

A VARIAÇÃO ENTRE *QUE* E *O QUE* EM ORAÇÕES INTERROGATIVAS DIRETAS NA HISTÓRIA DA LÍNGUA PORTUGUESA: PRIMEIROS RESULTADOS

Vitor Morelli SILVA

Orientadora: Prof.^a Dr.^a Charlotte Marie Chambelland Galves

Resumo: Esta pesquisa investiga a variação entre *que* e *o que* como pronomes interrogativos em orações interrogativas diretas ao longo da história da língua portuguesa. O objetivo é encontrar hipóteses para o surgimento de *o que* no português, bem como analisar os contextos sintáticos e textuais em que esse pronome aparece mais frequentemente ao longo do tempo, fazendo uso de dois corpora sintaticamente anotados, o Corpus Histórico do Português Tycho Brahe e o P.S. Post Scriptum. Através do programa *Corpus Search*, elaboraram-se buscas para encontrar as orações. A análise dos dados mostra que o pronome *o que* começa a aparecer somente no século XVII nas diretas; que a forma *que* se cristalizou em algumas estruturas formadas por verbos impessoais no decorrer do tempo; e que não há uma relação entre a ordem do sujeito e do verbo e a seleção de *que* ou *o que*.

Palavras-chave: Sintaxe histórica; Pronomes interrogativos; Corpora anotados sintaticamente; Língua portuguesa.

Introdução

Pouco se encontra de literatura sobre a variação entre *que* e *o que* enquanto pronomes interrogativos na língua portuguesa. Tal variação pode ser constatada na literatura em língua portuguesa, como nestas orações de *Falar a verdade a mentir*, de Almeida Garrett (1845/1904), retiradas do Corpus Histórico do Português Tycho Brahe (Galves, Andrade e Faria, 2017):

- “Que diz elle?” (G_004_PSD,0.26)
- “Que é isto?” (G_004_PSD,0.1702)
- “o parvo do remendão o que dizia?” (G_004_PSD,0.1975)
- Pois organizar o paiz o que é, pateta, senão repartir a gente por si e pelo seus amigos? (G_004_PSD,0.2309)

Como se observa, é possível encontrar dentro de um mesmo texto, em orações interrogativas diretas, tanto *que* como *o que*. A partir dessa observação, é objetivo desta pesquisa¹ estudar a variação entre *que* e *o que* em orações interrogativas ao longo da história da língua portuguesa para então elaborar e testar hipóteses acerca da emergência da forma *o que*. Além disso, pretende-se analisar os contextos sintáticos e textuais em que esse pronome aparece mais frequentemente ao longo do tempo.

Dada a escassez de bibliografia sobre o tema, recorreu-se a gramáticas dos séculos XVI, XIX, XX e XXI e a uma análise minuciosa dos dados, confrontando-os com o que

¹ Este trabalho baseia-se em nossa pesquisa de Iniciação Científica (em execução) financiada pela Fundação de Amparo à Pesquisa de São Paulo (FAPESP) “A história de *o que* em orações interrogativas na língua portuguesa: de periférico à regra geral” (processo nº 2021/01980-5), orientada pela Prof.^a Dr.^a Charlotte Marie Chambelland Galves. Portanto, e por se tratar de uma apresentação dos primeiros resultados, parte da metodologia e análise dos dados estão no Relatório I enviado à FAPESP. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do autor e não necessariamente refletem a visão da FAPESP.

dizem as gramáticas sobre *que/o que* interrogativo. Neste trabalho, apresentam-se os corpora, a metodologia aplicada e os primeiros resultados da pesquisa.

Metodologia

Foi feita, a princípio, uma revisão do que dizem as gramáticas sobre a variação entre *que* e *o que* em orações interrogativas. Foram consultadas, ao todo, dez gramáticas: Barros (1540), Sousa (1804), Grivet (1881), Ribeiro (1881), Dias (1894), Azevedo Filho (1986), Mateus e outros autores (Mateus et al, 2006), Cegalla (2008), Haury (2015) e Bechara (2019)². Dessa revisão, constatou-se que:

- Em Barros (1540), mostra-se uma relação entre pronomes relativos e pronomes interrogativos, uma vez que o gramático considera os nomes interrogativos como um tipo de nomes relativos. No caso acusativo (do objeto) do grupo dos interrogativos (onde o pronome interrogativo *que* deveria aparecer), o autor só considera os pronomes *quem* (escrito *que* com til) e “quál”, não aparecendo as formas *que* e *o que*.
- Grivet (1889) considera *que* um pronome “indefinido” que não implica nenhuma referência específica (ao contrário do pronome relativo). Nesta gramática, afirma-se que o pronome *que* é indefinido quando significa *que cousa, cousa que, ou que quantidade, que porção*. Além disso, diz que o uso do artigo definido pode ocorrer para “firmar” o sentido de pronome indefinido de *que* interrogativo, significando *que cousa* ou *cousa que*, diferentemente de *o* em *o que* com o significado de *aquilo* ou *aquele*, portanto, *aquilo que* ou *aquele que*;
- Ribeiro (1889) aponta a forma *o que* como incorreta;
- Dias (1894) apresenta *o que* entre parênteses depois de mencionar *que* como um pronome substantivo interrogativo;
- Azevedo Filho (1968) o vê o uso do artigo como um reforço fonético;
- Sousa (1804) e Cegalla (2008) não mencionam a variação;
- Mateus e outros autores (Mateus et al, 2006) diz que a forma “que” como sintagma nominal interrogativo é cada vez menos utilizada, “sendo *o que* a forma mais produtiva” (p. 464).
- Haury (2015) interpreta *o* como eufônico e expletivo;
- Bechara (2019) enxerga *o que* como uma forma enfática que pode substituir *que*;

² As gramáticas de Barros (1540), Sousa (1804), Dias (1894) e Mateus e outros autores (Mateus et al, 2006) são portuguesas; as gramáticas de Grivet (1881), Ribeiro (1881), Azevedo Filho (1896), Cegalla (2008), Haury (2015) e Bechara (2019) são brasileiras. Foram consultadas gramáticas dos dois países para fins de comparação do fenômeno de variação *que/o que* nas variedades portuguesa e brasileira. No entanto, tendo em vista que o material brasileiro com que se trabalhou era menor e apresentou poucos dados de orações interrogativas, a comparação não foi possível.

Como se observa, não há um consenso sobre o motivo da existência de *o* antes de *que* interrogativo. No entanto, pode-se afirmar que há uma tendência de generalização do uso da forma *o que*, ao menos no português europeu, como apontam Mateus e outros autores (Mateus et al, 2006). Feita a revisão das gramáticas, parte-se então para a coleta de dados.

Esta pesquisa usa duas bases de dados: o Corpus Histórico do Português Tycho Brahe (doravante Corpus Tycho Brahe) e o P.S. Post Scriptum (doravante Corpus Post Scriptum). Os dois corpora são sintaticamente anotados consoante as convenções do *Portuguese Syntactic Annotation Manual* (Magro e Galves, 2019). Foi utilizado nesta pesquisa o programa *Corpus Search*³: uma vez baixados os textos dos corpora, é possível rodar buscas no computador em uma velocidade muito rápida com esse programa. No terminal de texto, deve-se realizar a busca localizando os arquivos em suas pastas no computador. Contudo, é preciso antes escrever as buscas de acordo com a linguagem de *Corpus Search*⁴.

O Corpus Tycho Brahe possui, em sua última versão, 88 textos — sendo 27 desses anotados sintaticamente —, contemplando autores nascidos entre 1380 e 1976.⁵ É com base nesses 27 textos, que possuem, ao todo, 62.153 *tokens* (unidades sintáticas, geralmente orações) que a análise é feita, abrangendo os séculos XVI a XIX. Os autores usados, cujos textos anotados sintaticamente estão disponíveis no *website*, e os gêneros de seus textos são:

Tabela 1 - Os autores do Corpus Tycho Brahe e os gêneros de seus textos

Século	Nome e ano de nascimento do(a) autor(a)	Código no corpus	Gênero textual
XVI	Fernão Mendes Pinto (n.1510)	p_001	Narrativa
XVI	Pero Magalhães de Gandavo (n.1540)	g_008	Narrativa; Crônica
XVI	Diogo do Couto (n.1542)	c_007	Narrativa; Crônica
XVI	Luis de Sousa (n.1556)	s_001	Narrativa
XVI	F. Rodrigues Lobo (n.1579)	l_001	Narrativa
XVII	Manuel de Galhegos (n.1597)	g_001	Notícias
XVII	Antonio Vieira (n.1608)	v_002	Cartas
XVII	Padre A. Vieira (n.1608)	v_004	Dissertação (Sermão)

³ Disponível em: <<http://corpussearch.sourceforge.net/CS.html>>. Acesso em 6 nov. 2021.

⁴ Em computadores com sistema operacional Windows, pode-se utilizar o Bloco de Notas para realizar tal tarefa.

⁵ Cf. apresentação do corpus em: <<http://www.tycho.iel.unicamp.br/~tycho/corpus/index.html>>. Acesso em 18 nov. 2021.

XVII	Francisco Manuel de Melo (n.1608)	m_003	Cartas
XVII	Antonio das Chagas (n.1631)	c_003	Cartas
XVII	Manuel Bernardes (n.1644)	b_003	Narrativa
XVII	José da Cunha Brochado (n.1651)	b_008	Cartas
XVII	Maria do Céu (n.1658)	c_002	Narrativa; Biografia
XVII	Francisco Xavier de Menezes (n.1673)	m_008	Notícias
XVII	André de Barros (n.1675)	b_001	Narrativa
XVIII	Cavaleiro de Oliveira (Fco Xavier) (n.1702)	c_001	Cartas
XVIII	António José da Silva (o Judeu) (n.1705)	s_004	Teatro
XVIII	Matias Aires (n.1705)	a_001	Dissertação
XVIII	L. Antonio Verney (n.1713)	v_001	Gramática
XVIII	Marquesa de Alorna (n.1750)	a_004	Cartas
XVIII	Jose Daniel Rodrigues da Costa (n.1757)	c_005	Teatro
XVIII	Vários (n.~1700-1799)	va_009	Notícias
XIX	J. B. da Silva L. de Almeida Garrett (n.1799)	g_004	Teatro
XIX	Marquês de Fronteira e d'Alorna (n.1802)	a_003	Narrativa
XIX	Camilo Castelo	b_005	Narrativa

	Branco (n.1826)		
XIX	Ramalho Ortigão (n.1836)	o_001	Cartas
XIX	Vários (n.~1830)	va_002	Ata

Foi utilizado também um conjunto de cartas brasileiras que também pertence ao Corpus Tycho Brahe, mas o arquivo ainda não está disponível no *site* do corpus. Este arquivo contém, ao todo, 697 tokens.

O Corpus Post Scriptum possui um acervo de cartas escritas no período da Idade Moderna em Portugal e na Espanha. Como se explica na apresentação da página inicial,

[...] Estes documentos são escritos epistolares quase todos eles inéditos, feitos por autores de diferentes proveniências sociais. Podiam ser amos ou criados, adultos ou crianças, homens ou mulheres, ladrões, soldados, artesãos, padres, militantes políticos e outros tipos de agentes sociais. Em grande parte, a sua epistolografia sobreviveu por razões excepcionais, quando os seus percursos se cruzaram com os meios de perseguição da Inquisição e dos tribunais civis, eclesiásticos e militares, instituições que costumavam fazer uso da correspondência privada como prova de delitos. Em outros casos, já mais raros, as cartas foram preservadas em contexto não criminal, mas foram também trocadas numa interação de bastidores e são enquadráveis em termos situacionais. Estas fontes textuais apresentam frequentemente uma retórica (quase) oral, tematizando assuntos do quotidiano que até hoje não tem sido fácil estudar senão a partir de pequenos exemplos. O projeto P.S. não só reúne uma ampla coleção destas cartas privadas, como as apresenta em dois formatos pesquisáveis: o da edição crítica digital e o do corpus linguisticamente anotado.⁶

Interessam a esta pesquisa apenas os textos de autores portugueses, que, dentre os textos anotados sintaticamente, totalizam 354 arquivos — são, ao todo, 4.922 *tokens*. O alcance temporal do Corpus Post Scriptum é o mesmo do Corpus Tycho Brahe.

Elaboraram-se, então, as seguintes buscas na linguagem de *Corpus Search*:

Tabela 2 - Buscas por *que* e *o que*

<i>BUSCA 1: QUE</i>	<i>BUSCA 2: O QUE</i>
define: port.def print_indices: t node: CP* query: (CP-QUE* IsRoot) AND (CP-QUE* idoms WNP*) AND (WNP* idomonly WPRO) AND (WPRO idoms Que)	define: port.def print_indices: t node: CP* query: (CP-QUE* IsRoot) AND (CP-QUE* idoms WNP*) AND (WNP* idoms D) AND (D idoms o) AND (D HasSister WPRO) AND (WPRO idoms que)

Em ambas as buscas o cabeçalho (que começa em *define* e termina em *node*) é o mesmo. A linha “*define: port.def*” remete ao arquivo de definições utilizado pelo pesquisador

⁶ Retirado de: <<http://teitok.clul.ul.pt/postscriptum/pt/index.php?action=home>>. Acesso em 18 nov. 2021.

em *Corpus Search* — basicamente, é uma mensagem para que o programa siga as definições estabelecidas no arquivo, onde se pode definir categorias que não existem no sistema de anotação ou reagrupá-las sob uma nova etiqueta, caso seja necessário; O comando “*print_indices: t*” pede a *Corpus Search* que um índice numérico esteja presente em cada nó no arquivo de saída⁷, o que facilita a leitura do resultado da busca; e *node*, por sua vez, corresponde ao nó sintático pertinente para a busca. No caso das buscas apresentadas acima, o nó pertinente é CP*, ou seja, qualquer (*) *complementizer phrase* (CP).

A partir de “*query*”, começa-se a descrever as condições da busca. Para *que*, tem-se que um CP-QUE* é a raiz (*IsRoot*) — o que permite filtrar orações interrogativas subordinadas; na segunda linha, diz-se que CP-QUE* domina imediatamente (*idoms*) qualquer sintagma nominal *que* (WNP*); da terceira linha, depreende-se que o WNP* apenas domina imediatamente (*idomsonly*) um pronome *que* (WPRO) e este domina imediatamente um *que*.

Para *o que*, as duas primeiras linhas são iguais; a terceira linha indica que o WNP* domina imediatamente um determinante (D) e, na quarta linha, diz-se que esse D domina imediatamente *o* — ou seja, o determinante é o artigo masculino *o*; na quinta linha entende-se que o determinante tem relação de irmandade (*HasSister*) com um WPRO e, na última linha, é indicado que WPRO domina imediatamente um *que* (i.e., o WPRO é um *que*).

Ficaram excluídos dessas buscas apresentadas — porque não foram pedidos — alguns dados que estavam em orações interrogativas em discurso direto (marcados como CP-QUE-SPE na anotação). A fim de encontrar mais interrogativas, portanto, foram rodadas duas novas buscas para *que* e outra para *o que*:

Tabela 3 - Buscas por *que* e *o que* em orações interrogativas em discurso direto

BUSCA 3: QUE em CP-QUE-SPE	BUSCA 4: O QUE em CP-QUE-SPE
define: port.def print_indices: t node: CP*	define: port.def print_indices: t node: CP*
query: (CP-QUE-SPE* idoms WNP*) AND (VB* HasSister CP-QUE-SPE*) AND (WNP* idomsonly WPRO) AND (WPRO idoms que)	query: (CP-QUE-SPE* idoms WNP*) AND (VB* HasSister CP-QUE-SPE*) AND (WNP* idoms D) AND (D idoms o) AND (D HasSister WPRO) AND (WPRO idoms que)

Em ambas as buscas, houve a retirada de “CP-QUE is root” e a inserção de “CP-QUE-SPE*” em lugar de “CP-QUE*”, sendo “-SPE” a sub-etiqueta que indica discurso direto. Adicionou-se também a linha VB* HasSister CP-QUE-SPE*, que reforça a relação de complemento do CP-QUE ao verbo.

Foi necessário também criar mais duas buscas para o estudo das orações interrogativas com sujeito nulo. As novas buscas foram inspiradas nas quatro primeiras, porém com alterações restritivas que excluem casos de sujeito lexicalmente realizado, ou, de acordo com

⁷ Cf. *Corpus Search 2 Users Guide*, seção II.4. Disponível em: <http://corpussearch.sourceforge.net/CS-manual/CommandFile.html#print_indices>. Acesso em 6 nov. 2021.

a nomenclatura do corpus, casos que têm as categorias vazias *NP-SBJ* **pro** ou *NP-SBJ* **exp** anotadas⁸. São elas:

Tabela 4 - Buscas por *que* e *o que* em orações interrogativas com sujeito nulo

<i>BUSCA 5: QUE - sujeito nulo</i>	<i>BUSCA 6: O QUE - sujeito nulo</i>
define: port.def print_indices: t node: CP* query: (CP-QUE* IsRoot) AND (CP-QUE* idoms WNP*) AND (WNP* idomsonly WPRO) AND (WPRO idoms Que) AND (WNP* HasSister IP*) AND (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms *pro****exp***)	define: port.def print_indices: t node: CP* query: (CP-QUE* IsRoot) AND (CP-QUE* idoms WNP*) AND (WNP* idoms D) AND (D idoms o) AND (D HasSister WPRO) AND (WPRO idoms que) AND (WNP* HasSister IP*) AND (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms *pro****exp***)

A diferença destas duas buscas para as primeiras é menção de que *WNP** é irmão de um *IP** (geralmente um *IP-SUB*) e é dentro deste *IP** que está anotado o sujeito nulo (*NP-SBJ* idoms *pro****exp****). Como as buscas específicas para encontrar dados de interrogativas em discurso direto (buscas 3 e 4, tabela 3) não apresentaram muitos resultados, não foi necessário escrever mais duas buscas para encontrar os sujeitos nulos nesses casos.

Análise dos dados

A princípio, tencionava-se fazer uma comparação entre dados portugueses e brasileiros, mas foram utilizados apenas dois arquivos de textos brasileiros — as Atas dos Brasileiros (arquivo va_002 no Corpus Tycho Brahe) e o corpus de cartas — com muito menos dados, sendo encontrado somente um resultado nas cartas, na busca por orações interrogativas indiretas com *o que*. No caso das Atas dos Brasileiros, há poucos dados — são apenas duas interrogativas diretas —, e por essa razão não foi possível fazer uma análise específica com arquivos brasileiros. Esta pesquisa trabalha, portanto, com dados do português europeu.

Dito isso, encontram-se os seguintes valores totais, isto é, o número de dados do Corpus Tycho Brahe e do Corpus Post Scriptum somados, de *que* e *o que* por séculos:

Tabela 5 - Ocorrências de *que* e *o que* e porcentagem de *o que* por séculos⁹

Século	<i>Que</i>	<i>O que</i>	Porcentagem aproximada de <i>o que</i> (N <i>o que</i> / N <i>o que</i> + N <i>que</i>)

⁸ No caso de **pro** e **exp**, os asteriscos entre *pro* e *exp* não significam “qualquer”. Trata-se de uma convenção de anotação de sujeitos nulos referenciais (*pro*) e nulos expletivos (*exp*).

⁹ Alguns valores podem ter sofrido pequenas variações, que não consistem em mudanças que afetam a análise apresentada no 17º SePeG, por conta de uma revisão dos dados realizada durante a elaboração do Relatório I enviado à Fapesp.

XVI	2	0	0% (0/2)
XVII	35	2	5,40% (2/37)
XVIII	90	2	2,17% (2/92)
XIX	66	9	13,63% (9/77)

Nos textos dos autores Francisco Manuel de Melo, António José da Silva, Jose Daniel Rodrigues da Costa e J. B. da Silva L. de Almeida Garrett, *que* e *o que* aparecem em variação. Abaixo (tabela 6), apresenta-se a porcentagem e *o que* em relação ao número total de ocorrências.

Tabela 6 - Ocorrências e porcentagem de *o que* em textos de autores que utilizam ambas as formas

Autor	Gênero textual	Século	<i>Que</i>	<i>O que</i>	Porcentagem aproximada de <i>o que</i> (N <i>o que</i> / N <i>o que</i> + N <i>que</i>)
Francisco Manuel de Melo	Cartas	XVII	7	2	22,22% (2/9)
António José da Silva (o Judeu)	Teatro	XVIII	61	1	1,61% (1/62)
Jose Daniel Rodrigues da Costa	Teatro	XVIII	16	1	5,88% (1/17)
J. B. da Silva L. de Almeida Garrett	Teatro	XIX	46	7	13,20% (7/53)

Os valores da tabela 5 mostram que *o que* começou a ser utilizado no século XVII, mas de forma marginal; e ainda se mantém marginal no século XVIII, pois o número de *o que* é o mesmo do século anterior e há um número maior de *que*. No século XIX, há um aumento do uso de *o que* e conseqüentemente um aumento da porcentagem dessa variante. Como foram obtidos poucos dados em corpora grandes, não é possível fazer uma análise estatística muito apurada; não obstante, tentou-se encontrar características que pudessem influenciar a variação.

Os dados mostram uma tendência de aumento do uso de *o que*, e, ainda que não se tenha dados do século XX, isso vai de acordo com o que diz a gramática contemporânea de

Mateus e outros autores (Mateus et al, 2006) sobre *que* estar cada vez menos sendo utilizado para fazer perguntas. Na tabela 6, é notório o aumento da porcentagem de *o que* entre os séculos XVIII e XIX, quando há mais dados de orações interrogativas, que se encontram essencialmente em peças de teatro. No caso de Francisco Melo, a alta porcentagem de *o que* está relacionada com o número menor de resultados com *que*.

No Corpus Post Scriptum, apenas foram encontradas duas orações interrogativas diretas no século XIX:

1. Que será se chegar à real presença de Sua Majestade que se vende e se compra a justiça? (CARDS0094,.14)

2. Estimando eu um homem casado, o que dirás tu: « tolinha, caída que é! »? (CARDS0093,.6)

A oração (1) pertence à carta de Bartolomeu Luís a um elemento da Intendência Geral da Polícia da Corte e Reino¹⁰; a oração (2) pertence à carta de um autor anônimo¹¹.

Como primeira abordagem qualitativa dos dados, foi feita uma análise da estrutura das orações interrogativas diretas tendo como inspiração a pesquisa que Sikansi (1994) realizou quanto à estrutura das orações interrogativas com pronome interrogativo, especificamente, no que diz respeito à ordem sujeito-verbo (VS vs. SV). Ao analisar os dados do corpus NURC/SP e atentar-se para a variação entre *que* e *o que*, a autora chama atenção para o fato de que, nesse corpus, *que* favorece a ordem VS e *o que* apresenta um comportamento quase neutro no que diz respeito à ordem do sujeito e do verbo, mas conclui, ao comparar os dados com os de outros corpora, que esses pronomes apresentam um comportamento inconstante sobre a ordem dos constituintes. Observando isso nos textos do Corpus Tycho Brahe, que são portugueses, percebe-se que todas as interrogativas que começam por *que*, sem exceção, apresentam a ordem VS. São, no total, 66 orações. Seguem alguns exemplos:

3. “¿Que quiere Vossa Excelência dos felizes?” (A_004_PART_PSD, 50.705) — século XVIII;

4. “— Que faria de ti a política, meu querido, meu poeta da Pátria e da alma?” (B_005_PSD, 3.11) — século XIX;

5. “— Que tens tu, mulher? — exclamou a mãe.” (B_005_PSD, 25.687) — século XIX.

As orações com *o que* (e com sujeito presente) também são todas VS. Como não há casos de ordem SV para *o que* com sujeito realizado, não se pode afirmar que o que constatou Sikansi (1994) para dados do corpus NURC/SP se aplique aos textos do Corpus Tycho Brahe.

Contudo, tendo em mãos esses dados do Corpus Tycho Brahe, percebeu-se que alguns tipos de orações interrogativas diretas parecem ter-se cristalizado com a variante *que*. É o que se averigua com as buscas 5 e 6 (tabela 4), que filtram apenas as interrogativas com sujeito nulo. Nos arquivos de saída dessas buscas, constatou-se que algumas orações interrogativas com sujeito nulo expletivo só aparecem com *que*. Trata-se de orações com as expressões *que importa*, *que há*, *que + se* indeterminador e *que é de*.

Computando-se os valores, tem-se um total de 98 interrogativas com sujeito nulo com o uso de *que*, sendo 66 do tipo NP-SBJ **pro**, isto é, sujeitos nulos referenciais, e 32 do tipo NP-SBJ **exp**, sujeitos nulos expletivos, no Corpus Tycho Brahe.

¹⁰ Cf. Corpus Post Scriptum. Disponível em: <<http://teitok.clul.ul.pt/postscriptum/index.php?action=edit&cid=CARDS0094>>. Acesso em 4 jan. 2022.

¹¹ Cf. Corpus Post Scriptum. Disponível em: <<http://teitok.clul.ul.pt/postscriptum/index.php?action=edit&cid=CARDS0093>>. Acesso em 4 jan. 2022.

Tabela 7 - Ocorrências da variante *que* com sujeito nulo expletivo por século no Corpus Tycho Brahe

Período	Total de sujeitos nulos	Sujeitos nulos expletivos (*exp*)	Porcentagem de sujeitos nulos expletivos em relação ao total
Século XVI	1	0	0%
Século XVII	7	6	85,71%
Século XVIII	56	16	28,57%
Século XIX	34	10	29,41%

O resultado da busca de *o que* apresenta uma discrepância de valores ao se comparar com a quantidade de perguntas com *que*. Apenas foi encontrado um caso de sujeito nulo, no século XVIII:

6. “Ó mana, o que fez?” (C_005_PSD,211.1904)

Em (6), trata-se de uma oração interrogativa com sujeito nulo referencial. Nessa oração, é possível fazer a paráfrase do pronome interrogativo *o que* pela expressão *que coisa*. Uma análise dos casos que aqui são chamados de cristalizados mostra que essa paráfrase nem sempre funciona bem com *que*. São casos como os de abaixo:

7. “Que importa que sejam diversos os lugares se nos sujeitos não há diversidade?” (A_001_PSD, 68.1459) — século XVIII

8. “Que há de ser?” (S_004, 11.75) — século XVIII

9. “— Que é de tua mãe rapariga? — perguntou de fora, porque a língua da chave estava corrida.” (B_005_PSD, 27.717) — século XIX

10. “— Que se diz de meu amo?” (G_004_PSD, 0.2819) — século XIX.

No que diz respeito aos sujeitos nulos expletivos, que correspondem nos séculos XVII, XVIII e XIX, respectivamente, a 85,71%, 28,57% e 29,41% do total de sujeitos nulos (tabela 7) para a variante *que*, percebe-se que, dentre eles, as expressões cristalizadas são a maior parte das ocorrências, como se evidencia na tabela 8.

Tabela 8 - Estruturas cristalizadas em relação ao total de casos com sujeito nulo expletivo no Corpus Tycho Brahe

Século	Total expletivos na variante <i>que</i>	<i>Que importa, que há, que é de, que se...</i>	% estruturas cristalizadas com <i>que</i> em relação ao total de expletivos nesta variante
XVII	6	5	83,3%
XVIII	16	14	87,5%
XIX	10	8	80%

É curioso notar que, na oração (1), encontrada no Corpus Post Scriptum, há sujeito nulo expletivo e também não se pode parafrasear *que* por *que coisa* na estrutura *que + ser* com

sentido de *acontecer*: “Que será se chegar à real presença de Sua Majestade que se vende e se compra a justiça?” (CARDS0094, 14).

Os fenômenos associados à variação *que/o que* parecem indicar que *o que* teria uma função referencial mais forte, uma vez que não se encontram casos de sujeitos nulos expletivos com *o que*. Analisar esse aspecto referencial de *o que* é o próximo passo desta pesquisa de iniciação científica, atentando-se para o discurso. Além disso, serão adicionados dados de orações interrogativas indiretas à análise, verificando-se também quando ocorre a entrada de *o que* nessas orações.

Conclusão: primeiros resultados

Descobriu-se que, nas orações interrogativas diretas, a variante *o que* apenas começou a aparecer no século XVII; aliás, *o que* mostra-se menos recorrente até o século XIX. Isso permite afirmar que *o que* é a forma inovadora na língua. O que se percebeu até então é que não há uma relação entre a realização lexical ou não do sujeito e a ocorrência de uma forma ou outra, mas que algumas estruturas com verbos impessoais foram cristalizadas com a variante *que*.

A cristalização permite hipotetizar que a origem da forma *o que* esteja ligada com a referencialidade do elemento interrogativo. Essa hipótese se fortalece quando se percebe que a paráfrase para *que/oque* indicada por Grivet (1881) — *que coisa* — nem sempre funciona bem com *que*.

A ordem das orações interrogativas que têm sujeito lexical no corpus Tycho Brahe é V-S, o que demonstra que não há uma relação entre a ordem dos constituintes da oração e a seleção de uma ou outra forma do pronome interrogativo.

A análise se direciona, a partir deste ponto, para o aspecto referencial de *o que*; esse novo rumo permitirá verificar se existe uma relação entre o discurso e a sintaxe na emergência de *o que*, com base na proposta de Pesetsky (1987) de que certos elementos interrogativos são D-linked, ou seja ligados ao discurso. Além disso, serão incluídos dados de orações interrogativas indiretas, o que permitirá verificar, nessas orações, se a entrada de *o que* ocorreu em um momento diferente e também se o discurso pode ter uma relação com o uso de *o que*.

Referências

- AZEVEDO FILHO, L. A. de. (1968). Gramática básica da língua portuguesa, Ed. Fundo de Cultura, RJ.
- BARROS, J. de. (1540). Grammatica da lingua portuguesa, Olyssipone: apud Lodouicum Rotorigiu[m], Typographum. Disponível em: <<https://purl.pt/12148>>. Acesso em 13 nov. 2021.
- BECHARA, E. (2019). Moderna gramática portuguesa. 39. ed. rev. e ampl., Ed. Nova Fronteira, RJ.
- CEGALLA, D. (2008). Novíssima gramática da língua portuguesa. 48. ed., Ed. Nacional, SP.
- CLUL. (2014). P.S. Post Scriptum. Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época Moderna. Disponível em: <<http://ps.clul.ul.pt>>. Acesso em 24 Jan. 2021.
- DIAS, A. E. da S. (1984). Grammatica portugueza elementar. 9ª ed. rev., A. Ferreira Machado, Lisboa. Disponível em: <<https://purl.pt/29127>>. Acesso em 13 nov. 2021.
- GALVES, C.; ANDRADE, A.; FARIA, P. (2017). Tycho Brahe Parsed Corpus of Historical Portuguese. Campinas: Instituto de Estudos da Linguagem, Universidade de Campinas. Disponível em: <<http://www.tycho.iel.unicamp.br/corpus/>>. Acesso em: 24 Jan. 2021.
- MATEUS, M. H. M. et al. (2006). GRAMÁTICA da língua portuguesa. 7. ed., Ed. Caminho, Lisboa. (Coleção universitária. Série linguística).
- GRIVET, C. A. O. (1881). Nova grammatica analytica da lingua portugueza, Typ. De G. Leuzinger & Filhos, RJ. Disponível em: <<https://digital.bbm.usp.br/handle/bbm/7507>> Acesso em: 4 Mai. 2020.
- HAUY, A. B. (2015). Gramática da língua portuguesa padrão. 1. ed. 1. reimpr., Editora da Universidade de São Paulo, SP.

- MAGRO, C.; GALVES, C. (2019). Portuguese Syntactic Annotation Manual. Disponível em: <<https://sites.google.com/site/portuguesesyntacticannotation/>>. Acesso em: 24 Jan. 2021.
- PESETSKY, D. (1987). Wh-in-Situ: Movement and unselective binding. In: REULAND, E.; TER MEULEN, A. The representation of (in)definiteness. MIT Press, Cambridge, p. 98-129.
- RIBEIRO, J. (1881). Grammatica portugueza, Typ. De Jorge Seckler, SP. Disponível em: <<https://digital.bbm.usp.br/handle/bbm/2097>>. Acesso em 4 mai. 2020.
- SIKANSI, N. S. (1994). A estrutura das sentenças com pronome interrogativo no português brasileiro atual. Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP. Disponível em: <<http://www.repositorio.unicamp.br/handle/REPOSIP/271035>>. Acesso em 3 Mai. 2020.
- SOUSA, M. D. de. (1804). Grammatica portugueza, Real Imp. da Universidade, Coimbra. Disponível em: <<https://purl.pt/17363>>. Acesso em 13 nov. 2021.