

Estado da publicação: O preprint foi publicado em um periódico como um artigo
DOI do artigo publicado: <https://doi.org/10.1590/1678-460x202238248508>

A redução da preposição para em relação à palavra precedente

Ana Paula Huback

<https://doi.org/10.1590/1678-460x202148508>

Submetido em: 2021-12-16

Postado em: 2021-12-16 (versão 1)
(AAAA-MM-DD)



<https://doi.org/10.1590/1678-460x202148508>

A redução da preposição *para* em relação à palavra precedente

The reduction of the preposition *para* (*to, for*) in Brazilian Portuguese according to the previous word

Ana Paula Huback¹

Resumo

Neste artigo, analisa-se a redução da preposição *para* a *pra* no português brasileiro em sequências de (palavra anterior + *para* ou *pra*). O Corpus LAEL de língua falada foi adotado para fazer buscas por essas sequências. Considerando-se as 30 sequências mais frequentes de (palavra anterior + *para* ou *pra*), análises estatísticas foram realizadas para observar que fatores interferiam nessa redução. Os conceitos de frequência da sequência (Krug 1998, 2003, Bush 2001), frequência relativa (Torres Cacoullos 2006) e frequência de ocorrência (Bybee 2002a, 2002b, 2006) foram utilizados para verificar se a repetição das palavras em alguma dessas formas desempenhava algum papel nas sequências em que *para* se reduz a *pra*. Observou-se que a frequência relativa favorece a redução de *para* a *pra*, juntamente com classe de palavra, número de sílabas e tonicidade. A partir desse fenômeno, discute-se a possível estocagem de sequências de duas ou mais palavras no léxico mental dos falantes, bem como as consequências disso para processos de variação e mudança linguística.

Palavras-chave: sequências de palavras, estocagem lexical, efeitos de frequência, redução fonética

Columbia University. Nova York – Estados Unidos. <https://orcid.org/0000-0002-4467-2031>. E-mail: aphuback@gmail.com.

Abstract

This article analyzes the reduction of the preposition *para* (*to, for*) to *pra* in Brazilian Portuguese. Sequences of (previous word + *para* or *pra*) selected from a corpus of spoken language were statistically analyzed to understand which factors could be playing a role in this reduction. A statistical analysis of the 30 most frequent sequences of (previous word + *para* or *pra*) was carried out to identify the factors that seem to influence this reduction. The concepts of string frequency (Krug 1998, 2003, Bush 2001), relative frequency (Torres Cacoullos 2006), and token frequency (Bybee 2002a, 2002b, 2006) were adopted to verify whether frequency of usage in any of these forms would influence the reduction. Results show that relative frequency favors the reduction of *para* to *pra*, alongside word class, number of syllables, and stressed syllable. Based on this analysis, the lexical storage of multiword sequences is discussed, along with its effects for processes of linguistic variation and change.

Keywords: chunking, lexical storage, multiword sequences, phonetic reduction, frequency effects

1. Introdução

Nas línguas do mundo, um dos fenômenos bastante recorrentes é a fusão de palavras que são frequentemente usadas juntas na mesma sequência. Quando as palavras têm poucos contextos alternantes, elas podem ser analisadas lexicalmente como uma só unidade e, a partir disso, são suscetíveis a fenômenos que vão desde a redução de segmentos fonéticos até mudanças na própria ortografia. Exemplos desses casos no inglês são: *God be with you* > *goodbye*, *may be* > *maybe*. Em espanhol, podemos citar *en hora buena* > *enhorabuena*. No português brasileiro (doravante PB), temos casos como *em boa hora* > *embora*, *a pesar de* > *apesar de*, *vossa mercê* > *você*, dentre vários outros. Em todos esses exemplos, houve fusão de duas ou mais palavras e, posteriormente, mudança ortográfica. Alguns desses itens apresentam, também, apagamento de segmentos, como *God be with you* e *vossa mercê*, para citar apenas dois. Existem, igualmente, casos em que sequências muito frequentes na fala são pronunciadas como uma só unidade, ainda

que sua ortografia não tenha mudado: *pode deixar* > *podexá*, *deixa eu ver* > *xovê*, *vamos embora* > *vãbora*. Todos esses processos típicos da língua não-monitorada, seja ela falada ou escrita, demonstram que, quando duas ou mais palavras são frequentemente usadas juntas, fenômenos de fusão e redução de segmentos podem ocorrer.

Este artigo analisa a redução da preposição *para* a *pra* em relação à palavra que a antecede. No PB, sobretudo na língua falada, *para* é frequentemente pronunciado como *pra*. Além dessa variação, outras formas de combinação de preposição com o artigo seguinte também são encontradas: *pro* (*para* + *o*), *pros* (*para* + *os*), *pra* (*para* + *a*), *pras* (*para* + *as*), *prum* (*para* + *um*), *pruns* (*para* + *uns*), *pruma* (*para* + *uma*), *prumas* (*para* + *umas*). Embora as gramáticas prescritivas e dicionários ainda não tenham incorporado essas formas, elas são correntes na língua falada. Ocorrências de *pra* + artigo feminino singular *a* também são frequentes, mas, devido à fusão desses dois itens na forma *pra*, não há como identificar ocorrências da preposição por si mesma ou em contração com o artigo *a*. Análises feitas com o Corpus LAEL (Banco do Português – Linguística Aplicada a Estudos da Linguagem) mostram que os artigos definidos são frequentemente usados depois da preposição *para*. Isso, provavelmente responde por esse alto grau de fusão entre a preposição e o artigo subsequente.

O Corpus do Português (<https://www.corpusdoportugues.org/x.asp>), que contém mais de dois bilhões de palavras de fontes diacrônica e sincrônica, mostra que *para* começou a ser usado como *pra* no século XV. Por outro lado, a contração com palavras subsequentes a *para* começou no século XIX, com o artigo definido feminino *a*, gerando a forma *pra*. Outras estruturas com artigo definido (*pro*, *pros*, *pras*) começam a ser registradas no século XX, bem como contrações com o artigo indefinido singular (*prum*, *pruma*), mas estas em menor ocorrência que aquelas. Os artigos definidos ou indefinidos são usados em posição seguinte à preposição *para*. O presente estudo, no entanto, se propõe a analisar o papel da palavra anterior na redução de *para*. Nosso objetivo é verificar se a frequência de co-ocorrência da palavra anterior + *para* ou *pra* exerce alguma influência nessa redução. A partir disso, discutimos as implicações dessa análise para a forma como os seres humanos estocam e acessam itens lexicais. Diversas análises anteriores para

outros fenômenos linguísticos (Phillips, 2001; Bybee, 2004, 2006; Bybee & Scheibman, 1999) já apontaram para a possibilidade de fusão de material linguístico adjacente e o papel da repetição de palavras em fenômenos de apagamento e redução de segmentos. Na subseção seguinte, vamos fazer uma análise mais detalhada das pesquisas linguísticas que norteiam este trabalho.

Frequência como determinante de variação linguística

A discussão sobre como nossa mente armazena e acessa informações de caráter linguístico está há tempos presente na literatura linguística. De acordo com modelos de uso da língua (Bybee, 1985, 2001, 2010), nosso cérebro lida com informações linguísticas a partir dos mesmos critérios com que gerencia dados de outros domínios cognitivos. Sendo assim, os mesmos parâmetros usados para categorizar, estocar e recuperar conhecimentos gerais são adotados para gerir informações linguísticas. A afirmação de que operações mentais linguísticas e não linguísticas são análogas pressupõe que os mesmos princípios que governam atividades do nosso dia a dia também determinam a estocagem linguística e podem desencadear fenômenos de variação e mudança linguística. Em relação a atividades cotidianas, a frequência com que indivíduos repetem certas rotinas é provavelmente um determinante para que essas se tornem automáticas e, conseqüentemente, requeiram menos esforço físico ou mental. O mesmo ocorre com rotinas linguísticas. De acordo com modelos baseados no uso, a repetição é crucial para explicar como certas estruturas se modificam ao longo do tempo. Em termos de palavras individuais, é conhecida na literatura linguística a hipótese de que a frequência de ocorrência de itens individuais pode ser responsável por processos de redução fonética (Schuchardt, 1885; Leslau, 1969; Krishnamurti, 1978; Fidelholtz, 1975; Shen, 1990). Os efeitos de frequência não são, no entanto, categóricos no sentido de afetarem somente as palavras muito usadas: Phillips (1984), por exemplo, mostra que variações que interferem na fisiologia da fala e causam apagamento de segmentos geralmente afetam os itens mais frequentes primeiro, ao passo que mudanças analógicas afetam as palavras menos frequentes primeiro. Com base nessa afirmação, podemos

entender que palavras mais frequentes, por terem uma forte representação no léxico mental dos falantes, geralmente são mais resistentes a processos de analogia.

Em relação a sequências de palavras, a literatura linguística tem mostrado que a frequência de coocorrência, ou seja, o uso de duas ou mais palavras contíguas, pode afetar o apagamento de segmentos. Gahl e Garnsey (2004) analisam o apagamento de /t, d/ no inglês norte-americano e afirmam que “language processing appears to be sensitive to syntactic probabilities” e “probabilities are an inherent part of grammar, and of syntax in particular” (Gahl & Garnsey, 2004, p. 748). Segundo os autores, a redução de segmentos é mais recorrente em sequências de palavras em que o segundo item tem probabilidade alta de ocorrer em comparação ao primeiro. Esses linguistas também enfatizam que as palavras mais frequentes tendem a ser reduzidas, mas somente se isso não comprometer o significado da expressão como um todo. Por outro lado, com relação às palavras menos frequentes, é comum que os falantes as alonguem para que sejam propriamente entendidas, independentemente de sua baixa probabilidade de ocorrência em um certo contexto. Todas essas afirmações sugerem que “probabilities are capable of affecting pronunciation” (Gahl & Garnsey, 2004, p. 751). Os autores apresentam evidências linguísticas significativas para demonstrar a hipótese de que palavras são mais apagadas em contextos sintáticos em que sejam altamente previsíveis. De acordo com o mesmo princípio, contextos de baixa probabilidade sintática são mais propensos a palavras com pronúncia plena, sem apagamentos de segmentos. Apesar de argumentos controversos (Newmeyer, 2006), Gahl e Garnsey (2006) afirmam que “in order to account for the form of language, the grammar needs to include probabilities.” (Gahl & Garnsey, 2006, p. 405).

Algumas pesquisas linguísticas reúnem evidências de que informações probabilísticas são estocadas juntamente com material linguístico e podem, portanto, desempenhar um papel na variação. Palavras que ocorrem em posição contígua com frequência começam a ser lexicalmente analisadas como uma só unidade, o que pode causar redução e apagamento de segmentos. Browman e Goldstein (1992) e Pagliuca e Mowrey (1987) apresentam evidências de que os gestos articulatórios são reduzidos e se sobrepõem quando uma

palavra é repetida muitas vezes. O mesmo pode ser dito em relação a sequências de palavras. Quando itens lexicais têm uma probabilidade alta dentro de um contexto sintático, podem ser mais suscetíveis a sofrer redução fonética. Nesta análise da redução da preposição *para* em relação à palavra que a antecede, três medidas de frequência foram consideradas: frequência de ocorrência, frequência relativa e frequência da sequência. Abaixo fornecemos explicações para cada uma delas:

- A *frequência de ocorrência* é a contagem de quantas vezes uma palavra (ou expressão) ocorre em um corpus de língua (Bybee 1985, 2001, 2010). Para nossa pesquisa, foram contadas as ocorrências de cada uma das palavras pertencentes a sequências de (palavra precedente + *para* ou *pra*);
- A *frequência relativa* é “the frequency of a collocation with respect to occurrences of the lexical item outside the collocation. Relative frequency is important because it promotes the autonomy of the new fused unit from its erstwhile lexical component.” (Torres Cacoullous, 2006, p. 37). Nesta análise, a frequência relativa será definida como a soma de (palavra precedente + *pra*) dividida pelo somatório de (palavra precedente + *para* ou *pra*).
- A *frequência da sequência* pode ser calculada a partir da divisão de (palavra precedente + *para* ou *pra*) pelo número total de palavras do corpus (Krug, 1998, 2003; Bush, 2001).

Neste artigo, a redução de *para* a *pra* vai ser analisada a partir da perspectiva de modelos linguísticos baseados no uso, considerando o papel da frequência e da palavra precedente a *para* ou *pra*.

2. A presente análise

Esta análise investiga a redução da preposição *para* em relação à palavra que a precede. Nesse contexto, analisa-se essa preposição como uma classe anafórica, ou seja, que

retoma a palavra anterior. Conforme mencionado anteriormente, os dados analisados por esta pesquisa vêm do Corpus LAEL, que contém 7.021.948 palavras¹ extraídas de amostras de língua falada compostas por conversas ao telefone, entrevistas, debates políticos, reuniões de negócios e narrações de jogos de futebol. As 30 sequências mais frequentes de (palavra precedente + *para* ou *pra*) foram selecionadas para serem investigadas neste artigo. A Tabela 1 abaixo mostra a lista completa dessas sequências, em ordem descendente, começando pela sequência mais frequente como um todo (forma não-reduzida e reduzida contadas em conjunto).

Tabela 1 – As 30 sequências mais frequentes de (palavra precedente + *para* ou *pra*)

Sequência	Frequência da sequência	Redução de <i>para</i> a <i>pra</i>	
		Sim	Não
1. foi para	1.170	28% (333/1.170)	72% (837/1.170)
2. lá para	1.097	58% (632/1.097)	42% (465/1.097)
3. ir para	1.001	40% (397/1.001)	60% (604/1.001)
4. fui para	917	24% (217/917)	76% (700/917)
5. e para	902	47% (419/902)	53% (483/902)
6. dá para	856	79% (675/856)	21% (181/856)
7. vai para	743	51% (376/743)	49% (367/743)
8. ia para	722	41% (296/722)	59% (426/722)
9. veio para	687	43% (296/687)	57% (391/687)
10. só para	636	56% (359/636)	64% (277/636)
11. é para	622	65% (404/622)	35% (218/622)
12. né para	615	90% (552/615)	10% (63/615)
13. era para	565	48% (271/565)	52% (294/565)
14. aqui para	539	54% (293/539)	46% (246/539)
15. assim para	523	81% (425/523)	19% (98/523)

¹ Parte desse corpus está disponível para o público no website <http://www2.lael.pucsp.br/corpora/bp/>. Para este artigo, no entanto, tivemos acesso ao corpus inteiro graças à ajuda de Tony Berber Sardinha, a quem somos extremamente gratos.

16. mais para	499	51% (254/499)	49% (245/499)
17. que para	417	46% (193/417)	54% (224/417)
18. disse para	399	63% (251/399)	37% (148/399)
19. dinheiro para	388	38% (148/388)	62% (240/388)
20. coisa para	385	60% (228/385)	40% (157/385)
21. vim para	355	44% (158/355)	56% (197/355)
22. então para	324	55% (178/324)	45% (146/324)
23. dava para	312	53% (165/312)	47% (147/312)
24. isso para	309	35% (109/309)	65% (200/309)
25. tudo para	308	57% (176/308)	43% (132/308)
26. mas para	284	40% (113/284)	60% (171/284)
27. bom para	266	60% (159/266)	40% (107/266)
28. fomos para	262	38% (99/262)	62% (163/262)
29. foram para	240	34% (83/240)	66% (157/240)
30. também para	231	38% (89/231)	62% (142/231)
Total		50,4% (8.348/16.574)	49,6% (8.226/16.574)

Fonte: Elaborada pela autora.

Selecionamos as 30 sequências mais frequentes porque, na lista de sequências com frequência mais baixa, havia centenas delas com uma ou duas ocorrências, então seria impossível fazer uma seleção a partir desses itens de baixíssima frequência. Essas 30 sequências foram submetidas à Regressão Logística do programa SPSS para Mac. As variáveis escolhidas como possivelmente relevantes para as palavras que antecediam *para* ou *pra* foram:

- *Classe de palavra*: adjetivo, advérbio, conjunção, substantivo, pronome e verbo. Não adicionamos outras classes (artigo, numeral etc.) porque essas não apareceram na lista que apresentamos na Tabela 1;

- *Tonicidade*: palavras oxítonas ou paroxítonas. Não houve palavras proparoxítonas na lista das 30 sequências mais frequentes de palavra anterior + *para* ou *pra*;
- *Número de sílabas*: monossílabas, dissílabas ou trissílabas. Nas 30 sequências mais frequentes com a preposição *para* (ou *pra*) não houve palavras com mais de três sílabas;
- *Segmento precedente*: o segmento final da palavra anterior à preposição poderia ser consoante ou vogal;
- *Frequência de ocorrência*: palavras de baixa frequência ou de alta frequência;
- *Frequência relativa*: frequência baixa ou alta;
- *Frequência da sequência*: frequência alta ou baixa.

3. Análise de dados

Usando as 30 sequências mais frequentes, nossa busca no Corpus LAEL selecionou 16.574 ocorrências de (palavra precedente + *para* ou *pra*). Os resultados gerais com uso de *para* ou sua redução a *pra* se encontram na Tabela 2.

Tabela 2 – Resultados gerais

Variável dependente	%/Ocorrências
PRA	(50,4%) 8.348/16.574
PARA	(49,6%) 8.226/16.574

Fonte: Elaborada pela autora

Os resultados gerais mostram que a diferença entre o uso de *para* ou *pra* nessas sequências é mínima (50,4% versus 49,6%). O uso da forma reduzida *pra* na língua

falada (base do Corpus LAEL utilizado aqui) já é tão disseminado que apresentada praticamente a mesma porcentagem de uso que as ocorrências de *para*. Na língua escrita monitorada, a forma preferida provavelmente ainda é *para*, de modo que *pra* é restrito a contextos informais, tais como mensagens de texto e posts em redes sociais.

Quando submetemos esses dados à Regressão Logística do SPSS, os fatores considerados relevantes foram *frequência relativa*, *classe de palavras*, *tonicidade* e *número de sílabas*. Segmento precedente, frequência da sequência e frequência de ocorrência foram as variáveis descartadas. A Tabela 3 mostra os valores para cada um desses fatores.

Tabela 3 – Variáveis na equação

Variáveis	Wald	df	Sig.	Exp(B)
Frequência relativa	846.020	1	.000	.354
Classe de palavra	12.373	1	.000	.969
Tonicidade	6.270	1	.012	1.196
Número de sílabas	4.387	1	.036	.872
Segmento anterior	3.648	1	.056	1.111
Frequência da sequência	.440	1	.507	.971
Frequência de ocorrência	.069	1	.793	1.011
Constant	83.141	1	.000	4.097

Total de ocorrências = 16.574

Fonte: Elaborada pela autora

Frequência relativa foi a variável que mais determinou a redução de *para*. Classe de palavra, tonicidade e número de sílabas também foram selecionadas, mas seu impacto na variação não é tão significativo quando frequência relativa. A seguir, cada uma dessas variáveis será discutida, a começar por frequência relativa, na Tabela 4.

Tabela 4 – Frequência relativa

Faixas de frequência	<i>pra</i>	<i>para</i>
Frequência relativa baixa	39,7% (3.851/9.692)	60,3% (5.841/9.692)
Frequência relativa alta	65,3% (4.497/6.882)	34,7% (2.385/6.882)

Total de ocorrências = 16.574, $\chi^2 = 1055.900$, $df = 1$, $p = 0.000$

Fonte: Elaborada pela autora

A Tabela 4 mostra que a frequência relativa pode ser considerada o fator mais relevante para a redução de *para* a *pra*. Sequências altamente frequentes apresentaram mais redução fonética (65,3%) do que sequências de baixa frequência relativa (39,7%). É importante discutir por que, dentre as três medidas de frequência adotadas (frequência de ocorrência, da sequência e relativa), somente a última foi considerada relevante. Para que duas ou mais palavras sejam processadas lexicalmente como uma só unidade é importante que elas sejam usadas juntas tão frequentemente que adquiram um certo grau de fusão. Quando uma palavra tem alta frequência de ocorrência, ela geralmente ocorre em contextos alternantes, o que impede que seja categorizada dentro de uma sequência fixa com uma outra palavra. Uma grande quantidade de palavras diferentes precede ou segue uma palavra altamente frequente. Essa alternância de contextos impede que uma palavra de alta frequência de ocorrência e *para* sejam processadas como uma só unidade. O grau de fusão entre uma palavra frequente e *para* é muito baixo e isso, por sua vez, inibe a redução de *para* nesse contexto. A alta frequência de ocorrência geralmente é relevante para mudanças redutivas, mas não em um fenômeno em que sequências de palavras são consideradas.

Com relação à frequência da sequência, essa é uma medida da ocorrência da palavra precedente + *para* ou *pra* juntos em relação ao tamanho do corpus inteiro. Como esse conceito abarca as ocorrências de *para* e *pra* em um só cálculo, não foi relevante para esse caso.

A frequência relativa, que analisa cada uma das variáveis separadamente, mostrou-se a mais relevante para a análise da redução de *para*. Essa medida se refere à frequência da

(palavra precedente + *pra*) (somente na forma reduzida), dividida por (palavra precedente + *para*) (somente na forma não-reduzida). Essa é uma medida, portanto, que se refere basicamente à frequência da palavra precedente e *pra*, sendo assim, por lidar somente com a versão reduzida da preposição, não é uma surpresa que esse tenha sido o fator mais relevante para a redução dessa preposição. De um modo geral, o que esses resultados de frequência demonstram é que duas ou mais palavras podem ser lexicalmente processadas como uma única unidade, desde que a sequência (não as palavras individuais que a compõem) tenha uma alta frequência como um todo. Como acontece também com outros fenômenos linguísticos, a formação de sequências de palavras ocorre com itens cuja co-ocorrência seja alta, com o objetivo de diminuir o esforço dos gestos articulatórios (Pagliuca; Mowrey, 1987). A frequência relativa pode, portanto, ser entendida como um reflexo da sequenciação de palavras desencadeada por poucos contextos alternativos. A seguir vamos analisar os resultados para classe de palavra, apresentados na Tabela 5.

Tabela 5 – Classe de palavras

Classes de palavras	<i>pra</i>	<i>para</i>
Adjetivos	59,8% (159/266)	40,2% (107/266)
Advérbios	62,9% (2.604/4.140)	37,1% (1.536/4.410)
Conjunções	46,9% (903/1.927)	53,1% (1.024/1.927)
Pronomes	46,2% (285/617)	53,8% (332/617)
Substantivos	48,6% (376/773)	51,4% (397/773)
Verbos	45,4% (4.021/8.851)	54,6% (4.830/8.851)

Total de ocorrências = 16.574, $\chi^2 = 370.493$, $df = 5$, $p = 0.000$

Fonte: Elaborada pela autora

A Tabela 5 acima mostra que os advérbios são a classe de palavras mais propensa à redução de *para* a *pra* (62,9%). Adjetivos vêm a seguir com uma percentagem bastante próxima (59,8%), mas é importante esclarecer que havia apenas um adjetivo (*bom*), então essa alta redução pode ser um efeito da palavra individual, não da classe como um todo. É interessante notar que verbos são a classe de palavra mais frequente antes de *para*, mas

esses são os que causam a menor redução (45,4%). A explicação que propomos para esse resultado é a de que advérbios não se flexionam, mas os verbos do PB têm um sistema de flexão com muitas formas conjugadas. O verbo *ir*, por exemplo, foi um dos mais usados nas 30 sequências listadas na Tabela 1. No entanto, em vez de ter uma forma única como um advérbio teria, esse verbo apresenta diversas formas conjugadas, tais como *foi, fui, vai, ia, fomos, foram* etc. Essa grande variabilidade de formas alternantes do mesmo verbo impede que haja fusão com *para*. A fusão pode ocorrer com formas conjugadas individuais (como *vai*, um dos líderes do uso de *pra*), mas o efeito de frequência na classe como um todo não aparece por causa da alternância de formas. Essa pode ser uma justificativa para o fato de que verbos desfavorecem a redução de *para*.

Outro ponto que vale a pena mencionar em relação ao fato de que advérbios favorecem a redução de *para*, enquanto verbos a desfavorecem, é a constatação de que, de acordo com o Novo Dicionário Houaiss da Língua Portuguesa (2001), existe, na nossa língua, uma quantidade maior de verbos que de advérbios, então, por causa disso, haverá muito mais variação em relação a qual verbo ocupa a posição anterior a *para*. Os advérbios, por serem uma classe com menos possibilidades de palavras, estarão sempre com contextos mais fixos antes de *para*, o que pode favorecer a fusão com *para*. Essa diferença em termos de frequência de tipo das duas classes de palavras fará com que haja muita competição entre verbos antes de *para* e menos competição entre os advérbios na mesma posição.

Além disso, é importante investigar, também, as possibilidades sintáticas de classes de palavras que podem vir depois de verbos e advérbios, grupos que causaram, respectivamente, menor e maior redução de *para* a *pra*. A Tabela 6 mostra que classes de palavras podem preencher os espaços sintáticos depois de verbos ou advérbios:

Tabela 6 – Classes de palavras que podem vir depois de verbos ou advérbios

Classes de palavras que podem vir depois de <u>verbos</u>	Classes de palavras que podem vir depois de <u>advérbios</u>
advérbios	advérbios

adjetivos	adjetivos
artigos	conjunções
conjunções	numerais
numerais	preposições
preposições	verbos
pronomes	
substantivos	
verbos	

Fonte: Elaborada pela autora

Os verbos podem ser seguidos por quase qualquer classe de palavras, o que aumenta a possibilidade de competidores de outras classes que possam ocupar o mesmo espaço que *para*, como preposição, preenche. Os advérbios, por outro lado, têm uma gama menor de classes de palavras possíveis na posição subsequente, o que diminui a competição com itens que possam ocupar o mesmo espaço que *para*. Essas possibilidades de combinações sintáticas nos mostram que o grau de fusão de um advérbio com *para* pode ser maior que o de um verbo com essa mesma palavra. A menor alternância de contextos sintáticos faz com que as conexões entre um advérbio e *para* sejam mais coesas, o que faz com que as duas palavras envolvidas comecem a ser analisadas como uma unidade lexical que pode, potencialmente, sofrer redução fonética. Como foi mencionado anteriormente, a lista de advérbios do PB é muito menor que a de verbos. Então, juntando-se essa fusão sintática das duas categorias com o fato de que dentro da classe de advérbios em si há menos itens que na de verbos para preencher essa lacuna, pode-se entender melhor por que os advérbios são os grandes favorecedores da redução de *para* a *pra* e por que os verbos estão no extremo oposto desse *continuum*. A Tabela 7 apresenta os resultados para tonicidade.

Tabela 7 – Tonicidade

Tonicidade	<i>pra</i>	<i>para</i>
Oxítone	51,8% (5.933/11.458)	48,2% (5.525/11.458)

Paroxítona	47,2% (2.415/5.116)	52,8% (2.701/5.116)
-------------------	---------------------	---------------------

Total de ocorrências = 16.574, $x^2 = 29.620$, $df = 1$, $p = 0.000$

Fonte: Elaborada pela autora

A Tabela 7 mostra que palavras oxítonas interferem mais na redução de *para* (51,8% versus 47,2% para palavras paroxítonas). Conforme mencionamos anteriormente, não havia palavras proparoxítonas na lista das 30 seqüências mais frequentes de palavra precedente + *para* ou *pra*. O resultado para as palavras oxítonas e paroxítonas talvez possa ser explicado em termos de padrões suprasegmentais de pronúncia, tonicidade e prosódia do PB, mas, como não temos acesso à gravação do Corpus LAEL, não seria seguro fazer afirmações dessa natureza. Nossa explicação sugere uma análise do fator tonicidade em conjunto com classe de palavra. Como foi mostrado na Tabela 5, advérbios apresentam maior taxa de redução de *para*. Havia um total de oito palavras nessa classe, dentre as quais quatro eram oxítonas e as demais eram monossílabas, que foram consideradas também como oxítonas. Parece ser, portanto, que existe uma sobreposição entre esses dois fatores e talvez a classe de palavra tenha uma significância maior que a tonicidade. A seguir discutimos os resultados para número de sílabas.

Tabela 8 – Número de sílabas

Número de sílabas	<i>pra</i>	<i>para</i>
Monossílabos	50,5% (5.241/10.380)	49,5% (5.139/10.380)
Dissílabos	51% (2.959/5.806)	49% (2.847/5.806)
Trissílabos	38,1% (148/388)	61,9% (240/388)

Total de ocorrências = 16.574, $x^2 = 24.081$, $df = 2$, $p = 0.000$

Fonte: Elaborada pela autora

Com relação ao número de sílabas, *para* é reduzido quase igualmente em palavras com uma ou duas sílabas. Palavras com três sílabas, por outro lado, inibem a redução. É importante esclarecer, no entanto, que havia apenas uma palavra com três sílabas (*dinheiro*), por isso não se pode afirmar que a categoria como um todo desfavoreça a

redução de *para*. Por outro lado, na literatura linguística, há evidência de que a sequenciação e a fusão de palavras ocorrem com mais frequência com palavras curtas do que com palavras longas. Conforme Zipf (1935) aponta, palavras curtas são mais frequentes na língua, veiculam menos informação semântica e, portanto, são mais propensas a reduções sem comprometer o sentido dos enunciados. Palavras mais longas, por outro lado, são menos frequentes e sofrem menos apagamento de segmentos. Parece, portanto, que palavras mais curtas (com uma ou duas sílabas) são mais propensas à fusão com *para*, o que, por sua vez, pode causar a redução a *pra*.

4. Conclusões

Este artigo analisa a redução da preposição *para*, frequentemente pronunciada como *pra*, em relação à palavra que a precede. Três conceitos diferentes de frequência (de ocorrência, de sequência e relativa) são adotados para analisar se a redução dessa preposição pode ser explicada por efeitos de frequência. A análise probabilística realizada mostra que, na fala, *pra* já é tão usado quanto *para*, o que demonstra seu alto grau de expansão na língua. Com relação aos fatores que interferem nessa redução, a regressão logística realizada pelo programa SPSS mostrou que a frequência relativa é a variável mais determinante, seguida por classe de palavra. Quanto ao primeiro fator, observamos que a alta frequência relativa favorece a redução de *para*, o que nos fornece evidência para a estocagem de sequências de mais de uma palavra no léxico mental nos falantes. Também se observou que os advérbios são a classe de palavras que mais favorece a redução de *para*, mas há uma sobreposição entre número de sílabas, tonicidade e classe de palavras e a interação entre esses três fatores talvez seja a melhor forma de justificar essa redução. Finalmente, este artigo contribui para a discussão sobre efeitos de frequência e estocagem lexical com riqueza de detalhes fonéticos e probabilísticos.

Conflito de interesses

(X) *A autora declara não ter qualquer conflito de interesse, em potencial, neste estudo e assume responsabilidade total pelo conteúdo do artigo.*

Referências

Browman, C., & Goldstein., L. (1992). Articulatory phonology: an overview. *Phonetica*, 49(3-4), 155-80. <https://doi.org/10.1159/000261913>

Bush, N. (2001). Frequency effects and word-boundary palatalization. In J. Bybee, & P. Hopper (Eds.). *Frequency and the emergence of linguistic structure* (pp. 255-80). John Benjamins Publishing Company.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins Publishing Company.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.

Bybee, J. (2002a) Mechanisms of change in grammaticization: the role of frequency. In R. Janda, & J. Brian (Eds). *Handbook of historical linguistics* (pp. 602-623). Blackwell.

Bybee, J. (2002b). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, 24(2), 215-221.

<https://doi.org/10.1017/S0272263102002061>

Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82(4), 529-551. <http://www.jstor.org/stable/4490266>

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, 37, 575-96.

<https://doi.org/10.1515/ling.37.4.575>

Fidelholtz, J.L. (1975). Word frequency and vowel reduction in English. In *Regional meeting Chicago Linguistics Society*, 7 (pp. 200-213.). Chicago Linguistic Society.

Corpus Banco de Português (Linguística Aplicada a Estudos da Linguagem – LAEL). <http://www2.lael.pucsp.br/corpora/bp/conc/>. (acessado 11 de dezembro, 2021).

Corpus do Português (<https://www.corpusdoportugues.org/x.asp>). (acessado 11 de dezembro, 2021).

Gahl, S., & Garnsey, S.M. (2004). Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language*, 80(4), 748-775. <http://linguistics.berkeley.edu/~gahl/GahlGarnsey2004.pdf> (acessado 15 de dezembro, 2021).

Houaiss, A. (2001). *Novo dicionário Houaiss da língua portuguesa*. Objetiva.

Krishnamurti, B. (1978). Areal and lexical diffusion of sound change: evidence from Dravidian. *Language*, 54(1), 1-20. <https://doi.org/10.2307/412996>

Krug, M. (1998). String frequency: a cognitive motivation factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics*, 26(4), 286-320. <https://doi.org/10.1177/007542429802600402>

Krug, M. (2003). Frequency as a determinant in grammatical variation and change. In: G. Rohdenburg, & B. Mondorf (Eds.), *Determinants of Grammatical Variation* (pp. 7-67). Berlin.

Leslau, W. (1969). Frequency and change in the Ethiopian languages. *Word*, 25, 180-189. <https://doi.org/10.1080/00437956.1969.11435567>

Newmeyer, P. (2006). On Gahl and Garnsey on grammar and usage. *Language*, 82(2), 399-404. <https://www.jstor.org/stable/4490162> (acessado 15 de dezembro, 2021)

Pagliuca, W., & Mowrey, R. (1987). Articulatory evolution. In A.G. Ramat, O. Carruba, & G. Bernini (Eds.), *Papers from the 7th International Conference on Historical Linguistics* (pp. 459-72). John Benjamins Publishing Company.
<https://doi.org/10.1075/cilt.48.34pag>

Phillips, B. (1984). Word frequency and the actuation of sound change. *Language* 60(2), 320-42. <https://doi.org/10.2307/413643>

Phillips, B. (2001). Lexical diffusion, frequency, analysis. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 123-36). John Benjamins Publishing Company.

Schuchardt, H. (1885). On sound laws: against the Neogrammarians. In H. Schuchardt (Ed.), *Schuchardt, the Neogrammarians, and the transformational theory of phonological change: four essays* (p. 39-72). Athenaeum.

Shen, Z. (1990). Lexical diffusion: a population perspective and a mathematical model. *Journal of Chinese Linguistics*, 18, 159-291.

Torres Cacoullós, Rena. (2006). Relative frequency in the grammaticization of collocations: nominal to concessive *a pesar de*. In T.L. Face, & C.A. Klee (Eds.), *Selected proceedings of the 8th Hispanic Linguistics Symposium* (pp. 37-49), Cascadilla Proceedings Project. <https://www.lingref.com/cpp/hls/8/paper1253.pdf> (acessado 15 de dezembro, 2021).

Zipf, G.K. (1935). *The psycho-biology of language: an introduction to dynamic philology*. MIT Press.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores concordam que caso o manuscrito venha a ser aceito e postado no servidor SciELO Preprints, a retirada do mesmo se dará mediante retratação.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.