

O EFEITO DE VARIÁVEIS SOCIAIS SOBRE O COMPORTAMENTO VARIÁVEL DA PREPOSIÇÃO “PARA” EM SEABRA-BA

THE EFFECT OF SOCIAL VARIABLES ON THE VARIABLE BEHAVIOR OF THE PREPOSITION “PARA” IN SEABRA-BA

Elias de Souza Santos*

Letícia Souza Silva**

RESUMO: Objetivamos com este estudo examinar a influência de preditores sociais sobre o uso da preposição “pra” na variedade do português falado em Seabra (BA). Deprendemos os dados de análise de uma amostra de gravações de 16 informantes, pertencentes ao banco de dados do Projeto *Se abra à Chapada: coletando, explorando e mapeando dados sociolinguísticos* (CEMEDADOS), cuja estratificação se deu conforme o sexo/gênero (masculino e feminino), a faixa etária (20-33, 35-48 e 50+anos) e a escolaridade (fundamental, média e universitária) do falante entrevistado. Após termos tabulado os dados, submetemo-los a testes de qui-quadrado e a uma abordagem de modelagem logística binária no R (R CORE TEAM, 2020). A fim de atingirmos o objetivo proposto, calculamos, selecionamos e interpretamos o modelo de regressão logística, testando, por fim, as pressuposições principais dos testes. Constatamos com os resultados que a faixa etária e a escolaridade influenciam significativamente o uso da variante “pra” na comunidade de fala analisada.

PALAVRAS-CHAVE: Redução da preposição “para”. Variação. Preditores sociais.

ABSTRACT: The aim of this study was to examine the influence of social predictors on the use of the preposition “pra” (“to”) in the variety of Portuguese spoken in Seabra (BA). We deduced the analysis data from a sample of recordings of 16 informants, belonging to the database of the Project *Se abra à Chapada: coletando, explorando e mapeando*

* Doutorando em Estudos Linguísticos no Programa de Pós-Graduação em Estudos Linguísticos da Universidade Estadual de Feira de Santana (PPGEL/UEFS). Professor de Linguística/Língua Portuguesa no Departamento de Ciências Humanas e Tecnologias da Universidade do Estado da Bahia (UNEB), Campus XXIII, Seabra. E-mail: eliasantos@uneb.br.

** Graduanda em Letras Língua Portuguesa e Literaturas no Departamento de Ciências Humanas e Tecnologias da Universidade do Estado da Bahia (UNEB), Campus XXIII, Seabra. E-mail: leticiiaaaa.souza@gmail.com.

dados sociolinguísticos (‘Open yourself to Chapada: collecting, exploring and mapping sociolinguistic data’, CEMEDADOS), whose stratification took place according to sex/gender (male and female), age (20-33, 35-48 and 50+ years) and education (elementary, medium and university) of the interviewed speaker. After having tabulated the data, we submitted them to chi-square tests and a binary logistic modeling approach in R (R CORE TEAM, 2020). In order to reach the proposed objective, we calculated, selected and interpreted the logistic regression model, finally testing the main assumptions of the tests. We found with the results that age and education significantly influence the use of the “pra” variant in the analyzed speech community.

KEYWORDS: Reduction of the preposition “para” (‘to’). Variation. Social predictors.

“PARA” OU “PRA” INÍCIO DE CONVERSA?

A preposição “para”, nas variedades faladas do Português Brasileiro (daqui por diante PB), apresenta três formas em variação (FELGUEIRAS, 1993; MAYA, 2004; LUCENA, 2006; SILVA, 2010; FERREIRA, 2014, 2018), a exemplo de (1) Eu vou PARA Seabra; (2) Eu vou PRA Seabra; e (3) Eu vou PA Seabra. Conforme resultados dos trabalhos dos pesquisadores antes assinalados, verificamos que o uso das formas reduzidas “pra” e “pa”, na ordem em que se apresentam, são as preferidas dos brasileiros, que pouco fazem uso da variante padrão “para”.

O estudo de Felgueiras (1993 apud Maya, 2004), por exemplo, apontou que a variante prototípica usada na variedade carioca é o “pra” (73%), seguida de “pa” (18%) e “para” (9%), apresentando um padrão semelhante de uso nas demais variedades do PB investigadas, como veremos na seção da revisão teórica deste estudo. Esses dados sugerem que o uso variável da preposição “para” não diferencia variedades do português falado em nosso país. Contudo, é preciso que estudos sejam empreendidos em diferentes comunidades de fala no Brasil para que essa sugestão seja validada.

Diante das premissas apresentadas, podemos assumir a hipótese de que os falantes de língua portuguesa, na comunidade de Seabra, apresentam um comportamento linguístico bastante parecido com o de outras variedades brasileiras quanto ao uso variável da preposição “para”. Se este for o fato, inquirimos, então: quais são os fatores sociais que mais favorecem o emprego da variante “pra” em relação a “para”? Com o interesse em respondermos tal questionamento, objetivamos, neste estudo, examinar (i) se existem padrões de similitudes entre a variedade do PB por nós analisada e outras já investigadas; e (ii) se existem correlações entre as variáveis previsoras sociais quanto à expressão da preposição “para” na comunidade de Seabra.

Em razão disso, capturamos as construções preposicionais em uma amostra de dados sociolinguísticos do tipo DID (Diálogo entre informante e documentador), pertencente ao banco de dados do Projeto CEMEDADOS, sediado no Departamento de Ciências Humanas e Tecnologias da Universidade do Estado da Bahia, Campus XXIII, Seabra, coordenado pelo

primeiro autor deste texto. Os dados obtidos foram submetidos a uma abordagem de modelagem logística binária no R (R CORE TEAM, 2020), com a inclusão de variáveis previsoras sociais (sexo, faixa etária e escolaridade).

Sem pormenorizar, os resultados indicaram haver um padrão no uso das variantes “para” e “pra” em contraste com os resultados de pesquisas aqui revisados. Além disso, apontam que o aumento da proporção de uso da forma prototípica “pra” em Seabra ocorre mais entre: os falantes menos escolarizados (nível fundamental) e com faixa etária entre os 35 e 48 de idade (intermediária), em que quanto mais o falante avança no nível de escolaridade, mais ele favorece o uso da variante padrão. No que concerne à variável sexo, ela não se mostrou significativa na correlação com a variável de resposta nem em interação com as demais variáveis previsoras.

As discussões dessas informações são alargadas ao longo deste texto, cuja organização está estruturada da seguinte maneira: na segunda seção, apresentamos uma revisão dos estudos do fenômeno em perspectiva em diferentes variedades do PB e os principais fatores de condicionamento das variantes; na seção 3, examinamos os dados quantitativos do falar seabrense, apontando os fatores sociais relevantes no uso da forma reduzida da preposição “para”; e, na seção 4, traçamos as considerações a que chegamos em relação à variante eleita pelos falantes da comunidade em cena.

REALIZANDO UM PASSEIO POR REGIÕES DO BRASIL: EXPLICITANDO AS VARIANTES

Nesta seção, apresentamos sucintamente resultados de pesquisas sobre o comportamento variável da preposição “para” em *corpora* de amostras de variedades do português falado em diferentes regiões do país, notadamente, os de Felgueiras (1993), Maya (2004), Lucena (2006), Silva (2010) e Ferreira (2014, 2018). O delineamento desses resultados permitirá compreendermos, de maneira mais abrangente, o quadro das variantes (padrão e não padrão) em uso no Brasil.

Felgueiras (1993) empreendeu um estudo sobre a variação no uso da preposição “para”, investigando possíveis condicionamentos na escolha de variantes relacionados ao fenômeno em foco na fala de cariocas, como estilísticos, prosódicos, fonológicos e discursivos. Os dados da pesquisa foram extraídos dos bancos de dados dos projetos Norma Urbana Culta (NURC) e Programa de Estudos sobre o Uso da Língua (PEUL). Daquele primeiro, a autora utilizou 14 entrevistas, estratificadas conforme o sexo (sete homens e sete mulheres), a faixa etária (oito deles com idades entre 26 e 49 anos e seis com mais de 50 anos) e formalidade da fala (seis de elocução formal e oito de diálogo entre informante e documentador). Já deste segundo, fez uso de 12 entrevistas, estratificadas por sexo (seis homens e seis mulheres) e faixa etária (quatro de 15 a 25; quatro de 26 a 49 anos; quatro com mais de 50 anos). A pesquisa obteve 1954

dados no total, sendo 176 (9%) de “para”, 1436 (73%) de “pra” e 342 (18%) de “pa”, não inserindo esta última variante em seu texto, visto que com a mesma não obteve resultados conclusivos. Não contrastaremos os resultados do estudo de Felgueiras (1993) com os nossos, pois não conseguimos ter acesso direto à pesquisa.

Maya (2004) desenvolveu um estudo acerca da variação da preposição “para” na fala dos porto-alegrenses, no Rio Grande do Sul, na perspectiva da Teoria da Variação e Mudança linguísticas. Sua pesquisa contou com dados depreendidos de 24 entrevistas pertencentes ao Projeto Variação Linguística da Região Sul (VARSUL), gravadas na década de 1990. O autor controlou variáveis linguísticas (contexto morfológico precedente e seguinte, contexto fonológico seguinte, tonicidade da sílaba seguinte, número de sílabas do item seguinte, posição em relação a pausas e paralelismo formal) e sociais (sexo, escolaridade e idade). Ele levantou um total de 2034 dados, sendo 62% correspondentes à variante “pra”, 36% a “pa” e 2% a “para”. Pela baixa ocorrência do uso da preposição “para”, dispôs-se a realizar uma análise binária da variante “pa” em relação a “pra”.

A ferramenta estatística usada por Maya (2004), o GOLDVARB, selecionou as variáveis escolaridade, contexto morfológico seguinte, sexo, contexto fonológico seguinte e tonicidade da sílaba seguinte como possuidoras de significância. Quanto à escolaridade, a hipótese levantada pelo autor não se confirma, a de que a variante “pa” seria favorecida pelos indivíduos com baixo nível de escolarização, pois atestou significância para o nível secundário de escolarização (P.R. 0,61). Para o contexto morfológico seguinte, a hipótese é confirmada, dado que a forma “pa”, quando seguida por palavra lexical, favorece o uso dessa variante (P.R. 0,55), em oposição à palavra gramatical. O contexto fonológico seguinte (consoantes dentais, 0,63; palatais, 0,57; e bilabiais, 0,57), a tonicidade da sílaba seguinte (átona, 0,56) e o sexo (homens, 0,57) também favoreceram o uso da variante “pa”. Maya (2004) concluiu que a preposição “para” é pouco utilizada pelos porto-alegrenses, sendo “pra” a variante elegida em detrimento de “pa”, variante restrita à fala e desconhecida das gramáticas prescritivas.

Em *Elementos para o estudo da variação linguística na Paraíba*, Lucena (2006) se propôs a analisar o comportamento variável da preposição “para” no falar pessoense, na Paraíba. Fundamentado no modelo teórico-metodológico da Sociolinguística Quantitativa, o estudo contou com a seleção de 24 entrevistas pertencentes ao banco de dados do Projeto Variação Linguística no Estado de Paraíba (VALPB), estratificadas de acordo com as variáveis sociais sexo, faixa etária e nível de escolarização. Além das variáveis sociais, foram controladas variáveis linguísticas (contexto fonológico seguinte, posição em relação a pausas, paralelismo formal, número de sílabas do item seguinte e presença de vibrante no item seguinte).

Lucena (2006) fez um levantamento de 1445 dados no total, sendo 87 (6%) das aplicações de “para”, 655 (45%) de “pra” e 703 (49%) de “pa”. Devido à baixa ocorrência de “para”, o autor optou por realizar uma primeira rodada binária, juntando as variantes “pra” e “pa”, confrontando-as com “para”. As variantes linguísticas que se mostraram mais relevantes para

o comportamento da variante não padrão (amálgama de “pra” e “pa”) foram o paralelismo formal (“pra” precedida de forma semelhante, 0,68; e “pa” precedida de forma semelhante, 0,84) e o contexto fonológico seguinte (labiodentais, 0,81; alveolares, 0,59; e dentais, 0,59), confirmando a hipótese levantada para o paralelismo de que as variantes tenderiam a ocorrer mais quando precedidas de formas semelhantes na frase, e refutando aquela relacionada ao contexto fonológico seguinte de que as formas “pra” e “pa” seriam favorecidas por vogal central e por consoantes bilabiais, dentais e alveolares.

Quanto às variáveis sociais, os falantes com baixa escolarização favoreceram o uso das variantes “pra” e “pa” (0,86), sendo a hipótese levantada confirmada. Já para a faixa etária, o autor observa que os mais jovens tendem a usar mais a variante padrão “para” (0,63), cuja explicação poderia estar relacionada à imposição da norma padrão feita pela escola. Assim, Lucena (2006) conclui sua pesquisa constatando que as variantes linguísticas “pra” e “pa” são as preferidas no português falado em João Pessoa (PB) e a forma padrão “para” está gradualmente limitada à língua escrita.

Silva (2010) efetuou um estudo sobre a preposição “para” e suas variantes no falar araguanense, em Tocantins, com dados extraídos do Projeto Variação Linguística no Estado do Tocantins (VALTINS), sob a perspectiva da Teoria Variacionista. A análise quantitativa dos dados dessa pesquisa foi realizada com o auxílio da ferramenta computacional denominada GOLDVARB 2001. A autora fez uso de 36 entrevistas, estratificadas por sexo (homens e mulheres), faixa etária (15-25; 26-49; e 49+ anos) e escolarização (baixa, média e alta). As variáveis linguísticas controladas foram contexto fonológico seguinte, paralelismo formal, pausa e presença de vibrante no item seguinte. As hipóteses principais adotadas foram as de que: a utilização da preposição “para” seria a mais usada entre pessoas com escolaridade alta e as variantes “pra” e “pa” seriam mais utilizadas por falantes com média ou baixa escolaridade. Essas hipóteses não foram confirmadas no estudo.

De um total de 3210 ocorrências, 35 (1%) foram da forma padrão “para”, 1852 (54%) de “pra” e 1323 (45%) de “pa”. Dada a baixa ocorrência de “para”, a variante “pra” foi considerada a variável dependente, sendo confrontada com “pa”. Os fatores linguísticos e extralinguísticos que exerceram papel significativo foram: o contexto fonológico seguinte (dorsal, 0,53; vogais anteriores, 0,60; e vogais posteriores, 0,66), a escolaridade (alta, 0,61), a faixa etária (+49 anos, 0,56) e o sexo (mulher, 0,53).

Ferreira (2014) realizou um estudo sobre a preposição “para” e suas variantes no falar de Londrina, no Paraná. A análise realizada foi baseada na Teoria da Variação e Mudança Linguísticas. As amostras coletadas partiram de oito entrevistas do banco de dados do Projeto Variação Linguística no Sul do Brasil (VARSUL), estratificadas por sexo, faixa etária e escolaridade. As hipóteses adotadas foram as de que a utilização da forma padrão seria mínima em relação às variantes “pra” e “pa”, sendo estas relacionadas a um nível de escolaridade mais baixo, e a variante “para”, com acento na penúltima sílaba, quando reduzida para as formas

“pra” ou “pa”, altera seu estatuto prosódico, tornando-se uma sílaba sem acento. A preposição em sua variante reduzida juntar-se-ia a uma palavra que se iniciasse com vogal átona.

Dado o baixo índice de ocorrências da variante “para”, a autora decidiu por realizar uma rodada binária com as variantes “pra” e “pa”. Logo a primeira hipótese se confirma, a de que a forma padrão teria baixo índice de uso em comparação com as demais formas. Já o fator escolaridade não se confirma totalmente e deve ser relativizado em relação ao sexo de cada informante, a terceira hipótese também não se confirma, pois houve um relativo equilíbrio entre os fatores tônica (0,54) e átona (0,45). De mais a mais, outras variáveis linguísticas se mostram significantes, a exemplo do contexto seguinte com pausa (0,72) e com palavras não léxicas (0,57), bem como social, com a idade, em que falantes mais velhos favoreceram o uso de “pra”.

Ferreira (2018) realizou uma investigação acerca da variação da preposição “para” na fala de curitibanos e de florianopolitanos, utilizando-se de dados extraídos do Projeto VARSUL. A amostra de análise foi constituída de 32 entrevistas, estratificadas conforme o sexo, a faixa etária e a escolaridade do informante. Dado o baixo número de ocorrências da variante padrão “para” (49, 1,9%), a autora constituiu uma variável binária, cujas variantes consideradas foram as formas reduzidas “pra” (2086, 80,5%) e “pa” (456, 17,6%). Devido a algumas eliminações por conta dos contextos, os dados de “pra” foram reduzidos a 1913 (81,4%), selecionando na nova rodada o contexto morfológico seguinte (palavra gramatical, 0,65), o contexto fonológico seguinte (vogal coronal, 0,65; e consoante dorsal, 0,56), paralelismo (antecedida de “pra”, 0,64), idade (+ de 50 anos, 0,56), cidade (Curitiba, 0,55), escolaridade (até 11 anos de estudo, 0,53) e número de sílabas do item seguinte (duas ou mais sílabas, 0,53) como significativos.

De acordo com os trabalhos revisitados, no que diz respeito ao comportamento variável da preposição “para”, constatamos que o uso das variantes reduzidas “pra” e “pa” é maior do que o da variante padrão “para”, sendo a forma “pra” a preferida dos falantes, apesar das especificidades de cada amostra estudada em diferentes variedades do português falado no Brasil. Frente ao exposto, em nossa análise, atentaremos para esse fato, de sorte que o ponto seja levado em consideração quando da comparação com os estudos revisitados nesta seção.

ANÁLISE MULTIVARIADA DOS DADOS

Nesta seção, apresentamos os resultados das análises quantitativas dos dados do comportamento variável da preposição “para”, realizadas no R (R CORE TEAM, 2020) a partir de modelos de regressão logística, sem a inclusão de variáveis aleatórias como o informante, por exemplo. Antes, observemos, na Tabela 1, a distribuição das variantes “para” e “pra” na amostra analisada.

Tabela 1 - Distribuição geral de “para” e “pra”

VARIANTES	PARA	PRA	TOTAL
Ocorrências	165	609	774
%	21	79	100

Fonte: Elaboração própria.

Foram depreendidas 165 ocorrências da forma “para” e 609 ocorrências de “pra” na fala dos seabrenses, os dados de “para” representam 21% enquanto os de “pra” 79%, conforme tabela antes exposta. A hipótese que assumimos para a distribuição dos dados é a de que a variante “pra” ocorre na fala dos informantes da amostra sob análise em função de fatores sociais como o sexo, a faixa etária e a escolaridade. Ao notarmos que as variáveis sociais investigadas apresentaram proporções diferentes para o fenômeno variável em estudo, realizamos testes estatísticos para averiguarmos se tais proporções eram significativamente distintas.

Quando queremos verificar correlações entre uma determinada variável nominal com outras variáveis nominais, podemos realizar testes de qui-quadrado, levando em conta que estes se relacionam a uma regressão logística com um preditor binário. Ao aplicarmos tais testes aos dados em análise, observamos que todas as variáveis sociais controladas, com exceção do sexo, indicaram correlação com a variável de resposta, de acordo com os resultados reportados na Tabela 2.

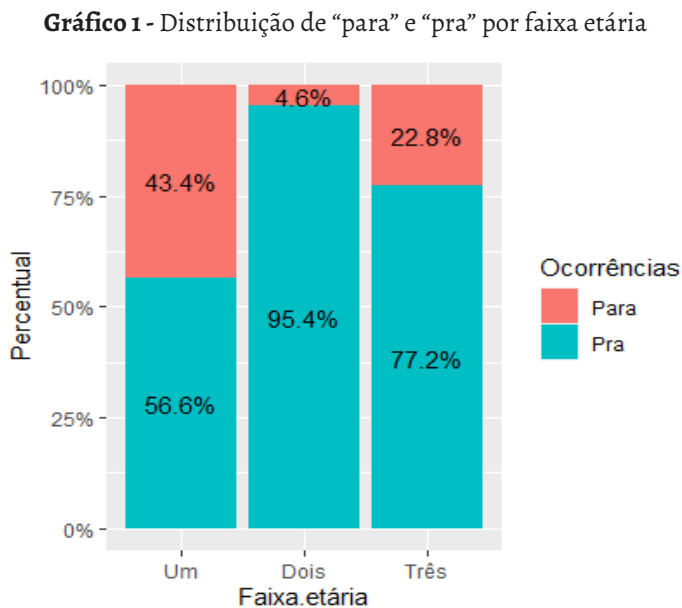
Tabela 2 - Resultado dos testes de qui-quadrado para as variáveis sociais

VARIÁVEIS		PREP. “PARA”		PREP. “PRA”		VALOR-P
		N	%	N	%	
Sexo	Homem	68	09	231	30	0.498
	Mulher	97	12	378	49	
Faixa Etária	20-33 anos	75	10	98	12	2.2e-16
	35-48 anos	12	02	247	32	
	50+ anos	78	10	264	34	
Escolaridade	Fundamental	06	01	180	23	2.367e-12
	Média	100	13	226	29	
	Universitária	59	08	203	26	

Fonte: Elaboração própria.

Os resultados da Tabela 2 indicam que não há correlação entre a variável de resposta e a variável preditora sexo, pois o Valor-p, neste caso, é maior do que 0.05, o que significa que não houve nenhuma diferença significativa, dado que a probabilidade de obtermos um valor da estatística de teste como o observado é muito improvável. No entanto, foi possível coletar

informações relevantes para indicar os padrões de uso das preposições “para” ~ “pra”. Em relação às faixas etárias controladas na amostra em perspectiva, pretendemos responder a seguinte inquirição: há diferença no uso da variante “pra” em relação ao perfil dos falantes que constituem cada faixa etária? Vejamos no Gráfico 1, a seguir, de barras empilhadas, a distribuição proporcional dos dados por cada faixa etária.



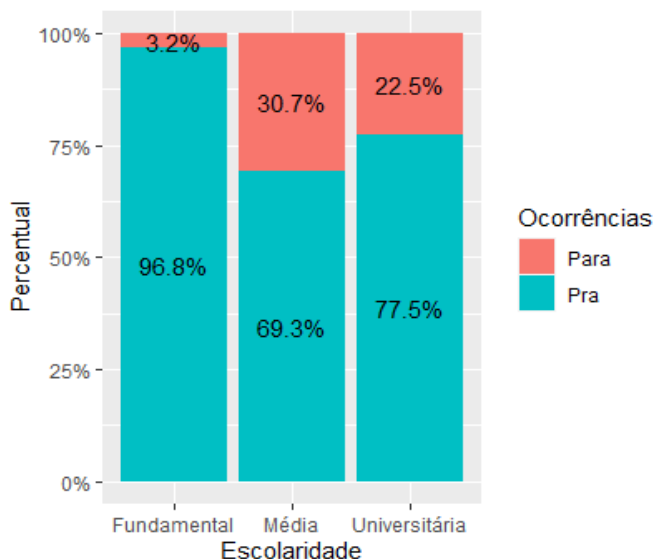
Fonte: Elaboração própria.

Como já dito, a faixa etária é composta por falantes com idades entre 20-33 anos, 35-48 anos e 50 anos ou mais. Para a faixa etária um, foram depreendidos 75 dados da preposição “para” e 98 dados da variante “pra”, apresentando um percentual de 43,4% para a preposição “para”, em oposição a 56,6% da variante “pra”. Já para os falantes da faixa etária dois, as ocorrências de “para” foram 12 (4,6%), e “pra” ocorreu 247 (95,4%) vezes na amostra. Por último, os falantes da faixa etária três apresentaram uma cifra de 78 usos da preposição “para”, referentes a 22,8%, e 264 usos de “pra”, 72,2%. Esse grupo de fatores demonstrou uma diferença estatisticamente significativa, como observado na Tabela 2, em que $\chi^2 = 93,514$, $df = 2$, $p < 0,001$.

Havíamos hipotetizado que os falantes mais jovens favoreceriam o uso da forma reduzida da preposição “para”, indicando um efeito da faixa etária. A distribuição percentual (cf. Gráfico 1) dos usos em função desse preditor não confirma essa hipótese, pois as faixas etárias dois e três foram as que mais exibiram a realização da variante “pra”. Todavia, em todas as variantes que compreendem o preditor, a preposição “pra” apresentou ocorrências, mostrando uma simetria no uso, pelos falantes correspondentes às três variantes, com uma leve diferença proporcional para a faixa etária um.

No que concerne à escolaridade, tínhamos controlado o nível fundamental, médio e universitário, cuja diferença na distribuição é estatisticamente significativa, conforme visualizamos na Tabela 2, em que $\chi^2 = 53,539$, $df = 2$, $p < 0,001$. Observemos, no Gráfico 2, de barras empilhadas, a seguir, a distribuição de “para” e “pra” por nível de escolaridade.

Gráfico 2 - Distribuição de “para” e “pra” por escolaridade



Fonte: Elaboração própria.

O Gráfico 2 expõe a distribuição das ocorrências das preposições “para” e “pra” na amostra em análise. Os falantes com nível de escolaridade fundamental foram os que mais favoreceram o uso da variante “pra”, com uma percentagem de 96,8% dos eventos, ao passo que a preposição “para” alcançou 3,2%; os da escolaridade média foram os que menos favoreceram o uso de “pra”, com 69,3% dos eventos, ao tempo em que atestaram para a preposição “para” um percentual de 30,7% das ocorrências; e, por fim, a escolaridade universitária obteve 77,5% dos eventos de “pra”, em oposição a “para”, que exibiu uma percentagem de 22,5%.

A hipótese que tínhamos aventado foi a de que a preposição “pra” seria favorecida pelos falantes que possuíssem um baixo nível de escolaridade, isto é, fundamental, tendo sido a mesma confirmada, pois o uso da variante “pra” foi mais produtivo com falantes com baixo nível de escolaridade, um realidade distinta da maioria dos trabalhos revisitados neste estudo, cujos resultados parecem indicar que o uso de “pra” não está sendo muito influenciado pela escolaridade, já que as diferenças entre os níveis de escolaridade não são tão distantes.

Na Tabela 3, a seguir, apresentamos os resultados da análise multivariada dos fatores estatisticamente significativos. Vale sublinhar que, antes de termos chegado ao modelo aqui exportado, checamos se não houve violações quanto aos pressupostos básicos para a validação

de modelos logísticos, a exemplo da multicolinearidade entre as variáveis predictoras. A fim de verificarmos tal pressuposto, usamos a função *vif* do pacote *car*:: (GRIES, 2013; LEVISHINA, 2015; OUSHIRO, 2017) no modelo, sem possíveis interações, obtendo o resultado de que as variáveis sexo, faixa etária e escolaridade não são multicolineares entre si, pois seus valores de GVIF estão abaixo de 5. Na sequência, excluímos a variável predictoradora sexo dos modelos realizados, com ou sem interação, aqui não exportados, pois eles não convergiram, mantendo, apenas, no modelo validado, as variáveis faixa etária e escolaridade.

Tabela 3 - Estimativas, erros padrão e valores de significância para o modelo

	Estimativa	Erro Padrão	Valor-Z	Significância P
Intercept	2.4333	0.4430	5.493	3.96e-08 ***
Faixa.etáriaDois	2.7185	0.3426	7.934	2.11e-15 ***
Faixa.etáriaTrês	0.7879	0.2189	3.600	0.000319 ***
EscolaridadeMédia	-2.3771	0.4373	-5.436	5.44e-08 ***
EscolaridadeUniversitária	-2.4132	0.4577	-5.273	1.34e-07 ***

$R^2 = 0.284$

$C = 0.793$

Modelo = glm (Ocorrências ~ Faixa.etária + Escolaridade, data = ds, family = binomial
O valor de *Intercept* corresponde aos fatores: Faixa.etáriaUm; EscolaridadeFundamental
. p > 0.05; * p < 0.05; ** p < 0.01; *** p < 0.001

Fonte: Elaboração própria.

Antes de apresentarmos a interpretação dos dados da Tabela 3, vamos entender primeiro o que ela nos apresenta de informações. O modelo é uma função de primeiro grau, cuja notação matemática é $y = a + bx$, em que o *intercept* está relacionado ao coeficiente linear do modelo, ou seja, ao valor de y no momento em que x é igual a zero. A estimativa é tida como o coeficiente angular do modelo, que corresponde ao “[...] quanto muda o valor do eixo y a cada unidade de x ” (OUSHIRO, 2017, p. 135). A segunda coluna referente ao erro padrão apresenta “[...] a medida da precisão das previsões: quanto menor esse valor, maior é o grau de precisão do modelo” (OUSHIRO, 2017, p. 135). O valor- z representa o resultado da razão entre a estimativa e o erro padrão e a significância ou valor- p é a probabilidade de ter observado um assentado resultado se, porventura, a hipótese nula for verdadeira, ou seja, se não houver diferenças significativas entre as variantes que compõem cada uma das variáveis predictoras do modelo; em caso contrário, teríamos uma hipótese alternativa, resultado da negação da hipótese nula.

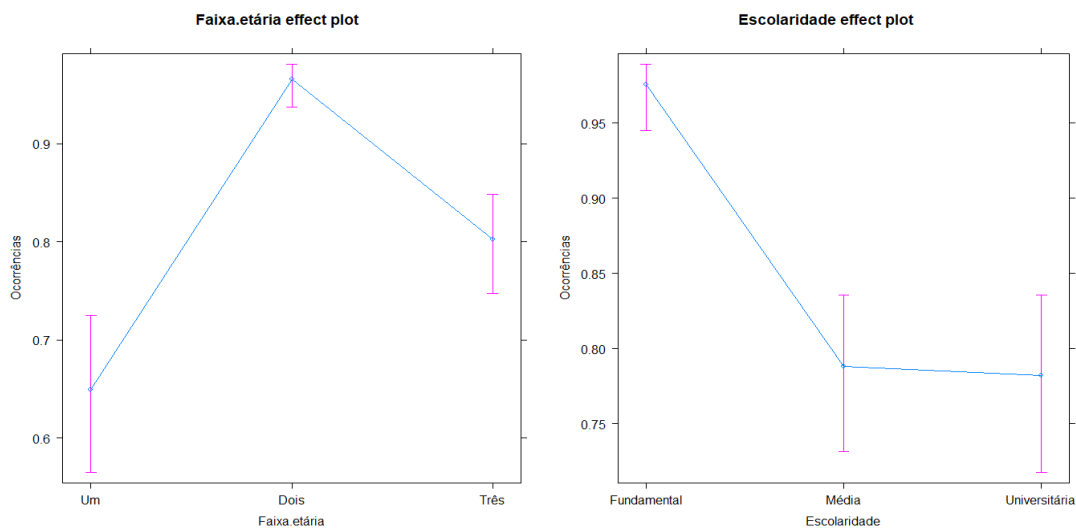
Ao aplicarmos a função *lrm* do pacote *rms* para rodar um modelo de regressão logística, obtemos outras medidas que não são geradas com a função *glm*, a exemplo de R^2 e C , “[...] medidas estatísticas de qualidade do ajuste, ou seja, de quão bem o modelo é capaz de

explicar a variação encontrada nos dados” (OUSHIRO, 2017, p. 191). Com efeito, obtemos com o modelo um $C = 0.793$, cujo valor apresenta um poder aceitável de discriminação, pois está mais próximo de 1, conseguindo explicar a variação encontrada nos dados em análise. Quanto ao valor de $R^2 = 0.284$, isso significa dizer que o modelo explica 28% da variação encontrada para a variável de resposta.

Vale anotar que usamos as funções *drop1* e *step* (com as direções *forward*, *backward* e *both*), com o objetivo de chegar a um modelo satisfatório dos dados. Além disso, verificamos se as variáveis predictoras interagiam entre si, dado que modelos com a inclusão de interações nos permitem obter estimativas mais precisas para a variável de resposta, no entanto, como assinalado, os modelos com interação que propusemos não convergiram.

Voltando aos resultados do modelo apresentado na Tabela 3, é possível afirmarmos que a variante “pra” correlaciona-se com a faixa etária e com a escolaridade, sendo desfavorecida pela faixa etária um e pela escolaridade média e universitária. Observemos o gráfico de efeitos, a seguir, com o intuito de melhor visualizarmos os resultados encontrados para as variáveis faixa etária e escolaridade.

Gráfico 3 - Efeitos das variáveis faixa etária e escolaridade



Fonte: Elaboração própria

O gráfico à esquerda mostra os resultados com as medidas de probabilidade. Vemos que os falantes da faixa etária dois, seguidos dos da faixa etária três, tendem a empregar mais a preposição “pra”, em oposição aos da faixa etária um. Já o gráfico à direita mostra claramente que, quanto menos é escolarizado o falante, menor a tendência a empregar a variante “para”. Pelo gráfico, também parece haver uma diferença significativa entre falantes com níveis de escolaridade médio e universitário.

Considerando os resultados obtidos com o modelo reportado, podemos afirmar que as variáveis sociais faixa etária e escolaridade se correlacionam com a variável de resposta: as faixas etárias dois e três e a escolaridade fundamental favorecem o uso da preposição “pra”, em oposição à faixa etária um e às escolaridades média e universitária, que desfavorecem o emprego dessa variante, preferindo o uso da forma preposicional “para”, dita padrão. Desses resultados podemos inferir que o fato de os mais velhos favorecerem o uso de “pra” parece estar indicando que essa forma não esteja se comportando como a mais inovadora, algo a ser verificado em trabalho futuro, ao contrastar esses resultados em relação ao uso de “pa”, uma terceira variante também identificada em nossa amostra de análise.

Dos estudos revisados nesta pesquisa, notamos que os resultados alcançados por Ferreira (2014, 2018) e Maya (2004), quanto à faixa etária, coadunam com os encontrados, aqui, por nós, em que os falantes mais velhos também são favorecedores do emprego da preposição “pra”, depois dos informantes da faixa etária dois, os mais favorecedores desse uso, assim como o resultado apontado por Lucena (2006), segundo o qual os falantes da faixa etária intermediária demonstraram preferência moderada pela variante “pra”. No que diz respeito à escolaridade, nossos resultados dialogam com os estudos de Lucena (2006), ainda que o autor tenha dividido essa variável entre analfabetos e universitários, de maneira extrema, demonstrando que os menos escolarizados são os principais promovedores do uso da variante “pra”. Vale sublinhar que a variável escolaridade parece não estar influenciando muito o uso de “pra”, dado que a diferença entre os níveis de escolaridade, nos trabalhos revisados, têm apresentado comportamentos distantes, casos opostos ao observado, neste estudo, para a comunidade de fala seabrense.

UM “PARA” OU UM “PRA” FECHAR O CERCO? ALGUMAS CONSIDERAÇÕES NÃO FINAIS

Nesta pesquisa, focamos no comportamento variável da preposição “para” no falar seabrense, com base nos pressupostos teórico-metodológicos da Teoria da Variação e Mudança Linguísticas (WEINREICH; LABOV; HERZOG, 2006 [1968], LABOV, 2008 [1972]), cujo condicionamento variável está atrelado a aspectos linguísticos e extralinguísticos, sendo que os primeiros aspectos exploraremos em estudos futuros, visto que aqui verificamos apenas a atuação das variáveis sociais sobre o uso da variante prototípica na comunidade de fala analisada, isto é, do uso de “pra” em relação a “para”.

Os resultados por nós alcançados se configuram da seguinte maneira:

1) Com a aplicação dos testes de qui-quadrado, observamos que apenas as variáveis sociais faixa etária e escolaridade se mostraram significativas na amostra analisada, cuja hipótese relacionada à faixa etária não foi confirmada, dado que são os falantes mais velhos que tendem a favorecer o uso da variante não padrão “pra”, no mesmo tempo em que confirmamos

a hipótese referente ao nível de escolaridade, de que os falantes menos escolarizados (ensino fundamental) prefeririam usar a forma “pra”;

2) Com a análise de regressão logística, averiguamos que a faixa etária e a escolaridade apresentaram significância estatística com valores-p menores do que 0,001, para os quais as estimativas demonstraram que o uso da preposição “pra” é desfavorecido pelos falantes com idades entre 20 e 33 anos e por falantes com nível de escolaridade médio e fundamental, como visto no gráfico de efeitos antes exposto.

Diante disso, é possível dizermos que o comportamento variável da preposição “para” apresenta-se de maneira abrangente, alcançando, quem sabe, a totalidade do território brasileiro. De tal modo, esperamos que os resultados obtidos neste estudo possam contribuir para outras pesquisas que vêm sendo empreendidas na área, com o intuito de ampliar o conhecimento das variedades faladas no PB. Em etapas futuras, buscaremos explorar o efeito das variáveis linguísticas sobre o fenômeno, de modo a compreendermos melhor a influência desses condicionadores na amostra de fala aqui investigada por nós.

REFERÊNCIAS

- FELGUEIRAS, C. M. **Análise da variação no uso da preposição para**. 1993. 120f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFRJ, Rio de Janeiro, 1993.
- FERREIRA, M. O. **A variação da preposição para na fala de Curitiba e de Florianópolis pelos dados do VARSUL**. 2018. 116f. Dissertação (Mestrado em Letras) – Instituto de Letras, UFRGS, Rio Grande do Sul, 2018.
- FERREIRA, M. O. **A variação da preposição para na fala de Londrina pelos dados do VARSUL**. 2014. 71f. Monografia (Graduação em Letras) – Instituto de Letras, UFRGS, Porto Alegre, 2014.
- GRIES, S Th. **Statistics for linguistics with R: a practical introduction**, Berlin, Boston: De Gruyter Mouton, 2013.
- LABOV, W. **Padrões sociolinguísticos**. São Paulo: Parábola, 2008 [1972].
- LEVISHINA, N. **How to do linguistics with R: data exploration and statistical analysis**. Amsterdam: John Benjamins, 2015.
- LUCENA, R. M. de. **Elementos para o estudo da variação lingüística na Paraíba**. João Pessoa: Ed. do Autor, 2006.
- MAYA, L. Z. A. **Variação da preposição para na fala de Porto Alegre**. 2004. 121f. Dissertação (Mestrado em Letras) – Faculdade de Letras, PUCRS, Porto Alegre, 2004.

OUSHIRO, L. **Introdução à estatística para linguistas**. Zenodo, 2017. Disponível em: <http://doi.org/10.5281/zenodo.822070>. Acesso em: 20 jul. 2021.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2020.

SILVA, N. de A. **A preposição para e suas variantes no falar araguanense**. 2010. 74f. Dissertação (Mestrado em Letras) – Centro de Ciências Humanas, Letras e Artes, UFPB, João Pessoa, 2010.

WEINREICH, U.; LABOV, W.; HERZOG, M. **Fundamentos empíricos para uma teoria da mudança lingüística**. Tradução de Marcos Bagno; revisão técnica de Carlos Alberto Faraco; posfácio de Maria da Conceição Paiva e Maria Eugênia L. Duarte. São Paulo: Parábola, 2006 [1968].

Recebido para publicação em: 20 ago. 2021.

Aceito para publicação em: 4 nov. 2021.