.

Assim, na ocorrência (4) acima, tanto a falante quanto sua interlocutora, bem como o analista da língua, reconhecem o sentido pretendido – que a adolescente não iria arrumar um namorado – devido a um fundo comum cultural e, mais especificamente, da interação comunicativa. Assume-se que a mesma ocorrência, caso houvesse sido proferida como uma *interrogativa qu-que* (5-b), *é-que* (5-c) ou *qu-in-situ* (5-d) naquele contexto, pressuporia o mesmo:

² Texto original: “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 the participants in the conversation.”

³ Stalnaker (2002) não apresenta uma definição do conceito de fundo comum, dado que ela presumivelmente seria circular com o conceito de pressuposição do falante: pressuposição é aquilo que é assumido como fundo comum, e fundo comum é aquilo que é pressuposto.

- (5) a. Marina: sabe aquela adolescente que não sabe o que fala... depois se arrepende? falei... “**onde** cê vai arrumar um namorado?” (F2G-INF)
b. (...) falei... “**onde que** cê vai arrumar um namorado?”
c. (...) falei... “**onde é que** cê vai arrumar um namorado?”
d. (...) falei... “cê vai arrumar um namorado **onde**?”
Pressuposição: Você não vai arrumar um namorado em lugar algum.

O critério com base no conceito de pressuposição do falante também se aplica a asserções, como o exemplo (1), aqui rerepresentado em (6):

- (6) a. Marco: aí ele pegava um livro... e via **qual** era os problemas que aquele antibiótico ia dar... (M3G-INF)
b. ele (...) via **qual que** era os problemas que aquele antibiótico ia dar
c. ele (...) via **qual é que** era os problemas que aquele antibiótico ia dar
Pressuposição: Aquele antibiótico ia dar alguns problemas.

Diante de um mesmo “estado de coisas” (Labov, 1978), as sentenças interrogativas têm equivalência semântica na medida em que têm uma mesma pressuposição. Assim, o conceito de pressuposição do falante é empregado como critério de equivalência semântica para todas as Interrogativas de constituinte, sejam perguntas ou asserções.

3.2 Os envelopes de variação

Neste estudo, a possibilidade de realizações alternativas define-se através do exame do comportamento das formas na língua em uso. Uma estrutura é considerada variante a partir do conceito de competência comunicativa (Hymes, 1991 [1979]), ou seja, ela é considerada opcional se de fato for (a) possível; (b) factível; (c) apropriada; e (d) empregada, em um mesmo conjunto de contextos.

Tais atributos são cumulativos, cada qual sendo um subconjunto do grupo anterior: uma forma pode ser possível gramaticalmente, mas não factível; ela pode ser possível e factível, mas não apropriada; ela pode ser possível, factível e apropriada, mas pode não ser, de fato, empregada. Desse modo, o universo de análise se define pelo subconjunto de formas empregadas, que

está contido em todos os demais (ver Figura 3.1).

Figura 3.1 Definição do envelope de variação a partir do conceito de competência comunicativa

3.2.1 Formas possíveis: análise de gramaticalidade

Em testes de gramaticalidade, verifica-se que nem todas as formas de Interrogativas-Q são possíveis em alguns contextos sintáticos.

Pode-se fazer uma primeira divisão entre sentenças com e sem verbos finitos. Sentenças sem verbo finito não permitem o emprego de *interrogativas qu-que* e *interrogativas é-que*, como mostram os exemplos (7) e (8) – com verbo não finito – e (9) e (10) – sem verbo:

- (7)
 - a. **O que** fazer?
 - b. Fazer o quê?
 - c. *O que que fazer?
 - d. *O que é que fazer? (Mito, s/d)
- (8)
 - a. Eduardo: se ele cumpriu com o que ele prometeu... **por que** não ter a reeleição? (M2S-INF)
 - b. não ter a reeleição por quê?
 - c. *por que que não ter a reeleição?
 - d. *por que é que não ter a reeleição?
- (9)
 - a. Clara: taxa do lixo **por quê?** cê já tá pagando... (F3G-INF)
 - b. por que taxa do lixo?
 - c. *por que que taxa do lixo?

- d. *por que é que taxa do lixo?
- (10) a. Aline: **qual** o futuro dessa criança? nenhum... (F1G-INF)
b. *o futuro dessa criança qual?
c. *qual que o futuro dessa criança?
d. *qual é que o futuro dessa criança?

Interrogativas qu-que e *é-que* ocorrem sem verbo finito quando se trata de truncamentos ou hesitações, como em (11) abaixo. Estes casos, no entanto, foram descartados da análise quantitativa, pois não seria possível determinar a possibilidade de sua forma *in situ*, tampouco sua função sintática.

- (11) a. Suzana: a senhora que estudou lá... **que que-** quando a senhora saiu “ah isso poderia ser diferente” (F1S-DOC)
b. Maíra: **como é que** assim... na época deles... que que cê achava da escola pública que eles fizeram? (F1S-DOC)

Dentre as formas com verbo finito, pode-se fazer uma divisão entre as interrogativas com constituinte pré-verbal (*-qu*, *qu-que* e *é-que*) e aquelas com constituinte pós-verbal (*qu-in-situ*). É claro que *qu-in-situ*, em princípio, significa que o constituinte interrogativo se encontra na posição de sua função sintática, que não é necessariamente pós-verbal. No entanto, o contexto em que o constituinte interrogativo é ao mesmo tempo pré-verbal e *in situ* corresponde aos casos em que ele exerce a função sintática de sujeito. Nesse contexto, há uma ambiguidade estrutural entre *qu-in-situ* e *-qu*,⁴ trata-se um contexto de neutralização em que não é possível determinar que forma interrogativa foi produzida pelo falante, como mostram os exemplos abaixo:

- (12) a. Claudia: **o que** mais marcou na minha infância?... (F1G-INF)
b. Jefferson: e **quem** paga essa conta? (M2G-INF)

Desse modo, ocorrências como (12) foram descartadas da análise quantitativa e *interrogativas qu-in-situ* acabam se definindo pela posição pós-verbal do constituinte interrogativo.

Em orações encaixadas sem complementizador (demarcadas por [] nos

⁴ A ambiguidade não ocorre em casos de *interrogativas qu-que* e *é-que*, pois a presença do complementizador garante que o constituinte interrogativo não está *in situ*.

exemplos), o constituinte interrogativo não pode ocorrer *in situ*, diferentemente das outras três formas:

- (13) a. Renata: você já sabe [**em quem** (que/ é que) vai votar nas próximas eleições]? (F2S-DOC)
 b. *você já sabe [vai votar em quem nas próximas eleições]?

A alternância é possível quando a oração encaixada é introduzida por um complementizador *que*:

- (14) a. Suzana: e **por que** (que/ é que) a senhora acha [que o público é tão diferente]? (F1S-DOC)
 b. e a senhora acha [que o público é tão diferente **por quê**]?⁵

Em contextos de ilhas sintáticas, ao contrário, apenas a forma *in situ* é possível, enquanto as formas com constituinte pré-verbal não são gramaticais. Tal é caso de constituintes interrogativos dentro de orações relativas (15) ou orações subordinadas finais (16):

- (15) a. Zélia: então é uma teoria [que vai lidar **com o quê**]? com o holístico... com o todo... (F3S-INF)
 b. *então é uma teoria [que com o que vai lidar]?
 c. *então com o que é uma teoria [que vai lidar]?
 (16) a. Zélia: usei exemplo clássico né [pra mostrar **o quê**]? (F3S-INF)
 b. *usei exemplo clássico [pra o que mostrar]?
 c. *o que usei exemplo clássico [pra mostrar]?

Por fim, em casos de sentenças com mais de um constituinte interrogativo argumento do verbo, cada qual permanece *in situ*:

- (17) a. Joseane: **quem** vai trocar **com quem**? (F1S-DOC)
 b. *Quem com quem vai trocar?
 c. *Com quem quem vai trocar?⁶

⁵ A alternância também parece possível se a oração encaixada é introduzida pelo complementizador *se*, embora não ocorra nos dados: “Ele não sabia se ia aonde?”; “Aonde (que/ é que) ele não sabia se ia?”.

⁶ Se houver mais de um constituinte interrogativo e pelo menos um deles não for argumento do verbo, a alternância parece ser possível:
 (i) O que você fez quando?
 (ii) Você fez o quê quando?
 mas: (iii) ??? O que quando você fez?
 (iv) ??? Quando o que você fez?.

Os exemplos expostos acima mostram, então, contextos em que nem todas as formas interrogativas são possíveis. Isso é resumido pela Tabela 3.2, que elenca os contextos sintáticos em que cada forma de Interrogativa-Q pode ocorrer.

Tabela 3.2 Possibilidades de estruturação sintática das Interrogativas-Q

Contextos quanto à estrutura da oração	Interrogativas-Q			
	<i>-qu</i>	<i>qu-que</i>	<i>é-que</i>	<i>qu-in-situ</i>
A principal	✓	✓	✓	✓
B com verbo finito	✓	✓	✓	✓
C encaixada com complementizador	✓	✓	✓	✓
D encaixada sem complementizador	✓	✓	✓	X
E sem verbo finito	✓	X	X	✓
F ilha sintática	X	X	X	✓

Três contextos sintáticos, representados nas linhas A-C acima, permitem o emprego das quatro estruturas de Interrogativas-Q: orações principais, orações com verbo finito e orações encaixadas com complementizador; o contexto D, orações encaixadas sem complementizador, permite a alternância entre as três formas com constituinte interrogativo pré-verbal, mas não permite *qu-in-situ*; o contexto E, orações sem verbo finito, engloba apenas *interrogativas-qu* e *qu-in-situ*; o contexto F, por fim, não permite variação.

Desse modo, a partir do critério de gramaticalidade, é possível definir mais de um conjunto de contextos em que pelo menos duas formas se alternam. Dois desses conjuntos são mais representativos do fenômeno variável sob análise: (i) aquele que engloba os contextos A-C; e (ii) aquele que engloba os contextos A-D. No primeiro caso, as quatro formas de Interrogativas-Q podem ser reorganizadas de acordo com a *posição do constituinte interrogativo na sentença*: pré-verbal (*-qu*, *qu-que* e *é-que*) ou pós-verbal (*in-situ*).

No segundo caso, com a inclusão do contexto D, devem ser excluídas as ocorrências de *interrogativas qu-in-situ*, visto que elas não são possíveis em

(v) ??? Quando você fez o quê?

De todo modo, não houve casos desse tipo no *corpus* analisado.

orações encaixadas sem complementizador – contexto bastante produtivo para as três formas interrogativas com constituinte pré-verbal, como se evidenciará adiante (ver Tabela 3.3, p. 67).

A possibilidade de alternância sintática, no entanto, não garante a opcionalidade na língua em uso. A seguir, discutem-se a viabilidade, a adequação pragmática e a produtividade das Interrogativas-Q.

3.2.2 Formas factíveis: análise de viabilidade

Certas formas linguísticas, ainda que sejam gramaticais, não são factualmente possíveis. A aplicação sem fim de uma regra recursiva como o encaixamento de orações relativas (“O menino que viu a vizinha que comprou um carro que foi roubado pelo homem que...”) é um bom exemplo: ainda que se saiba que tal sentença pode continuar indefinidamente, nenhum falante o faz. Hymes (1991 [1979]: 22) informa que a questão da viabilidade está relacionada com fatores psicolinguísticos tais como a memória e a saliência.

Para as Interrogativas-Q, a extensão dos enunciados parece ser mais restrita para as *interrogativas qu-in-situ*. Observe-se a ocorrência (18-a), formulada como uma *interrogativa-qu*:

- (18) a. Cecília: **por que** você vai ser contra a transformação de uma língua que é algo tão... tão... enraizado no indivíduo né?... (F2S-INF)
b. Você vai ser contra a transformação de uma língua que é algo tão enraizado no indivíduo por quê?

A forma equivalente (18-b) é perfeitamente gramatical, mas pode-se prever que, quanto mais longa a sentença, mais inviável ela se torna com a estrutura de uma *interrogativa qu-in-situ*.

A definição de um “limite” no número de palavras ou constituintes da sentença não deixa de ser arbitrária pois, como se apontou, não se trata de uma questão de agramaticalidade. Neste estudo, toma-se como critério de extensão a sentença mais longa encontrada no *corpus* com a estrutura de uma *interrogativa qu-in-situ* (19), que conta com 20 palavras:

- (19) Frederico: eles iriam realizar greve etc. e tal pressionaram o governo

de uma certa forma com relação à reforma previdenciária **por quê?**
(M3S-INF)

3.2.3 Formas adequadas: análise pragmática

Diversos trabalhos (Lopes-Rossi, 1993, 1996; Ambar *et al.*, 2001; Mito & Kato, 2005) notam maior produtividade da construção *in situ* no PB em relação ao PE; esse fato se explicaria por uma restrição pragmática ao emprego dessa forma interrogativa na variedade europeia, em que ela pode ocorrer somente em situações de pergunta-eco, ao passo que tal restrição não existe no PB.

Entretanto, Pires & Taylor (2007) questionam a livre opcionalidade da construção *in situ* no PB e argumentam que haveria, sim, restrições semânticas e pragmáticas para o emprego dessa estrutura: elas só poderiam ocorrer em contextos em que a pressuposição da interrogativa já faz parte do fundo comum (Stalnaker, 1978) entre os interlocutores. Desse modo, segundo os autores, as *interrogativas qu-in-situ* ocorreriam no PB somente nas seguintes situações:

- (20) perguntas [+específicas]:
A: Eu fiz sobremesa.
B: Você fez **que tipo de sobremesa?**

- (21) perguntas esperadas, como em um interrogatório judiciário:
Advogado: Você pode me dizer o que aconteceu no dia 1º de janeiro de 2005, às 4 da tarde?
Réu: Eu estava dirigindo na Avenida dos Andradas.
Advogado: E você estava dirigindo **em que direção?**
Réu: Eu estava indo para o sul, na direção da biblioteca.
Advogado: E o policial disse que você estava dirigindo **a que velocidade?**

- (22) perguntas de referência, que buscam a repetição ou paráfrase de um antecedente imediato:
A: Eu não vendi aquelas pinturas estranhas.
B: Você não vendeu **que pinturas estranhas?**

- (23) em que certas características do contexto extralinguístico satisfazem o requisito de fundo comum:
B vê seu amigo A lendo algo (informação compartilhada extralinguística):

B: Você (es)tá lendo **o quê**?

(Exemplos extraídos de Pires & Taylor 2007:3-4. Grifos acrescentados.)

Perguntas “de sopetão”, em que a pressuposição não fazia parte do fundo comum, seriam consideradas pragmaticamente infelizes (#) se construídas com o constituinte interrogativo *in situ* (24)-(25-a). No entanto, os autores afirmam que as mesmas perguntas seriam adequadas se fossem construídas com constituinte movido, como em (25-b):

- (24) Você aborda um colega de trabalho e pergunta, de sopetão:
B: # Você conhece **quem** em São Paulo?
- (25) Você está conversando com um estudante que disse ter interesse em viajar (mas não disse se já viajou para qualquer lugar). Você pergunta:
- a. B: # Você (já) visitou **que/quais países europeus**?
 - b. B: **Que/quais países europeus** você (já) visitou?

(Exemplos extraídos de Pires & Taylor 2007:6-7. Grifos acrescentados.)

Pires & Taylor (2007) fazem a ressalva de que suas intuições se referem a um “dialeto mais restritivo no PB”, embora não informem qual. Para alguns falantes do PB, no entanto, ambas as construções (25-a) e (25-b) são um tanto inadequadas, pelo fato de a pressuposição da interrogativa não fazer parte do fundo comum entre os interlocutores.⁷ De um modo ou de outro, esses exemplos demonstram como as intuições podem ser falíveis (Labov, 1996); nem sempre se pode confiar em julgamentos dos falantes, sobretudo quando se trata de avaliações de adequação discursivo-pragmática.

Nesta pesquisa, decidiu-se testar a hipótese de Pires & Taylor (2007) sobre a restrição pragmática das *interrogativas qu-in-situ* através de um grupo de fatores, partindo-se da premissa de que, se essa regra for verdadeira, é possível que ela se configure como uma regra variável. A proposta era a de classificar as ocorrências de interrogativas de acordo com dois tipos de contexto: (a) sem pressuposição, correspondente às perguntas “de sopetão”; (b) com pressuposição. Todavia, já a análise qualitativa demonstrou a impossibilidade

⁷ Fez-se uma consulta informal com alguns falantes paulistanos, do interior do estado de São Paulo e do estado de Minas Gerais.

de testar essa hipótese quantitativamente: não houve nenhuma ocorrência de pergunta “de sopetão” – nem no *corpus* de língua oral, nem no de língua escrita.

Por um lado, é possível que esse fato decorra da natureza dos *corpora*, ou seja, que a própria situação de entrevista sociolinguística nos moldes labovianos ou textos de revista, jornal e redações escolares não sejam propícios à ocorrência de perguntas “de sopetão”. No entanto, é difícil imaginar contextos em que, de fato, não há qualquer grau de fundo comum: como vivemos em sociedade, há muitas crenças mútuas que os falantes podem projetar sobre seus interlocutores pelo simples fato de pertencerem a uma determinada comunidade, serem de determinado sexo/gênero, idade, classe social. Da perspectiva da língua em uso por uma comunidade real, é mais adequado considerar que sempre há algum grau de compartilhamento de crenças e informações entre seus membros (Clark, 1996).

Desse modo, a observação feita por Pires & Taylor (2007) não é tratada aqui como uma restrição categórica – e, portanto, não tem consequências restritivas para o envelope de variação. A questão do fundo comum foi reformulada em grupos de fatores (Grau de relação entre informante e documentador, Conjunto de respostas previstas e Grau de ativação do fundo comum), que serão apresentados na seção 3.3.

3.2.4 Formas empregadas: análise de produção

A análise qualitativa também verificou se certas formas, ainda que gramaticais, factíveis e pragmaticamente adequadas, de fato ocorrem no *corpus* analisado ou, ao contrário, se ocorrem com muita frequência. Na literatura sociolinguística, um tipo similar de análise tem sido chamado de questão *type-token* (Tagliamonte, 2006:95): a inclusão de todas as ocorrências de um tipo lexical frequente pode enviesar os resultados da análise quantitativa. Por exemplo, em estudos sobre o apagamento de /-t, -d/ no inglês (Cf. p.ex. Wolfram 1993), costuma-se excluir ou limitar o número de ocorrências de itens lexicais como *and* ‘e’, *n’t* (morfema de negação), *just* ‘somente’ e *kind of* ‘mais ou menos, tipo’, em que /-t, -d/ muito frequentemente não são

realizados.

No caso das interrogativas de constituinte, há certas expressões bastante cristalizadas que, ainda que possam apresentar variação, dificilmente ocorrem com outra forma interrogativa no *corpus*. É o caso das expressões grifadas abaixo, que foram, portanto, desconsideradas na análise quantitativa:

- (26)
- a. Valquíria: mas não é? **onde já se viu...** cê pagar cinco reais num pacotinho de feijão! o arroz nove dez... (F3G-INF)
 - b. Gisele: não aprovo essas coisa não... **quem somos nós** pra fazer alguma coisa? nada... pra fazer alguma coisa só Deus (F2G-INF)
 - c. Lucia: bom **quem nasceu primeiro o ovo ou a galinha?** (F3S-INF)⁸
 - d. Aline: tá mas na próxima Olimpíadas... **quem sabe?** (F1G-INF)

De modo semelhante, não se incluíram certos tipos de dados que ocorrem com muita frequência com a mesma estrutura. Embora se observe variação, as construções abaixo tendem a ser empregadas com uma mesma estrutura na fala de um mesmo informante e, portanto, podem ser consideradas formas semilexicalizadas ou fórmulas discursivas:

- (27)
- a. Beatriz: e a lei mudou:: então **o que aconteceu?** eu fui visitar o Fort Orange... (F3S-INF)
 - b. Gustavo: hoje eu até eu penso um pouco diferente mas **vou fazer o quê?** eu não gostava... **vou fazer o quê?** (M1G-INF)
 - c. Gisele: me lembro é do meu avô... chamava minha avó de... **como que chama?** ele falava uma língua lá... (F2G-INF)
 - d. Karina: ah e **o que que tem?** ah não tem nada demais (F1S-DOC)
 - e. Silvia: o meu marido vai lá e fala “**não que que é isso** meu filho! sei lá o quê”... (F3S-INF)

Os exemplos acima concernem principalmente à variação entre as três estruturas com constituinte pré-verbal *vs.* a construção *in situ*. Das 41 ocorrências de enunciados do tipo (27-a), apenas uma foi de *interrogativa qu-in-situ* (“Aconteceu o quê?”); das 17 do tipo (27-b), ao contrário, apenas uma teve constituinte pré-verbal (“O que eu vou fazer?”);⁹ todos os 101 enunciados

⁸ As expressões em (26-c) e (26-d) são excluídas da análise também pelo fato de o constituinte interrogativo ter a função de sujeito (ver subseção 3.1.1).

⁹ Foram excluídas apenas as ocorrências em que a expressão “(vai/vou) fazer o quê” era uma pergunta retórica, com o sentido de “não há nada a se fazer”. Apenas nestes casos a expressão parece estar

do tipo (27-c) (incluindo variações como “como fala?”, “como se diz?” etc.) ocorreram com constituinte pré-verbal, assim como as 9 ocorrências do tipo (27-d) e (27-e). Houve, portanto, 168 ocorrências semelhantes àsquelas em (27) que foram excluídas da análise quantitativa.

Embora os casos em (27) possam apresentar variação entre as três formas com constituinte pré-verbal, houve pouca variação na fala de cada informante – por exemplo, um dos falantes tende a falar apenas “como que chama?”, enquanto outro emprega apenas “como se diz?”. Esse fato evidencia o caráter de fórmula discursiva dessas construções.

Além de examinar distribuições quase categóricas, como é o caso de expressões cristalizadas ou semilexicalizadas apresentadas acima, a análise de produção também evidencia a relevância dos contextos sintáticos que permitem a variação entre as estruturas. A Tabela 3.2 (p. 60), apresentada na discussão sobre o critério de gramaticalidade, esboçou dois conjuntos de contextos que permitem a variação entre formas interrogativas de constituinte. No segundo deles, que engloba os contextos A-D daquela tabela, argumentou-se que as *interrogativas qu-in-situ* deveriam ser excluídas em favor da inclusão do contexto em que o constituinte interrogativo ocorre em orações encaixadas sem complementizador (em vez de, por exemplo, considerar apenas um conjunto de dados em que as quatro formas são possíveis). Esse argumento agora se complementa com a apresentação da produtividade da construção interrogativa em orações encaixadas sem complementizador em relação ao conjunto total de dados (ver Tabela 3.3): esse contexto representa 649 (39,4%) do total de 1648 ocorrências de Interrogativas-Q. Isso leva a crer que de fato há um envelope de variação que congrega apenas as três formas com constituinte interrogativo pré-verbal.

A Tabela 3.3 também permite melhor visualização dos contextos em que os dois conjuntos de dados se interseccionam: aqueles da linha “principais e encaixadas com complementizador”, que totalizam 999 ocorrências, correspondem aos contextos A-C da Tabela 3.2; aqueles da coluna “Q pré-verbal”, ou 1421 dados, correspondem às interrogativas com constituinte pré-verbal dos contextos A-D.

semilexicalizada na forma *qu-in-situ*.

Tabela 3.3 Distribuição geral de ocorrências de Interrogativas-Q no *corpus* de entrevistas sociolinguísticas de acordo com os contextos sintáticos relevantes

Orações	Q pré-verbal -qu, qu-que, é-que	Q pós-verbal qu-in-situ	Total
principais/encaixadas (com compl.)	772	227	999
encaixadas (sem compl.)	649	N/A	649
Total	1421	227	1648

Os critérios de gramaticalidade, viabilidade, adequação pragmática e produtividade podem remeter à crítica de Lavandera (1978) a respeito dos envelopes de variação “penosamente definidos” (Labov, 1978: 10) nos casos de variáveis além do nível fonológico. No entanto, como afirma Labov (1978), é necessário isolar o contexto em que se observa a variação relevante, a fim de garantir a relevância das análises estatísticas.

Desse modo, a partir desses quatro critérios, definem-se dois envelopes de variação, assim formalizados:

Envelope de Variação I
Variável (Posição do constituinte interrogativo)

Sentenças completas com até 20 palavras que contêm apenas um constituinte interrogativo em uma oração finita principal ou em uma oração encaixada introduzida por um complementizador, excetuando-se expressões cristalizadas ou semilexicalizadas e os casos de *interrogativas-qu/qu-in-situ* em que o constituinte interrogativo exerce a função sintática de sujeito.

Envelope de Variação II
Variável (Constituinte interrogativo pré-verbal)

Sentenças completas que contêm apenas um constituinte interrogativo pré-verbal em uma oração finita principal ou encaixada, excetuando-se expressões cristalizadas ou semilexicalizadas.

A definição de dois envelopes de variação para as Interrogativas-Q implica que se postula a existência de duas variáveis dependentes: uma delas, (Posição

do constituinte interrogativo), é de natureza sintática; a outra, (Constituinte interrogativo pré-verbal), é de natureza morfossintática. Nos estudos sociolinguísticos, que se iniciaram pelo estudo de variáveis fonéticas, Labov (2006 [1966], 1972) propôs a notação das variáveis dependentes entre parênteses, a fim de diferenciá-las de fones (representados entre colchetes []) e fonemas (representados entre barras / /). Seguindo a convenção variacionista, adota-se aqui a notação das variáveis entre parênteses, que já se estendeu a outros níveis de análise.

No Envelope de Variação I, conforme o próprio nome da variável leva a entender, a construção *in situ* naturalmente se opõe às outras três, em que o constituinte interrogativo é pré-verbal. Na análise dessa variável, portanto, a pergunta mais geral que se coloca é: nos contextos em que as quatro formas se alternam, o que favorece a ocorrência de *interrogativas qu-in-situ*?

Já no caso do Envelope de Variação II, o conjunto de dados se define pelas três estruturas com constituinte interrogativo pré-verbal (-*qu*, *qu-que* e *é-que*); desse modo, a questão geral aqui é diferente da anterior. Neste recorte, a análise é feita a partir da perspectiva das *interrogativas qu-que*, com vistas a mapear uma possível mudança linguística em progresso no emprego dessa estrutura: nos contextos em que as três formas com constituinte interrogativo pré-verbal se alternam, o que favorece a ocorrência de *interrogativas qu-que*?

3.3 Grupos de fatores

Para responder às perguntas gerais de natureza variacionista, são levantadas hipóteses, que se estabelecem sob a forma de grupos de fatores. Alguns deles são relevantes para ambas as análises propostas, enquanto outros referem-se apenas à análise da variável (Posição do constituinte interrogativo) (Envelope de Variação I) ou à análise da variável (Constituinte interrogativo pré-verbal) (Envelope de Variação II). Há ainda alguns grupos de fatores que não se aplicam a todos os dados, mas a um subconjunto de ocorrências de um envelope de variação; tais casos serão esclarecidos oportunamente.

3.3.1 Grupos de fatores extralinguísticos

As hipóteses de natureza extralinguística incluem, além das variáveis usadas para a composição da amostra (Sexo/gênero, Faixa etária e Escolaridade – ver seção 2.1, p. 29), outros três grupos de fatores que podem ter uma influência no emprego variável de Interrogativas-Q: Quem fala, Grau de relação entre documentador e informante, e Espontaneidade da pergunta (ver Tabela 3.4).

Tabela 3.4 Grupos de fatores extralinguísticos (Envelopes de Variação I e II)

Variável	Variantes
Sexo	masculino feminino
Faixa etária	de 20 a 30 anos de 35 a 45 anos mais de 50 anos
Escolaridade	até Ensino Fundamental II curso superior
Quem fala	documentador informante
Grau de relação entre documentador e informante	(1) Bastante próximo. O entrevistado fazia parte do meu círculo imediato de amigos/familiares e conversávamos frequentemente. (2) Próximo. Conversávamos frequentemente, mas o entrevistado não fazia parte de meu círculo imediato de amigos/familiares. (3) Próximo, mas não conversávamos frequentemente. (4) Neutro. Ele era meu conhecido, mas não nos falávamos com frequência. (5) Distante. Não o conhecia anteriormente e só conversamos na ocasião da entrevista. (?) informação não disponível
Espontaneidade da pergunta ^a	obrigatória sugerida temática de dados pessoais espontânea (N/A não é dado de documentador)

^a Aplica-se somente a dados de documentadores.

Quem fala: Como esta pesquisa também analisa os dados da fala de documentadores, é importante que a variável “Quem fala”, i.e. documentador ou informante, seja controlada, com vistas a verificar se há diferenças significativas no emprego de formas interrogativas a depender do papel desempenhado pelo falante na situação de entrevista sociolinguística. Observa-se que esse grupo de fatores não é totalmente independente de outros grupos, do modo como se desejaria em uma análise de covariação: 16 dos 19 dos documentadores pertencem à primeira faixa etária, dado que se trata de alunos universitários que coletaram as entrevistas durante o curso de Sociolinguística; ao mesmo tempo, todos os documentadores têm nível superior de escolaridade. Alguns cuidados são tomados, portanto, na análise quantitativa para o teste dessa hipótese.¹⁰

Grau de relação entre documentador e informante: Este grupo de fatores tem o objetivo de testar indiretamente a hipótese de Pires & Taylor (2007) a respeito das restrições no emprego de *interrogativas qu-in-situ* em relação ao fundo comum entre os interlocutores: quanto mais próximo o grau de relação entre os interlocutores, é de se esperar que o conjunto de informações e crenças compartilhadas entre eles também seja maior. Haveria, então, maior tendência de emprego de *interrogativas qu-in-situ* quanto mais próximo o grau de relação entre os falantes? Para testar essa hipótese, o formulário enviado aos documentadores incluiu uma pergunta sobre a relação que ele tinha com o informante à época da entrevista: o documentador deveria assinalar uma de cinco opções em um contínuo de proximidade-distância de relação (ver Tabela 3.4 e Anexo A, p. 149); assumiu-se que o grau assinalado pelo documentador era o mesmo para o seu informante. Para dados de entrevistas feitas por alunos que não responderam ao questionário, criou-se um sexto fator, “informação não disponível”. Esse grupo de fatores também foi codificado para os dados da variável (Constituinte interrogativo pré-verbal), embora não houvesse uma expectativa inicial de correlação.

¹⁰ Na descrição de cada grupo de fatores, serão sinalizados os casos em que o grupo não é totalmente independente de outros. Na seção 3.3.3 adiante, esses casos serão reunidos, explicitando-se os motivos das interações e as soluções empregadas para a análise desses grupos.

Espontaneidade da pergunta: Este grupo de fatores aplica-se ao subconjunto de dados de perguntas de constituinte produzidas pelos documentadores – há, portanto, interação com o grupo de fatores “Quem fala” e “Sinceridade pragmática da pergunta” (ver p. 79), de modo que esses grupos de fatores não foram analisados nas mesmas rodadas do GoldVarb X. Dados dos informantes foram codificados como “N/A não se aplica” e foram retirados da análise dessa variável. Considerando-se que a fala do documentador é ensaiada ou menos espontânea pelo fato de haver um roteiro de entrevista, esse grupo de fatores investiga se as perguntas de constituinte do roteiro, todas formuladas como *interrogativas-qu* (ver Anexo C, p. 154), têm uma tendência menor a apresentar variação.¹¹ Em outras palavras, em uma fala claramente monitorada, as *interrogativas qu-in-situ* e *interrogativas qu-que* seriam menos empregadas? No roteiro, havia um grupo de perguntas sobre percepções linguísticas, chamadas de *obrigatórias*, que deveriam ser feitas compulsoriamente ao fim da entrevista e em uma determinada ordem, sem alterações (28-a); outras partes da entrevista foram divididas em temas, com algumas perguntas *sugeridas*, mas não obrigatórias (28-b); outras perguntas feitas pelos documentadores podem ser consideradas *temáticas*, já que se enquadram em temas arrolados no roteiro (28-c); além dessas perguntas, os alunos deveriam preencher uma ficha com *dados pessoais* do informante (nome completo, idade, data de nascimento etc. – ver Anexo C, p. 160), que poderiam ser obtidos ou não durante a gravação (28-d); por fim, há perguntas que surgem da própria interação, a partir de informações dadas pelos informantes durante a entrevista e que, assume-se, não foram formuladas previamente (28-e). Assim, esses fatores se organizam em um contínuo de espontaneidade, desde as perguntas obrigatórias, para as quais os alunos foram explicitamente instruídos a seguir tal e qual, até perguntas que não eram previstas no roteiro. Alguns exemplos de cada tipo de pergunta são fornecidos abaixo:

- (28) Espontaneidade da pergunta
a. obrigatória

¹¹ Ver, p.ex., perguntas em (5b) e (6c) do roteiro e perguntas 5, 6, 13, 19 e 21 da lista de Perguntas Obrigatórias.

- (i) Ligia: tá... e quando você tá conversando com alguém... **qual que** você fala mais... “nós” ou “a gente”? (F1S-DOC)
- (ii) Carlos: entendi... é... quando você está falando sobre você junto com mais alguém... você usa mais **que palavra** “nós”... ou “a gente”? (M3S-DOC)
- b. sugerida
 - (i) Helena: pensando em São Paulo... na cidade... **qual que** é o maior problema que cê acha que é [hes.] tem... a cidade de São Paulo? segurança... roubo trânsito? (F1S-DOC)
 - (ii) Amanda: cê pensa que tem que ser mais importante **o quê?** segurança transporte educação... (F1S-DOC)
- c. temática
 - (i) Mariana: e **o que que** você acha que seria uma solução... pro transporte? (F1S-DOC)
 - (ii) Rafael: aqui perto tem **quantas escolas** aqui... na região... assim? (M1S-DOC)
- d. de dados pessoais
 - (i) Renata: **que escola** você estudou? **que faculdade** você fez? (F1S-DOC)
 - (ii) Rafael: e:: o senhor estudou **em que escolas?** (M1S-DOC)
- e. espontânea
 - (i) Fernanda: e **como que** é esse curso? explica um pouquinho (F1S-DOC)
 - (ii) Regina: você aluga mais ou menos **quantos filmes** assim por... por mês? (F1S-DOC)

3.3.2 Grupos de fatores linguísticos

Os grupos de fatores linguísticos (ver Tabela 3.5) encerram hipóteses de natureza prosódica (Número de palavras na oração, Tonicidade da sílaba seguinte), morfológica (Constituinte interrogativo), sintática (Tipo de verbo, Função sintática, Tipo de Oração (1) e (2), Que elemento ocorre antes da oração principal, Ordem SV-VS) e discursivo-pragmática (Sinceridade pragmática da pergunta, Conjunto de respostas “previstas”, Grau de ativação do fundo comum).

Constituinte interrogativo: Todas as ocorrências de Interrogativas-Q foram codificadas de acordo com o constituinte interrogativo. Inicialmente, foi feita uma classificação bastante detalhada. *Que* foi separado de *o que* e de

Tabela 3.5 Grupos de fatores linguísticos (Envelopes de Variação I e II)

Variável	Variantes
Constituinte interrogativo	que o que que + NP qual(is) qual(is) + NP quanto(-a, -os, -as) quanto(-a, -os, -as) + NP quem como quando onde por que pra que quão
Função sintática	constituente de <i>small clause</i> sujeito objeto direto objeto indireto objeto de preposição adjunto adverbial de modo adjunto adverbial de tempo adjunto adverbial de causa adjunto adverbial de lugar adjunto adverbial de finalidade
Tipo de verbo	ligação (ser, estar, parecer, ficar) modal (poder, dever, ter que) intransitivo transitivo direto transitivo indireto bitransitivo

que + NP por dois motivos: por um lado, a análise qualitativa dos dados indicou que a forma *que* parecia ser mais empregada em construções *qu-que* (i.e. era mais frequente uma estrutura como “Que que ele comprou?” do que “O que que ele comprou?”); por outro lado, separar os casos de *que* e de *que + NP* justifica-se na medida em que, no PE, cada estrutura admite diferentes sequências entre sujeito e verbo – a primeira requer a inversão do verbo em perguntas (*que V S*), enquanto a segunda admite a não-inversão (*que + NP S V*) (Ambar *et al.*, 2001), de modo que é plausível aventar a hipótese de que no PB tais estruturas também tenham comportamentos distintos. Assim, de modo semelhante, foram codificados separadamente os constituintes *qual* e *qual + NP*, *quanto(-a,-os,-as)* e *quanto(-a,-os,-as) + NP*. Ainda é necessário apontar que, embora se trate de um grupo de fatores diferente de Função sintática (p.ex., *o que* pode ser sujeito, objeto direto, objeto indireto...), há pouca ortogonalidade entre as duas variáveis (p.ex., *por que* sempre coincide com adjunto adverbial de causa).

Função sintática: De forma semelhante ao grupo de fatores Constituinte interrogativo, Função sintática foi inicialmente codificado de forma bastante detalhada, separando-se diferentes tipos de argumentos verbais e adjuntos adverbiais. Esse grupo de fatores objetiva testar a hipótese de que a “distância” do movimento do constituinte ao início da oração, relativa à sua posição canônica na sentença, pode estar correlacionada ao emprego de *interrogativas qu-in-situ* e *qu-que*. Esse grupo de fatores interage com Constituinte interrogativo, assim como Tipo de verbo (ver Tipo de verbo, p. 76).

- (29) Função sintática
- a. constituinte de *small clause*
 - (i) Kenji: eu não tenho essa informação ... **que- qual seria a fonte?**... agora eu fiquei curioso ... (M3S-INF)
 - (ii) Carlos: você fez uma diferenciação entre isso e saúde né? **e a saúde seria o quê?** mais ligada a alimentação? (M3S-DOC)
 - b. sujeito
 - (i) Gisele: **que que faltou?** (F2G-INF)
 - (ii) Gustavo: a era Fernando Henrique de novo de jeito nenhum aí **sobra quem?** a Heloísa Helena né? (M1G-INF)

- c. objeto direto
 - (i) Vanessa: porque o::... **o que nós ouvíamos...** há um tempo atrás? viver num país como Estados Unidos é tudo de bom (F2S-INF)
 - (ii) Paula: pra mim continuou a mesma coisa... **eles queriam o quê?** mais vaga no CRUSP? (F1S-INF)
- d. objeto indireto
 - (i) Marcio: **qual sotaque no Brasil... você mais gosta?**... o gaúcho?... o carioca?... o paulistano?... ou o nordestino? (M1S-DOC)
 - (ii) Maíra: agora fiquei curiosa **essas escolas equivaliam a quê?** equivaliam também a um curso técnico ensino médio? (F1S-DOC)
- e. objeto de preposição
 - (i) Regina: e... **com quantos anos vocês se casaram?** (F1S-DOC)
 - (ii) Carolina: e **ela estudou até que série mais ou menos?** (F1S-DOC)
- f. adjunto adverbial de modo
 - (i) Jorge: e na hora de falar assim... você acha que em São Paulo **como que o pessoal pronuncia o “r”?** (M1S-DOC)
 - (ii) Eduardo: o que que eu vou fazer? **vou intensificar de que maneira?** mais exercícios menos tempo de recuperação versão mais rápida mais carga tudo isso (M2S-INF)
- g. adjunto adverbial de tempo
 - (i) Amanda: mas **que horas cê acorda** se cê sai seis e... (F1S-DOC)
 - (ii) Carlos: **cê tá há quanto tempo** na Bela Vista? (M3S-DOC)
- h. adjunto adverbial de causa
 - (i) Renata: mas **por que você fez na época turismo?** (F2S-DOC)
 - (ii) Maíra: e **não autorizou por quê?** que que ele disse? (F1S-DOC)
- i. adjunto adverbial de lugar
 - (i) Carlos: e **onde é que se faz moda?** (M3S-DOC)
 - (ii) Valter: tem próximo aqui na avenida Cangaíba é:: ... **você desceu em qual ponto?** (M2G-INF)
- j. adjunto adverbial de finalidade
 - (i) Marta: ele fala “nossa **pra que cê gastou** por que que você comprou...” entendeu? (F2G-INF)
 - (ii) Beatriz: aí eu falei “não mas eu só quero lhe ver” “mas **vai me ver pra quê?** pra me fotografar e mostrar por aí? não”... (F3S-INF)

Tipo de verbo: Este é considerado um grupo de fatores sintático pois objetiva investigar se as diferentes estruturas de sentença, de acordo com os argumentos e os adjuntos do verbo, têm um peso sobre o uso de diferentes formas interrogativas. A diátese de cada verbo é determinada a partir do modo como foi empregado em cada ocorrência; assim, um mesmo verbo como “trabalhar” foi considerado intransitivo em (30-a), transitivo em (30-b) e transitivo indireto em (30-c):

- (30) a. Rafael: e cês trabalhavam **em quantas pessoas?** (M1S-DOC)
 b. Fabiana: o que que vocês vão- né?... **o que que** vocês vão- a gente vai trabalhar... (F1S-INF)
 c. Jorge: é... então... seus irmãos também trabalharam **com o quê?** (M1S-DOC)

Os grupos de fatores na Tabela 3.6 se referem apenas aos dados do Envelope de Variação I, cujas hipóteses focalizam o emprego de *interrogativas qu-in-situ*.

Tipo de oração (1): Dadas as duas possibilidades de estruturação sintática em que podem ocorrer as *interrogativas qu-in-situ* (oração principal ou oração encaixada introduzida por complementizador *que*), fez-se uma separação dos dados nesses dois contextos. Contudo, apesar de se observar variação, houve apenas 20 ocorrências como (31-b), de modo que não foi possível testar essa hipótese.

- (31) Tipo de oração (1)
 a. principal
 (i) Marcel: deu vontade de perguntar “pô você tem tanto dinheiro... **por que** não arruma esses dentes aí?” (M1S-INF)
 (ii) Karina: aí cê tentou **quantas vezes** fazer a quinta? (F1S-DOC)
 b. encaixada introduzida por complementizador
 (i) Maíra: **quem** você acha [que fala mais assim]? empregada doméstica... caixas de banco atendentes de loja operadoras de telemarketing ou seus amigos? (F1S-DOC)

Tabela 3.6 Grupos de fatores linguísticos (Envelope de Variação I)

Variável	Variantes
Tipo de oração (1) ^a	principal encaixada introduzida por complementizador
Número de palavras na oração	2 a 6 palavras 7 a 20 palavras
Que elemento ocorre antes da oração principal	nenhum tópico coindexado com sujeito tópico coindexado com objeto direto tópico coindexado com objeto indireto outros elementos topicalizados oração subordinada
Sinceridade pragmática da pergunta	“sincera” retórica de estruturação do discurso
Conjunto de respostas “previstas” ^b	resposta aberta múltipla escolha com um “chute” (N/A Não é pergunta “sincera”)
Grau de ativação do fundo comum ^c	pressup. ativada na sentença anterior pressup. ativada na 2ª sentença anterior pressup. ativada na 3ª, 4ª ou 5ª sentença anterior pressup. mencionada na conversa, mas não nas últimas 5 sentenças referente ativado em uma das últimas 5 sentenças pressup. não ativada na conversa (N/A pergunta retórica)

^a Não testado devido à má distribuição dos dados. ^b Aplica-se somente a perguntas pragmaticamente “sinceras”. ^c Aplica-se somente a perguntas pragmaticamente “sinceras” e de estruturação do discurso.

- (ii) Fabio: é... você quer [que eu diga **o quê** mais]?... quer que comece a falar delas agora? ((risos))... (M1G-INF)

Número de palavras na oração: Na análise qualitativa, percebeu-se que as *interrogativas qu-in-situ* pareciam ser mais curtas (“Você nasceu onde?”; “Você faz o quê?”). As ocorrências foram codificadas de acordo com o número de palavras na sentença em que se encontra o constituinte interrogativo a partir dos seguintes critérios: (a) a variante interrogativa conta como uma única palavra, i.e. “onde” (-*qu* ou *in situ*), “onde que” ou “onde é que” contam igualmente como uma palavra; (b) expressões lexicalizadas (“a gente”, “o(a) senhor(a)”, “a respeito” etc.) contam como uma palavra; (c) excluem-se marcadores discursivos (“né”, “assim”, “então”, “daí” etc.), hesitações (é:..... ahh...) e vocativos. A questão que se coloca é se as *interrogativas qu-in-situ* teriam uma tendência a serem mais curtas, a fim de facilitar a entonação de pergunta. É possível que a correlação seja a inversa: não seriam as *interrogativas qu-in-situ* que tenderiam a ser mais curtas, mas sim as interrogativas com constituinte pré-verbal que favoreceriam o “alongamento” da sentença. A ambivalência da hipótese em teste, em princípio, não é desejável na análise quantitativa, pois dualiza as possibilidades de interpretação. Decidiu-se, no entanto, testar esse grupo de fatores como um primeiro exame dessas hipóteses.

Que elemento ocorre antes da oração principal: Este grupo de fatores permite investigar se a presença de elementos movidos para o início da sentença (orações subordinadas, tópicos) tem uma influência na estrutura das interrogativas de constituinte. A hipótese concerne à organização da informação no fluxo conversacional, que pode ter motivações discursivo-pragmáticas: quando há outros elementos movidos para o início da sentença, haveria maior tendência de emprego de *interrogativas qu-in-situ*, a fim de “distribuir” constituintes cujas funções se ligam ao discurso? Para codificar os dados de acordo com esse grupo de fatores, foi necessário verificar a entonação de cada sentença para que se diferenciasssem elementos topicalizados de outros constituintes oracionais (sobretudo sujeitos). Decidiu-se também codificar

orações subordinadas condicionais e temporais antepostas, pois sua entonação é semelhante à entonação ascendente de tópico sentencial.

- (32) Que elemento ocorre antes da oração principal
- a. nenhum
 - (i) Iara: ah mas **como você fez segundo grau** se você não sabe essa matéria? essa matéria é de segundo grau (F1G-INF)
 - (ii) Rafael: **o senhor tinha que idade em setenta?** (M1S-DOC)
 - b. tópico coindexado com sujeito
 - (i) Paula: e [os outros professores]_i **por que t_i não ofereciam?** (F1S-INF)
 - (ii) Karina: e [a sua mãe]_k é... **t_k estudou até que:... que série?** (F1S-DOC)
 - c. tópico coindexado com objeto direto
 - (i) Jorge: e [latim]_i **por que que aprendia t_i?** (M1S-DOC)
 - (ii) Helena: uhn... e [seu aniversário]_k... **cê faz aniversário_k quando?** (F1S-DOC)
 - d. tópico coindexado com objeto indireto
 - (i) Maíra: [a educação pública em São Paulo]_i... **que que cê acha dela_i?** como é que tá boa ruim? (F1S-DOC)
 - (ii) Mariana: e [o francês]_k **você se interessou t_k por quê?** (F1S-DOC)
 - e. outros elementos topicalizados
 - (i) Suzana: e [aqui na USP] **o que que a senhora acha da greve?** (F1S-DOC)
 - (ii) Maíra: [São Vicente] **cê morou quanto tempo?** (F1S-DOC)
 - f. oração subordinada
 - (i) Jefferson: [se é uma democracia] **por que que o voto é obrigatório?** (M2G-INF)
 - (ii) João: você:: [se pudesse financeiramente... por numa escola particular]... **você optaria por qual?** pública ou privada? (M2S-DOC)

Sinceridade pragmática da pergunta: Na seção 2.3 (p. 43), observou-se que nem todas as perguntas buscam uma informação de fato. Oushiro & Nasser (2010) propõem uma tipologia de perguntas de acordo com o mecanismo de troca de turnos (Sacks *et al.*, 1974) e a atualização do fundo comum entre os interlocutores (Clark, 1996; Stalnaker, 2002), que foi utilizada na codificação desse grupo de fatores. Perguntas pragmaticamente “sinceras”

são aquelas mais próximas do protótipo do par pergunta-resposta (Schegloff, 1972a; Schegloff & Sacks, 1973; Sacks *et al.*, 1974): elas passam o turno, impõem uma obrigação de resposta no interlocutor e têm a função de atualizar o fundo comum (33-a); perguntas retóricas funcionam como asserções (Rohde, 2006), podem ser respondidas por “é verdade” e não atualizam o fundo comum, dada a obviedade da resposta (33-b);¹² perguntas de estruturação do discurso são aquelas para as quais o falante corrente fornece uma resposta: elas não passam o turno, impõem uma obrigação de “resposta” no próprio falante e têm a função de atualizar o fundo comum (33-c).

Não se trata de uma tipologia de “funções discursivas” das perguntas; no dia-a-dia, as perguntas se prestam a inúmeras funções, cujo inventário pode apenas ser iniciado: saudações (“Tudo bem? Como vai?”), pedidos (“Você pode fechar a porta?”), convites (“Vamos sair pra jantar no sábado?”), reparo (“Como assim? O que você quer dizer? Ahn?”), perguntas de informação (“Que horas você chega amanhã?”), hesitações (“Como se diz?”), charadas (“O que é o que é...?”) etc. Essa tipologia examina as ações tomadas pelos falantes que, a partir de sua competência comunicativa (Hymes, 1991 [1979]), sabem reconhecer os modos adequados de agir diante delas (tomar o próximo turno, ouvir a continuidade da fala de seu interlocutor, fornecer ou não uma resposta).¹³

(33) Sinceridade pragmática da pergunta

- a. perguntas pragmaticamente “sinceras”
 - (i) Carla: **há quanto tempo ele tá lá?** (F1S-DOC)
Pedro: faz... há quatro anos...
 - (ii) Fabiana: **sua fita tem quantos minutos?** (F1S-INF)
Suzana: tem quarenta e cinco- não! noventa
- b. perguntas retóricas
 - (i) José: então a gente tá gastando... com duas eleições... **quanto é que você gasta aí nesses dois turnos? quanto que vai em dinheiro?** (M3S-INF)
Jorge: é verdade...
 - (ii) Clara: eu tenho que ter um carro... dá alguma coisa nela de

¹² Rohde (2006) afirma que perguntas retóricas não podem ser caracterizadas por sua forma sintática; a obviedade da resposta só pode ser determinada a partir de cada situação interacional. Note-se que em (33-b) ambas as perguntas, em outro contexto, poderiam ser consideradas pragmaticamente “sinceras”.

¹³ Cf. Oushiro & Nasser (2010) para uma discussão mais detalhada.

- noite **eu vou aonde?** (F3G-INF)
- c. perguntas de estruturação do discurso
- (i) Cecília: vale a pena... é:: **por que que vale a pena?** porque nós raramente vamos ao cinema durante o ano... (F2S-INF)
 - (ii) Marco: sim a televisão e **a televisão te traz o quê?**... só te traz coisa errada... é **novela... ensina o quê?**... ensina o camarada a beber ensina a prostituição... (M3G-INF)

Conjunto de respostas “previstas”: Este grupo de fatores aplica-se somente às perguntas pragmaticamente “sinceras”. Enquanto algumas perguntas deixam as possibilidades de resposta em aberto (34-a), outras são formuladas já com uma expectativa de resposta, seja com múltiplas opções (34-b) ou apenas uma (34-c). A hipótese aqui levantada tem a ver com o fundo comum (Clark, 1996; Stalnaker, 2002): quanto menor o conjunto de respostas “previstas” pelo falante, i.e. quanto maior a previsibilidade da resposta, pode-se supor que há um maior grau de informações compartilhadas; haveria, então, maior tendência de emprego de *interrogativas qu-in-situ* nesse caso?

- (34) Conjunto de respostas “previstas”
- a. aberto
 - (i) Fernanda: e **em que rua que você mora?** (F1S-DOC)
 - (ii) Carolina: e **você mora em que lugar de São Paulo?** (F1S-DOC)
 - b. múltipla escolha
 - (i) Jefferson: **do que que você tem mais medo... da polícia ou do ladrão?** (M2G-INF)
 - (ii) Amanda: **cê pensa que tem que ser mais importante o quê? segurança transporte educação...** (F1S-DOC)
 - c. com um “chute”
 - (i) Carlos: e **o que que cê faz lá? cê pega onda?** (M3S-DOC)
 - (ii) Jorge: **sua mãe trabalhava com o quê?** assim **ela... era dona-de-casa?** (M1S-DOC)

Grau de ativação do fundo comum: Weiner & Labov (1983 [1977]), em seu estudo sobre a alternância entre voz ativa e voz passiva sem agente no inglês, investigam se a ativação de referentes no discurso têm uma influência no emprego das diferentes estruturas sintáticas. De modo semelhante, este

grupo de fatores tem o objetivo de verificar se pressuposições e referentes recentemente ativados no discurso influenciam o emprego da estrutura com constituinte pré-verbal ou *in situ*. Por tratar de pressuposições, essa hipótese também está relacionada com a atualização do fundo comum (Clark, 1996; Stalnaker, 2002) entre os interlocutores. Diferentemente de Pires & Taylor (2007), a questão aqui não é se a pressuposição faz parte ou não do fundo comum – pois ela sempre faz –, mas quão acessível ela está no discurso. Os fatores exemplificados abaixo organizam-se em diferentes graus de ativação, desde a sentença anterior (máximo grau de ativação) até a não menção até aquele momento da entrevista (mínimo grau de ativação). Nesse grupo de fatores, excluem-se as ocorrências de perguntas retóricas, dado que nelas a pressuposição sempre está ativada; a hipótese aplica-se, portanto, a perguntas pragmaticamente sinceras e de estruturação do discurso. Os números entre parênteses no interior dos exemplos abaixo indicam a contagem de sentenças anteriores.

- (35) Grau de ativação do fundo comum
- a. pressuposição ativada na sentença anterior
 - (i) Pedro: eu acho horrível... (1) acho horrível
Carla: **por que você acha horrível?** (F1S-DOC)
 - (ii) Aline: assim... (1) eles tão falando muito da saúde...
Marcio: é né?
Aline: mas **tão falando da saúde por quê?** (F1G-INF)
 - b. pressuposição ativada na 2ª sentença anterior
 - (i) 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? (F1S-DOC)
 - (ii) Paulo: e (2) 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
Paulo: quando você...
Lucia: mas **é considerado aonde?** (F3G-INF)
 - c. pressuposição ativada na 3ª, 4ª ou 5ª sentença anterior
 - (i) 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?**” (F3S-INF)

- (ii) 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?** (M1S-DOC)
- d. pressuposição mencionada na conversa, mas não nas últimas 5 sentenças
 - (i) Carolina: **aonde que é o colégio que cê dá aula mesmo?** (F1S-DOC)
 - (ii) Rafael: agora **tem que idade mesmo?** (M1S-DOC)
- e. referente ativado em uma das últimas 5 sentenças
 - (i) Carla: eu já tinha esquecido o nome **da sua irmã... quantos anos ela tem?** (F1S-DOC)
 - (ii) Amanda: não sabe ainda mas **ele** gosta de computador... Joaquim: **ele tá com quantos anos?** tá com dezoito ou dez-...? (M3G-INF)
- f. pressuposição não ativada previamente na conversa
 - (i) Carlos: **que que cê faz como lazer assim?** (M3S-DOC)
 - (ii) Marina: **você vem de que canto pra cá?** (F1G-INF)

Por fim, há três grupos de fatores, resumidos na Tabela 3.7, que traduzem hipóteses acerca do emprego das variantes do Envelope de Variação II (análise de *interrogativas qu-que*).

Tipo de oração (2): Diferentemente das restrições no Envelope de Variação I, as estruturas do Envelope de Variação II, além de poderem ser empregadas em orações principais ou encaixadas com complementizador, podem ser empregadas em orações encaixadas sem complementizador (subordinadas a verbos, substantivos e adjetivos) ou, ainda, podem integrar uma oração-constituente de *small clause*. Na análise quantitativa, no entanto, esse grupo de fatores foi redefinido em uma divisão binária, “oração principal” *vs.* “oração subordinada”, devido à má distribuição dos dados pelos fatores inicialmente previstos.

- (36) Tipo de oração (2)
 - a. principal

Tabela 3.7 Grupos de fatores linguísticos (Envelope de Variação II)

Variável	Variantes
Tipo de oração (2)	principal encaixada introduzida por complementizador subordinada a verbo subordinada a substantivo subordinada a adjetivo oração-constituente de <i>small clause</i>
Ordem SV-VS ^a	sujeito-verbo verbo-sujeito
Tonicidade da sílaba seguinte	tônica átona pausa

^a Não testado devido ao uso quase categórico da estrutura SV ou VS a depender do verbo.

- (i) Marcel: deu vontade de perguntar “pô você tem tanto dinheiro... **por que** não arruma esses dentes aí?” (M1S-INF)
- (ii) Simone: e os jaleco deles é pura bactéria... então **como que** eles pode sair na rua... e atender os paciente com o mesmo jaleco?... é uma coisa assim absurda (F2G-INF)
- (iii) Joaquim: **onde é que** tá? Vou ver se eu acho aqui... (M3G-INF)
- b. encaixada introduzida por complementizador¹⁴
 - (i) Maíra: **quem** você acha [que fala mais assim]? empregada doméstica... caixas de banco atendentes de loja operadoras de telemarketing ou seus amigos? (F1S-DOC)
- c. subordinada a verbo
 - (i) Beatriz: não precisa falar **o que** você é **o que** você faz... (F3S-INF)
 - (ii) Claudia: ela nunca falou pra gente **por que que** ela veio pra cá... (F1G-INF)
 - (iii) Valquíria: mas algum- alguns já tão se formando... quer saber direitinho **como é que** é...
- d. subordinada a substantivo
 - (i) Marta: porque muita gente fala errado... não tem noção de **como** a gente fala errado (F2G-INF)
 - (ii) Iara: francês eu não sei... não tenho nem noção... **como que**

¹⁴ Não houve ocorrências de *interrogativas qu-que* e *é-que* em orações encaixadas introduzidas por complementizador.

- é a língua francesa (F1G-INF)
- (iii) João: tem uma idéia mais de perto... **como é que** acontece as coisas lá e tal... (M2S-DOC)
- e. subordinada a adjetivo¹⁵
 - (i) Jefferson: ela nunca tá segura **do que** ela tá falando... (M2G-INF)
- f. oração-constituente de *small clause*¹⁶
 - (i) Fabrício: eu gosto do PROUNE eu acho uma boa oportunidade... o problema é **quanto** o aluno custa... (M1S-INF)
 - (ii) Jefferson: a primeira coisa que os cara te perguntam quando você entra no hospital não é **o que que** você tá sentindo (M2G-INF)

Ordem SV-VS: Alguns estudos dentro do Programa Gerativista explicam o surgimento de *interrogativas qu-que* no PB como decorrência direta ou indireta da perda da ordem verbo-sujeito (VS) em orações interrogativas (Kato, 1987; Duarte, 1992; Kato *et al.*, 1996; Sikansi, 1998; Ambar *et al.*, 2001). Em línguas V2 (p.ex. o alemão) ou de V2 residual (p.ex. o PE), o verbo se moveria para o núcleo de CP devido ao traço [+wh], resultando na inversão do verbo em relação ao sujeito; no PB, a perda da ordem VS deixaria o núcleo de CP “disponível” para ser preenchido, daí a possibilidade de haver um complementizador *que* após o constituinte interrogativo. A partir dessa argumentação, foi proposto um grupo de fatores que pudesse testar tal hipótese, mas já a análise qualitativa demonstrou a sua inviabilidade devido ao emprego quase categórico da ordem SV ou VS a depender do tipo de verbo.¹⁷

Tonicidade da sílaba seguinte: Esse é um grupo de fatores de natureza prosódica. Segundo o Princípio de Alternância Rítmica de Selkirk (1984),

¹⁵ Não houve ocorrências de *interrogativas qu-que* e *é-que* subordinadas a adjetivos.

¹⁶ Não houve ocorrências de *interrogativas é-que* constituintes de *small clause*.

¹⁷ Dois estudos já haviam tentado investigar essa hipótese quantitativamente: Duarte (1992) realizou um levantamento de frequências; Sikansi (1998), por sua vez, realizou testes de covariação com dados sincrônicos de diversos contextos de uso (entrevistas de televisão, redações de vestibular, peças teatrais, fala espontânea, entre outros). A autora aponta que certos tipos de dados foram excluídos da análise – sentenças sem sujeito expreso e aquelas em que o constituinte interrogativo exercia a função de sujeito – e conclui que: (a) os verbos transitivos são inibidores da ordem VS, pois o sujeito, nesse contexto, ocorre quase sempre antes do verbo; (b) os verbos copulares, ergativos e intransitivos tendem a aparecer antes do sujeito. A se pautar por essas conclusões quase categóricas, a autora dá a entender que sua análise incluiu sentenças que não necessariamente permitem a variação entre a ordem SV-VS no PB atual.

todas as línguas naturais tendem à isocronia de sílabas acentuadas, ou seja, tende-se a evitar sequências de sílabas tônicas ou de sílabas átonas e a preferir-se a alternância entre esses dois tipos de sílabas. Esse grupo de fatores investiga se o complementizador *que* das *interrogativas qu-que*, sempre átono, pode funcionar como uma estratégia prosódica para que se evite o choque de acento de duas sílabas tônicas (ver ex. (37-a-ii): por QUE que VAle a PEEna?) Desse modo, os segmentos seguintes ao constituinte interrogativo (grifados nos exemplos abaixo) foram codificados como sílabas tônicas (37-a), átonas (37-b) ou pausa (37-c).

- (37) a. tônica
- (i) Jefferson: **qual FOI a pior cena de violência que você já viu?** (M2G-INF)
 - (ii) Cecília: (...) **por que que VAle a PEEna?** porque nós raramente vamos ao cinema durante o ano... (F2S-INF)
 - (iii) Kenji: **como é que POde um cara casar com uma separada com filhos** não sei o quê? (M3S-INF)
- b. átona
- (i) Beatriz: então **o que me-aconseLHAram?** “a senhora entra... não desça do carro... (F3S-INF)
 - (ii) Carlos: esta história de educação... é... **o que que cê TÁ sentindo assim?** (M3S-DOC)
 - (iii) ah é? **como é que chaMAva?**
- c. pausa
- (i) Paula: eles sabem **onde ... fica o Brasil no mapa** (F1S-INF)
 - (ii) Luis: vamos ver né **que que ... vai melhorar...** (M2G-INF)
 - (iii) Jefferson: e sabe **como é que ... um traficante ganha... um:: avião?**... (M2G-INF)

3.3.3 Interações entre grupos de fatores

Alguns dos grupos de fatores acima estabelecidos não são totalmente independentes entre si, conforme já se indicou anteriormente. Isso representa um problema para a análise multivariada com o programa GoldVarb X, uma vez que esse tipo de análise pressupõe a ortogonalidade dos fatores de cada grupo: cada fator deve “coocorrer livremente e não ser sub- ou supercategorias uns dos outros” (Guy & Zilles, 2007:52). A lista abaixo explicita os grupos de fatores que interagem entre si e os motivos para isso.

- **Quem fala e Sexo/gênero:** A maioria dos documentadores (13 de 19) é do sexo feminino.
- **Quem fala e Faixa etária:** A maioria dos documentadores (16 de 19) pertence à primeira faixa etária.
- **Função sintática e Constituinte interrogativo :** “Por que” corresponde a adjunto adverbial de causa; “quando” corresponde a adjunto adverbial de tempo.
- **Função sintática e Tipo de verbo:** Todos os constituintes de *small clause* ocorrem com verbo de ligação.
- **Sinceridade pragmática da pergunta e Espontaneidade da pergunta:** Todas as perguntas analisadas em Espontaneidade da pergunta são pragmaticamente “sinceras”.
- **Sinceridade pragmática da pergunta e Conjunto de respostas previstas:** Todas as perguntas de Conjunto de respostas previstas são pragmaticamente sinceras.
- **Sinceridade pragmática da pergunta e Grau de ativação do fundo comum:** Todas as perguntas de Grau de ativação do fundo comum são pragmaticamente sinceras ou de estruturação do discurso.
- **Sinceridade pragmática da pergunta e Quem fala:** A maioria das perguntas de documentadores é pragmaticamente sincera.

Em todos esses casos, as interações se devem a características intrínsecas de cada grupo e, portanto, elas dificilmente seriam sanadas através do aumento do número de dados ou da reorganização da amostra. Duas soluções foram empregadas: (i) os pares de grupos de fatores que interagem não foram analisados nas mesmas rodadas, ou seja, realizaram-se duas rodadas distintas, cada qual com um dos grupos; ou (ii) os pares de grupos de fatores foram cruzados em um único grupo – por exemplo, os grupos de fatores Quem fala (documentador ou informante) e Sexo/gênero (masculino

ou feminino) formaram um novo grupo com quatro fatores (documentador do sexo masculino, documentador do sexo feminino, informante do sexo masculino e informante do sexo feminino) – e compararam-se as rodadas com e sem o cruzamento dos grupos de fatores.

À parte dessas interações, há algumas remanescentes que parecem ser acidentais, sobretudo devido à distribuição desequilibrada de dados em alguns grupos com muitos fatores. Nesses casos, realizaram-se amalgamações com base em justificativas tanto teóricas quanto quantitativas (Guy & Zilles, 2007:188). Como nem todas as interações puderam ser resolvidas, elas serão destacadas nos capítulos 4 e 5.

4

A variável (Posição do constituente interrogativo)

O capítulo anterior especificou os contextos em que as quatro estruturas de Interrogativas-Q são factualmente alternantes, de acordo com os quatro critérios baseados no conceito de competência comunicativa. Este capítulo se debruça sobre a variável (Posição do constituente interrogativo), cujas variantes se definem pelo constituente interrogativo pré-verbal ou *in situ*. A análise quantitativa é feita a partir da perspectiva desta última, com o objetivo de evidenciar os contextos que favorecem ou desfavorecem o emprego de *interrogativas qu-in-situ*.

Para tanto, foram testados seis grupos de fatores extralinguísticos (Sexo/gênero, Faixa etária, Escolaridade, Quem fala: documentador ou informante, Grau de relação entre documentador e informante, e Espontaneidade da pergunta) e oito grupos de fatores linguísticos (Número de palavras na oração, Constituinte interrogativo, Tipo de verbo, Função sintática, Que outro elemento ocorre antes da oração principal, Conjunto de respostas previstas, Grau de ativação do fundo comum e Sinceridade pragmática da pergunta). Os grupos de fatores que se revelaram mais significativos para o emprego de *interrogativas qu-in-situ* são aqueles de natureza discursivo-pragmática e morfossintática. Os grupos de fatores extralinguísticos, embora sejam menos significativos estatisticamente, também demonstram correlação; em especial, demonstra-se que Sexo/gênero revela-se significativo quando cruzado com um grupos de fatores discursivo-pragmático, Sinceridade pragmática da pergunta.

4.1 Envelope de variação e Distribuição geral dos dados

A análise quantitativa da variável (Posição do constituinte interrogativo) refere-se ao conjunto de dados compreendidos no Envelope de variação I:

Envelope de Variação I Variável (Posição do constituinte interrogativo)

Sentenças completas com até 20 palavras que contêm apenas um constituinte interrogativo em uma oração finita principal ou em uma oração encaixada introduzida por um complementizador, excetuando-se expressões cristalizadas ou semilexicalizadas e os casos de *interrogativas-qu/qu-in-situ* em que o constituinte interrogativo exerce a função sintática de sujeito.

Nesse envelope, ocorreram 999 dados de Interrogativas-Q nas entrevistas sociolinguísticas, dos quais 227 (22,7%) representam *interrogativas qu-in-situ* (ver Tabela 4.1). A partir dessa primeira distribuição, definida em termos de frequência, percebe-se que as *interrogativas qu-in-situ* são bastante produtivas na língua oral quando se consideram os contextos em que a estrutura de fato é opcional: embora sua frequência seja de 13,8% em uma distribuição global de Interrogativas-Q, o seu emprego, quando factualmente possível, corresponde a quase um quarto dos dados.

Tabela 4.1 Distribuição de Interrogativas-Q no *corpus* de entrevistas sociolinguísticas no Envelope de Variação I e no Quadro Geral

	Envelope de Variação I		Quadro Geral	
	N	%	N	%
<i>Interrogativas-qu</i>	289	28,9	721	43,8
<i>Interrogativas qu-que</i>	408	40,8	579	35,1
<i>Interrogativas é-que</i>	75	7,5	121	7,3
<i>Interrogativas qu-in-situ</i>	227	22,7	227	13,8
Total	999	100,0	1648	100,0

A análise quantitativa testou se há correlações entre a variável dependente – *interrogativas qu-in-situ* vs. interrogativas com constituinte interrogativo pré-verbal (-*qu*, *qu-que* e *é-que*) – e grupos de fatores extralinguísticos e linguísticos, cujos resultados se apresentam a seguir.

4.2 Grupos de fatores extralinguísticos

Foram testados seis grupos de fatores extralinguísticos: (i) Sexo/gênero; (ii) Faixa etária; (iii) Escolaridade; (iv) Quem fala: documentador ou informante; (v) Grau de relação entre documentador e informante; e (vi) Espontaneidade da pergunta.

Dentre eles, não são selecionados Quem Fala, Escolaridade e Sexo/gênero. A Tabela 4.2 apresenta a distribuição dos dados nesses grupos de fatores e seus respectivos pesos relativos em rodadas *one-level*.

Tabela 4.2 Distribuição de ocorrências de *interrogativas qu-in-situ* em relação a Quem Fala, Sexo/Gênero e Escolaridade^a

	P.R.	%	N <i>in-situ</i>	N total
Quem fala^b				
informante	[.50]	23,0	117	509
documentador	[.50]	22,4	110	490
		22,7	227	999
Escolaridade^c				
até ensino fundamental II	[.54]	25,4	67	264
ensino superior	[.49]	21,8	160	735
		22,7	227	999
Sexo/gênero^d				
masculino	[.54]	25,7	97	378
feminino	[.48]	20,9	130	621
		22,7	227	999

^a [] indica a não-seleção dos grupos de fatores como significativos. Pesos relativos correspondem a rodada *one-level*. ^b *Input*: 0,227. *p* = 0,849.

^c *Input*: 0,227. *p* = 0,239. ^d *Input*: 0,226. *p* = 0,089.

O fato de o grupo de fatores **Quem fala** não ser selecionado indica

que o papel desempenhado pelo falante na entrevista sociolinguística não tem uma influência no emprego variável de Interrogativas-Q. Diante desse resultado, esse grupo de fatores foi excluído das rodadas subsequentes (exceto quando cruzado com outras variáveis) mas manteve-se a análise conjunta dos dados de documentadores e de informantes, em vez de separá-los em arquivos de dados distintos. **Escolaridade**, por sua vez, provavelmente não foi selecionado devido ao fato de nenhuma das formas investigadas ser negativamente avaliada, nem tampouco ter prestígio encoberto. Embora as *interrogativas qu-in-situ* sejam pouquíssimo produtivas na língua escrita, como se viu na seção 2.1, é improvável que haja um monitoramento consciente por parte dos falantes no sentido de evitá-las. Esse resultado reforça a necessidade de investigar grupos de fatores discursivo-pragmáticos, a fim de identificar que aspectos da língua oral ou da fala-em-interação condicionam o emprego de *interrogativas qu-in-situ* tão fortemente em relação à fala e à escrita, uma vez que essa diferença não parece depender de fatos de estigma ou prestígio.

A neutralidade do grupo de fatores **Sexo/gênero** também segue um padrão esperado para variáveis sintáticas, à semelhança de outros estudos (p.ex. [Weiner & Labov 1983 \[1977\]](#); [Mendes 2005](#)). Esse resultado não surpreende sobretudo quando se levam em conta os resultados dos grupos de fatores Escolaridade, não selecionado, e Faixa etária, que aponta para um caso de variação estável na comunidade (ver Tabela 4.4 adiante). Em muitos estudos sociolinguísticos prévios (Cf. [Labov 2001](#)), constatou-se que mulheres tendem a empregar as formas socialmente prestigiadas e a liderar processos de mudança linguística; como nenhum desses parâmetros parece estar em jogo aqui, a não-seleção do grupo de fatores Sexo/gênero se alinha com padrões mais gerais de variação verificados em outras comunidades.

Como há um número diferente de informantes homens e mulheres devido à adição de dados documentadores, cuja maioria é do sexo feminino, fez-se uma análise em que se cruzaram Quem fala e Sexo/gênero a fim de verificar se o desequilíbrio da amostra tem um efeito nos resultados quantitativos. O novo grupo de fatores ainda assim não é selecionado, tanto em rodadas *one-level* quanto multivariada (ver Tabela 4.3).

Contudo, como afirma [Cheshire \(2005\)](#), o encaixamento social da variação

Tabela 4.3 Distribuição de ocorrências de *interrogativas qu-in-situ* em relação ao cruzamento de Quem fala e Sexo/gênero^a

	P.R.	%	N <i>in-situ</i>	N total
Quem fala e Sexo/gênero				
documentadores do sexo masculino	[.59]	29,5	54	183
documentadores do sexo feminino	[.44]	18,2	56	307
informantes do sexo masculino	[.49]	22,1	43	195
informantes do sexo feminino	[.52]	23,6	74	314
		22,7	227	999

a [] indica a não-seleção do grupo de fatores como significativo. Pesos relativos correspondem a rodadas *one-level*.

além do nível fonológico pode se dar de maneira mais complexa (ver seção 1.3, p. 17). Em seu estudo sobre expressões sinalizadoras de novas entidades no discurso,¹ em três cidades inglesas (Hull, Milton Keynes e Reading), [Cheshire \(2005\)](#) demonstra que homens e mulheres empregam diferentes estratégias discursivas e sugere que a correlação de variáveis sintáticas e discursivo-pragmáticas com grupos de fatores sociais se dá de maneira indireta. No presente estudo, essa hipótese foi testada através do cruzamento de Sexo/gênero com grupos de fatores discursivo-pragmáticos, cujos resultados serão comentados mais adiante (ver p. 103).

Os demais grupos de fatores extralinguísticos, Faixa etária, Grau de relação entre documentador e informante, e Espontaneidade da pergunta são selecionados como significativos.

O resultado do grupo de fatores **Faixa etária** (ver Tabela 4.4) aponta para um caso de variação estável na comunidade paulistana. Diferentemente do que se esperaria em um caso de mudança em progresso em tempo aparente, em que os falantes mais jovens favoreceriam a forma inovadora, as *interrogativas qu-in-situ* são favorecidas tanto pelos falantes mais idosos (P.R. 0,54) quanto pelos mais jovens (P.R. 0,53), e são fortemente desfavorecidas pelos falantes

¹ Por exemplo, “There’s a drug car in the village square. It’s parked near the bus shelter” *vs.* “A drug car’s in the village square. It’s parked near the bus shelter.” (Exemplo extraído de [Cheshire 2005](#): 482.)

Tabela 4.4 Correlação entre *interrogativas qu-in-situ* e Faixa etária^a

	P.R.	%	N <i>in-situ</i>	N total
Faixa etária				
de 20 a 30 anos	.53	21,4	108	504
de 35 a 45 anos	.37	15,2	30	198
mais de 50 anos	.54	30,0	89	297
	<i>Range: 17</i>	22,7	227	999

^a Input: 0,147. $p < 0,01$. Log likelihood = -442,635.

da faixa etária intermediária (P.R. 0,37). De certa forma, esse resultado está de acordo com a frequência relativamente estável no emprego de *qu-in-situ* da década de 1970 até os dias de hoje (ver Tabela 1, p. 7, Introdução). No entanto, esse padrão curiosamente se assemelha àquele de formas consideradas estigmatizadas na língua, uma vez que os falantes da faixa etária intermediária, em geral plenamente inseridos no mercado de trabalho e mais sensíveis a pressões normativas, parecem evitar o emprego de *interrogativas qu-in-situ*.²

Antes de interpretar esses resultados, é necessário verificar se o desequilíbrio da amostra em relação à faixa etária dos falantes é responsável pelo padrão verificado. Como essa amostra inclui os dados de documentadores que, em sua grande maioria, são jovens universitários com menos de 30 anos, houve uma quantidade bastante díspar de ocorrências para cada faixa etária: 504 para a primeira, 198 para a segunda, 297 para a terceira. Desse modo, realizou-se alternativamente uma análise em que se cruzaram Quem fala e Faixa etária em um único grupo de fatores.

Na Tabela 4.5, é possível perceber que as tendências de emprego de *interrogativas qu-in-situ* são semelhantes tanto para documentadores quanto para informantes, seja da terceira faixa etária (P.R.s 0,57 e 0,56 respectivamente), seja da primeira (P.R.s 0,51 e 0,55 respectivamente). Embora a

² Nascimento (Ms.) verifica um padrão semelhante em sua análise da variação entre as formas pronominais de 2ª pessoa *você(s)* e *cê(s)*, que se baseia, de modo geral, no mesmo *corpus* de entrevistas. Embora nenhuma dessas formas pronominais seja estigmatizada na língua falada, seus resultados indicam que os falantes da faixa etária intermediária evitam o emprego da forma inovadora *cê(s)*. A explicação aqui proposta, passível de investigações mais aprofundadas, também se assemelha à de Nascimento (Ms.).

Tabela 4.5 Correlação entre *interrogativas qu-in-situ* e o cruzamento de Quem fala e Faixa etária^a

	P.R.	%	N <i>in-situ</i>	N total
Quem fala e Faixa etária				
documentadores da 3ª faixa etária	.57	32,3	30	93
informantes da 3ª faixa etária	.56	28,9	59	204
informantes da 1ª faixa etária	.55	22,2	32	144
documentadores da 1ª faixa etária	.51	21,1	76	360
informantes da 2ª faixa etária	.40	16,1	26	161
documentadores da 2ª faixa etária	.25	10,8	4	37
	<i>Range: 31</i>	22,7	227	999

a Input: 0,158 . p < 0,04. Log-likelihood: -442,019.

diferença seja mais drástica entre os pesos relativos para documentadores e informantes de segunda faixa etária (P.R.s 0,25 e 0,40 respectivamente), o teste de qui-quadrado entre as análises com e sem separação dos dados de documentadores e informantes por faixa etária (ou seja, contrastando os resultados da Tabela 4.4 e da Tabela 4.5) indica que a diferença entre os pesos relativos, de acordo com o papel do falante na entrevista, não é, de fato, significativa ($\chi^2 = 1,23(3)$, $p > 0,70$). A Figura 4.1 representa melhor a semelhança nas tendências de cada faixa etária, independentemente do papel do falante na entrevista sociolinguística.

Desse modo, os pesos relativos da Tabela 4.4 devem corresponder de fato às tendências de variação na comunidade. O padrão de desfavorecimento de *qu-in-situ* para os falantes da faixa etária intermediária talvez se explique pelo “estilo”³ – mais dissertativo do que narrativo – em que os falantes de segunda faixa etária tendem a se expressar. Embora o roteiro de entrevista tivesse a preocupação de controlar os estilos e ainda que os documentadores houvessem sido instruídos a dirigir suas perguntas de modo a obter igualmente tanto amostras de fala mais narrativas quanto mais dissertativas (ver Anexo C,

³ A noção de “estilo” está longe de ser um conceito pacífico nos estudos linguísticos, uma vez que não se restringe aos binômios formal-informal ou narrativo-dissertativo. De modo geral, entende-se por “estilo” a variação de registro na fala de um mesmo indivíduo (Schilling-Estes, 2002).

Figura 4.1 Pesos relativos para *interrogativas qu-in-situ* por Faixa etária e Quem fala

p. 154), os informantes da segunda faixa etária preferem discorrer sobre os problemas da cidade (segurança, saúde, transporte público) de modo mais distanciado e objetivo, enquanto aqueles da terceira faixa etária preferem relatar causos e histórias de suas vidas. Essas constatações estão de acordo com as observações qualitativas traçadas a respeito do *corpus* de língua escrita, em que os pouquíssimos casos de *interrogativas qu-in-situ* foram propiciados por um estilo mais narrativo (ver seção 2.1, p. 29). Ademais, o fato de a maioria dos documentadores pertencer à primeira faixa etária parece fazer com que as interações conversacionais entre documentadores e informantes com menos de 30 anos sejam mais “descontraídas”. Desse modo, é possível que a estrutura de *interrogativas qu-in-situ*, ainda que não seja estigmatizada nem prestigiada na comunidade, seja associada a um falar mais “informal”.

O grupo de fatores **Grau de relação entre documentador e informante** (ver Tabela 4.6) busca verificar se o grau de fundo comum ou de informações compartilhadas entre os falantes tem influência no emprego de *interrogativas qu-in-situ*. Nesse grupo, os graus foram codificados de acordo com as respostas fornecidas pelos documentadores no questionário a eles enviado (ver Anexo A, p. 148), que tinha como um dos objetivos descobrir

o grau de proximidade que cada um tinha com seu informante. Na análise quantitativa, foram amalgamados os graus de relação mais próximos (1, 2 e 3), por um lado, e os graus de relação mais distante (4 e 5), por outro ($\chi^2 = 6,04(3)$, $p > 0,10$). Dos 31 documentadores que responderam ao questionário, 20 indicaram ter uma relação distante com o entrevistado. Decidiu-se também amalgamar os dados daqueles para os quais essa informação não estava disponível juntamente com os dados de relação distante (em vez de descartá-los), já que era mais provável que, se os documentadores houvessem respondido ao questionário, tais seriam os graus de relação assinalados. O teste de qui-quadrado ($\chi^2 = 0,66(1)$, $p > 0,30$) confirma que essa amalgamação é plausível.

Tabela 4.6 Correlação entre *interrogativas qu-in-situ* e grupos de fatores extralinguísticos^a

	P.R.	%	N <i>in-situ</i>	N total
Grau de relação entre documentador e informante				
+distante	.54	24,1	186	771
+próximo	.36	18,0	41	228
	<i>Range: 18</i>	22,7	227	999
Espontaneidade da pergunta				
de dados pessoais	.80	52,6	51	97
espontânea	.56	22,6	47	208
+do roteiro	.30	6,8	8	117
temática	.22	5,9	4	68
	<i>Range: 58</i>	22,4	110	490

^a *Input*: 0,115. $p < 0,05$. *Log likelihood* = -388,695. + indica que houve amalgamação de fatores.

Os resultados obtidos revelam uma correlação inversa àquela esperada inicialmente: a hipótese previa, com base em Pires & Taylor (2007), que graus de relação mais próximos – ou seja, a existência de maior fundo comum entre os interlocutores – levariam a uma maior tendência de emprego de *interrogativas qu-in-situ*. A Tabela 4.6, no entanto, indica justamente o oposto: desfavorecimento de *interrogativas qu-in-situ* (P.R. 0,36) para graus

mais próximos e pequeno favorecimento (P.R. 0,54) para graus mais distantes.

Desse modo, faz-se necessário um olhar mais detalhado aos dados e à situação discursiva: quando os interlocutores têm um grau de relação mais distante, o fundo comum entre eles depende mais diretamente da situação interacional em que se encontram, diferentemente do que pode ocorrer em interações conversacionais entre pessoas mais próximas, cujo fundo comum engloba outras conversações prévias. Adiante, argumenta-se que o emprego de *interrogativas qu-in-situ* depende fundamentalmente do fluxo de informações no aqui-e-agora da conversa. Outros grupos de fatores (ver Espontaneidade da pergunta, p. 98, Conjunto de respostas previstas, p. 105, e Grau de ativação do fundo comum, p. 106) têm justamente o objetivo de examinar mais detalhadamente fatos da fala-em-interação, que podem elucidar por que graus de relação mais próximos desfavorecem o emprego de *qu-in-situ*.

O grupo de fatores **Espontaneidade da pergunta** (ver Tabela 4.6), que se aplica somente às perguntas feitas pelos documentadores, tem por base o roteiro de entrevistas que eles deveriam seguir e investiga se o grau de espontaneidade da pergunta (obrigatória, sugerida, temática, de dados pessoais e espontânea) tem influência na variação entre formas interrogativas. O menor número total de dados para esse grupo (490) deve-se ao fato de que se aplica a um subconjunto de dados. Aqui, foram amalgamadas as perguntas obrigatórias e sugeridas ($\chi^2 = 0,63(1)$, $p > 0,30$) no novo fator “perguntas do roteiro”. As perguntas espontâneas, como se esperava, favorecem o emprego de *interrogativas qu-in-situ* (P.R. 0,56), enquanto perguntas relacionadas ao roteiro as desfavorecem consideravelmente (P.R.s abaixo de 0,30). Em princípio, esse resultado pode sugerir que o emprego variável de interrogativas é condicionado pelo grau de monitoramento ou planejamento da fala. No entanto, verifica-se que as perguntas de dados pessoais (“Você mora onde?”, “Você faz o quê?”) são aquelas que mais favorecem o emprego de *interrogativas qu-in-situ* (P.R. 0,80). Se se tratasse de uma questão de monitoramento ou planejamento da fala, era de se esperar que perguntas de dados pessoais também desfavorecessem o emprego as *interrogativas qu-in-situ*, pois a obtenção desses dados também era prevista no roteiro utilizado para proceder às entrevistas (ver Anexo C, p. 160).

Uma interpretação alternativa se faz necessária. Perguntas de dados pessoais e espontâneas, que mais favorecem *qu-in-situ*, têm em comum o fato de serem mais previsíveis: em situação de entrevista é natural que se pergunte sobre a profissão, local de residência, família. Aliás, mesmo que não se tratasse de uma entrevista gravada, esses seriam tópicos bastante propícios para a condução de uma conversa “natural”. De modo semelhante, as perguntas espontâneas, que surgem da interação conversacional, estão mais diretamente ancoradas no fluxo de informações da conversação e, portanto, também podem ser consideradas mais esperadas. Assim, o grupo de fatores Espontaneidade da pergunta, cujo objetivo inicial era testar a influência do grau de planejamento da fala, acaba revelando mais propriamente o papel da previsibilidade da pergunta: quanto mais previsível, maior a tendência de se empregar *qu-in-situ*.

É possível que decorra daí a intuição de Pires & Taylor (2007) sobre a inadequação pragmática de *interrogativas qu-in-situ* em perguntas “de sopetão”, dado que elas teriam um grau mínimo de previsibilidade. Entretanto, é difícil imaginar situações em que não há qualquer grau de fundo comum entre interlocutores, uma vez que os falantes vivem em comunidades e que compartilham, culturalmente, uma série de informações sobre outras pessoas (ver discussão da subseção 3.2.3, p. 62). A adequação pragmática de perguntas deve ser vista de acordo com a sua pertinência no contexto discursivo em que os falantes se encontram.

Os resultados quantitativos dos grupos de fatores extralinguísticos apontam, de maneira geral, para a importância do aqui-e-agora da conversa para o emprego de *interrogativas qu-in-situ*. Enquanto categoriais sociais amplas como Sexo/gênero e Escolaridade, além do papel dos falantes na entrevista sociolinguística, não são selecionados como significativos para a variação na posição do constituinte interrogativo, outros grupos de fatores extralinguísticos mostram influência: para Faixa etária, sugere-se que o favorecimento de *qu-in-situ* entre falantes mais jovens e mais velhos deve-se a um estilo de fala mais narrativo do que dissertativo; para os grupos de fatores Grau de relação entre documentador e informante e Espontaneidade da pergunta, aponta-se para características da fala-em-interação (construção de fundo

comum e previsibilidade da pergunta) como preponderantes na variação. Essas questões também são pertinentes na discussão de grupos de fatores de natureza discursivo-pragmática, pela qual se iniciará a próxima seção.

4.3 Grupos de fatores linguísticos

Foram testados oito grupo de fatores linguísticos: (i) Que elemento ocorre antes da oração principal; (ii) Sinceridade pragmática da pergunta; (iii) Conjunto de respostas previstas; (iv) Grau de ativação do fundo comum; (v) Tipo de verbo; (vi) Constituinte interrogativo; (vii) Função sintática; e (viii) Número de palavras na oração. Todos estes são selecionados nas análises multivariadas como estatisticamente relevantes.

Grupos de fatores discursivo-pragmáticos

O grupo de fatores **Que elemento ocorre antes da oração principal** (ver Tabela 4.7), inicialmente composto de seis tipos de fatores que podem ocorrer antes do constituinte interrogativo (ver Tabela 3.6, p. 77), foi redefinido para uma distinção binária, uma vez que na maioria dos dados figuraram sentenças sem qualquer elemento antes da oração principal (847 ocorrências). Amalgamaram-se então todos os fatores que se referiam a elementos topicalizados e a orações subordinadas (1-a) em contraste com nenhum elemento (1-b), e o grupo de fatores foi renomeado como “Ocorre outro elemento antes da oração principal?” (sim ou não) para refletir a reformulação da questão colocada.

- (1) a. Mariana: e [o francês]_k **você se interessou t_k por quê?** (F1S-DOC)
b. Rafael: **o senhor tinha que idade em setenta?** (M1S-DOC)

A Tabela 4.7 indica que, na presença de outros elementos antes da oração principal (tópico coindexado com sujeito, tópico coindexado com objeto direto ou indireto, ou uma oração subordinada), é maior a tendência ao emprego de *interrogativas qu-in-situ* (P.R. 0,66). Essa tendência pode ser entendida como

Tabela 4.7 Correlação entre *interrogativas qu-in-situ* e a ocorrência de outro elemento antes da oração principal^a

	P.R.	%	N <i>in-situ</i>	N total
Ocorre outro elemento antes da oração principal?				
+sim	.66	28,9	44	152
não	.47	21,6	183	847
	<i>Range: 19</i>	22,7	227	999

^a *Input*: 0,115. $p < 0,05$. *Log-likelihood*: - 388.695. + indica que houve amalgamação de fatores

uma estratégia de organização do discurso, a fim de manter o constituinte interrogativo em uma posição de proeminência prosódica na sentença: quando há outro elemento à esquerda, tende-se a realizar o constituinte interrogativo à direita, em posição pós-verbal.

O grupo de fatores **Sinceridade pragmática da pergunta** investiga se diferentes tipos de perguntas, em relação à troca de turnos e à atualização do fundo comum entre os interlocutores, têm uma influência no emprego de diferentes estruturas de interrogativas de constituinte. Os dados foram codificados em perguntas de estruturação do discurso, para as quais o próprio falante fornece uma resposta (2-a); perguntas “sinceras”, que esperam uma resposta do interlocutor (2-b); e perguntas retóricas, cuja resposta é óbvia e não requerida (2-c).

- (2)
- a. 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? (M3G-INF)
 - b. Carla: **há quanto tempo ele tá lá?** (F1S-DOC)
Pedro: faz... há quatro anos...
 - c. Clara: eu tenho que ter um carro... dá alguma coisa nela de noite **eu vou aonde?** (F3G-INF)

Os resultados da Tabela 4.8 indicam que as perguntas de estruturação do discurso favorecem o emprego de *interrogativas qu-in-situ* (P.R. 0,68), enquanto perguntas “sinceras” e retóricas as desfavorecem levemente (P.R.s

Tabela 4.8 Correlação entre *interrogativas qu-in-situ* e Sinceridade Pragmática da Pergunta^a

	P.R.	%	N <i>in-situ</i>	N total
Sinceridade pragmática da pergunta				
de estruturação do discurso	.68	34,7	51	147
“sincera”	.47	21,6	129	597
retórica	.45	18,4	47	255
	<i>Range: 23</i>	22,7	227	999

a Input: 0,165. p < 0,02. Log-likelihood: - 426,574.

0,47 e 0,45 respectivamente). Embora os pesos desses dois últimos tipos de perguntas não sejam muito diferentes, nota-se também que as perguntas retóricas são aquelas que menos favorecem a forma *in situ*, o que indica que perguntas de estruturação do discurso e perguntas retóricas, em geral classificadas como um mesmo tipo de pergunta, comportam-se de forma algo distinta em relação à estruturação da sentença e merecem ser caracterizadas diferentemente (Oushiro & Nasser, 2010).

O favorecimento de *in situ* em perguntas de estruturação do discurso pode ser manifestação de uma estratégia de manutenção de turno por parte do falante que o detém. Segundo Sacks *et al.* (1974:706-7), os falantes projetam o fim de cada turno conversacional através da identificação de unidades linguísticas (que podem ser uma palavra ou toda uma sentença); o emprego de um constituinte interrogativo no início da oração pode criar a expectativa no interlocutor de que seu próximo turno deva ser uma resposta à pergunta colocada; como as perguntas de estruturação do discurso não têm o objetivo de obter uma informação do interlocutor – ao contrário, impõem uma obrigatoriedade de resposta pelo falante corrente – as *interrogativas qu-in-situ* podem servir melhor ao propósito de manutenção de turno.⁴

O emprego de *interrogativas qu-in-situ* em perguntas de estruturação do discurso também pode ser uma estratégia cognitiva para o processamento de

⁴ Não se pretende sugerir que essa estratégia seja implementada conscientemente.

informações. Observe-se o exemplo (3):

- (3) Marco: sim a televisão e **a televisão te traz o quê?**... só te traz coisa errada... é **novela... ensina o quê?**... ensina o camarada a beber ensina a prostituição... (M3G-INF)

O falante Marco poderia ter dito “a televisão... só te traz coisa errada... novela ensina o camarada a beber ensina prostituição...”. No entanto, ao empregar a palavra *o quê* no lugar de sua “resposta” correspondente, o falante atualiza o fundo comum em dois tempos: primeiro com o constituinte interrogativo e depois com a informação nova propriamente.

De fato, estudos sobre a organização do discurso indicam a preferência por postergar a introdução de novas informações para a última parte da sentença (Cf. p.ex. Chafe 1974, Weiner & Labov 1983 [1977], Langacker 1991). Como o objetivo de perguntas de estruturação do discurso é a atualização do fundo comum pelo falante corrente, a preferência por *interrogativas qu-in-situ* nesse contexto também pode ser reflexo de tendências mais gerais de organização do discurso.

O cruzamento de Sinceridade pragmática da pergunta com Sexo/gênero do falante traz resultados interessantes (ver Tabela 4.9). Embora Sexo/gênero não seja selecionado como significativo quando analisado como variável independente, os resultados apontam para um encaixamento social mais complexo da variação na posição do constituinte interrogativo.

Na Tabela 4.9, verifica-se que tanto homens quanto mulheres favorecem o emprego de *interrogativas qu-in-situ* em perguntas de estruturação do discurso (P.R.s 0,73 e 0,67 respectivamente), resultado semelhante ao da análise anterior, da Tabela 4.8. No entanto, há diferenças de comportamento linguístico entre homens e mulheres no caso de perguntas “sinceras” e perguntas retóricas; enquanto homens favorecem o emprego de *interrogativas qu-in-situ* em perguntas “sinceras” (P.R. 0,56) e o desfavorecem em perguntas retóricas (P.R. 0,26), as mulheres exibem um resultado inverso, visto que favorecem *qu-in-situ* em perguntas retóricas (P.R. 0,55) e o desfavorecem em perguntas “sinceras” (P.R. 0,42). Tal diferença é melhor visualizada na Figura 4.2, que revela que esses grupos de fatores não são de fato independentes entre si.

Tabela 4.9 Correlação entre *interrogativas qu-in-situ* e o cruzamento de Sinceridade pragmática da pergunta e Sexo/gênero^a

	P.R.	%	N <i>in-situ</i>	N total
Sinceridade pragmática da pergunta e Sexo/gênero				
estr. discurso / masculino	.73	35,5	27	76
estr. discurso / feminino	.67	33,8	24	71
“sincera” / masculino	.56	28,4	63	222
retórica / feminino	.55	22,9	40	175
“sincera” / feminino	.42	17,6	66	375
retórica / masculino	.26	8,8	7	80
	<i>Range =</i> .47	22,7	227	999

^a *Input* = 0,160. *p* < 0,05. *Log likelihood* = -419,499.

Figura 4.2 Pesos relativos para *interrogativas qu-in-situ* em relação ao Sexo/gênero e Sinceridade pragmática da pergunta

Quando dois grupos de fatores são verdadeiramente independentes, os resultados de pesos relativos formam um padrão de retas paralelas. No entanto, a Figura 4.2 mostra uma intersecção entre as curvas de pesos relativos quanto ao emprego de perguntas retóricas: homens e mulheres parecem empregar diferentes estruturas sintáticas em relação à mesma estratégia

discursiva; para as mulheres, a estrutura de *interrogativas qu-in-situ* parece estar mais associada a perguntas retóricas do que para os homens, que a desfavorecem fortemente no mesmo contexto. Um teste de qui-quadrado que contrasta essa análise com aquela em que os sexos não são separados (Tabela 4.8) indica que a diferença entre homens e mulheres é significativa ($\chi^2 = 14,15(3)$, $p < 0,01$); em outras palavras, é muito pouco provável (chances menores de 1%) que a distribuição observada na Tabela 4.9 e na Figura 4.2 seja aleatória e, portanto, as diferenças de comportamento entre homens e mulheres quanto ao emprego de *interrogativas qu-in-situ* em perguntas retóricas não devem ser ignoradas.

A análise do cruzamento desses fatores sugere que diferentes estruturas sintáticas em alternância podem não se distribuir igualmente em diferentes grupos sociais. É possível que a variação não seja diretamente associada a categorias sociais – diferentemente do que em geral ocorre com variáveis fonético-fonológicas – pelo fato de a correlação ser mediada por diferentes estratégias de organização do discurso. Esse resultado traz evidências favoráveis à hipótese de Cheshire (2005) a respeito do encaixamento social indireto da variação além do nível fonológico e do emprego de diferentes estratégias discursivas por homens e mulheres.⁵

Os grupos de fatores apresentados na Tabela 4.10 referem-se a subconjuntos de dados do Envelope de variação I. O Conjunto de respostas ‘previstas’ aplica-se somente às perguntas pragmaticamente “sinceras”, e Grau de ativação do fundo comum aplica-se a perguntas pragmaticamente “sinceras” e de estruturação do discurso. Ambos os grupos de fatores investigam mais detalhadamente as questões levantadas na interpretação de Espontaneidade da pergunta e Grau de relação entre documentador e informante, pois concernem à previsibilidade da resposta e à ativação do fundo comum.

Em **Conjunto de respostas ‘previstas’**, foram amalgamados os dois fatores em que havia um “fechamento” das possibilidades de resposta – múltipla escolha (4-a) e com um “chute” (4-b) ($\chi^2 = 1,28(1)$, $p > 0,20$) –, em

⁵ Assim como para Sinceridade pragmática da pergunta, também se realizaram cruzamentos entre Sexo/gênero e os demais grupos de fatores discursivo-pragmáticos (Conjunto de respostas previstas, Grau de ativação do fundo comum e Ocorre outro elemento antes da oração principal). Todavia, apenas Sinceridade pragmática da pergunta se mostrou dependente de grupo de fatores Sexo/gênero.

Tabela 4.10 Correlação entre *interrogativas qu-in-situ* e grupos de fatores de subconjuntos de dados^a

	P.R.	%	N <i>in-situ</i>	N total
Conjunto de respostas ‘previstas’				
+‘fechado’	.84	37,7	20	53
‘aberto’	.46	20,0	109	544
	<i>Range: .38</i>	21,6	129	597
Grau de ativação do fundo comum				
+pressuposição ativada na 1 ^a ou 2 ^a sentença anterior	.66	35,2	56	159
referente ativado em uma das últimas 5 sentenças	.59	27,8	42	151
+pressuposição ativada na 3 ^a sentença anterior ou antes	.49	22,5	20	89
pressuposição não ativada	.39	18,0	62	345
	<i>Range: .27</i>	24,2	180	744

^a Input: 0,115. $p < 0,05$. *Log likelihood* = -388,695. + indica que houve amalgamação de fatores

contraste com as perguntas de resposta aberta (4-c). Os resultados indicam que as perguntas de resposta “fechada” favorecem fortemente o emprego de *qu-in-situ* (P.R. 0,84), o que reforça a interpretação de que, quanto mais previsível a resposta, maior a tendência de emprego dessa forma interrogativa.

- (4) a. Amanda: **cê pensa que tem que ser mais importante o quê? segurança transporte educação...** (F1S-DOC)
 b. Jorge: **sua mãe trabalhava com o quê? assim ela... era dona-de-casa?** (M1S-DOC)
 c. Carolina: **e você mora em que lugar de São Paulo?** (F1S-DOC)

O grupo de fatores **Grau de ativação do fundo comum** investiga a hipótese de que pressuposições e referentes recentemente ativados no discurso, de acordo com a última sentença em que foram mencionados, influenciam o emprego de *interrogativas qu-in-situ*. Nesse grupo, amalgamaram-se os dois graus em que a pressuposição da interrogativa está maximamente ativada (na última ou penúltima sentença) e os dois graus de ativação intermediária (na

3^a, 4^a, 5^a sentença anterior e ativada anteriormente na conversa mas não nas últimas 5 sentenças) ($\chi^2 = 0,77(2)$, $p > 0,50$). Os resultados se alinham com a hipótese inicial, que previa maior tendência ao emprego de *qu-in-situ* quando a pressuposição ou um dos referentes havia sido recentemente ativado (P.R.s 0,66 e 0,59 respectivamente). Os demais fatores se conformam à hierarquia de ativação, pois pressuposições ativadas a partir da terceira sentença anterior apresentam um peso relativo menor para o emprego de *qu-in-situ*, próximo ao ponto neutro (P.R. 0,49), enquanto pressuposições não ativadas anteriormente desfavorecem a estrutura ainda mais (P.R. 0,39). Aqui, nota-se novamente um fato que aponta para a importância do aqui-e-agora na conversa para o emprego de *interrogativas qu-in-situ*: há uma diferença relevante para a ativação de pressuposições e referentes em uma das últimas duas sentenças. Isso sugere que o “gerenciamento” da conversação (nos termos de [Sacks et al. 1974](#)) ocorre localmente e que os falantes são bastante sensíveis ao fluxo de informações, fato que reflete na escolha que os falantes fazem de diferentes estruturas sintáticas.

É possível ainda refinar a análise apresentada na Tabela 4.10. Na primeira codificação de Grau de ativação do fundo comum, a ativação de referentes foi classificada indistintamente caso houvesse ocorrido em uma das últimas cinco sentenças. Contudo, o valor para o seu peso relativo (P.R. 0,59), que é intermediário ao valor dos pesos relativos de pressuposições ativadas na última ou penúltima sentença (P.R. 0,66) e de pressuposições ativadas na terceira sentença ou antes (P.R. 0,49), sugere que também pode haver diferenças significativas quanto ao grau de ativação de referentes no fluxo da conversação, à semelhança das pressuposições. Desse modo, esse fator foi recodificado em dois, com a separação de referentes ativados na última/penúltima sentença ou na terceira até quinta sentença anterior (ver Tabela 4.11).⁶ Verificou-se então que a ativação tanto de pressuposições quanto de referentes tem efeitos semelhantes ($\chi^2 = 1,46(1)$, $p > 0,20$).

Os resultados de Conjunto de respostas previstas e Grau de ativação do fundo comum reforçam o papel das informações compartilhadas entre os interlocutores no emprego variável de interrogativas de constituinte: as

⁶ Essa análise foi sugerida pelo professor Gregory R. Guy.

Tabela 4.11 Correlação entre *interrogativas qu-in-situ* e Grau de ativação do fundo comum recodificado^a

	P.R.	%	N <i>in-situ</i>	N total
Grau de ativação do fundo comum				
+pressup. ou referente ativado na 1 ^a ou 2 ^a sentença anterior	.63	34,0	89	262
+pressup. ou referente ativado na 3 ^a sentença anterior ou antes	.51	21,2	29	137
pressuposição não ativada	.38	18,0	62	345
	<i>Range = 25</i>	24,2	180	744

^a Input 0,115. $p < 0,05$. $\text{Log likelihood} = -389,425$. $\chi^2 = 1,46(1)$, $p > 0,20$. + indica que houve amalgamação de fatores.

interrogativas qu-in-situ são mais favorecidas quanto mais previsível ou mais ativado está o fundo comum. Além disso, fica demonstrado que uma visão estática do conceito de fundo comum, tal como adotada por Pires & Taylor (2007), não está de acordo com o modo em que a língua é usada cotidianamente por pessoas que vivem em comunidades. Na perspectiva desta análise, o fundo comum deve ser examinado em sua dinamicidade na fala-em-interação, pois ele se atualiza constantemente ao longo da conversa.

Grupos de fatores morfossintáticos

Os grupos de fatores Tipo de Verbo, Constituinte interrogativo e Função sintática foram analisados em rodadas distintas, uma vez que interagem entre si. No entanto, seus resultados são apresentados na mesma tabela (4.12), para que possam ser comparados.

Em **Tipo de verbo**, foram amalgamados os verbos transitivo diretos e bitransitivos, e os verbos de ligação e modais – agora chamados de “verbos auxiliares” ($\chi^2 = 1,27(2)$, $p > 0,50$). Os verbos transitivos indiretos e os verbos intransitivos favorecem o emprego de *interrogativas qu-in-situ* (P.R. 0,67 e 0,62 respectivamente), enquanto os verbos auxiliares o desfavorecem (P.R. 0,40). De modo geral, esse padrão indica que o emprego de *interrogativas*

Tabela 4.12 Correlação entre *interrogativas qu-in-situ* e grupos de fatores morfossintáticos^a

	P.R.	%	N <i>in-situ</i>	N total
Tipo de verbo^b				
transitivo indireto	.67	39,1	43	110
intransitivo	.62	34,6	36	104
+transitivo direto e bitransitivo	.50	21,2	107	505
+auxiliar	.40	14,6	41	280
	<i>Range: 27</i>	22,7	227	999
Constituinte interrogativo^c				
quando	.85	62,5	5	8
onde	.80	47,4	36	76
o que	.74	33,7	85	252
quem	.72	29,2	14	48
+quanto/quanto+NP	.71	39,2	20	51
que+NP	.63	28,6	24	84
+por que/pra que	.61	22,3	27	121
que	.27	7,3	9	124
+qual/qual+NP	.21	4,5	4	89
como	.09	2,1	3	146
	<i>Range: 76</i>	22,7	227	999
Função sintática^d				
+outros adjuntos adverbiais	.83	55,0	72	131
objeto indireto	.79	48,1	26	54
objeto direto	.58	24,4	81	332
+adj. adv. de causa e de finalidade	.53	22,0	27	123
constituente de <i>small clause</i>	.29	8,6	14	162
adj. adv. de modo	.19	4,5	5	110
sujeito	.16	3,6	2	56
	<i>Range: 67</i>	23,5	227	968

^a Divergência entre P.R.s e porcentagens nos valores sombreados indica interação. + indica que houve amalgamação de fatores ^b *Input*: 0,147. *p* < 0,01. *Log likelihood* = -431,565. ^c *Input*: 0,112. *p* < 0,05. *Log likelihood*: -385,845. ^d *Input*: 0,115. *p* < 0,05.

qu-in-situ está relacionado com a estrutura da sentença quanto à presença de argumentos e adjuntos adverbiais, relação que será discutida com mais detalhes para o grupo de fatores Função sintática.

Para a variável **Constituinte interrogativo**, devido ao número reduzido de dados em determinadas células, amalgamaram-se os fatores *quanto(-a, -os, -as)* com *quanto(-a, -os, -as) + NP*, *qual(-is)* com *qual(-is) + NP* e *por que* com *pra que* ($\chi^2 = 0,71(3)$, $p > 0,80$). Outras amalgamações, no entanto, não são desejáveis, dado que essas provavelmente juntariam fatores com função sintática semelhante – hipótese para qual há outro grupo de fatores.

Os resultados aqui devem ser examinados em paralelo aos de Função sintática, já que os constituintes interrogativos seguem as tendências de sua função sintática mais comum: *quando* (P.R. 0,85) e *onde* (P.R. 0,80) são os fatores que mais favorecem o emprego de *qu-in-situ*, assim como os adjuntos adverbiais (P.R. 0,83); *por que* e *pra que* (P.R. 0,61) encontram-se em uma posição hierárquica intermediária, assim como os adjuntos adverbiais de causa e de finalidade (P.R. 0,53); *como* (P.R. 0,09) é o constituinte interrogativo que menos favorece o emprego de *qu-in-situ*, o que está de acordo com o baixo peso relativo para o adjunto adverbial de modo (P.R. 0,19).

Em **Função sintática**, também foi necessário amalgamar dados para que se pudessem realizar testes quantitativos. Foram amalgamados alguns dos adjuntos adverbiais (tempo, lugar, companhia, outros) que exibiram distribuição semelhante ($\chi^2 = 1,06(3)$, $p > 0,70$), mas os adjuntos adverbiais de causa, de finalidade e de modo encontram-se separados por apresentarem pesos relativos bastante distintos dos demais adjuntos adverbiais. Foram excluídos 31 dados de sentenças com “Como + verbo-ser” (p.ex. “Como (que/ é que) foi trabalhar fora?”) por ocorrerem categoricamente com as três construções interrogativas pré-verbais.

Dentre os três grupos de fatores morfossintáticos, Função sintática pode ser considerado o mais importante, ainda que o seu *range* (diferença entre maior e menor pesos relativos) seja menor do que aquele para Constituinte interrogativo (67 *vs.* 76). Além do fato de que a hierarquia de restrições em Constituinte interrogativo segue aquela de Função sintática, o *range* mais amplo do primeiro provavelmente é decorrência de seu grande número

de fatores (10 no total, contra 7 para Função sintática). Ademais, alguns fatores de Constituinte interrogativo, tais como *quando* e *quanto*, possuem um número reduzido de dados, o que dificulta a verificação de seu efeito na variação.

Uma análise com ambos os grupos Constituinte interrogativo e Função sintática na mesma rodada pode ser mais elucidativa. Para tanto, foram mantidos apenas os dados de constituintes interrogativos que podem exercer mais de uma função sintática, para que houvesse ortogonalidade entre os grupos de fatores. Assim, incluíram-se os dados de *quem*, *que*, *o que*, *que* + *NP*, *qual* e *qual* + *NP*, e excluíram-se aqueles para os quais a relação era (quase) biunívoca: *por que* (= adjunto adverbial de causa), *pra que* (= adjunto adverbial de finalidade), *quando* (= adjunto adverbial de tempo), *onde* (muitas vezes adjunto adverbial de lugar), *como* (muitas vezes adjunto adverbial de modo) e *quanto/quanto* + *NP* (muitas vezes objeto direto) (ver Tabela 4.13). Nessa análise, Função sintática tem *range* de maior do que Constituinte interrogativo (77 e 56 respectivamente) e é o primeiro grupo de fatores a ser selecionado.

De modo geral, observa-se uma tendência geral de emprego que corresponde à “obrigatoriedade” ou não do constituinte na sentença: enquanto os adjuntos adverbiais, argumentos “acessórios” na sentença, favorecem fortemente o emprego de *interrogativas qu-in-situ* (P.R. 0,87), as funções sintáticas de sujeito e de constituinte de *small clause* têm a tendência oposta, de forte desfavorecimento (P.R.s 0,10 e 0,24 respectivamente). Os demais argumentos do verbo, objeto direto e objeto indireto, encontram-se em posição intermediária e também obedecem a uma hierarquia de “obrigatoriedade” na sentença; pode-se dizer que objetos indiretos, ainda que argumentos do verbo, podem ser mais facilmente omitidos do que objetos diretos – por exemplo, é bastante natural dizer “Ele vendeu o carro”, sem menção do objeto indireto, mas normalmente só se diz “Ele vendeu pro João” se o objeto direto já está implícito. A hierarquia verificada também diz respeito à “distância” entre a posição canônica do constituinte em relação ao início da sentença. Tanto a hierarquia de “obrigatoriedade” quanto de “distância” estão representadas no contínuo abaixo:

Tabela 4.13 Correlação entre *interrogativas qu-in-situ* e grupos de fatores Função sintática e Constituinte interrogativo recodificados^a

	P.R.	%	N <i>in-situ</i>	N total
Função sintática				
adjuntos adverbiais	.87	55,0	72	131
objeto indireto	.84	48,1	26	54
objeto direto	.51	24,4	81	332
constituente de <i>small clause</i>	.24	7,3	14	193
sujeito	.10	3,6	2	56
	<i>Range</i> = 77	25,5	195	766
Constituinte interrogativo				
o que	.76	33,7	85	252
quem	.74	29,2	14	48
que + NP	.41	28,6	24	84
que	.21	7,3	9	124
+qual	.20	4,5	4	89
	<i>Range</i> = 56	22,8	136	597

^a Input = 0,121. $p < 0,05$.

sujeito < constituinte de *small clause* < objeto direto < objeto indireto < adjuntos adverbiais

Percebe-se que esse resultado é paralelo ao que se verificou em Tipo de verbo, visto que diferentes tipos de verbos permitem a presença de diferentes argumentos ou adjuntos.

verbos auxiliares < verbos transitivo diretos e bitransitivos < verbos intransitivos < verbos transitivos indiretos

Os verbos auxiliares desta análise englobam os verbos de ligação, que coocorrem com constituintes de *small clause*; verbos transitivos coocorrem com objetos diretos e indiretos; e verbos intransitivos, por sua vez, muitas vezes coocorrem com adjuntos adverbiais. Assim, os efeitos verificados no grupo de fatores Tipo de verbo podem ser atribuídos a Função sintática, já que a função desempenhada pelo constituinte interrogativo cria relações necessárias

com o tipo de verbo. Explica-se: o uso de um verbo transitivo indireto, por exemplo, não necessariamente obriga que o constituinte interrogativo daquela sentença seja o objeto indireto correspondente; a presença de um verbo transitivo direto não obriga que o constituinte interrogativo seja o seu objeto direto; e a presença de um verbo auxiliar (de ligação ou modal) não obriga que o constituinte interrogativo seja constituinte de *small clause*. Quando o direcionamento da relação entre os grupos de fatores é revertido, criam-se relações necessárias: um constituinte interrogativo com função de objeto indireto implica na presença de um verbo transitivo indireto ou bitransitivo; a presença de um constituinte interrogativo com função de objeto direto implica na presença de um verbo transitivo direto ou bitransitivo; e a presença de um constituinte interrogativo com função de constituinte de *small clause* implica na presença de um verbo de ligação (subconjunto dos verbos auxiliares). Desse modo, os resultados verificados para Tipo de verbo podem ser consequência dos efeitos verificados para a Função sintática.

O fato de adjuntos adverbiais e objetos indiretos favorecerem o emprego de *interrogativas qu-in-situ* revela que o constituinte interrogativo tende a ficar próximo de sua posição canônica na sentença: quanto maior seria o movimento, maior a tendência de realizá-lo *in situ*.

Grupo de fatores prosódico

O **Número de palavras na oração** é considerado um grupo de fatores prosódico uma vez que, quanto mais longa a sentença, pode-se prever que a estrutura de *interrogativas qu-in-situ* seja desfavorecida dada a dificuldade de manutenção de uma prosódia interrogativa. De fato, esse é o resultado que se verifica na Tabela 4.14: a forma *in situ* é desfavorecida quanto maior for a sentença (P.R. 0,42).

No entanto, como já se indicou na análise qualitativa, é possível que aqui a correlação seja inversa: a correlação observada pode significar que é a construção com constituinte pré-verbal que favorece o “alongamento” da sentença (p.ex. com o encaixamento de outras orações) ou, colocado de outro modo, que é a estrutura *in situ* que favorece a concisão da pergunta.

Tabela 4.14 Correlação entre *interrogativas qu-in-situ* e Número de palavras na oração^a

	P.R.	%	N <i>in-situ</i>	N total
Número de palavras na oração				
de 2 a 6 palavras	.54	27,3	183	670
de 7 a 20 palavras	.42	13,4	44	329
	<i>Range: 12</i>	22,7	227	999

^a *Input*: 0,115. $p < 0,05$.

Em uma análise que inverte o direcionamento da correlação – Número de palavras como variável dependente e Posição do constituinte interrogativo como variável independente – o efeito parece ser mais significativo ($range = 21$), visto que há um forte desfavorecimento de sentenças mais extensas com as *interrogativas qu-in-situ* (ver Tabela 4.15).

Tabela 4.15 Correlação entre Número de palavras na oração e Posição do constituinte interrogativo^a

	P.R.	%	N sentenças longas ^b	N total
Posição do constituinte interrogativo				
Q movido	.55	36,9	285	772
Q <i>in situ</i>	.34	19,4	44	227
	<i>Range = 21</i>	32,9	329	999

^a *Input*: 0,323. $p < 0,001$. Rodada *one-level*. ^b Sentenças com mais de 7 palavras.

A Tabela 4.16 organiza os grupos de fatores linguísticos e não-linguísticos selecionados como significativos para o emprego de *interrogativas qu-in-situ*, segundo a força de seu efeito.

Tabela 4.16 Grupos de fatores correlacionados ao emprego de *interrogativas qu-in-situ*

Grupo de fatores	Range	Natureza ^a	Fatores que favorecem qu-in-situ
Função sintática ^b	77	M	adjuntos adverbiais e objetos indiretos
Espontaneidade da pergunta	58	E/DP	perguntas de dados pessoais e espontâneas
Constituinte interrogativo ^c	56	M	o que / quem
“Sinceridade” e Sexo/gênero	47	DP/E	perg. estr. discurso(M/F), perg. sinceras(M), perg. retóricas(F)
Conjunto de respostas	38	DP	perguntas ‘fechadas’
Tipo de verbo	27	M	transitivo indireto e intransitivo
Grau de ativação do fundo comum	25	DP	pressuposição ou referente ativado na 1 ^a ou 2 ^a sentença anterior
Sinceridade pragmática da pergunta	23	DP	perguntas de estruturação do discurso
Número de palavras ^d	21	P	sentenças curtas
Ocorre outro elemento...?	19	DP	presença de tópico/oração subordinada
Grau de relação doc-inf	18	E	graus mais distantes
Faixa etária	17	E	3 ^a e 1 ^a faixa etária
Número de palavras ^e	12	P	sentenças curtas

^a M – morfossintática; E – extralinguística; DP – discursivo-pragmática; P – prosódica. ^bSem interação com Constituinte interrogativo.

^cSem interação com Função sintática. ^dNúmero de palavras como variável dependente. ^eNúmero de palavras como variável independente.

Os grupos de fatores morfossintáticos e discursivo-pragmáticos são aqueles que têm maior influência na variação. Quanto à morfossintaxe, a variação está correlacionada à distância de movimento do constituinte interrogativo (quanto maior seria o movimento, maior a tendência de o constituinte permanecer *in situ*) e à sua “obrigatoriedade” ou não na estrutura argumental (quanto menor a “obrigatoriedade”, maior a tendência de emprego de *qu-in-situ*).

Os grupos de fatores discursivo-pragmáticos mostram a importância do fluxo de informações no aqui-e-agora da conversação para a alternância entre diferentes estruturas sintáticas: *qu-in-situ* é favorecido em contextos em que há maior previsibilidade da pergunta e, de modo semelhante aos resultados de Weiner & Labov (1983 [1977]), em que a pressuposição ou um dos referentes está mais ativado; nesta pesquisa, atribui-se a seleção desses fatores ao papel do fundo comum entre os interlocutores, que deve ser examinado em sua dinamicidade: os interlocutores se baseiam em seu conjunto de crenças compartilhadas para organizar o discurso, o que evidencia o caráter cooperativo da conversação face-a-face (Clark, 1996).

Outros fatores da oralidade também se mostram relevantes. *Qu-in-situ* é favorecido por sentenças curtas, embora, neste caso, a correlação provavelmente seja inversa: a estrutura *in situ* é que favorece a concisão da oração.

Os limites entre fatores discursivo-pragmáticos (ou da fala-em-interação) e fatores extralinguísticos se diluem ao se examinar o fluxo da conversação: os resultados dos grupos de fatores Espontaneidade da pergunta, Grau de relação entre documentador e informante e Faixa etária podem ser atribuídos a questões de estilo (formal *vs.* informal, dissertativo/descritivo *vs.* narrativo) e ao fundo comum entre os interlocutores. Embora categorias sociais clássicas como Escolaridade e Sexo/gênero não sejam selecionados como significativos, o exame desse último em cruzamento com grupos de fatores discursivo-pragmáticos revela que o encaixamento social da variação sintática pode ocorrer de modo indireto (Cheshire, 2005). Esses resultados trazem novas questões à Teoria da Variação sobre a relação entre fatores linguísticos e não linguísticos para variáveis de diferentes níveis de análise.

5

A variável (Constituinte interrogativo pré-verbal)

O capítulo anterior apresentou os resultados da análise dos dados que integram o envelope de variação definido pela posição do constituinte interrogativo.

Neste capítulo, a análise se volta para o Envelope de Variação II, que engloba contextos em que *qu-in-situ* nem sempre é uma forma possível, mas em que todas as outras três formas com constituinte pré-verbal o são. A variável, portanto, compõe-se das variantes *-qu*, *qu-que* e *é-que*. Na análise quantitativa, deu-se especial atenção ao emprego de *interrogativas qu-que*, de modo que este capítulo tem o objetivo de explicitar os fatores favorecedores de seu emprego. Foram testados seis grupos de fatores extralinguísticos (Sexo/gênero, Faixa etária, Escolaridade, Quem fala: documentador ou informante, Espontaneidade da pergunta e Grau de relação entre documentador e informante) e cinco grupos de fatores linguísticos (Tipo de verbo, Tipo de oração (2), Função sintática, Constituinte interrogativo e Tonicidade da sílaba seguinte). Os resultados apontam para uma mudança linguística em progresso em favor das *interrogativas qu-que*, cujo uso é favorecido em orações principais e contextos em que haveria choque de acento entre sílabas tônicas.

5.1 Envelope de variação e Distribuição geral dos dados

A variável (Constituinte interrogativo pré-verbal) engloba o conjunto de dados do Envelope de Variação II:

Envelope de Variação II Variável (Constituinte interrogativo pré-verbal)

Sentenças completas que contêm apenas um constituinte interrogativo pré-verbal em uma oração finita principal ou encaixada, excetuando-se expressões cristalizadas ou semilexicalizadas.

Foram extraídas 1421 ocorrências das três formas que têm o constituinte interrogativo pré-verbal (*interrogativas-qu*, *interrogativas qu-que* e *interrogativas é-que*); dentre elas, as *interrogativas qu-que* são representadas por 579 dados (40,7%) e são a segunda forma mais frequente (ver Tabela 5.1).

Tabela 5.1 Distribuição de Interrogativas-Q no *corpus* de entrevistas sociolinguísticas no Envelope de Variação II e no Quadro Geral

	Envelope de Variação II		Quadro Geral	
	N	%	N	%
<i>Interrogativas-qu</i>	721	50,8	721	43,8
<i>Interrogativas qu-que</i>	579	40,7	579	35,1
<i>Interrogativas é-que</i>	121	8,5	121	7,3
<i>Interrogativas qu-in-situ</i>	N/A	N/A	227	13,8
Total	1421	100,0	1648	100,0

A alta produtividade de *interrogativas qu-que* na língua oral é bastante contrastante com a sua quase ausência na língua escrita, como se verificou nas amostras de jornal, de revista e de redações escolares (ver distribuição na Tabela 2.4, p. 35). A análise quantitativa desses dados, que investiga grupos de fatores extralinguísticos e linguísticos, busca dar conta desse fato e de outros fatores correlacionados ao seu emprego.

5.2 Grupos de fatores extralinguísticos

Foram testados seis grupos de fatores extralinguísticos: (i) Sexo/gênero; (ii) Faixa etária; (iii) Escolaridade; (iv) Quem fala: documentador ou informante; (v) Grau de relação entre documentador e informante; e (vi) Espontaneidade da pergunta.

Tabela 5.2 Distribuição de ocorrências de *interrogativas qu-que* em relação a Quem fala, Escolaridade, Sexo/gênero e Espontaneidade da pergunta^a

	P.R.	%	N qu-que	N total
Quem fala^b				
documentador	[.57]	37,6	213	448
informante	[.47]	47,5	366	973
			579	1421
Escolaridade^c				
até o ensino fundamental II	[.49]	39,8	192	483
ensino superior	[.51]	41,3	387	938
			579	1421
Sexo/gênero^d				
feminino	[.52]	42,6	367	862
masculino	[.47]	37,9	212	559
		40,7	579	1421
Espontaneidade da pergunta^e				
temática	[.57]	54,9	39	71
obrigatória	[.57]	54,6	53	97
sugerida	[.53]	50	11	22
espontânea	[.47]	44,9	92	205
de dados pessoais	[.36]	34,0	18	53
		47,5	213	448

^a [] indica a não-seleção dos grupos de fatores como significativos. Pesos relativos referem-se a rodadas *one-level*. ^b *Input* = 0,407. *p* < 0,01. ^c *Input* = 0,407. *p* = 0,601. ^d *Input*: 0,407. *p* = 0,085. ^e *Input* = 0,475. *p* < 0,01.

Assim como na análise da variável (Posição do constituinte interrogativo) (ver capítulo 4), os grupos de fatores Quem fala, Escolaridade e Sexo/gênero não são selecionados como significativos para a variação entre as formas

com constituinte interrogativo pré-verbal. Além desses, Espontaneidade da pergunta também não é selecionado nessa análise. A Tabela 5.2 apresenta a distribuição geral dos dados entre os fatores desses grupos e os pesos relativos de rodadas *one-level*.

O fato de o grupo de fatores **Quem fala** não ser selecionado indica que o papel desempenhado pelo falante na situação de entrevista sociolinguística não tem uma influência significativa para a alternância. A não-seleção de **Escolaridade**, **Sexo/gênero** e **Espontaneidade da pergunta**, por sua vez, deve ser reflexo do fato de nenhuma das formas *-qu*, *qu-que* ou *é-que* ser especificamente avaliada positiva ou negativamente no português paulistano atual. Em outras palavras, nenhuma dessas formas parece estar sujeita a um monitoramento por parte dos falantes no sentido de evitá-las. Note-se, no entanto, que apesar de Sexo/gênero não ser selecionado em uma análise multivariada, as mulheres revelam um leve favorecimento da estrutura *qu-que* em rodada *one-level*.

Os grupos de fatores Grau de relação entre documentador e informante e Faixa etária, por outro lado, são selecionados como significativos.

Tabela 5.3 Correlação entre *interrogativas qu-que* e grupos de fatores extralinguísticos^a

	P.R.	%	N qu-que	N total
Grau de relação entre documentador e informante				
+distante	.53	42,4	459	1082
+próximo	.41	35,4	120	339
	<i>Range = 12</i>	40,7	579	1421
Faixa etária				
de 20 a 30 anos	.56	45,8	291	635
de 35 a 45 anos	.44	35,2	127	361
mais de 50 anos	.47	37,9	161	425
	<i>Range = 12</i>	40,7	579	1421

^a Input: 0,400. $p < 0,02$. Log likelihood = -795,089. + indica que houve amalgamação de fatores.

Em **Grau de relação entre documentador e informante** (ver Tabela

5.3), foram amalgamados os graus de relação mais distantes, de um lado, e os graus de relação mais próximos, de outro ($\chi^2 = 4,77(4)$, $p > 0,30$). As *interrogativas qu-que* são favorecidas em interações conversacionais entre falantes com graus de relação mais distantes (P.R. 0,53), de modo semelhante ao que ocorre com o emprego de *interrogativas qu-in-situ* (ver capítulo 4, seção 4.2, p. 91). Esse resultado, no entanto, contraria a expectativa de que esse grupo de fatores não seria selecionado no caso da variável (Constituinte interrogativo pré-verbal). Em princípio, esse grupo foi proposto para avaliar o efeito do fundo comum (Clark, 1996) no emprego de Interrogativas-Q e a correlação observada diria respeito ao fluxo de informações no aqui-e-agora da conversa. Não parece plausível aplicar a mesma explicação para o emprego de *interrogativas qu-que*; todavia, é interessante notar o mesmo padrão de correlação para ambas as variáveis.

Em **Faixa etária**, o padrão de emprego de *interrogativas qu-que* aponta para uma possível mudança linguística em progresso, já que os falantes mais jovens tendem a favorecê-las (P.R. 0,56) em relação aos falantes mais velhos (P.R.s 0,44 e 0,47 para a segunda e terceira faixa etária, respectivamente). A proximidade de pesos relativos para ambos os grupos de falantes mais velhos permite uma análise mais concisa que contrasta apenas duas faixas etárias: aqueles com menos ou mais de 35 anos ($\chi^2 = 0,31(1)$, $p > 0,50$. Ver Tabela 5.4).

Tabela 5.4 Correlação entre *interrogativas qu-que* e Faixa etária recodificado^a

	P.R.	%	N qu-que	N total
Faixa etária				
de 20 a 30 anos	.56	45,8	291	635
+mais de 35 anos	.45	36,6	288	786
	<i>Range = 11</i>	40,7	579	1421

^a *Input* = 0,400. $p < 0,02$. *Log likelihood*: -795,246. + indica que houve amalgamação de fatores. $\chi^2 = 0,31(1)$, $p > 0,50$.

Pelo fato de que esta análise contrasta o emprego de *interrogativas qu-que*

com duas outras variantes, *-qu* e *é-que*, também interessa examinar se a provável mudança em progresso está ocorrendo em detrimento de ambas as formas, ou apenas uma delas. Para isso, foram feitas duas novas análises: na primeira, a estrutura *qu-que* foi contrastada apenas a forma *-qu*; no segunda, apenas com a forma *é-que* (ver Tabela 5.5).¹ Os resultados de ambas as rodadas indicam igualmente uma mudança em curso em favor das *interrogativas qu-que*.

Tabela 5.5 Correlação entre *interrogativas qu-que* e Faixa etária em duas análises contrastivas

	<i>vs. -qu</i> ^a		<i>vs. é-que</i> ^b	
	P.R.	%	P.R.	%
Faixa etária	<i>N</i> = 1300		<i>N</i> = 700	
de 20 a 30 anos	.55	48,6	.61	89,0
de 35 a 45 anos	.46	38,7	.40	79,4
mais de 50 anos	.46	43,2	.40	75,6
<i>Range</i>	9	44,5	21	82,7

a Input: 0,438. *p* < 0,02. *b* Input: 0,843. *p* < 0,02.

Dentre essas duas análises, as *interrogativas qu-que* são mais fortemente favorecidas por falantes mais jovens no contraste com as *interrogativas é-que* (P.R. 0,61) do que no contraste com as *interrogativas-qu* (P.R. 0,55). Esse resultado pode explicar a retração mais rápida de *interrogativas é-que* do que as *interrogativas-qu* em comparação com as frequências de uso da década de 1970 (ver Tabela 1, p. 7, Introdução). Embora os *corpora* utilizados naquele levantamento de frequências sejam diferentes daquele usado neste estudo por serem menores e só abarcarem falantes mais escolarizados, eles revelavam uma queda da taxa de *é-que* (21,8% para 8,1%, ou cerca de um terço) maior do que na taxa de *-qu* (de 56,0% para 40,5%, ou cerca de um décimo).

¹ Todas as rodadas foram feitas em análises multivariadas que incluíram outros 5 grupos de fatores (Sexo/gênero, Escolaridade, Grau de relação entre documentador e informante, Tipo de oração (2) e Tonicidade da sílaba seguinte). Não se incluíram todos os grupos de fatores devido às interações que surgem ao analisar um subconjunto do total de dados. Em todas as análises, manteve-se o mesmo arquivos de condições, exceto pelas mudanças na variável dependente.

É importante lembrar que o modelo ideal para o estudo de mudança em progresso é a conjugação de observações no tempo aparente, como se faz neste estudo, e observações no tempo real, com *corpora* de diferentes períodos (Labov, 1994); contudo, a verificação de aumento de frequência no emprego de *interrogativas qu-que* entre os anos 1970 e 2000 e das tendências de uso na comunidade de fala paulistana (verificadas através de pesos relativos) traz boas evidências de uma mudança em curso.

Assim como na análise do capítulo anterior, foram feitos cruzamentos dos grupos de fatores Faixa etária e Sexo/gênero com Quem fala, já que a inclusão de dados de documentadores torna a amostra desbalanceada quanto ao perfil sociolinguístico dos falantes. No entanto, tais rodadas apresentam resultados semelhantes à análise sem o cruzamento de fatores: tanto documentadores quanto informantes de primeira faixa etária favorecem o emprego de *interrogativas qu-que*, e o cruzamento dos grupo de fatores Quem fala e Sexo/gênero não é selecionado como significativo para a variação (ver Tabela 5.6).

Os resultados de Faixa etária, aliados aos da análise de Sexo/gênero, Escolaridade e Espontaneidade da pergunta – esses três últimos, não selecionados como significativos –, indicam que se trata de uma mudança linguística de baixo (Labov, 2001), inconsciente para os falantes: as *interrogativas qu-que* vêm se implementando progressivamente na língua falada e tal mudança não parece estar sujeita à avaliação (positiva ou negativa), tampouco ao monitoramento dos falantes. Embora Sexo/gênero não seja selecionado como significativo, a porcentagem mais alta de emprego de *interrogativas qu-que* pelas mulheres (42,6% *vs.* 37,9%; ver Tabela 5.2, p. 119) aproxima o padrão de variação das interrogativas de constituinte pré-verbais daquele já reportado consistentemente na literatura sociolinguística para variáveis fonológicas. Esse padrão é resumido por Labov (2001:292) em um de seus princípios de mudança linguística: “Em mudanças linguísticas de baixo, as mulheres empregam formas inovadoras com frequências mais altas do que os homens.” (Tradução própria.²)

² Texto original: “In linguistic change from below, women use higher frequencies of innovative forms than men do.”

Tabela 5.6 Correlação entre *interrogativas qu-que* e os cruzamentos de Quem fala-Faixa etária e Quem fala-Sexo/gênero^a

	P.R.	%	N qu-que	N total
Quem fala e Faixa etária^b				
documentadores da 1ª faixa etária	.56	51,1	170	333
informantes da 1ª faixa etária	.56	40,1	121	302
informantes da 3ª faixa etária	.47	37,6	130	346
informantes da 2ª faixa etária	.46	35,4	115	325
documentadores da 3ª faixa etária	.41	39,2	31	79
documentadores da 2ª faixa etária	.35	33,3	12	36
		40,7	579	1421
Quem fala e Sexo/gênero^c				
documentadores do sexo masculino	[.57]	48,1	74	154
documentadores do sexo feminino	[.57]	47,5	139	294
informantes do sexo feminino	[.50]	40,1	228	568
informantes do sexo masculino	[.43]	34,1	138	405
		40,7	579	1421

^a [] indica a não-seleção do grupo de fatores. Divergência entre P.R.s e porcentagens nos valores sombreados indica interação. ^b Input = 0,391. $p < 0,05$. ^c Input = 0,406.

5.3 Grupos de fatores linguísticos

Foram testados cinco grupos de fatores linguísticos: Tipo de verbo, Tipo de oração (2), Função sintática, Constituinte interrogativo e Tonicidade da sílaba seguinte. Dentre eles, **Tipo de verbo** não é selecionado como significativo (ver Tabela 5.7), diferentemente do que havia ocorrido com a variável (Posição do constituinte interrogativo).

Em **Tipo de oração (2)**, foram amalgamados os subtipos de oração subordinada (oração subordinada a verbo, a substantivo, a adjetivo e oração-constituente de *small clause*) devido ao número reduzido de dados para cada fator, conforme já se assinalou no capítulo 3: não houve qualquer ocorrência de *interrogativas qu-que* e *é-que* subordinadas a adjetivos, tampouco de *interrogativas é-que* em oração-constituente de *small clause*.³

³ Aqui não se apresenta um teste de qui-quadrado, pois o teste da hipótese não era possível sem a

Tabela 5.7 Distribuição de ocorrências de *interrogativas qu-que* em relação a Tipo de verbo^a

	P.R.	%	N qu-que	N total
Tipo de verbo				
transitivo direto	[.55]	45,7	283	619
intransitivo	[.48]	38,4	66	172
de ligação	[.46]	36,8	161	438
bitransitivo	[.46]	36,7	11	30
transitivo indireto	[.45]	35,9	42	117
modal	[.45]	35,6	16	45
		40,7	579	1421

^a *Input* = 0,407. [] indica a não-seleção do grupo de fatores. Pesos relativos correspondem a rodada *one-level*.

Os resultados da Tabela 5.8 mostram que *interrogativas qu-que* são favorecidas quando se encontram em orações principais (P.R. 0,61) e, inversamente, desfavorecidas quando em orações subordinadas (P.R. 0,38). Bybee (2002), em um artigo cujo título já é bastante sugestivo (“Orações principais são inovadoras, orações subordinadas são conservadoras”), fornece pistas para dar conta desse resultado. Tendo em vista que o emprego de *interrogativas qu-que* parece estar se expandindo em tempo aparente, o que se verifica na Tabela 5.8 é justamente a liderança das orações principais no provável processo de mudança.

Através da análise de uma série de casos em que se verifica maior conservadorismo em orações subordinadas, tanto em questões sintáticas como morfológicas (ordem SVO-SOV em inglês e alemão, subjuntivo em *-ra vs.* construção perifrástica do mais-que-perfeito em espanhol, presente sintético *vs.* perifrástico em armênio, entre outros), Bybee (2002) apresenta explicações de natureza discursivo-pragmática e cognitiva para a diferença no ritmo da mudança entre orações principais e subordinadas – em alguns casos, aliás, a mudança não prossegue paralelamente e orações subordinadas se especializam

amalgamação.

Tabela 5.8 Correlação entre *interrogativas qu-que* e Tipo de oração (2)^a

	P.R.	%	N qu-que	N total
Tipo de oração (2)				
principal	.61	52,8	408	772
+subordinada	.38	26,3	171	649
	<i>Range = 23</i>	40,7	579	1421

^a *Input* = 0,400. $p < 0,02$. + indica que houve amalgamação de fatores.

com a forma mais antiga. A autora sugere que orações principais tenderiam a apresentar relações pragmáticas mais complexas por ser o local privilegiado para topicalizações, ênfases e construções presentativas. Oações subordinadas, por outro lado, seriam processadas em “blocos” relativamente maiores, o que torna seus constituintes menos independentes e, portanto, menos suscetíveis à mudança; além disso, orações subordinadas seriam cognitivamente mais complexas por conter informações mais detalhadas, fato que diminuiria a probabilidade de emprego de inovações linguísticas. Essas considerações têm consequência para uma teoria gramatical na medida em que, segundo Bybee (2002), orações principais e subordinadas podem seguir princípios relativamente independentes entre si.

Os grupos de fatores Função sintática e Constituinte interrogativo foram analisados em rodadas distintas, dada a já mencionada interação entre eles. Em **Função sintática**, os adjuntos adverbiais de causa e finalidade foram amalgamados, assim como os de lugar e modo a outros adjuntos adverbiais ($\chi^2 = 0,49(3)$, $p > 0,90$). Os resultados da Tabela 5.9, no entanto, são difíceis de interpretar: não parece haver um padrão de emprego em relação a argumentos ou adjuntos adverbiais semelhantes entre si; as *interrogativas qu-que* são favorecidas por adjuntos adverbiais de causa, finalidade e tempo, assim como por objetos diretos, ao mesmo tempo em que são desfavorecidas por outros tipos de argumentos (sujeitos, constituintes de *small clause*, objetos indiretos) e outros adjuntos adverbiais. Ademais, o número extremamente

reduzido de adjuntos adverbiais de tempo (total de 16, dos quais 10 são *qu-que*) prejudica a interpretação do peso verificado para esse fator. De fato, uma análise que faz um contraste simples entre argumentos e adjuntos adverbiais resulta na não-seleção de Função sintática como significativo.

Tabela 5.9 Correlação entre *interrogativas qu-que* e Função sintática

	P.R.	%	N <i>qu-que</i>	N total
Função sintática – detalhado^a				
+adj. adv. causa e finalidade	.75	70,9	90	127
adj. adv. tempo	.67	62,5	10	16
objeto direto	.58	47,2	210	445
sujeito	.49	35,4	46	130
constituente de <i>small clause</i>	.45	37,4	128	342
objeto indireto	.36	25,3	25	99
+outros adjuntos adverbiais	.35	26,7	70	262
	<i>Range = 40</i>	40,7	579	1421
Função sintática – argumentos <i>vs.</i> adjuntos^b				
adjuntos adverbiais	[.50]	40,8	179	439
argumentos	[.50]	40,7	400	982
		40,7	579	1421

^a *Input* = 0,398. $p < 0,02$. + indica que houve amalgamação de fatores. $\chi^2 = 0,49(3)$, $p > 0,90$. Divergência entre P.R.s e porcentagens nos valores sombreados indica interação. ^b [] indica a não-seleção do grupo de fatores em rodada multivariada. Pesos relativos referem-se a rodada *one-level*. *Input* = 0,407. $p = 0,99$.

A análise do grupo de fatores **Constituente interrogativo** pode explicar a aparente falta de padrão acima observada. Na Tabela 5.10, verifica-se que *interrogativas qu-que* são favorecidas pelos constituintes (*o*) *que*, *por que* e *pra que* e são desfavorecidas pelos demais. Os pesos relativos elevados para esses constituintes já podem explicar o favorecimento de adjuntos adverbiais de causa (= *por que*) e finalidade (= *pra que*) e possivelmente explicam o favorecimento de objeto direto (muitas vezes na forma (*o*) *que*). Esses quatro constituintes interrogativos têm em comum o fato de serem oxítonos, diferentemente dos demais: *como*, *quanto*, *quando* e *onde* são paroxítonos, assim como a maioria dos constituintes interrogativos seguidos de um NP,

já que a maioria dos nomes em português é paroxítona. A hierarquia de constituintes interrogativos na Tabela 5.10 revela uma forte divisão entre palavras oxítonas e paroxítonas. Seria o efeito verificado nos grupos de fatores Função sintática e Constituinte interrogativo prosódico?

Tabela 5.10 Correlação entre *interrogativas qu-que* e grupos de fatores linguísticos^a

	P.R.	%	N qu-que	N total
Constituinte interrogativo				
que	.90	87,8	122	139
+por que/pra que	.77	72,4	89	123
o que	.50	36,4	156	428
+qual/qual + NP	.47	43,5	50	115
que + NP	.40	33,7	29	86
como	.39	30,6	85	278
quanto/quanto + NP	.33	23,5	16	68
quem	.31	21,2	21	99
onde/quando	.17	12,9	11	85
	Range = 60	40,7	579	1421
Tonicidade da sílaba seguinte				
tônica	.53	41,7	402	964
+átona/pausa	.45	38,7	177	457
	Range = 8	40,7	579	1421

^a Input = 0,400. p < 0,02. + indica que houve amalgamação de fatores.

O grupo de fatores **Tonicidade da sílaba seguinte** tem justamente o objetivo de testar se o emprego de *interrogativas qu-que* – ou, mais especificamente, a presença de um complementizador *que* átono após o constituinte interrogativo – pode funcionar como uma estratégia de evitar o choque de acento entre duas sílabas tônicas, seguindo o Princípio de Alternância Rítmica (Selkirk, 1984). Os resultados conduzem a uma confirmação da hipótese, visto que constituintes interrogativos seguidos de sílabas tônicas favorecem o emprego de *interrogativas qu-que* (P.R. 0,53) (1-a) em relação àqueles seguidos de sílabas átonas (P.R. 0,45) (1-b).

- (1) a. Cecília: (...) **por QUE que VAle a PEEna?** porque nós raramente vamos ao cinema durante o ano... (F2S-INF)
 b. Jefferson: mas **QUEM que manTÉM o Rio de JaNEIro?** (M2G-INF)

Todavia, uma análise mais detalhada também deve levar em conta a tonicidade da sílaba anterior, uma vez que nem todos os constituintes interrogativos terminam em sílaba tônica e, portanto, favoreceriam a presença de uma sílaba átona intermediária a fim de evitar o choque de acento. Assim, realizou-se um cruzamento entre os grupos de fatores Constituintes interrogativo e Tonicidade da sílaba seguinte, que reagrupou as ocorrências de acordo com quatro possibilidades combinatórias: (i) constituintes interrogativos oxítonos ((o) *que*, *por que*, *pra que*, *quem* e *qual*) seguidos de sílaba tônica (2-a); (ii) constituintes interrogativos oxítonos seguidos de sílaba átona ou pausa (2-b); (iii) constituintes interrogativos paroxítonos (*como*, *quanto*, *quando*, *onde*, *que + NP*, *qual + NP*, *quanto + NP*) seguidos de sílaba tônica (2-c); e (iv) constituintes interrogativos paroxítonos seguidos de sílaba átona ou pausa (2-d). Os resultados dessa análise se encontram na Tabela 5.11.

- (2) a. Fernanda: **QUE que POde te chaMAR a atenÇÃO?** (F1S-DOC)
 b. Maíra: **por QUE que as pesSOAS COMpram MUIto CARro MESmo pasSANDo FOme?** (F1S-DOC)
 c. Valter: esse eu não sei **COmo que FAZ** mas um um copo já dava pra fazer... (M3G-INF)
 d. Dora: **COmo que o goVERno não TOMa conheciMENTo de uma COIsa DESsa** né? (F3S-INF)

Em contraste com Tonicidade da sílaba seguinte (*range* = 8), o novo grupo de fatores é selecionado com um *range* muito maior, de 36. Verifica-se que *interrogativas qu-que* são favorecidas por constituintes interrogativos oxítonos seguidos de sílaba tônica (P.R. 0,62), contexto que em si representa um choque de acento; no outro extremo, *interrogativas qu-que* são fortemente desfavorecidas por constituintes interrogativos paroxítonos seguidos de sílaba átona ou pausa (P.R. 0,26), que representam uma sequência de sílabas átonas. Os fatores com pesos relativos intermediários, constituintes interrogativos

Tabela 5.11 Correlação entre *interrogativas qu-que* e o cruzamento de Constituinte Interrogativo e Tonicidade da Sílabas Seguinte^a

	P.R.	%	N qu-que	N total
Constituinte interrogativo e Tonicidade da sílabas seguinte				
Oxítono + sílaba tônica	.62	48,8	293	601
Oxítono + sílaba átona/pausa	.55	48,8	138	283
Paroxítono + sílaba tônica	.40	30,0	109	363
Paroxítono + sílaba átona/pausa	.26	22,4	39	174
<i>Range = 36</i>		40,7	579	1421

^a *Input* = 0,419. $p < 0,01$.

oxítonos seguidos de sílaba átona ou pausa (P.R. 0,55) e constituintes interrogativos paroxítonos seguidos de sílaba tônica (P.R. 0,40), indicam que a tonicidade da sílaba *anterior* (oxítona) é mais importante do que a tonicidade da sílaba seguinte (tônica) para a presença do complementizador *que*.

A Tabela 5.12 organiza os grupos de fatores linguísticos e não-linguísticos selecionados como significativos para o emprego de *interrogativas qu-que*, de acordo com o seu *range*.

A análise do grupo de fatores Faixa etária aponta para uma possível mudança linguística em progresso, observada em tempo aparente, em favor das *interrogativas qu-que*. Aqui não se desenvolve uma análise diacrônica a fim de confirmar essa tendência; entretanto, é possível entrever algum dos fatores relevantes para a variação. O padrão verificado segue uma tendência geral identificada por Bybee (2002) no curso de mudanças linguísticas, em que a implementação da variante inovadora tende a ocorrer mais rapidamente em orações principais do que em orações encaixadas.

A covariação com grupos de fatores morfossintáticos (Constituinte interrogativo e Função sintática) não demonstra uma hierarquia de restrições coerente, em razão do efeito indireto da tonicidade da última sílaba do

Tabela 5.12 Grupos de fatores correlacionados ao emprego de *interrogativas qu-qu*

Grupo de fatores	Range	Natureza^a	Fatores que favorecem qu-qu
Constituinte interrogativo	60	M/P	(o) <i>que, por que, pra que</i>
Função sintática	40	M/P	adj. adv. causa, finalidade, tempo; objeto direto
Const. interrogativo/Tonicidade	36	P	Q oxítono+sílaba tônica e Q oxítono+sílaba átona ou pausa
Tipo de oração (2)	23	M	orações principais
Grau de relação doc-inf	12	E	graus mais distantes
Faixa etária	11	E	mais jovens
Tonicidade da sílaba seguinte	8	P	sílabas tônicas

^a M – morfossintática; P – prosódica; E – extralinguística.

constituente interrogativo (oxítono *vs.* paroxítono). O cruzamento entre Constituinte interrogativo e Tonicidade da sílaba seguinte mostra que o emprego de *interrogativas qu-que* é favorecido em contextos em que haveria choque de acento (constituíntes oxítonos seguidos de sílabas tônicas) e desfavorecido em sequências de sílabas átonas (constituente paroxítono seguido de sílaba átona). Esse resultado sugere que a presença de um complementizador *que* átono obedece ao Princípio de Alternância Rítmica (Selkirk, 1984).

Conclusão

*But different folk have different views:
I know a person small–
She keeps ten million serving-men,
Who get no rest at all!
She sends ‘em abroad on her own affairs,
From the second she opens her eyes–
One million Hows, two million Wheres,
And seven million Whys!*

R. Kipling. *The Elephant’s Child*.

Esta pesquisa investiga o emprego variável de quatro formas interrogativas de constituinte de acordo com a teoria e os métodos da Sociolinguística Variacionista. Trata-se de uma análise que se volta para a língua em uso pela comunidade de fala paulistana e que parte da premissa de que o sistema linguístico é, a um tempo, heterogêneo e ordenado.

Após mais de 40 anos de estudos sociolinguísticos, é possível traçar amplas generalizações a respeito da distribuição e encaixamento linguístico e social de variáveis fonológicas (Labov, 1994, 2001; Labov *et al.*, 2006). O mesmo ainda não se pode afirmar de variáveis morfossintáticas, discursivas e pragmáticas,

uma vez que a análise de variáveis além do nível fonológico levanta uma série de questões de ordem teórica e metodológica para sua investigação e interpretação. Há muitas perguntas a serem respondidas; espera-se que este estudo contribua para o encaminhamento de respostas a algumas dessas questões.

No capítulo 1, apresentaram-se quatro problemáticas pertinentes ao estudo da variação morfossintática: a questão da equivalência semântica, a definição do envelope de variação, a relativa escassez de dados e a falta de correlação com grupos de fatores sociais. Essas questões são agora discutidas à luz das propostas e resultados desta pesquisa.

Para as Interrogativas-Q, a existência de pelo menos quatro estruturas em alternância, assim como o fato de elas desempenharem diferentes funções discursivas (perguntas e asserções), apresenta desafios analíticos. Nesta pesquisa, estabeleceu-se um critério formal como ponto de partida – o estudo de *formas interrogativas* – e empregou-se o conceito de *pressuposição do falante* como critério verificador da equivalência semântica entre as formas variantes. O conceito de Stalnaker (2002) revela-se particularmente frutífero para o estudo da língua falada por uma comunidade real, uma vez que leva em conta o ato comunicativo em seu contexto de uso, os seus participantes e a competência comunicativa dos falantes.

A análise qualitativa dessas estruturas definiu dois envelopes de variação, cada qual voltado ao exame de uma variante em particular (as *interrogativas qu-in-situ* e as *interrogativas qu-que*). Ambos os envelopes de variação foram delimitados a partir do conceito de *competência comunicativa* (Hymes, 1991 [1979]): as formas são consideradas opcionais se são possíveis, factíveis, adequadas e empregadas nos mesmos contextos.

Esses critérios, propostos aqui para o estudo de uma variável sintática e outra morfossintática, possivelmente são aplicáveis a qualquer estudo de variação. Se o objetivo do paradigma quantitativo é modelar a competência comunicativa dos falantes (Sankoff, 1980), os modelos quantitativos traduzidos em regras variáveis devem abarcar esse conhecimento em todas as suas dimensões, não apenas no nível gramatical. O modelo de análise qualitativa aqui proposto tem a vantagem de reunir, em um dos conceitos mais centrais

dos estudos sociolinguísticos, técnicas já empregadas mas esparsas na caixa de ferramentas variacionista: a análise de gramaticalidade (Labov, 1969; Weiner & Labov, 1983 [1977]); a análise de viabilidade (Kroch, 1987); a análise de adequação pragmática (Rickford *et al.* 2007 apud Buchstaller 2009); e a análise de produtividade (Wolfram, 1993; Tagliamonte, 2006). É certo que cada um desses critérios pode ser mais ou menos relevante a depender do fenômeno variável em estudo – por exemplo, questões de gramaticalidade podem ser mais prementes em variáveis sintáticas; questões de adequação pragmática em variáveis discursivas; questões de produtividade em variáveis fonológicas. A aplicabilidade ou pertinência dos quatro critérios em diferentes tipos de variáveis é uma questão de natureza empírica.

A relativa escassez de ocorrências das variantes foi sanada através do exame de um *corpus* robusto, composto de amostras de língua oral e escrita que somam cerca de um milhão de palavras. Para essa tarefa, a utilização de uma ferramenta de extração automática de dados, o programa R, foi fundamental. É evidente que quantidade não é sinônimo de qualidade, e que o emprego de ferramentas de automatização de certas tarefas da análise sociolinguística não exime o pesquisador de um olhar atento às características de seus dados, como se desenvolveu neste estudo. Entretanto, o emprego de rotinas automatizadas através de concordanceadores e outros programas como R, Python, Praat e Plotnik (Cf. p.ex. Labov *et al.* 2006; Hall 2008), além de uma preocupação em integrar as etapas do estudo sociolinguístico (coleta, transcrição, codificação etc.) (Cieri & Strassel, 2010), tem se tornado cada vez mais comum e abre a perspectiva de grandes avanços para o campo.⁴ Como lembra Chambers (2002), o estudo da variação se tornou possível em parte devido ao advento de ferramentas tecnológicas; novos avanços também são tributários desse mesmo fato.

As hipóteses levantadas neste estudo, estabelecidas na forma de grupos de fatores, procuraram investigar fatos extralinguísticos, discursivo-pragmáticos, sintáticos, morfológicos e prosódicos que explicassem o emprego variável de interrogativas de constituinte. Um resumo dos grupos de fatores analisados

⁴ Aqui também a análise de variáveis fonológicas parece estar à frente do exame de outros tipos de variáveis.

para cada uma das variáveis definidas nos envelopes de variação I e II se encontra na Tabela 5.13.

Grupos de fatores morfossintáticos (Constituinte interrogativo, Tipo de verbo e Função sintática) se encontram entre os mais relevantes para a posição do constituinte interrogativo. A variação é sensível à distância do movimento do constituinte (*qu-in-situ* é mais favorecido quanto maior seria o movimento) e à sua obrigatoriedade ou não na estrutural argumental (*qu-in-situ* é favorecido por adjuntos adverbiais). A extensão da sentença (Número de palavras na oração) também se mostra importante, embora a correlação neste caso provavelmente seja inversa: é a estrutura de *qu-in-situ* que favorece a concisão da sentença, e não necessariamente o contrário.

A análise da variação entre formas de constituinte interrogativo pré-verbal revelou uma provável mudança linguística em progresso, de acordo com o construto de mudança em tempo aparente, em favor das *interrogativas qu-que*. A correlação verificada com grupos de fatores morfossintáticos (Constituinte interrogativo e Função sintática) pode estar, na verdade, condicionada a fatores prosódicos (Tonicidade da sílaba anterior e seguinte); o emprego de *qu-que* obedece a tendências mais gerais do Princípio de Alternância Rítmica (Selkirk, 1984). A provável mudança em progresso também segue um padrão mais geral, identificado por Bybee (2002), de preferência por orações principais em relação a orações encaixadas (Tipo de oração (2)) na implementação da variante inovadora.

Diferentemente daquilo que em geral se observa para variáveis fonológicas, grupos de fatores sociais não apresentaram fortes correlações com as variáveis sintática e morfossintática deste estudo. Antes de representar um impedimento para o desenvolvimento de análises sociolinguísticas de variáveis além do nível fonológico, esse fato coloca a Teoria da Variação diante da tarefa de comparar o comportamento de variantes de variáveis de diferentes níveis linguísticos. Embora amplas categoriais sociais como Sexo/gênero e Escolaridade não se correlacionem diretamente com as interrogativas de constituinte, outros fatores extralinguísticos, ligados a questões da fala-em-interação, revelaram influência no emprego de diferentes estruturas morfossintáticas. O Grau de relação entre o informante e o documentador mostrou-se relevante

Tabela 5.13 Resumo dos grupos de fatores analisados^a

Grupos de fatores	Envelope de variação I	Envelope de variação II
Extralinguísticos	Sexo/gênero	Sexo/gênero
	Faixa etária	Faixa etária
	Escolaridade	Escolaridade
	Quem fala	Quem fala
Discursivo-pragmáticos	Grau de relação DOC-INF	Grau de relação DOC-INF
	*Espontaneidade da pergunta	*Espontaneidade da pergunta
	Que outro elemento ocorre...?	
	Sinceridade pragmática da pergunta	
Morfofossintáticos	*Conjunto de respostas previstas	
	*Grau de ativação do fundo comum	
	Constituinte interrogativo	Constituinte interrogativo
	Tipo de verbo	Tipo de verbo
Prosódicos	Função sintática	Função sintática
	Tipo de oração (2)	Tipo de oração (2)
	Número de palavras	Tonicidade da sílaba seguinte
	Quem fala – Faixa etária	Quem fala – Faixa etária
Cruzamentos de grupos de fatores	Quem fala – Sexo/gênero	Quem fala – Sexo/gênero
	Sexo/gênero – Sinceridade pragmática	Constituinte interrogativo – Tonicidade
	Sexo/gênero – Grau de ativação	
	Sexo/gênero – Conjunto de respostas	
	Sexo/gênero – Que outro elemento ocorre...?	

^a **Negrito** indica GFs selecionados como significativos. * indica GFs que se referem a subconjuntos de dados.

para a variação em ambas as variáveis, e a Espontaneidade da pergunta para (Posição do constituinte interrogativo). Ainda que nenhuma das variantes seja estigmatizada na língua oral, o estilo de fala e a constante atualização do fundo comum entre os interlocutores são fatores importantes. Este estudo fornece exemplos de como aspectos da interação face-a-face e do aqui-e-agora da conversa podem ser operacionalizados e testados através de grupos de fatores (Sinceridade pragmática da pergunta, Conjunto de respostas previstas, Grau de ativação do fundo comum).

O cruzamento de Sexo/gênero com grupos de fatores discursivo-pragmáticos, sobretudo Sinceridade pragmática da pergunta, dá indícios de que o encaixamento da variação sintática pode se dar de maneira indireta na estrutura social, resultado que reitera as conclusões de [Cheshire \(2005\)](#) em seu estudo sobre variação em expressões sinalizadoras de novas entidades no discurso em três cidades inglesas. Se a relação com fatores extralinguísticos se dá de maneira indireta, não surpreende o fato de que tal padrão de variação não esteja disponível à percepção e avaliação dos falantes da língua. Ainda resta, no entanto, integrar explicações para a diferenciação de gênero em relação à alternância entre estruturas sintáticas.

Referências Bibliográficas

- ALMEIDA, Napoleão Mendes de. *Gramática Metódica da Língua Portuguesa*. São Paulo: Saraiva, 14 ed., 1962.
- AMBAR, Manuela, KATO, Mary A., MIOTO, Carlos & VELOSO, Rita. “Padrões de interrogativas-Q no português europeu e no português brasileiro: uma análise inter e intra-linguística.” *In: Boletim da Associação Brasileira de Linguística*, vol. 26, pp. 400–404, 2001.
- BAYLEY, Robert. “The quantitative paradigm.” *In: CHAMBERS, J.K., TRUDGILL, P. & SCHILLING-ESTES, N. (Eds.), The Handbook of Language Variation and Change*, pp. 117–141. Malden, MA: Blackwell, 2002.
- BECHARA, Evanildo. *Moderna gramática portuguesa*. Rio de Janeiro: Lucerna, 37 ed., 2005.
- BUCHSTALLER, Isabelle. “The quantitative analysis of morphosyntactic variation: constructing and quantifying the denominator.” *Language and Linguistic Compass*, vol. 3(4), 1010–1033, 2009.
- BYBEE, Joan. “Main clauses are innovative, subordinate clauses are conservative.” *In: Complex sentences in Grammar and Discourse. Essays in honor of Sandra A. Thompson.*, pp. 1–17. Amsterdam/Philadelphia: John Benjamins, 2002.
- CHAFE, Wallace. “Language and consciousness.” *Language*, vol. 50(50), 111–133, 1974.

- CHAMBERS, J.K. “Studying language variation: an informal epistemology.”
In: CHAMBERS, J.K., TRUDGILL, P. & SCHILLING-ESTES, N. (Eds.), The Handbook of Language Variation and Change, pp. 3–13. Malden, MA: Blackwell, 2002.
- CHENG, Lisa Lai-Shen. *On the Typology of Wh-questions*. Tese de doutorado. MIT, 1991.
- CHENG, Lisa Lai-Shen & ROORYCK, Johan. “Licensing wh-in-situ.” *Syntax*, vol. 3(1), 1–19, 2000.
- CHESHIRE, Jenny. “Syntactic variation and beyond: Gender and social class variation in the use of discourse-new markers.” *Journal of Sociolinguistics*, vol. 9(4), 479–508, 2005.
- CIERI, Christopher & STRASSEL, Stephanie. “Towards best practices in sociophonetics: Robust, digital, empirical, reproducible sociolinguistic methodology.”, 2010.
 URL http://papers.ldc.upenn.edu/NWAV2010/Robust_Digital_Empirical_Reproducible_Sociolinguistic_Methodology_Final.ppt
- CLARK, Herbert. *Using language*. Cambridge University Press, 1996.
- COVENEY, Aidan. *Variability in Spoken French: a sociolinguistic study of interrogation and negation*. Elm Bank, 2nd ed., 2002 [1996].
- CUNHA, Celso & CINTRA, Lindley. *A nova gramática do português contemporâneo*. Rio de Janeiro: Lexikon Informática, 3 ed., 2007.
- DUARTE, Maria Eugênia Lamoglia. “A perda da ordem V(erbo) S(ujeito) em interrogativas qu- no português do Brasil.” *DELTA*, vol. 8, 37–52, 1992.
- GARFINKEL, Harold. *Studies in ethnomethodology*. New Jersey: Prentice-Hall, 1967.
- GRIES, Stefan Th. *Quantitative Corpus Linguistics with R: a practical introduction*. Routledge, 2009.

- GROLLA, Elaine Bicudo. “Sobre a aquisição tardia de *qu in situ* em português brasileiro.” *DELTA*, vol. 21(1), 57–73, 2005.
- GUY, Gregory R. “The quantitative analysis of linguistic variation.” *In*: PRESTON, D. (Ed.), *American Dialect Research*, pp. 223–249. Amsterdam: Benjamins, 1993.
- . “Language variation and linguistic theory.” Ms.
- GUY, Gregory R. & ZILLES, Ana. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola, 2007.
- HALL, Damien John. *A sociolinguistic study of the Regional French of Normandy*. Tese de doutorado. University of Pennsylvania, 2008.
- HYMES, Dell. “On communicative competence.” *In*: BRUMFIT, C. J. & JOHNSON, K. (Eds.), *The Communicative Approach to Language Teaching*, pp. 5–26. Oxford: Oxford University Press, 1991 [1979].
- KATO, Mary A. “Inversão da ordem SV em interrogativas no português: uma questão sintática ou estilística?” *DELTA*, vol. 3, 243–252, 1987.
- . “Two types of *wh-in-situ* in Brazilian Portuguese.”, 2004. Trabalho apresentado no Georgetown Round Table, Washington DC.
- KATO, Mary A., BRAGA, Maria Luiza, CORRÊA, Vilma Reche, LOPES-ROSSI, Maria Aparecida & SIKANSI, Nilmara Soares. “As construções-Q no português brasileiro falado: perguntas, clivadas e relativas.” *In*: KOCH, I.G.V. (Ed.), *Gramática do português falado*, pp. 303–368. Campinas: Editora da Unicamp, 1996.
- KATO, Mary A. & MIOTO, Carlos. “A multi-evidence study of European and Brazilian Portuguese *wh*-questions.” *In*: KEPSEK, Stephan & REIS, Marga (Eds.), *Linguistic evidence: empirical, theoretical and computational perspectives*, pp. 307–328. Berlin, New York: Mouton de Gruyter, 2005.
- KROCH, Anthony S. “Limits on the human sentence generator.” *In*: *Theoretical Issues in Natural Language Processing*. New Mexico State University,

- Las Cruces, 1987.
 URL <http://acl.ldc.upenn.edu/T/T87/T87-1043.pdf>
- LABOV, Teresa & LABOV, William. “L’apprentissage de la syntaxe des interrogations.” *Langue française*, vol. 34, 52–80, 1977. Acesso em 24/01/2010.
 URL http://www.persee.fr/web/revues/home/prescript/article/lfr_0023-8368_1977_num_34_1_4817
- LABOV, William. “The social motivation of a sound change.” *Word*, vol. 19, 273–309, 1963.
- . “Contraction, deletion, and inherent variability of the English copula.” *Language*, vol. 45(4), 715–762, 1969.
- . *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.
- . “Where does the sociolinguistic variable stop? A response to Beatriz Lavandera.” In: *Working Papers in Sociolinguistics* 44. Austin: Southwest Educational Development Laboratory, 1978.
- . *Principles of Linguistic Change: internal factors*. Oxford UK & Cambridge USA: Blackwell, 1994.
- . “When intuitions fail.” In: MCNAIR, L., SINGER, K., DOLBRIN, L. & AUCON, M. (Eds.), *Papers from the Parasession on Theory and Data in Linguistics*, Chicago Linguistic Society 32:77-106, 1996. Acesso em 25/07/2008.
 URL <http://www.ling.upenn.edu/~wlabov/Papers/WIF.ok.pdf>
- . *Principles of Linguistic Change: external factors*. Oxford & Cambridge: Blackwell, 2001.
- . *The social stratification of English in New York City*. São Paulo: Cambridge University Press, 2006 [1966].

- LABOV, William, ASH, Sharon & BOBERG, Charles. *The Atlas of North American English*. Berlin, New York: Mouton de Gruyter, 2006.
- LANGACKER, Ronald. *Foundations of Cognitive Grammar*. California: Stanford University Press, 1991.
- LAVANDERA, Beatriz. “Where does the sociolinguistic variable stop?” *Language in Society*, vol. 7, 171–182, 1978.
- LESSA DE OLIVEIRA, Adriana Stella Cardoso. “Aquisição da linguagem e variação linguística.” *Estudos Linguísticos*, vol. 34, 409–414, 2005.
- LOPES-ROSSI, Maria Aparecida Garcia. “Estudo diacrônico sobre as interrogativas no português do Brasil.” In: ROBERTS, Ian & KATO, Mary A. (Eds.), *Português brasileiro: uma viagem diacrônica*, pp. 307–342. Campinas: Unicamp, 1993.
- . *A sintaxe diacrônica das Interrogativas-Q do português*. Tese de doutorado. IEL/Unicamp, 1996.
- LUFT, Celso P. *Gramática resumida*. São Paulo: Globo, s/d.
- MENDES, Ronald Beline. *Ter + particípio e Estar + gerúndio: aspecto verbal e variação no português do Brasil*. Tese de doutorado. IEL/Unicamp, 2005.
- MEYERHOFF, Miriam. “Towards a typology of linguistic variables.”, 1999. Paper apresentado no NWAV28. Acesso em 18/01/2010.
URL <http://www.ling.ed.ac.uk/~mhoff/Typology-of-variables.pdf>
- MILROY, Leslie & GORDON, Matthew. *Sociolinguistics: Method and Interpretation*. Malden, MA: Blackwell Pub, 2003.
- MIOTO, Carlos. “Construções interrogativas: Elementos para uma análise do português do Brasil.” *Caderno de Estudos Linguísticos, Campinas*, vol. 17, 39–64, 1989.
- . “Wh é que ≠ wh que.” *Estudos Linguísticos*, vol. 26, 648–654, 1997.

- . “A interface sintaxe-semântica: CP no PB.”, s/d. Ms.
- MIOTO, Carlos & FIGUEIREDO-SILVA, Maria Cristina. “Wh que = wh é que?” *DELTA*, vol. n. 11, vol.2, 301–311, 1995.
- MIOTO, Carlos & KATO, Mary A. “As interrogativas Q do português europeu e do português brasileiro atuais.” *Revista da ABRALIN*, vol. 4, no. 1 e 2, 171–196, 2005.
- MOLICA, Maria Cecília de Magalhães. “Relativas em tempo real no português brasileiro contemporâneo.” In: PAIVA, Maria da Conceição de & DUARTE, Maria Eugênia Lamoglia (Eds.), *Mudança linguística em tempo real*, pp. 129–138. Rio de Janeiro: Contra Capa, 2003.
- NASCIMENTO, Ivanete Belém do. *O uso variável de você e cê na cidade de São Paulo. Tempo aparente e tempo real*. Dissertação de Mestrado. Universidade de São Paulo, Ms.
- OMENA, Nelize Pires de. “A referência à primeira pessoa do plural: variação ou mudança?” In: PAIVA, Maria da Conceição de & DUARTE, Maria Eugênia Lamoglia (Eds.), *Mudança linguística em tempo real*, pp. 63–80. Rio de Janeiro: Contra Capa, 2003.
- OUSHIRO, Livia. “Condicionamento discursivo-pragmático no uso variável de Interrogativas-Q.” *Estudos Linguísticos*, vol. 39, 628–639, 2009.
- OUSHIRO, Livia & NASSER, Juliana Antunes. “Perguntas e respostas em entrevistas sociolinguísticas.” In: *Anais do Congresso Internacional Linguagem e Interação II*. São Leopoldo - RS: Casa Leiria, 2010.
- PAIVA, Maria da Conceição de. “O percurso da monotongação de [ey]: observações no tempo real.” In: PAIVA, Maria da Conceição de & DUARTE, Maria Eugênia Lamoglia (Eds.), *Mudança linguística em tempo real*, pp. 31–46. Rio de Janeiro: Contra Capa, 2003.
- PAIVA, Maria da Conceição de & DUARTE, Maria Eugênia Lamoglia (Eds.). *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa, 2003.

- PIRES, Acrísio & TAYLOR, Heather Lee. “The syntax of wh-in-situ and common ground.” In: MASULLO, Pascual (Ed.), *Romance Languages: Structure, interfaces, and microparametric variation*. Amsterdam: John Benjamins, 2007. Acesso em 30/08/2008.
 URL <http://74.125.45.104/search?q=cache:v63Xfg0x8fsJ:www.ling.umd.edu/~htaylor/papers/PiresTaylor07LSRL.pdf+Wh-in-situ+constructions:+Syntax+and/or+Phonolog%3F&hl=en&ct=clnk&cd=5&gl=br&client=firefox-a>.
- RICKFORD, John R., BALL, Arnetha, BLAKE, Renee, JACKSON, Raina & MARTIN, Nomi. “Rappin on the copula coffin: Theoretical and methodological issues in the analysis of copula variation in African-American Vernacular English.” *Language Variation and Change*, vol. 3, 103–132, 1991.
- ROHDE, Hannah. “Rhetorical questions as redundant interrogatives.” *San Diego Linguistics Papers*, vol. 2, 134–168, 2006. Acesso em 22/12/2009.
 URL <http://escholarship.org/uc/item/4xd7t5ww>
- SACKS, Harvey. “On the preferences for agreement and contiguity in sequences in conversation.” In: BUTTON, G. & LEE, J. R. E. (Eds.), *Talk and social organisation*, pp. 54–69. Clevedon, UK: Multilingual Matters, 1987 [1973].
 ———. *Lectures on conversation*. Malden, MA: Blackwell, 1992.
- SACKS, Harvey, SCHEGLOFF, Emanuel A. & JEFFERSON, Gail. “A simplest systematics for the organization of turn-taking for conversation.” *Language*, vol. 50, 696–735, 1974.
- SANKOFF, Gillian. “A quantitative paradigm for the study of communicative competence.” In: *The social life of language*, pp. 47–79. Philadelphia: University of Pennsylvania Press, 1980.
- SANKOFF, Gillian & THIBAUT, Pierre. “L’alternance entre les auxiliaires avoir et être en français parlé à Montréal.” *Langue française*, vol. 34, 81–108, 1977.

- SCHEGLOFF, Emanuel A. "Notes on a conversational practice: formulating place." *In*: SUDNOW, David (Ed.), *Studies in Social Interaction*, pp. 75–119. New York: The Free Press, 1972a.
- . "Sequencing in conversational openings." *In*: GUMPERZ, John J. & HYMES, Dell (Eds.), *Directions in sociolinguistics: the ethnography of communication*, pp. 346–380. New York: Holt, Rinehart & Winston, Inc., 1972b.
- SCHEGLOFF, Emanuel A. & SACKS, Harvey. "Opening up closings." *Semiotica*, vol. 8, 289–327, 1973.
- SCHILLING-ESTES, Natalie. "Investigating stylistic variation." *In*: CHAMBERS, J.K., TRUDGILL, P. & SCHILLING-ESTES, N. (Eds.), *Handbook of Language Variation and Change*. Malden, MA: Blackwell, 2002.
- SELKIRK, Elisabeth O. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Massachusetts: MIT Press, 1984.
- SIKANSI, Nilmara Soares. "Interrogativas-Q do português brasileiro moderno: quadro geral." *Caderno de Estudos Linguísticos, Campinas*, vol. 34, 119–129, 1998.
- SILVA, Luiz Antônio da. "Perguntas e respostas: oralidade e interação." *In*: PRETI, Dino Fioravante (Ed.), *Oralidade em diferentes discursos*, pp. 261–295. São Paulo: Humanitas, 2006.
- STALNAKER, Robert. "Assertion." *In*: COLE, P. (Ed.), *Pragmatics: Syntax and Semantics Volume 9*. New York: Academic Press, 1978.
- . "Common ground." *Linguistics and Philosophy*, vol. 25, 701–721, 2002.
- TAGLIAMONTE, Sali. *Analysing Sociolinguistic Variation*. São Paulo: Cambridge University Press, 2006.
- WEINER, Judith & LABOV, William. "Constraints on the agentless passive." *Journal of Linguistics*, vol. 19, 29–58, 1983 [1977].

- WEINREICH, Uriel, LABOV, William & HERZOG, Marvin I. “Empirical foundations for a theory of language change.” *In: LEHMANN, W.P. & MALKIEL, Y. (Eds.), Directions for Historical Linguistics: A Symposium.* Austin: University of Texas Press, 1968.
- WOLFRAM, Walt. “Identifying and interpreting variables.” *In: PRESTON, D. R. (Ed.), American Dialect Research*, pp. 193–221. Amsterdam/Philadelphia: John Benjamins, 1993.
- ZUBIZARRETA, Maria Luiza. *Focus, prosody, and word order*. Tese de doutorado. University of Southern California, Department of Linguistics, Los Angeles, CA, 1997. Ms.

Anexo A: Questionário enviado aos documentadores

O formulário abaixo foi elaborado com a ferramenta online de compartilhamento de arquivos Google Docs e enviado através de correio eletrônico aos alunos-documentadores do curso de Sociolinguística da USP, cujas entrevistas haviam sido selecionadas para compor uma das amostras utilizadas nesta pesquisa. Das 53 gravações selecionadas, foi possível obter o endereço eletrônico de 46 alunos através dos registros mantidos pelo professor do curso. O questionário foi respondido por 31 dos alunos contactados, 19 dos quais são paulistanos.

Formulário

- Nome:
- Data de nascimento:
- Local de nascimento:
- Telefones para contato:
- Ocupação/profissão:
- Relação com o entrevistado:
 - Amigo
 - Parente

- Amigo de amigo
 - Outro:
- Como você classificaria o seu grau de relação com o entrevistado na época da entrevista?
 - Bastante próximo. O entrevistado fazia parte do meu círculo imediato de amigos/familiares e conversávamos frequentemente.
 - Próximo. Conversávamos frequentemente, mas o entrevistado não fazia parte de meu círculo imediato de amigos/familiares.
 - Próximo, mas não conversávamos frequentemente.
 - Neutro. Ele era meu conhecido, mas não nos falávamos com frequência.
 - Distante. Não o conhecia anteriormente e só conversamos na ocasião da entrevista.
- Lugares em que você já morou e tempo aproximado que morou em cada local. Se em outra cidade, indicar também o estado (ex.: Sorocaba-SP-20 anos); se na cidade de São Paulo, colocar apenas o bairro (ex.: Tatuapé-5 anos).
- Nome do Pai: Colocar apenas iniciais. Ex.: P.R.M.
- Idade do Pai:
- Ocupação do Pai:
- Local de Nascimento do Pai:
- Grau de Escolaridade do Pai:
 - Fundamental I (até 4a. série)
 - Fundamental II (até 8a. série)
 - Ensino Médio (até 3o. colegial)
 - Superior (graduação completa ou incompleta, ou pós-graduação)
- Nome da Mãe: Colocar apenas iniciais. Ex.: A.D.B.
- Idade da Mãe:
- Ocupação da mãe:
- Local de Nascimento da Mãe:
- Grau de Escolaridade da mãe:
 - Fundamental I (até 4a. série)

- Fundamental II (até 8a. série)
- Ensino Médio (até 3o. colegial)
- Superior (graduação completa ou incompleta, ou pós-graduação)
- Nome, idade e escolaridade dos irmãos: Ex.: F.T., 35 anos, superior

Anexo B: *Script* para extração de dados de interrogativas de constituinte com o programa R

O objetivo de disponibilizar esse script é facilitar o desenvolvimento de outras pesquisas, para as quais ele pode ser adaptado. O livro de [Gries \(2009\)](#), escrito especificamente para linguistas, é uma boa introdução prática ao programa. Um exemplo do resultado final da extração automática de dados é fornecido na [Figura 1](#).

```
#define pasta padrão
setwd("C:/Users/...")

#carrega arquivo-fonte em um vetor
veja1Artes<-scan("081126Artes.txt", what="char", sep="\n",
quote="", comment.char="", encoding="UTF-8")

#junta todos os arquivos-fonte em um único vetor
all.in.one.list<-list(veja1Artes, veja1Blogosfera, veja1Brasil,
veja1Carta, veja1DiogoM, veja1Economia, veja1Entrevista, ...)

#atribui numeração a cada arquivo
number.of.files<-length(all.in.one.list)
all.veja <- character() # cria um vetor vazio para o arquivo completo
```

```

for (i in seq(number.of.files)) {
  all.veja <- append(all.veja, paste("<File_", i, "> ",
    all.in.one.list[[i]], sep=""))
}

all.veja # o número antes de tab indica o arquivo-fonte da sentença
cat(all.veja, file=choose.files("all.veja.txt"), sep="\n")

#concordancia
all.veja2<-unlist(strsplit(all.veja,"\\W+"))
all.veja2<-gsub("(\\W)","\\1", all.veja)
all.veja3<-unlist(strsplit(all.veja2," +"))

#especifica palavras a serem extraídas (que, quê, quem etc.)
(matches.in.veja<-grep("(^qu[e|ê][\\.\\"{0}?"{0}!\\.\",]?$|
  ^quem[\\.\\"{0}?"{0}!\\.\",]?$|^qual[\\.\\"{0}?"{0}!\\.\",]?$|
  ^quais[\\.\\"{0}?"{0}!\\.\",]?$|^quando[\\.\\"{0}?"{0}!\\.\",]?$|
  ^a?onde[\\.\\"{0}?"{0}!\\.\",]?$|^como[\\.\\"{0}?"{0}!\\.\",]?$|
  ^quant|^quão[\\.\\"{0}?"{0}!\\.\",]?$)", all.veja3, ignore.case=T))

#OU
##### alternativa para corpus de redações #####
(matches.in.F3Apatricia<-grep("(^[o|u]?qu[e|ê|i][\\.\\"{0}?"{0}!\\.\",]?$|^q[
  \\.\\"{0}?"{0}!\\.\",]?$|^quei?m[\\.\\"{0}?"{0}!\\.\",]?$|^qua[l|u][\\.\\"{0}?"{0}!\\.\",]?$|
  ^quais[\\.\\"{0}?"{0}!\\.\",]?$|^quand[o|u][\\.\\"{0}?"{0}!\\.\",]?$|^a?ond[e|i]
  [\\.\\"{0}?"{0}!\\.\",]?$|^com[o|u][\\.\\"{0}?"{0}!\\.\",]?$|^quant|^quão[\\.\\"{0}?"{0}!\\.\",]?$|
  ^pq[\\.\\"{0}?"{0}!\\.\",]?$|^p[o|u]rq[ê|e|i][\\.\\"{0}?"{0}!\\.\",]?$)", F3Apatricia3,
  ignore.case=T))
## inclui outras grafias possíveis de constituintes interrogativos ##

#define nomes das colunas
base<-paste("tkn no", "\t", "File number", "\t", "VARDEP", "\t",
  "estr sent", "\t", "constQ", "\t", "\t", "prec_cont", "\t",
  "NODE", "\t", "subs_cont", "\n", sep="")

#cria arquivo de dados com nome das colunas
cat(base, file="TOKENS_VEJA.txt", sep="", append=T)

#define o arquivo de destino
output.file <- file.choose("TOKENS_VEJA.txt")

#adiciona ocorrências de constituintes interrogativos em contexto

```

```
# (10 palavras antes/ 10 palavras depois) no arquivo especificado
for(i in matches.in.veja) {
  cat("\t", "\t", "\t", "\t", "\t", "\t",
    all.veja3[max(0,i-10):max(0,i-1)], "\t", all.veja3[i],
    "\t", all.veja3[(i+1):min(i+10,length(all.veja3)+1)],
    file=output.file, append=T, "\n")}
```

Figura 1 Planilha de codificação com dados extraídos através do programa R

B	C	D	E	F	G	H	I
1	VARDE	estr se	constQ				
1					<File_1> Ele começa pelo paradoxo característico dos tempos que correm: publicação de O Mundo Pós-Americano, tais sucessos não parecem reversíveis.]	como	é que, num mundo marcado por guerras, terrorismo
1					bárbaro, e isso corréi até mesmo o conhecimento de por	que	hoje se tornou impossível escrever poemas". Não
1					sem emoção. Entre ele e Ferris, não resta dúvida de	qual	é o homem mais perigoso. A mensagem desse con
1					esperado que chantagista e chantageado tenham alguma idéia sobre o	que	está em discussão entre eles. Mas, se não tiverem
1					cláustico, fica quase que ao gosto do freguês decidir o	que	os irmãos sentem de fato pelos personagens que
1					Giorno Perfetto, Ferzan Ozpetek volta no tempo para investigar o	que	sucedeu nas horas anteriores a um telefonema fe
3					regiões de operação e, ao fim e ao cabo, decide	quem	são os perdedores e os vencedores no setor. No c
3					dos lobistas. Oi e BrT travaram um duelo para ver	quem	conseguiria recrutá-lo para sua causa. A Telemar
3					da Funasa guarde uma lição é a de mostrar por	que	os políticos se digladiam por postos no governo.
4					do Sul: a Funasa deveria cuidar da saúde deles. Mas	quem	cuida da Funasa? <File_4> As disputas mais reñh
4					Afirmção gravíssima, vinda de alguém em posição de saber o	que	estava dizendo, e que, fosse o Brasil menos anest
5					imediatamente todas as minhas economias na bolsa de valores. O	que	fiz? Fechei os olhos e obedeci. <File_5> Eu já tinha
5					valores, pode ser uma boa oportunidade para entrar nela. O	que	eu fiz? Entrei. <File_5> "Penso, logo existo." Em la
6					e estaríamos mesmo a caminho da modernização tributária brasileira. Por	que	os legisladores não foram ousados o bastante pa
6					recursos. Como toda reforma tributária, a atual terá de arbitrar	quem	perde e quem ganha. Foi justamente o que derrot
6					reforma tributária, a atual terá de arbitrar quem perde e	quem	ganha. Foi justamente o que derrotou todas as tei
7					Seu trabalho representa um avanço, também, porque mostra, com clareza,	como	as universidades têm relação direta com a má qu
7					absurda diante de estudantes universitários tão pouco escolarizados. <File_7> O	que,	exatamente, se ensina aos futuros professores? <
7					a ideologia atrasada que domina as faculdades de pedagogia. <File_7>	Como	essa ideologia se manifesta? <File_7> Por exempl
7					analisar no Conselho Nacional de Educação. Perdi as contas de	quantas	vezes estive diante da palavra dialética, que, não
7					próprios não estão preparados para desempenhar a função. <File_7> Por	que	os professores são tão pouco autocríticos? <File
7					da responsabilidade. Nos sindicatos, todo esse corporativismo se exagera. <File_7>	Como	os sindicatos prejudicam a sala de aula? <File_7>
7					responde por apenas 2% das citações nas melhores revistas científicas.	Como,	a senhora explica essa ineficiência? <File_7> Sen
7					não faz: preparar para o mercado de trabalho. É estranho	que,	no meio acadêmico, uma formação voltada para
7					dizem, sem nenhum constrangimento, preferir "formar cidadãos". Cabe perguntar: o	que	o cidadão vai fazer da vida se ele não puder se in
7					típica de países em desenvolvimento, como o Brasil. <File_7> Por	que	num país em desenvolvimento o diploma universi
8					em breve. <File_8> 1 - Gêmeos em 3D <File_8> O	que	são: diversas técnicas que criam imagens em 3D
8					que são: diversas técnicas que criam imagens em 3D <File_8>	Quando	estarão disponíveis: 2010-2015 <File_8> Mark Th
8					<File_8> <File_8> 2 - Recarga à mão <File_8> O	que	é: sistema que transforma superfícies em fonte de
8					transforma superfícies em fonte de energia para carregar eletrônicos <File_8>	Quando	estará disponível: 2009 <File_8> Fotos divulgaçã
8					<File_8> 3 - A chave do freio <File_8> O	que	é: uma chave que, pré-programada, limita a veloc
8					uma chave que, pré-programada, limita a velocidade do carro <File_8>	Quando	estará disponível: 2009 <File_8> Não adianta ace
8					<File_8> 4 - Sob as ordens dos gestos <File_8> O	que	é: uma câmera que transforma movimentos em r

Anexo C: Roteiro da entrevista sociolinguística

GRAVAÇÃO DAS ENTREVISTAS

1. Tente escolher um local neutro para fazer a gravação. Por neutro, entenda-se, em princípio, um local em que seu informante se sinta à vontade. É melhor que não seja um local barulhento, pois isso dificultará seu trabalho de transcrição. Por outro lado, prefira um lugar que não tenha um clima muito formal (por exemplo, uma sala reservada de uma biblioteca, onde seria necessário manter um baixo volume de conversa, ou o escritório onde você trabalha). Não leve seu informante à sua casa, principalmente se você não o conhecer. Você pode ir até a casa do informante, mas procure saber se “é tranquilo” fazer a gravação em tal local.
 - (a) Note que esse requisito não é de observação obrigatória. Ele visa a facilitar o seu trabalho de transcrição.
 - (b) Não há problema (para a sua nota, inclusive) caso haja interrupções da sua conversa com o informante por um terceiro, que, de repente, aparece no local da gravação (por exemplo: um colega de trabalho, caso você grave seu informante no ambiente de trabalho dele; o garçom ou um amigo, caso você grave seu informante num café).
2. Explique ao seu informante que você precisa fazer um trabalho para a faculdade, mas não mencione as palavras linguagem, lingüística, gramática ou afins antes e durante a entrevista. Caso a pessoa insista, diga que o professor ainda não explicou que trabalho exatamente será feito, mas será alguma análise de natureza social. Explique que cada aluno terá que fazer uma gravação como essa que você está fazendo com ele/ela. De preferência, escolha um informante que não saiba que você faz Letras/Lingüística; sobretudo, escolha um informante que não saiba que você faz um curso de Sociolingüística, cujo professor “costuma pedir gravações de conversas”.

3. Após a gravação, preencha a declaração que lhe foi fornecida, e peça que seu informante a assine. Explique a ele/ela que esse é um mero procedimento de proteção aos seus dados pessoais, que não serão tornados públicos.
4. Comece a gravação com alguma questão de natureza pessoal. Parece uma boa idéia, começar perguntando há quanto tempo o informante mora onde mora. Pergunte também se ele gosta desse bairro; se gostaria de morar em outro lugar. Se houver espaço, pergunte porque ele mora nesse bairro; se ele escolheu morar ali, ou se não teve escolha.
 - (a) Procure, sempre, deixar que seu informante fale o máximo possível. Você pode interrompê-lo, lembrando coisas, emitindo opiniões, etc; apenas não fale mais que seu informante.
 - (b) Se seu informante tende a dar respostas curtas, por timidez, má-vontade ou algo parecido, saiba conduzir a conversa de modo a produzir seu engajamento/interesse nela.
5. Depois que você conversar sobre esse assunto introduza o seguinte tópico:

SEGURANÇA

- (a) Ao introduzir tal tópico, continue numa linha “pessoal”. Perguntas possíveis são: esse bairro onde você mora é seguro? Já aconteceu alguma coisa com você ou com algum conhecido seu, ou com alguém da sua família, no seu bairro? Se essas perguntas tiverem respostas afirmativas, desenvolva o assunto, nessa linha. Deixe o informante contar detalhes, ou conduza a conversa no sentido de que ele desenvolva uma narrativa pessoal de fatos (que dizem respeito a ele diretamente ou a alguma pessoa próxima). Se as respostas forem negativas, pergunte se ele conhece alguém que tenha sido assaltado; pergunte se ele conhece alguém que tenha sido vítima de algum ato de violência urbana; sempre preocupando-se com que o tom seja de NARRATIVA PESSOAL, em que o envolvimento do informante com o assunto é PRÓXIMO – de modo que ele possa esquecer-se de monitorar a própria fala, durante a gravação.
- (b) Depois de desenvolver esse assunto da maneira descrita acima, mude o tom da conversa: de um tom pessoal, para um tom impessoal (talvez “mais teórico”, político, jornalístico – que for possível na conversa). A questão aqui é estabelecer, na conversa, um distanciamento do informante em relação ao assunto. Exemplos de perguntas: você acredita que segurança deveria ser a prioridade número 1 no programa de governo de um prefeito ou de um governador? Você sabe qual

é o estado considerado mais seguro, para se viver, no Brasil? Você tem idéia de que país é considerado mais seguro, no mundo? Observe que essas perguntas podem ter resposta pontuais... Saiba conduzir a conversa de modo que seu informante fale, sem que narre fatos pessoais, mas adotando um tom mais DESCRITIVO/DISERTATIVO.

6. Em seguida, mude o rumo da conversa para um dos assuntos abaixo:

POLÍTICA, em geral

TRANSPORTE PÚBLICO

MEIO-AMBIENTE

ALIMENTAÇÃO (agricultura, fome no mundo, nutrição, hábitos alimentares)

- (a) Escolha apenas UM dos assuntos acima, conforme o andamento da conversa. Pode ser qualquer um deles, ou até mesmo outro, desde que se observem o sub-itens abaixo:
- (b) Você deve tratar desse assunto com os “tons” descritos acima, mas em ordem diferente:

- i. primeiro, um tom dissertativo, teórico, distanciado

- ii. depois, um tom pessoal, parcial, com narrativa de fatos da vida do próprio informante, ou da vida de uma pessoa próxima a ele.

Exemplos:

- (c) Agora há pouco você falou que acha que segurança deve ser prioridade. O que você acha do fato de que o governo federal do PT tenha priorizado a questão da fome?
Na mesma linha: o que você sabe sobre o programa bolsa-família? Você sabe de outro país em que haja algum programa semelhante?
- (d) Você conhece alguém que recebe o bolsa-família? Você conhece alguém que esteja desempregado há algum tempo? Você sabe se ele/ela recebeu algum auxílio do governo?

7. Finalmente, repita o procedimento acima (6) para o último assunto:

EDUCAÇÃO

- (a) Tom dissertativo, teórico, distanciado.

(b) Tom narrativo, pessoal.

8. Ao final da entrevista, você deve fazer as perguntas obrigatórias da lista em anexo. Mantenha a espontaneidade, apesar de essas perguntas serem “fechadas”, ou seja, você provavelmente terá um papel para ajudá-lo a lembrar-se dessas perguntas (diferentemente do que ocorreu até então).
9. A entrevista deve levar NO MÍNIMO UMA HORA. Dessa forma, a simples título de exemplificação, considere o seguinte planejamento:

Início da conversa (alguma coisa pessoal/ex: bairro) 6 minutos
SEGURANÇA narrativo/pessoal 8 minutos
teórico/distanciado 8 minutos
ASSUNTO X: teórico/distanciado 8 minutos
narrativo/pessoal 8 minutos
EDUCAÇÃO: teórico/distanciado 8 minutos
narrativo/pessoal 8 minutos
Questionário 6 minutos

10. Observações importantes:

- (a) ENTREVISTA é, para nós, um termo técnico. Portanto, não o leve ao pé-da-letra: seja espontâneo. Não se comporte como um entrevistador. Não trate seu informante (outro termo técnico) como um entrevistado...
- (b) Qualquer que seja o assunto, a regra mais básica – e de ouro – é que você, documentador, fale POUCO. Queremos os dados da fala de seu INFORMANTE, não da sua! Ou seja, cabe a você proporcionar as condições para que seu informante fale.
- (c) Por outro lado, se seu informante se empolgar e começar falar “demais” sobre um determinado assunto, saiba cortá-lo. Interrompa-o, naturalmente e da melhor forma que conseguir, e introduza o novo tópico. Atente para a necessidade de que os tópicos do plano acima devem aparecer, todos eles, na entrevista – ainda que não necessariamente com a mesma divisão de tempo.
- (d) Durante a entrevista, fale como você fala. Não force um sotaque. Não seja pedante (ou lingüisticamente tenso), nem relaxado (ou lingüisticamente distenso) demais.

- (e) Não fique mexendo no gravador. Tente agir naturalmente com ele, ou seja, como se você nem se lembrasse de que ele está ali. Não interrompa a gravação se alguém interromper momentaneamente a conversa. Obviamente, se seu informante tiver que se ausentar por alguns minutos, nesse caso, sim, interrompa a gravação e retome-a quando notar que seu informante está voltando.

Perguntas Obrigatórias

(para depois da conversa sobre educação). NÃO MUDAR A ORDEM!

1. Qual o sotaque no Brasil que você mais gosta? O gaúcho? O carioca? O paulistano? O nordestino?
2. Por quê?
3. Qual o sotaque que você mais detesta ou que te irrita mais?
4. Por quê?
5. O que você acha do jeito que as pessoas do interior falam?
6. Pra você, qual é a característica mais marcante da fala das pessoas do interior?
7. Quando você conversa com alguém, você presta atenção no jeito que a pessoa fala?
8. Tem alguma coisa na fala das pessoas que te chama a atenção especialmente?
9. Você acha que você fala mais porta ou porta, perto ou perto? [pronunciar primeiro o /r/ como tepe, e depois como retroflexo].
10. Você acha que um é melhor que o outro, pra falar? Por quê?
11. Você tem alguma opinião sobre esse jeito cada vez mais comum de falar, por exemplo, “você pode tá passando aqui amanhã”, “você pode tá entrando por essa porta?”, em vez de “você pode passar aqui amanhã”, etc?
12. Você fala desse modo?
13. Quem você acha que fala mais assim? Empregadas domésticas, caixas de banco, atendentes de lojas, operadoras de telemarketing ou seus amigos?
14. Quando você está falando sobre você junto com mais alguém, você fala mais “nós” ou fala mais “a gente”?

15. Você acha que é melhor usar “nós” ou “a gente”? Por quê?
16. Você sabia que, hoje em dia, “a gente” é considerado como um pronome pessoal, do mesmo jeito que “eu”, “você”, “tu”, “eles”?
17. Quando você estava no primeiro grau, você aprendeu “a gente” como pronome, na escola?
18. Você fala alguma língua estrangeira? Qual?
19. Que língua estrangeira você gostaria de aprender?
20. Por quê?
21. Que língua estrangeira você acha que é a mais bonita?
22. Por quê?

Ficha do Informante

Nome Completo:

Idade:

Data de Nascimento:

Endereço:

Telefones para contato:

Mora com a família? () sim () não

Bairros em que já morou, em SP (indicar aproximadamente quanto tempo morou em cada bairro):

Escolas em que estudou:

Fundamental 1:

Fundamental 2:

Ensino Médio:

Faculdade:

Número de irmãos: () mais velhos () mais novos

Nome, idade e escolaridade dos irmãos:

Nome do Pai:

Idade do Pai:

Grau de Escolaridade do Pai:

Ocupação do pai:

Natural de: () São Paulo (capital) () _____

Nome da Mãe:

Idade da Mãe:

Grau de Escolaridade da Mãe:

Ocupação da mãe:

Natural de: () São Paulo (capital) () _____