


# A INFLUÊNCIA DA AMOSTRA NA ORDENAÇÃO DE CONSTRUÇÕES CAUSAIS COM PORQUE E POR + INFINITIVO

**Mayra França Floret\***

 <https://orcid.org/0009-0000-2272-9124>

**Como citar este artigo:** FLORET, M. F. A influência da amostra na ordenação de construções causais com *porque* e *por + infinitivo*. *Todas as Letras – Revista de Língua e Literatura*, São Paulo, v. 26, n. 1, p. 1-16, jan./abr. 2024. DOI: <https://doi.org/10.5935/1980-6914/eLETL16259>.

**Submissão:** 29 de junho de 2023. **Aceite:** 11 de janeiro de 2024.

**Resumo:** A oração causal com *porque* e *por + infinitivo* pode ocorrer em diferentes posições dentro do período causal. Neste artigo, analisamos a atuação do princípio de iconicidade, do princípio de distribuição de informação e dos domínios da causalidade na ordenação de períodos causais com os dois conectores. Os resultados mostram que a oração causal tende a ser posposta independentemente das variáveis analisadas. Os resultados também mostram que as características particulares do texto *Atas dos brasileiros* se refletiram sobre os resultados encontrados para o século XIX de maneira geral, o que evidenciou a influência da seleção da amostra nos resultados da pesquisa.

**Palavras-chave:** Causa. Ordenação. Amostra. *Porque*. *Por + infinitivo*.

\* Universidade do Estado do Rio de Janeiro (UERJ), São Gonçalo, RJ, Brasil. E-mail: [mayrafloret@yahoo.com.br](mailto:mayrafloret@yahoo.com.br)

## INTRODUÇÃO

■ **A** relação entre segmentos que representam causa e efeito costuma se concretizar por meio do uso de conectores. No português, há diversos conectores que podem estabelecer essa relação, entre eles *porque* e *por + infinitivo*. Neste estudo, que é parte do trabalho mais amplo desenvolvido em Floret (2018)<sup>1</sup>, analisamos a ordenação de construções causais<sup>2</sup> ligadas por esses dois elementos, que foram escolhidos por quatro razões principais.

Em primeiro lugar, são altamente frequentes na língua, especialmente *porque*. Além disso, são derivados da mesma base, que é a preposição *por* (Barreto, 1999; Braga; Paiva, 2011a, 2011b), e estabelecem relações nos domínios referencial, epistêmico e interacional (Sweetser, 1990). Por fim, nas construções conectadas por esses conectores, a oração causal tem flexibilidade de posição, podendo ser anteposta, posposta ou interposta à oração núcleo. Os dados (1), (2) e (3) ilustram essas diferentes possibilidades.

- (1) Nessa Corte de D. João V., na qual as paixões e intrigas fervem e tumultuam, mas, **por ser em terra portugueza**, encontram-se peitos que batem sempre leaes (Século XX – *Provocações e debates*).
- (2) Meus versos arrebatados à força de os esticar, agradavam-lhe ainda assim, **porque no fim de contas eram um arremedo de poesia**; e porventura levavam um perfume da primavera da alma (Século XIX – *Como e porque sou romancista*).
- (3) O homem **porque** foi formado da terra, ainda que seja com dispendio da própria vida, e summa repugnancia da vontade, sempre vae buscar a terra, e só descança na sepultura (Século XVII – *Sermões do Pe. Antonio Vieira*).

Neste estudo, analisamos a atuação de dois princípios funcionais, o de iconicidade e o de distribuição de informação, na posição da oração causal em relação à oração núcleo. Para verificar a ordenação das construções causais, analisamos algumas de suas propriedades, entre elas o domínio da causalidade, a sequencialidade de eventos e o estatuto informacional das orações que compõem o período causal.

A coleta dos dados foi realizada em textos de dois períodos da história do português: período clássico (séculos XVII e XVIII) e período moderno/contemporâneo (séculos XIX, XX e XXI) (Castro, 2013). Por se tratar de uma pesquisa diacrônica, ficam evidentes alguns problemas envolvidos no estudo linguístico ao longo do tempo. Paiva e Duarte (2003), com base em Labov (1994), mostram que é impossível saber se uma amostra escrita é realmente representativa de uma comunidade de fala da época. Outro problema é que textos escritos podem preservar formas já não mais utilizadas na fala há algum tempo.

Textos escritos podem, ainda, ter sofrido alterações por copistas ou editores. Além disso, não é possível determinar a gramaticalidade das formas que não aparecem nos textos da amostra. Uma determinada forma pode não aparecer por ser agramatical ou simplesmente por uma escolha do autor. Na maioria das

1 Este artigo é baseado em parte da fundamentação teórica, dos resultados e da análise desenvolvida na dissertação de mestrado da autora (Floret, 2018).

2 Ao longo do artigo, referimo-nos à relação entre oração principal e oração causal como período causal, expressão adotada mais tradicionalmente em gramáticas do português, ou como construção causal (cf., por exemplo, Goldberg, 2006), independentemente da ordem em que ocorram.

vezes, não conhecemos a posição social do autor, nem a estrutura da comunidade em que determinado texto foi escrito. Some-se a isso o fato de que os textos sobrevivem ao acaso, ou seja, podem ser resultados de questões históricas imprevisíveis. Todas essas dificuldades apontadas colocam para o linguista a responsabilidade de fazer escolhas sobre como escolher os textos e até que ponto a amostra selecionada pode dar conta do fenômeno estudado.

A partir do estudo desenvolvido em Floret (2018), evidenciou-se a influência da seleção da amostra nos resultados encontrados. Dessa forma, o objetivo central deste estudo é mostrar de que maneira a seleção de determinados textos pode afetar e enviesar resultados mais gerais de um estudo, em especial de um estudo diacrônico.

A seguir, apresentamos brevemente os pressupostos teóricos que baseiam o estudo. Depois, indicamos a amostra e os procedimentos metodológicos adotados. Na sequência, apresentamos e discutimos os resultados encontrados, apontando os casos em que a seleção da amostra influenciou os resultados. Por fim, detalhamos alguns aspectos do texto que influenciaram os resultados encontrados e apresentamos as considerações finais.

## **O USO DA LÍNGUA COM BASE NA EXPERIÊNCIA**

O desenvolvimento deste estudo teve como base os pressupostos teóricos dos modelos baseados no uso – MBU (Kemmer; Barlow, 2000; Bybee, 2010; Bybee; Beckner, 2010). De acordo com os MBU, e como o próprio nome sugere, a experiência do falante no uso da língua é o principal elemento na organização e no processamento do sistema linguístico.

Entre os diversos princípios que norteiam os trabalhos com base nessa perspectiva teórica, destaca-se a relação entre forma e função. Por isso, analisamos a atuação de dois princípios funcionais, quais sejam, o de iconicidade e o de distribuição de informação. Nos dois casos, a forma como o falante usa a língua reflete questões ligadas à sua experiência extralingüística.

De acordo com Croft (1990, p. 102)<sup>3</sup>, por trás do princípio de iconicidade, está a ideia de que “a estrutura da língua reflete, de alguma forma, a estrutura da experiência”. Nesse sentido, segundo Haiman (1980, 1983), a estrutura reflete a forma como representamos a realidade. Assim, fica evidente que não usamos a língua de forma aleatória, mas motivada pelas nossas experiências não necessariamente linguísticas com o mundo.

Haiman (1980, 1983) destaca que o sequenciamento de eventos no texto é um caso comum de motivação icônica. Tendo em vista a relação entre causa e efeito, nossa experiência com a realidade mostra que causas precedem seus efeitos. Em outras palavras, a ocorrência do efeito está condicionada à ocorrência anterior da causa. Se a estrutura da língua reflete nossa experiência com a realidade, construções causais devem obedecer à ordem causa-efeito, o que significa a anteposição da oração causal, como no exemplo (1).

Além do princípio de iconicidade, o princípio de distribuição de informação também pode ajudar a explicar a ordenação de períodos causais. Segundo esse princípio, informações velhas precedem informações novas. Como a informação nova é, em geral, a parte mais importante da mensagem, tende a ser colocada ao

---

3 “The structure of language reflects in some way the structure of experience.”

final, ao passo que a informação velha seria um ponto de partida para a interpretação da informação nova (cf. Rafajlovicová, 2010).

Para Chafe (1976) e Clark e Haviland (1977), determinar se uma informação é velha ou nova depende do que o falante acredita que esteja disponível para o ouvinte. Se o falante acredita que determinada informação já é conhecida ou está disponível para o ouvinte, então será considerada velha. Contudo, se o falante considera que a informação em questão não está disponível ou não pode ser acessada pelo ouvinte, então essa informação será considerada nova. O falante deve organizar, portanto, as informações de acordo com o que avalia estar disponível para o ouvinte na situação comunicativa.

Por sua vez, Prince (1979) define o tipo de informação com base em critérios discursivos. Uma informação é considerada nova se está sendo inserida pela primeira vez no discurso. Se já foi mencionada anteriormente, é considerada velha. A autora propõe, ainda, um terceiro tipo, que é a informação inferível. Nesse caso, a informação não está diretamente mencionada no discurso anterior, mas pode ser inferida por meio de outras informações já disponíveis no contexto. Neste trabalho, consideramos a definição proposta por Prince (1979) e, assim, analisamos as informações encontradas nos dados como novas, velhas ou inferíveis, com base no que está no próprio texto.

Tendo em vista a organização de informações no texto, de acordo com Chafe (1984), o fluxo comunicativo tende a ser organizado a partir das informações velhas, disponíveis ou já conhecidas, para as informações novas. Isso reforça o que fora colocado anteriormente, ou seja, que informações velhas precedem informações novas. No período causal, esperamos que a primeira oração codifique informação velha ou inferível, ao passo que a segunda codifique informação nova. Portanto, a oração causal tende a ser anteposta se codifica informação velha ou inferível, e posposta se codifica informação nova.

Apesar de haver flexibilidade na posição da oração causal e de outras orações adverbiais, Diessel (2001) mostra que cada língua pode apresentar diferentes possibilidades de ordem não marcada. Com base no inglês falado, Diessel (2005) verificou que a oração adverbial é mais frequentemente posposta à oração núcleo. Essa tendência se confirma nos períodos causais, ainda que a oração causal possa, ocasionalmente, ser anteposta à oração núcleo (cf. Schiffrin, 1985; Silva, 2008; Floret, 2018).

Paiva (1991, 1996) mostra que princípios funcionais, como o de iconicidade, podem servir como motivação para a anteposição da oração causal. Se a ordem de ocorrência no mundo é causa-efeito, e o uso da língua reflete nossa experiência com a realidade, então podemos esperar que a oração causal seja anteposta à oração núcleo, como já mostramos anteriormente. Entretanto, a autora verifica que o princípio é pouco operante, já que o conector colabora para o estabelecimento da relação entre as orações independentemente da ordem em que ocorram.

Tendo em vista a distribuição de informação no período composto, de acordo com Diessel (2005), a maior tendência à posposição da oração com *porque* se deve ao seu conteúdo. Em geral, orações introduzidas pelo conector tendem a introduzir informação nova no discurso, o que explica sua maior frequência em posição final.

Como já mostramos anteriormente, as orações causais com *porque* e *por + infinitivo* instauram relações em diferentes domínios, referencial, epistêmico e interacional, o que será detalhado mais adiante. Podemos imaginar que o domínio codificado no período causal possa influenciar a posição da oração causal.

De acordo com Lopes (2004), o valor de causa real ou epistêmica pode ser parcialmente explicado pela posição da oração causal. Entretanto, ao analisar dados de fala, Silva (2008) verificou que a oração causal tende a ser posposta independentemente do domínio codificado. Já Santos (2016) verificou que a anteposição de orações causais é mais expressiva no domínio epistêmico, o que pode indicar uma possível correlação entre domínio e ordenação.

Na próxima seção, apresentamos a amostra utilizada e os procedimentos adotados no desenvolvimento deste estudo. Em seguida, indicamos os resultados encontrados, tendo em vista as questões discutidas, e mostramos como a seleção da amostra influenciou os resultados encontrados sobre a ordenação dos períodos causais ao longo do tempo.

## AMOSTRA E METODOLOGIA

A amostra utilizada é constituída por textos dos períodos clássico e moderno/contemporâneo. Dessa forma, coletamos os dados com construções causais conectadas por *porque* e *por + infinitivo* em textos do século XVII até XX/XXI. A fim de garantir o equilíbrio nos dados, a amostra de cada período conta com aproximadamente 300 mil palavras. O Quadro 1 mostra os textos escolhidos, os séculos a que pertencem e o número de palavras.

Texto	Século	Número de palavras
<i>Epanáforas de vária história portuguesa</i>	XVII	64.337
<i>História do futuro</i>	XVII	50.512
<i>Sermões do Pe. Antonio Vieira</i>	XVII	53.855
<i>Gazetas manuscritas da Biblioteca de Évora</i>	XVIII	56.771
<i>Reflexões sobre a vaidade dos homens</i>	XVIII	56.211
<i>Processos e crimes do século XVIII</i>	XVIII	3.125
<i>Atas dos brasileiros</i>	XIX	53.529
<i>Memórias do Marquês da Fronteira e d'Alorna</i>	XIX	60.919
<i>Como e porque sou romancista</i>	XIX	9.405
<i>Provocações e debates: contribuições para o estudo do Brazil social</i>	XX	92.617
<i>Capitalismo, trabalho e formação profissional: dilemas do trabalho dos assistentes sociais em Ribeirão Preto</i>	XXI	36.000
<i>Manual de desastres: desastres naturais</i>	XXI	49.675
Período clássico = 284.811 palavras		
Período moderno/contemporâneo = 302.145 palavras		

### Quadro 1 – Amostra selecionada

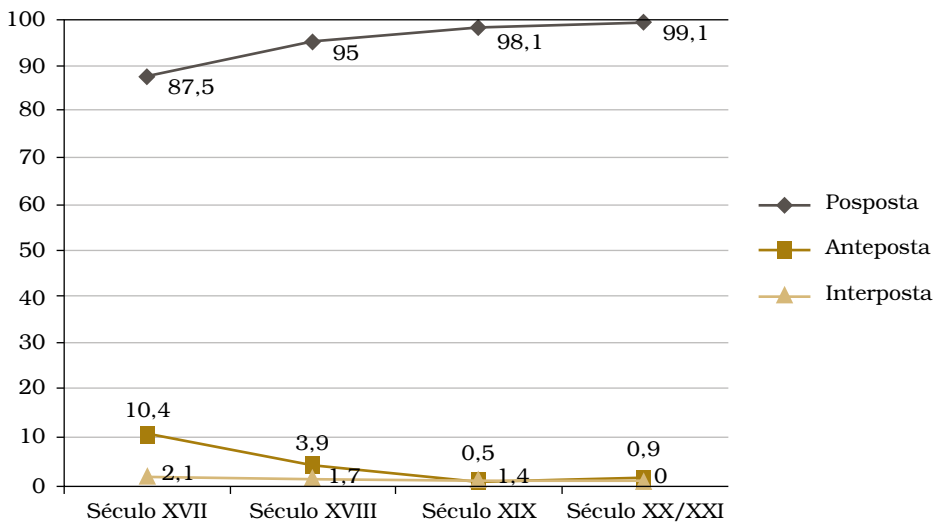
Fonte: Floret (2018, p. 46).

Os textos da amostra são de gêneros textuais diversos. Dessa forma, controlamos o próprio texto como uma variável independente, o que nos permitiu detectar a influência de um dos textos sobre os resultados do século, aspecto que será abordado mais adiante.

Os textos selecionados foram lidos, e as construções causais com *porque* e *por + infinitivo* foram coletadas e analisadas com base em alguns grupos de fatores. Para o objetivo deste artigo, selecionamos apenas o domínio da causalidade, a sequencialidade temporal e a estrutura informacional. Então, os dados foram tratados estatisticamente pelo programa GoldVarbX (Sankoff; Tagliamonte; Smith, 2005). Os resultados encontrados são analisados na seção a seguir.

### ORDENAÇÃO DE PERÍODOS CAUSAIS COM *PORQUE* E *POR + INFINITIVO*

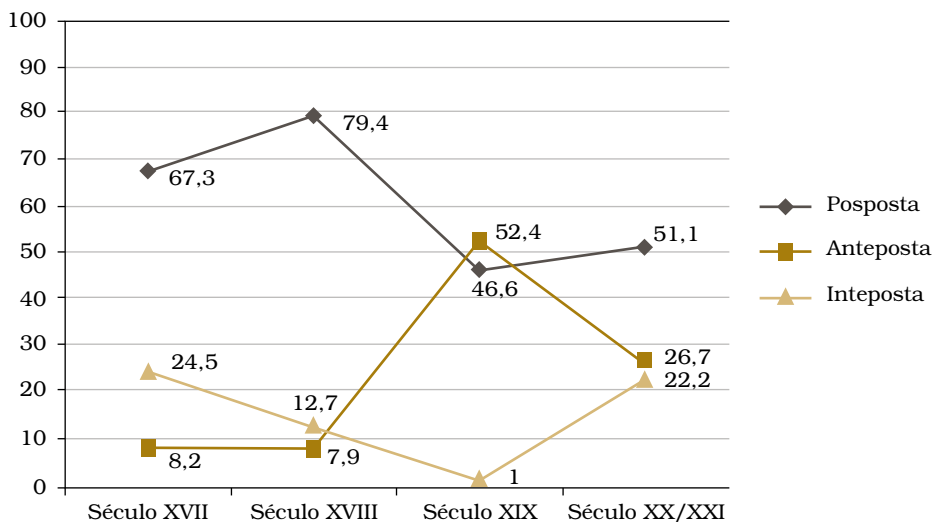
Antes de procedermos aos resultados relacionados às três variáveis mencionadas anteriormente, mostraremos a ordenação das construções analisadas ao longo dos séculos. Este é o primeiro indício de que um dos textos da amostra influenciou os resultados encontrados de maneira mais ampla. O Gráfico 1 mostra a ordenação das construções com *porque*.



**Gráfico 1** – Ordenação das construções com *porque* ao longo do tempo

Fonte: Floret (2018, p. 57).

De acordo com o Gráfico 1, a posposição da oração com *porque* já é muito mais frequente do que as demais desde o século XVII. Entretanto, a ocorrência de posposição vai se tornando cada vez mais frequente, atingindo 99,1% dos dados no século XX/XXI. Por sua vez, a anteposição vai perdendo frequência com o passar do tempo, chegando a atingir apenas 0,9% dos dados no século mais atual. A interposição mantém estabilidade desde o século XVII, correspondendo a apenas 2,1% dos dados, e chega a zero no século XX/XXI. O Gráfico 2 mostra a posição da oração com *por + infinitivo* ao longo do tempo.



**Gráfico 2** – Ordenação da construção com *por + infinitivo* ao longo do tempo

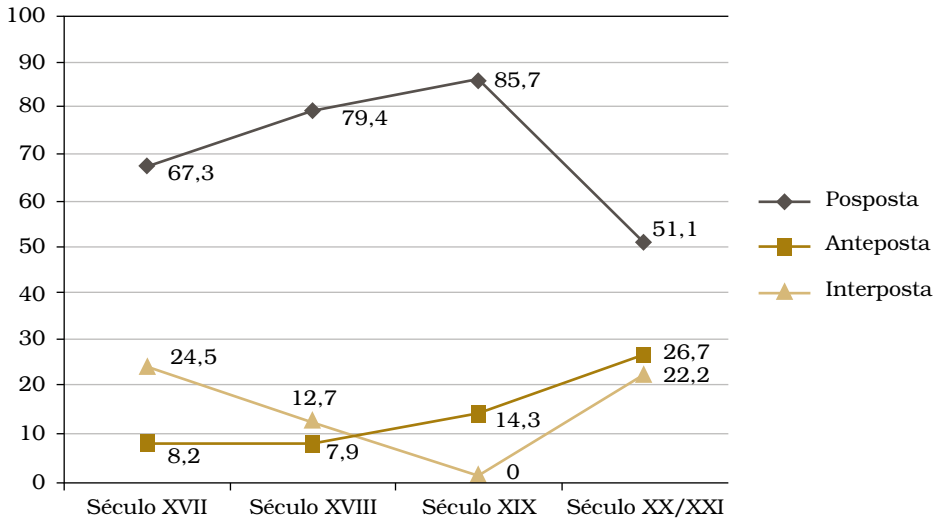
Fonte: Floret (2018, p. 58).

As linhas do Gráfico 2 mostram um padrão completamente diferente do encontrado para as construções com *porque*. A posposição é a posição mais frequente, ainda que com índices não tão expressivos. Entretanto, no século XIX, essa posição é menos frequente (46,6%) do que a anteposição (52,4%).

A anteposição, por sua vez, apresenta índices de frequência semelhantes nos séculos XVII (8,2%) e XVIII (7,9%). Entretanto, no século XIX, há um pico de frequência para essa posição, que chega a ser a mais frequente do século. A alta frequência de anteposição não se mantém no século XX/XXI, em que a posição atinge índice de 26,7%. Ainda assim, verificamos aumento de anteposição em relação aos séculos XVII e XVIII.

A interposição apresenta índices de frequência semelhantes nos séculos XVII (24,5%) e XX/XXI (22,2%). Entre esses dois séculos, porém, a frequência diminui, chegando a apenas 1% no século XIX.

Como fica evidente, o século XIX apresenta um comportamento inesperado em relação aos demais séculos e em relação à construção com *porque*. Por conta disso, foi necessário olhar novamente os dados encontrados para verificar um possível enviesamento dos dados. Assim, verificamos que o texto *Atas dos brasileiros* apresentava um grande número de construções com orações antepostas e conectadas por *por + infinitivo*. Como controlamos o próprio texto como uma variável, foi possível analisar os dados no programa GoldVarbX (Sankoff; Tagliamonte; Smith, 2005) novamente, excluindo os dados desse texto. Os resultados estão no Gráfico 3.



**Gráfico 3** – Ordenação das construções com *por + infinitivo* sem o texto *Atas dos brasileiros*

Fonte: Floret (2018, p. 59).

O Gráfico 3 mostra que a posposição é a posição mais frequente da oração conectada por *por + infinitivo* em todos os séculos. Entretanto, no século XX/XXI, essa posição perde frequência (51,1%). Já a anteposição vai ganhando frequência regularmente ao longo do tempo. No século XIX, corresponde a 14,3% dos dados. A trajetória de interposição não sofre alteração significativa ao desconsiderar o texto *Atas dos brasileiros*.

O Gráfico 3, em comparação com o Gráfico 2, mostra uma trajetória regular da construção em termos de ordenação. Até o século XIX, a anteposição e a posposição vão ganhando frequência, enquanto a interposição vai perdendo frequência. Já no século mais atual, a posposição perde frequência, enquanto a interposição ganha frequência.

A comparação entre os gráficos evidencia a influência de um texto sobre os dados mais gerais do estudo. Em um primeiro momento, poderíamos considerar um indício de mudança no século em questão. Entretanto, como essa possível mudança ocorre de forma abrupta, e esse comportamento não se mantém no século seguinte, fica claro que há uma influência pontual de um dos textos. O fato de ter controlado o próprio texto como variável ajudou nesse aspecto, já que ficou muito mais fácil medir a influência de um texto isoladamente.

Agora que sabemos da influência de um dos textos da amostra sobre os dados, precisamos de um olhar muito mais cauteloso para os resultados. A primeira variável a ser analisada é o domínio da causalidade. Como já mencionamos anteriormente, de acordo com Sweetser (1990), construções causais codificam relações em três domínios.

No domínio referencial, é estabelecida uma relação de causa e consequência, ou seja, a relação é possível e verificável no mundo real. O exemplo (4) ilustra essa possibilidade.



- (4) O senhor presidente declarou que a Sociedade continua o socorro ao sócio Manoel Cláudio **por estar** decrépito, e para este socorro marcou 2\$000 réis semanal em vista do artigo 10, §º 2º dos nosso estatuto (Século XIX – *Atas dos brasileiros*).

No exemplo (4), o fato de uma pessoa estar com idade avançada é a causa que leva ao efeito de que essa pessoa necessita de ajuda. São, portanto, dois estados de coisas que se relacionam no mundo real.

O domínio epistêmico, por sua vez, envolve algum tipo de raciocínio inferencial por parte do falante a partir de uma evidência. Nesse caso, fica mais evidente a presença do emissor na tomada de uma conclusão. O exemplo (5) ilustra esse domínio.

- (5) Esperamos que este livro contribua para a reflexão sobre o trabalho profissional dos assistentes sociais e sobre suas dificuldades para pensar seu referencial teórico-metodológico. [...] Entendemos necessária tal reflexão **porque ela promove para o profissional a crítica necessária para enfrentar as dificuldades que enfrenta no seu trabalho cotidiano e exige dele capacidade de leitura para decidir acerca do potencial do trabalho a realizar diante das mais diversas demandas** (Século XXI – *Capitalismo, trabalho e formação profissional*).

No exemplo (5), a causa sublinhada é um fato, uma evidência, que leva o autor a concluir que a reflexão é necessária. Nesse caso, fica clara a tomada de conclusão por parte do emissor, que é reforçada pelo uso de verbos como “entendemos” e “esperamos”.

Por fim, no domínio interacional, uma oração causal justifica um ato de fala desempenhado, como no exemplo (6).

- (6) Prendam-se pois as mulheres para que se evite o mal de que elas amem; sejam conduzidas por força para os Claustros, para que não suceda que as amemos nós; saiam do berço para aquelas sepulturas, **porque pode haver perigo na demora** (Século XVIII – *Reflexões sobre a vaidade dos homens*).

No exemplo (6), o emissor ordena que as mulheres “saiam do berço para aquelas sepulturas”, e justifica essa ordem por meio da oração causal sublinhada.

Tendo em vista que o principal objetivo deste artigo é mostrar a influência da seleção da amostra nos resultados mais gerais de uma pesquisa, detalharemos apenas os resultados referentes à construção *por + infinitivo*, contexto que evidencia essa influência. Em relação aos domínios da causalidade, podemos esperar que eles levem a diferentes formas de organização do período causal. Os resultados estão na Tabela 1.

Domínio	Século XVII		Século XVIII		Século XIX		Século XX/XXI	
<b>Referencial</b>	<b>P</b>	21 = 56,8%	<b>P</b>	37 = 92,5%	<b>P</b>	45 = 45,9%	<b>P</b>	16 = 43,3%
	<b>A</b>	4 = 10,8%	<b>A</b>	2 = 5%	<b>A</b>	52 = 53,1%	<b>A</b>	11 = 29,7%
	<b>I</b>	12 = 32,4%	<b>I</b>	1 = 2,5%	<b>I</b>	1 = 1%	<b>I</b>	10 = 27%
	<b>T</b>	37	<b>T</b>	40	<b>T</b>	98	<b>T</b>	37
<b>Epistêmico</b>	<b>P</b>	12 = 100%	<b>P</b>	13 = 56,5%	<b>P</b>	3 = 60%	<b>P</b>	7 = 87,5%
	<b>A</b>	0	<b>A</b>	3 = 13,1%	<b>A</b>	2 = 40%	<b>A</b>	1 = 12,5%
	<b>I</b>	0	<b>I</b>	7 = 30,4%	<b>I</b>	0	<b>I</b>	0
	<b>T</b>	12	<b>T</b>	23	<b>T</b>	5	<b>T</b>	8

Legenda: P = posposta; A = anteposta; I = interposta; e T = total.

**Tabela 1** – Domínio da causalidade nas construções com *por* + *infinitivo*

Fonte: Floret (2018, p. 62).

A Tabela 1 mostra que, independentemente do domínio em que se instaura a relação de causa, a oração causal introduzida por *por* + *infinitivo* tende a ser posposta. O mesmo acontece com as orações introduzidas por *porque*, mais frequentemente pospostas em todos os domínios. De acordo com a tabela, em todos os séculos, tanto no domínio referencial quanto no epistêmico, verifica-se predomínio da posposição sobre as demais possibilidades. Entretanto, isso não se confirma no século XIX.

Nesse século, a anteposição chega a ser mais frequente (53,1%) do que a posposição (45,9%) no domínio referencial, algo que está em desacordo com o que se poderia esperar, tendo em vista os demais contextos analisados. No domínio epistêmico, a posposição é mais frequente (60%), mas conta com apenas três dados. Não foram encontrados dados com *por* + *infinitivo* no domínio interacional. Por isso, esse domínio não aparece na tabela.

Como já dissemos anteriormente, o texto *Atas dos brasileiros*, produzido no século XIX, apresenta um grande número de construções com *por* + *infinitivo*, em que a oração causal é anteposta. Nesse caso, a ordem estabelecida é causa-efeito, o que estaria de acordo com o princípio de iconicidade. Além disso, a maioria dos dados com *por* + *infinitivo* encontrados no século XIX está nesse texto, o que faz com que sua organização e suas características acabem se refletindo nos resultados para o século.

A segunda variável que evidencia um comportamento não esperado para o século XIX é a sequencialidade temporal dos eventos no texto. Tendo em vista o princípio de iconicidade, esperamos que o uso da língua reflita nossa experiência com a realidade. Sendo assim, em períodos que codificam eventos sequenciais, ou seja, em que um precisa, necessariamente, começar antes do outro, a oração causal deve ser anteposta, visto que causas precedem seus efeitos. A Tabela 2 mostra os resultados encontrados.

Sequencialidade	Século XVII	Século XVIII	Século XIX	Século XX/XXI
<b>Sequencial</b>	<b>P</b> 20 = 55,6%	<b>P</b> 37 = 92,5%	<b>P</b> 41 = 43,6%	<b>P</b> 11 = 44%
	<b>A</b> 4 = 11,1%	<b>A</b> 2 = 5%	<b>A</b> 52 = 55,3%	<b>A</b> 8 = 32%
	<b>I</b> 12 = 33,3%	<b>I</b> 1 = 2,5%	<b>I</b> 1 = 1,1%	<b>I</b> 6 = 24%
	<b>T</b> 36	<b>T</b> 40	<b>T</b> 94	<b>T</b> 25
<b>Não sequencial</b>	<b>P</b> 1 = 100%	<b>P</b> 0	<b>P</b> 4 = 100%	<b>P</b> 5 = 41,7%
	<b>A</b> 0	<b>A</b> 0	<b>A</b> 0	<b>A</b> 3 = 25%
	<b>I</b> 0	<b>I</b> 0	<b>I</b> 0	<b>I</b> 4 = 33,3%
	<b>T</b> 1	<b>T</b> 0	<b>T</b> 4	<b>T</b> 12

**Tabela 2** – Sequencialidade temporal em construções com *por* + *infinitivo*

Fonte: Floret (2018, p. 68).

De acordo com a Tabela 2, as orações causais com *por* + *infinitivo* tendem a ser pospostas mesmo quando os eventos codificados são sequenciais. No caso das construções conectadas por *porque*, o mesmo ocorre; a posposição é muito mais frequente. Esse resultado não está de acordo com o esperado, visto que a ordenação efeito-causa não é icônica. Entretanto, confirma o que fora encontrado para os domínios da causalidade: a oração causal tende a ser posposta independentemente dos fatores analisados.

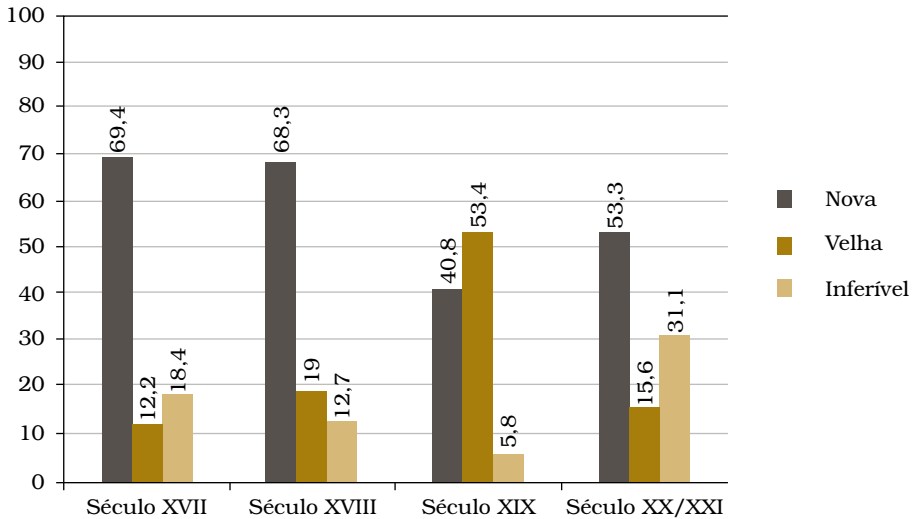
Mais uma vez, é necessário destacar que o século XIX não apresenta comportamento semelhante aos demais e à construção com *porque*. Nesse caso, a oração causal é mais frequentemente anteposta (55,3%), refletindo a ordenação icônica causa-efeito. Em princípio, esse resultado poderia ser interpretado como uma confirmação parcial da hipótese colocada. Porém, ao olharmos o resultado mais geral para as duas construções, verificamos que esse resultado é restrito a um único século, e com apenas uma das construções.

Como a grande maioria dos dados com *por* + *infinitivo* do século XIX foi encontrada no texto *Atas dos brasileiros*, evidentemente a forma como a construção é usada no texto – na ordem causa-efeito – tende a se refletir nos resultados de forma geral. Não podemos, portanto, interpretar esses resultados como algum tipo de mudança ou até mesmo como um indicativo de mudança em curso. Todos os resultados referentes à construção *por* + *infinitivo* no século XIX devem ser olhados com cautela.

Por fim, o estatuto informacional das orações que constituem o período causal reafirma a influência do texto *Atas dos brasileiros*. De acordo com o princípio de distribuição de informação, podemos esperar que o período causal se organize de forma que a primeira oração codifique informação velha ou inferível, ao passo que a segunda codifique informação nova. Portanto, a posição da oração causal no período depende do tipo de informação que ela carrega.

Com base nos resultados encontrados até aqui, já foi possível verificar a predominância da posposição sobre as demais possibilidades. Dessa forma, se a oração causal é mais frequentemente a segunda oração do período, então tende a codificar informação nova. A fim de verificar se as expectativas se confirmam,

contabilizamos a frequência de cada tipo de informação para as duas construções. Os resultados encontrados para a oração com *porque* confirmam o esperado: ela tende a codificar informação nova muito mais frequentemente do que informação velha ou inferível. Os resultados com *por + infinitivo*, por sua vez, confirmam esse resultado em parte, conforme o Gráfico 4.



**Gráfico 4** – Distribuição dos tipos de informação da oração com *por + infinitivo*

Fonte: Floret (2018, p. 70).

Os resultados encontrados para a construção com *por + infinitivo* confirmam a maior frequência de informação nova, o que está de acordo com a hipótese colocada. Se a oração causal tende a ser posposta, era de se esperar que codificasse mais frequentemente informação nova.

Entretanto, como já vimos nas outras variáveis, o resultado para o século XIX não está de acordo com os demais. Nesse caso, por ser um século com muitas construções com oração anteposta, a frequência de informação velha (53,4%) é maior do que a de informação nova (40,8%). Apesar de ser um resultado em desacordo com os demais, confirma a hipótese sobre a distribuição de informação. Se a oração com *por + infinitivo* é mais frequentemente anteposta, então tende a codificar informação velha.

Do que vimos até aqui, podemos chegar a algumas conclusões. Em primeiro lugar, verificamos que a oração causal com *porque* e *por + infinitivo* tende a ser posposta, independentemente dos domínios codificados ou da sequencialidade temporal. Dessa forma, não se confirmam as hipóteses colocadas, visto que a ordenação tende a ser efeito-causa independentemente da atuação dessas variáveis. É, portanto, uma ordenação não icônica, visto que não reflete a forma como experienciamos a relação de causa e efeito.

Por sua vez, o princípio de distribuição de informação atua da forma esperada. Considerando o fato de que as orações causais tendem a ser pospostas e que as informações tendem a ser organizadas da forma informação velha-informação nova, esperávamos encontrar maior frequência de informação nova na oração

causal. Esse resultado se confirmou: orações causais com *porque* e *por + infinitivo* codificam mais frequentemente informação nova.

É necessário destacar que todos esses resultados não se confirmam para as construções com *por + infinitivo* no século XIX. Como vimos, o texto *Atas dos brasileiros* possui algumas características particulares que acabam se refletindo nos resultados do século. Dessa forma, fica evidente que a seleção da amostra pode influenciar os resultados mais gerais encontrados em uma pesquisa.

Sendo assim, é necessário um olhar atento do pesquisador para as características de cada texto. Se o resultado encontrado para um contexto específico vai em uma direção oposta aos demais, é necessário verificar até que ponto as características particulares do texto afetam os resultados de forma geral, ou se há, de fato, um indicativo de mudança. Na seção a seguir, apresentamos algumas das particularidades do texto *Atas dos brasileiros*.

## O TEXTO *ATAS DOS BRASILEIROS*

O texto *Atas dos brasileiros* é formado por atas de reuniões que, de acordo com o gênero textual, reúnem descrições e informações sobre o que ocorreu em cada um desses eventos, além de nomes de participantes e informações sobre quando os eventos ocorreram. O texto foi escrito entre os anos 1860 e 1869.

Como já mostramos no Quadro 1, o texto possui 53.529 palavras entre as 123.853 palavras da amostra do século XIX. Em número de palavras, o texto representa 43,2% da amostra selecionada para o século, o que significa dizer que suas características particulares podem claramente se refletir nos resultados encontrados para o século, como ficou claro na seção anterior.

Além disso, foram encontrados nesse texto 46 dados de construção causal com *porque* e 96 com *por + infinitivo*, o que foge ao esperado, visto que, nos demais textos<sup>4</sup>, a construção com *porque* é mais frequente do que a com *por + infinitivo*. Dentre as 103 construções com *por + infinitivo* no século XIX, 96 estão em *Atas dos brasileiros*, mais uma vez confirmando que as características dos dados encontrados nesse texto tendem a enviesar os resultados.

Considerando esses 96 dados com *por + infinitivo*, em 53 (55,2%) a oração causal é anteposta à oração efeito, mais uma característica que particulariza o texto, tendo em vista o predomínio da posposição da oração causal independentemente da atuação das variáveis consideradas. Portanto, os dados mais frequentes são aqueles conectados por *por + infinitivo*, na ordem causa-efeito, como ilustra o exemplo (7).

(7) Foi mais deliberado que se fizesse aniversário na forma do costume. **Por não haver nada mais a tratar** o senhor presidente encerrou a sessão às 11 horas da noite (Século XIX – *Atas dos brasileiros*).

Assim, podemos dizer que o texto *Atas dos brasileiros* está na direção contrária de tudo o que encontramos como tendência nos demais textos e séculos. Primeiramente, tem mais construções com *por + infinitivo* do que com *porque*. Além disso, as orações com *por + infinitivo* são mais frequentemente antepostas, enquanto a grande maioria dos dados encontrados tem a oração causal em pospo-

4 Com exceção do texto *Manual de desastres*, em que também há mais construções com *por + infinitivo* do que com "porque".

sição. Some-se a isso o fato de o texto ter quase todos os dados com *por + infinitivo* encontrados no século XIX. Se o texto tivesse as mesmas características, mas apresentasse poucos dados, não influenciaria de forma tão evidente os resultados mais gerais.

Diante da situação colocada, ter considerado o próprio texto como uma variável foi fundamental. Assim, pudemos verificar que o que ocorre com o século XIX não é um indicativo de mudança, mas sim uma evidência de que o estilo do texto pode afetar a distribuição dos dados.

## CONSIDERAÇÕES FINAIS

Este artigo surgiu a partir de um estudo mais amplo sobre a ordenação das construções causais com *porque* e *por + infinitivo* (Floret, 2018). Neste estudo, encontramos alguns resultados que iam em direções opostas ao que se esperava encontrar. Um olhar mais atento permitiu verificar que as características particulares de um texto se refletiam nos resultados de todo o século do qual fazia parte.

Como dissemos ao longo do artigo, o próprio texto foi controlado como variável independente, o que permitiu comparar alguns resultados com e sem o texto *Atas dos brasileiros*. Dessa forma, ficou evidente que a seleção da amostra influencia os resultados encontrados em uma pesquisa e pode até mesmo levar à tomada de conclusões equivocadas. Se há resultados que fogem ao esperado e que não se confirmam nos demais contextos, é necessário analisar; pode ser que se trate de uma mudança em curso, ou seja, apenas o reflexo das particularidades de algum texto.

Além disso, ao longo deste artigo também verificamos que a oração causal com *porque* e *por + infinitivo* tende a ser posposta independentemente das variáveis que consideramos. O princípio de iconicidade não atua da forma esperada, já que a oração tende a ser posposta mesmo em períodos sequenciais, em que um evento precisa se iniciar antes do outro. Os domínios da causalidade também não levam a diferentes possibilidades de ordenação, visto que a oração causal tende a ser posposta em todas as situações. Por fim, encontramos evidências de que o princípio de distribuição de informação atua na organização do período causal, já que a oração causal, que normalmente é posposta, carrega mais frequentemente informação nova. Além disso, no século XIX, por influência do texto *Atas dos brasileiros*, as orações com *por + infinitivo* codificam mais frequentemente informação velha, o que está relacionado ao fato de essas orações serem predominantemente antepostas.

Dessa maneira, apesar de a oração causal ser flexível quanto à posição, verificamos que tende a ser posposta à oração efeito, o que leva à organização efeito-causa, que é diferente da forma como experienciamos a realidade. Além disso, ficou claro que é necessário ter cautela ao analisar os resultados, já que podem ser apenas reflexos do estilo do autor ou de características do texto. Mesmo assim, vimos que não é necessário descartar um texto que possa enviesar os resultados. Na verdade, ele pode ser uma oportunidade de mostrar que a língua em uso nem sempre segue em linhas retas da forma que imaginamos.

**THE INFLUENCE OF SAMPLE IN THE ORDERING OF CAUSAL CONSTRUCTIONS WITH *PORQUE* AND *POR + INFINITIVE***

**Abstract:** The causal clause with *porque* and *por + infinitive* can occur in different positions inside the causal sentence. In this article, we analyze the action of iconicity principle, the given-new principle and the domains of causality in the ordering of causal sentences with these two connectors. The results show that the causal clause tends to be postposed independently of the analyzed variables. The results also show that the particular features of the text *Atas dos brasileiros* were reflected in the results found for 19th century in general, which is an evidence of the influence of sample selection in the results of the research.

**Keywords:** Cause. Ordering. Sample. *Porque*. *Por + infinitive*.

**REFERÊNCIAS**

- BARRETO, T. M. *Gramaticalização das conjunções na história do português*. 1999. Tese (Doutorado em Letras) – Universidade Federal da Bahia, Salvador, 1999.
- BRAGA, M. L.; PAIVA, M. C. de. Gramaticalização e sociolinguística variacionista: o tratamento das construções introduzidas por *por*. In: LIMA, M. A. F. et al. *Colóquios linguísticos e literários: enfoques epistemológicos, metodológicos e descritivos*. Teresina: Editora da Universidade Federal do Piauí, 2011a. p. 125-153.
- BRAGA, M. L.; PAIVA, M. C. de. Gramaticalização e gramática de construções: estabilidade e instabilidade das construções complexas de causa em tempo real. *Letras & Letras*, Uberlândia, v. 27, p. 51-70, 2011b.
- BYBEE, J. *Language, usage and cognition*. New York: Cambridge University Press, 2010.
- BYBEE, J. L.; BECKNER, C. Usage-based theory. In: HEINE, B.; NARROG, H. (ed.). *The Oxford handbook of Linguistic analysis*. Oxford: Oxford University Press, 2010. p. 827-855.
- CASTRO, I. Formação da língua portuguesa. In: RAPOSO, E. et al. *Gramática do português*. Lisboa: Fundação Calouste Gulbenkian, 2013. p. 7-13.
- CHAFE, W. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In: LI, C. N. (ed.). *Subject and topic*. New York: Academic Press, 1976. p. 27-55.
- CHAFE, W. How people use adverbial clauses. In: *The proceedings of the tenth annual meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 1984.
- CLARK, H. H.; HAVILAND, S. E. Comprehension and the given – new contract. In: FREEDLE, R. O. (ed.). *Discourse production and comprehension*. New Jersey: Ablex Publishing Corporation, 1977. p. 1-40.
- CROFT, W. Grammatical categories: typological markedness, economy and iconicity. In: CROFT, W. *Typology and universals*. Cambridge: Cambridge University Press, 1990. p. 87-121.
- DIESSEL, H. The ordering distribution of main and adverbial clauses: a typological study. *Language*, v. 77, p. 345-365, 2001.

- DIESSEL, H. Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, v. 43, p. 449-470, 2005.
- FLORET, M. F. *A ordenação das construções causais com porque e por+infinitivo no Português clássico e contemporâneo*. 2018. 98f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFRJ, Rio de Janeiro, 2018.
- GOLDBERG, A. *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Express, 2006.
- HAIMAN, J. The iconicity of grammar. *Language*, v. 56, p. 515-540, 1980.
- HAIMAN, J. Iconic and economic motivation. *Language*, v. 59, p. 781-819, 1983.
- KEMMER, S.; BARLOW, M. Introduction: a usage-based conception of language. In: BARLOW, M.; KEMMER, S. (ed.). *Usage based models of language*. Stanford: CSLI Publications, 2000.
- LABOV, W. *Principles of linguistic change: internal factors*. Cambridge, Oxford: Blackwell Publishers, 1994.
- LOPES, M. H. C. C. *Aspectos sintáticos, semânticos e pragmáticos das construções causais: contributo para uma reflexão sobre o ensino da gramática*. 2004. Tese (Doutorado em Linguística) – Universidade do Porto, Porto, 2004.
- PAIVA, M. C. A. *Ordenação de cláusulas causais: forma e função*. 1991. Tese (Doutorado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1991.
- PAIVA, M. C. A. Aspectos semânticos e discursivos da relação de causalidade. In: MACEDO, A. T. de. *Variação e discurso*. Rio de Janeiro: Tempo Brasileiro, 1996. p. 63-74.
- PAIVA, M. C. A.; DUARTE, M. E. L. Mudança linguística: observações em tempo real. In: MOLLICA, M. M.; BRAGA, M. L. (org.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003. p. 179-190.
- PRINCE, E. F. On the given/new distinction. In: THE REGIONAL MEETING OF THE CHICAGO LINGUISTIC SOCIETY, 15., 1979, Chicago. *Papers [...]*. Chicago: Chicago Linguistic Society, 1979. p. 267-278.
- RAFAJLOVICOVÁ, R. Variation of clause patterns – reordering the information in a message. In: KACMÁROVÁ, A. (ed.). *English matters: a collection of papers by the Department of English language and Literature faculty*. Presov: University of Presov, 2010. p. 30-36.
- SANKOFF, D.; TAGLIAMONTE, S. S.; SMITH, E. Goldvarb X: a variable rule application for Macintosh and Windows. Toronto: University of Toronto, 2005.
- SANTOS, I. J. A. *Mecanismos de conexão frásica: a importância das variáveis sociais*. 2016. Tese (Doutorado em Ciências da Linguagem) – Universidade do Minho, Braga, 2016.
- SCHIFFRIN, D. Multiple constraints on discourse options: a quantitative analysis of causal sequences. *Discourse Processes*, 8, p. 281-303, 1985.
- SILVA, M. J. F. da. *Propriedades sintáticas e discursivas das orações com porque*. 2008. Dissertação (Mestrado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- SWEETSER, E. *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press, 1990.