

GEOSSOCIOLINGÜÍSTICA COGNITIVA: VARIAÇÃO SOCIOCULTURAL E ESPACIAL NA COGNIÇÃO E NO USO DA LINGUAGEM

COGNITIVE GEOSOCIOLINGUISTICS: SOCIOCULTURAL AND SPATIAL VARIATION IN COGNITION AND LANGUAGE USE

Elias de Souza Santos¹

RESUMO

A emergência de uma Geossociolinguística Cognitiva, um paradigma interdisciplinar que integra os princípios da Sociolinguística, da Geolinguística e da Linguística Cognitiva, desenvolvido na tese de doutorado de Santos (2021), representa um dos modelos mais recentes da Linguística Cognitiva. Seu objetivo é explorar as dimensões geosociocognitivas da variação e da mudança linguística, utilizando métodos empíricos quantitativos e robustos. Neste estudo, portanto, apresentam-se as razões da necessidade da emergente área, ainda que sinopticamente, mostrando sua inevitabilidade dentro do escopo da Linguística Cognitiva, os modos e os métodos de conciliar os aspectos “cognitivos”, “socioculturais” e “geoespaciais”. Como aplicação ao português apresenta-se um panorama de uma pesquisa sobre a variação onomasiológica conceptual dentro do domínio da sexualidade, mais especificamente dos órgãos genitais-sexuais feminino e masculino, fazendo usufruto de uma amostra sociodialetoológica representativa de 72 colaboradores inquiridos e residentes da Chapada Diamantina, Bahia, Brasil, desde uma perspectiva do modelo de descrição linguística em tema.

PALAVRAS-CHAVE: Variação onomasiológica. Conceptualização. Categorização. Prototipicidade. Geossociolinguística Cognitiva.

ABSTRACT

The emergence of Cognitive Geosociolinguistics, an interdisciplinary paradigm that integrates the principles of Sociolinguistics, Geolinguistics and Cognitive Linguistics, developed in Santos' doctoral thesis (2021), represents one of the most recent models of Cognitive Linguistics. Its objective is to explore the geosociocognitive dimensions of linguistic variation and change, using quantitative and robust empirical methods. This study, therefore, presents the reasons for the need for this emerging area, albeit synoptically, showing its inevitability within the scope of Cognitive Linguistics, the ways and methods of reconciling the “cognitive”, “sociocultural” and “geospatial” aspects. As an application to Portuguese, an overview of a research on conceptual onomasiological variation within the domain of sexuality is presented, more specifically of the female and male genital-sexual organs, making use of a representative sociodialectological sample of 72 employees surveyed and residents of Chapada Diamantina, Bahia, Brazil, from the perspective of the linguistic description model in theme.

KEYWORDS: Onomasiological variation. Conceptualization. Categorization. Prototypicality. Cognitive Geosociolinguistics.

1. Anunciando algumas nótulas

Weinreich, Labov e Herzog (2006 [1968]), percussores dos fundamentos empíricos para uma teoria da variação e mudança linguísticas em Sociolinguística, assumem que a variabilidade da

¹ Universidade do Estado da Bahia (UNEB), Departamento de Ciências Humanas e Tecnologias (DCHT-XXIII), helyasouza@gmail.com, <https://orcid.org/0000-0002-1809-8312>.

linguagem reflete a estrutura social. Tal axioma foi admitido em obras seminais como as de Labov (1963) acerca da variação dos ditongos /aw/ e /ay/ na Ilha de Martha's Vineyard, localizada na costa nordeste da América do Norte e ainda no trabalho de Labov (1966) no tocante à estratificação social do inglês na cidade de New York, de Labov (2008 [1972]) quanto a uma introdução sistemática à Sociolinguística e da obra de Trudgill (1974) a respeito da diferenciação social do inglês em Norwich. Desde então, no contexto da fonologia e da morfossintaxe, passou-se a examinar, de maneira regular, as variedades linguísticas, caracterizadas como estruturadas e correspondentes a sistemas e subsistemas apropriados às necessidades de seus utentes (Chambers *et al.*, 2002).

Para além dos contextos fonológicos e morfossintáticos, segundo Robinson (2010), aspectos onomasiológicos e semasiológicos da estrutura linguística pouco têm se constituído tema central da pesquisa sociolinguística. Conquanto, vale anotar que outras áreas do conhecimento humano vieram explorando estes aspectos, ainda que não tenham se voltado para as questões da variação e mudança em uma abordagem variacionista, a exemplo dos estudos léxicos na esfera da Geografia de palavras e dos empréstimos linguísticos apontados no parágrafo que se segue.

Embora no âmbito da Geografia das palavras, tendo em vista os trabalhos de Orton *et al.* (1962), Upton e Widdowson (1999), e dos empréstimos, com os trabalhos de Poplack *et al.* (1988), estudos léxicos tenham sido realizados, a variação e a mudança onomasiológica, no sentido “laboviano”, só posteriormente passou a ser explorada em investigações como a de Boberg (2004). Outras tentativas obtiveram êxito ao investigarem a variação do significado; podem ser citadas em outros campos do conhecimento sobre a linguagem aquelas relacionadas aos paradigmas funcionais com os estudos de Hasan (1989, 1992, 2009) e aquelas relacionadas às estruturas de análise do discurso com os estudos de Cheshire (2007), Macaulay (2005, 2006), Stenström (2000), Tagliamonte e D'Arcy (2004) e Wong (2002, 2008).

Os resultados dessas investigações, anotadas no parágrafo precedente, segundo Robinson (2010), sugerem que a construção de significado se relaciona com aspectos sociodemográficos e que a mudança semântica pode ser motivada pelo desejo dos falantes, como objetivo de indexarem diferentes posições de sua identidade. Esses estudos recomendam que uma maior exploração da variação semântica pode apresentar boas consequências para a compreensão da variação linguística, visto que ela não é gratuita, mas regularmente estruturada em termos de fatores como a idade, sexo/gênero, escolaridade, localidade e classe social dos falantes, por exemplo. Esse achado é, para Robinson (2010), uma contribuição para as observações de outros pesquisadores que demonstraram que o entrenchamento de categoriais conceptuais pode ser explicado em relação a fatores, a exemplo dos citados neste parágrafo.

Já a Linguística Cognitiva, por sua parte, como anotam Geeraerts e Cuyckens (2007), sempre concebeu o significado como o elemento mais importante da estrutura linguística. Ela considera que a estrutura semântica espelha de maneira flexível as percepções dos falantes, adaptando-se à interação deles com uma realidade física e cultural. Nessa conjuntura, era de se esperar que a variação

do significado estivesse regularmente na agenda de estudos da Linguística Cognitiva. Por certo, no domínio desta estrutura, de modo especial, no âmbito da Linguística de *Corpus*, a variação do significado léxico foi investigada por Gries (2006), Divjak (2006), Gries e Divjack (2009) e Beeching (2005), assim como no âmbito da pesquisa sociocognitivista, que buscava compreender a estrutura da variação léxica relacionada a fatores extralinguísticos, a exemplo do estudo de Geeraerts, Grondelaers e Bakema (1994) a respeito da variação onomasiológica no contexto de desenvolvimento dialetal.

Ambas disciplinas – a Sociolinguística e a Linguística Cognitiva – desde um viés teórico, concordam com a premissa de que o significado possui uma natureza variável, ainda que, na prática, os aspectos semânticos tenham sido explorados em extensões distintas em uma e outra vertente. Por um lado, na Sociolinguística, existem poucos métodos estabelecidos para poder tratar, de maneira específica, da socio-semântica. Por outro lado, na Semântica Cognitiva, afirmações são frequentemente feitas desde um viés qualitativo e introspectivo, ainda que “[...] só mais recentemente tem se voltado para a abordagem quantitativa de dados linguísticos reais” (Ferrari, 2016, p. 137). Do mesmo modo, “[...] abordagens sociolinguísticas baseadas no uso que considerem as diferenças individuais de fala ainda são relativamente infrequentes”². (Robinson, 2010, p. 87, tradução nossa).

A Sociolinguística e a Linguística Cognitiva pouco têm se beneficiado dos legados uma da outra. Assim, partindo do intercruzamento teórico-metodológico entre ambas áreas, este texto aborda a variação léxica de contextos em situações reais de uso, especificamente, a quantidade de variação que um conceito pode apresentar em articulação com fatores sociais, culturais e cognitivos. Esse tipo de variação é, para Geeraerts, Grondelaers e Bakema (1994), Geeraerts (2018), entre outros, concebida como o número de palavras ou expressões distintas que existem para se referir a um conceito particular. É importante anotar que a quantidade de variação léxica pode diferir, de maneira notável, entre os distintos conceitos. Para mais, investigaram-se os fatores que explicam tais diferenças entre conceitos.

Por outra parte, os estudos geolinguísticos têm mostrado a variação no âmbito do léxico. Estritamente, a variação léxica pode ser interpretada em termos de aspectos lectais³, como a localização territorial do falante. A título de exemplo, para alguns conceitos, outras variantes podem estar disponíveis dependendo da formação dos usuários da língua; falantes da região norte do Brasil, citando caso parecido, preferem o item léxico *peteca*, no tempo em que o item léxico *bola de gude*/ *bolinha de gude* é mais constantemente usado por falantes da região sul do país para se referirem às pequenas esferas de vidro com que os meninos gostam de brincar (Cardoso *et al.*, 2015). Ainda assim, estudos pilotos, como os de Geeraerts e Speelman (2010), Speelman e Geeraerts (2007, 2008), têm demonstrado que as características semânticas (a saber, características que dizem respeito à organização prototípica do léxico), concernentes ao significado do conceito para o qual as variantes são

² Do original: “[...] usage-based sociolinguistic approaches that consider individual speech differences are still relatively infrequent”.

³ O termo *lectais* é compreendido neste estudo como sendo o conjunto de “dialetos, variedades nacionais, socioletos, registros, estilos, idioletos” (Silva, 2012, p. 16).

utilizadas, influenciam o número de itens disponíveis, isto é, afetam significativamente a diversidade léxica. Como o escopo deste estudo tem limites definidos, não se pode sistematizar os resultados de outros aspectos da relação entre diversidade léxica e sentido, em sua forma mais ampla.

A face do exposto, Santos (2021) propôs a emergência de uma Geossociolinguística Cognitiva como uma proposta interdisciplinar que integra os princípios da Sociolinguística, da Geolinguística e da Linguística Cognitiva, visando a uma compreensão mais abrangente da variação e mudança linguísticas. A Sociolinguística, conforme os estudos pioneiros de Weinreich, Labov e Herzog (2006 [1968]), demonstra que a variação linguística é sistemática e reflete a estrutura social, sendo influenciada, exemplificadamente, por fatores como classe, gênero e localidade. Já a Geolinguística, com trabalhos como os de Orton *et al.* (1962) e Cardoso *et al.* (2015), evidencia como a distribuição espacial das variantes lexicais e fonéticas está ligada a identidades regionais e dinâmicas culturais. Ao incorporar a Linguística Cognitiva, que concebe o significado como flexível e enraizado em experiências culturais e corporificadas (Geeraerts; Cuyckens, 2007), essa abordagem amplia a análise para incluir não apenas como as variantes se distribuem social e geograficamente, mas também por que certos conceitos apresentam maior diversidade lexical, considerando fatores cognitivos como prototipicidade e saliência conceitual.

Essa perspectiva integradora permite explorar a variação onomasiológica – a escolha de diferentes itens léxicos para um mesmo conceito – como um fenômeno multideterminado, influenciado por fatores sociais (como estratificação e identidade), geográficos (como contato dialetal e fronteiras regionais) e cognitivos (como categorização e entrincheiramento semântico). Estudos como os de Geeraerts, Grondelaers e Bakema (1994) e Robinson (2010) mostram que a diversidade lexical não é aleatória, mas correlaciona-se com a interação entre percepções culturais, necessidades comunicativas e restrições cognitivas. Por exemplo, a preferência regional por peteca *versus* bola de gude no Brasil (Cardoso *et al.*, 2015), anteriormente assinalada, pode ser explicada não apenas por trajetórias históricas de contato linguístico (Geolinguística), mas também por processos de indexação identitária (Sociolinguística) e por graus de convencionalização cognitiva (Linguística Cognitiva). Assim, a Geossociolinguística Cognitiva propõe um modelo teórico-metodológico unificado, capaz de articular a variação linguística em suas dimensões espaciais, sociais e mentais, superando as lacunas deixadas por abordagens isoladas.

Na prática, com este estudo, examinou-se o efeito de determinados aspectos lexicais em uma variedade dialetológica do português falado no Território de Identidade denominado Chapada Diamantina, localizado entre a porção centro sul e centro oeste do interior da Bahia, Brasil, que serviram como ponto de partida, por serem lectos caracterizados por uma vasta quantidade de variantes léxicas estratificadas geograficamente.

Em outras palavras, tomou-se como objeto observacional de investigação conceitos léxicos pertencentes ao domínio da sexualidade, notadamente, conceitos que se referem aos órgãos genitais-sexuais feminino e masculino, a exemplo de pênis e vagina, com o intuito de compreender

como os falantes conceptualizam determinados conceitos com diferentes graus de prototipicidade, bem como as estratégias semânticas que elegem para tais conceitos, dado que as preferências que eles fazem no discurso não são aleatórias, mas que revelam padrões de usos conexos com aspectos nem sempre linguísticos, mas também contextuais, geossociais e culturais (Geeraerts; Grondelaers; Bakema, 1994). Isso significa dizer que a análise buscou determinar o poder indexal da variação semântica na expressão de itens léxicos através da interpretação de resultados alcançados a partir de uma abordagem quantitativa do *corpus*, como se verá, a seguir.

2. Aspectos teóricos, metodológicos e análise geossociolinguística-cognitiva dos resultados

Quando se pretende realizar análises estatísticas de dados, notadamente, multivariadas, de regressão logística, idealmente, elas devem ser precedidas de gráficos e análises exploratórias, a exemplo de testes de qui-quadrado, como assevera Oushiro (2017). O teste de qui-quadrado serve para aferir, de maneira quantitativa, o resultado de um experimento e a distribuição esperada para o fenômeno em estudo. À face disso, foram depreendidas da amostra de análise 480 ocorrências, 264 itens léxicos caracterizados por uma menor saliência, sendo entendidos, neste estudo, como ‘menos prototípicos’, o que resulta em 55% de seus usos, frente a 216 de itens léxicos tipificados como ‘mais prototípicos’, dado a uma maior saliência, o que equivale a um percentual de 45% de seus usos, como se observa na tabela 1, em seguida.

Tabela 1: Frequências das ocorrências de itens léxicos ‘menos prototípicos’ e itens léxicos ‘mais prototípicos’

Variantes	Itens léxicos menos prototípicos	Itens léxicos mais prototípicos	Total
N	264	216	480
%	55	45	100

Fonte: Elaboração do autor

Além da produção de gráficos, para cada uma das variáveis preditoras (independentes), aqui não exportados, submeteram-se os dados a testes de qui-quadrado, a fim de avaliar os resultados e a distribuição do fenômeno em investigação, como anteriormente assinalado. Destarte, os resultados obtidos, por meio dos valores-p, para cada preditor, com os testes aplicados, podem ser visualizados na tabela 2 a seguir:

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

Variáveis		Mais prototípico		Menos prototípico		Valor-p
		N	%	N	%	
Sexo	Homem	133	28	110	23	0.978
	Mulher	131	27	106	22	
Faixa etária	Um	101	21	108	23	0.02991
	Dois	89	19	63	13	
	Três	74	15	45	09	
Localidade	Central	96	20	93	19	0.02653
	Norte	87	18	80	17	
	Sul	81	17	43	09	
Sexo a que se refere o conceito	Feminino	96	20	134	28	3.596e-08
	Masculino	168	35	82	17	

Fonte: Elaboração do autor

Os resultados da tabela 2 indicam que não há correlação entre a variável resposta e a variável preditora sexo, pois o Valor-p, neste caso, é maior que 0.05, o que significa que não houve nenhuma diferença significativa, dado que a probabilidade de se obter um valor da estatística de teste como o observado é muito improvável. Conquanto, as variáveis faixa etária, localidade e sexo a que se refere o conceito apresentaram correlações com a variável resposta, exibindo valores-p significativos, anotados em negrito.

Além dos testes de qui-quadrado, foram também realizados testes de regressão logística. Estes testes são feitos quando se tem uma variável resposta (ou dependente) nominal binária e variáveis preditoras (ou independentes), podendo incluir mais de uma variável preditora no modelo, diferente dos testes de qui-quadrado que só permitem a inclusão de apenas uma por vez. Com efeito, o interesse nas análises de regressão logística estar em “[...] verificar o efeito simultâneo de múltiplas variáveis previsoras, a fim de chegar a um modelo para descrever, explicar e prevê o comportamento da variável resposta” (Oushiro, 2017, p. 182).

Dito de outro modo, a análise de regressão logística é o procedimento adequado para atender os requisitos postos. Ela é uma abordagem de modelagem matemática que pode ser usada para testar hipóteses sobre a relação de várias variáveis preditoras sobre variáveis respostas binárias (cf. Hosmer; Lemeshow, 1989; Kleinbaum, 1994; Tabachnick; Fidell, 2001; Baayen, 2008; Gries, 2013; Levshina, 2015; Oushiro, 2017; entre outros). Para mais, fornece informações sobre a variação – percentuais para os quais uma variável é explicada pelas variáveis resposta –, sendo também usada para determinar a importância de variáveis preditoras. Neste estudo, a regressão logística foi empregada para avaliar o efeito de fatores sociodemográficos e conceptuais sobre o uso de itens léxicos, cujos usos os caracterizam como ‘mais prototípicos’ ou ‘menos prototípicos’.

O trabalho com regressão logística compreende o *design* de uma infinidade de modelos distintos para comparar seu poder preditivo e a quantidade de variação que podem explicar. Esse exercício envolve revisão na codificação, por exemplo, recodificando variáveis, a fim de adicionar ou remover níveis, diferentes variáveis e interações, até que se chegue a um modelo satisfatório, conforme uma série de parâmetros de diagnóstico. Para as funções de regressão logística com efeitos fixos, como *glm* e *rlm* no R, há um número considerável de métodos automáticos que auxiliam na seleção do modelo mais explicativo; diferente da função de efeitos aleatórios *glmer*.

Salienta-se que a seleção de um modelo de efeitos mistos adequado para os dados estudados requer do desenho prévio de um dos modelos de efeitos fixos (principais e com interações). Inicialmente, é necessário verificar se as variáveis não incorrem em multicolinearidade, isto é, quando as variáveis preditoras do modelo estão relacionadas a um mesmo efeito. Uma vez comprovado o pressuposto da não multicolinearidade, em um modelo inicial, introduzem-se todas as variáveis principais e todas suas possíveis interações, submetendo-o a uma seleção automática, baseada no critério de Akaike (AIC). Esse procedimento é bastante útil quando se quer eliminar variáveis menos explicativas, gradualmente, até chegar a um modelo parcimonioso e que apresente o menor valor AIC.

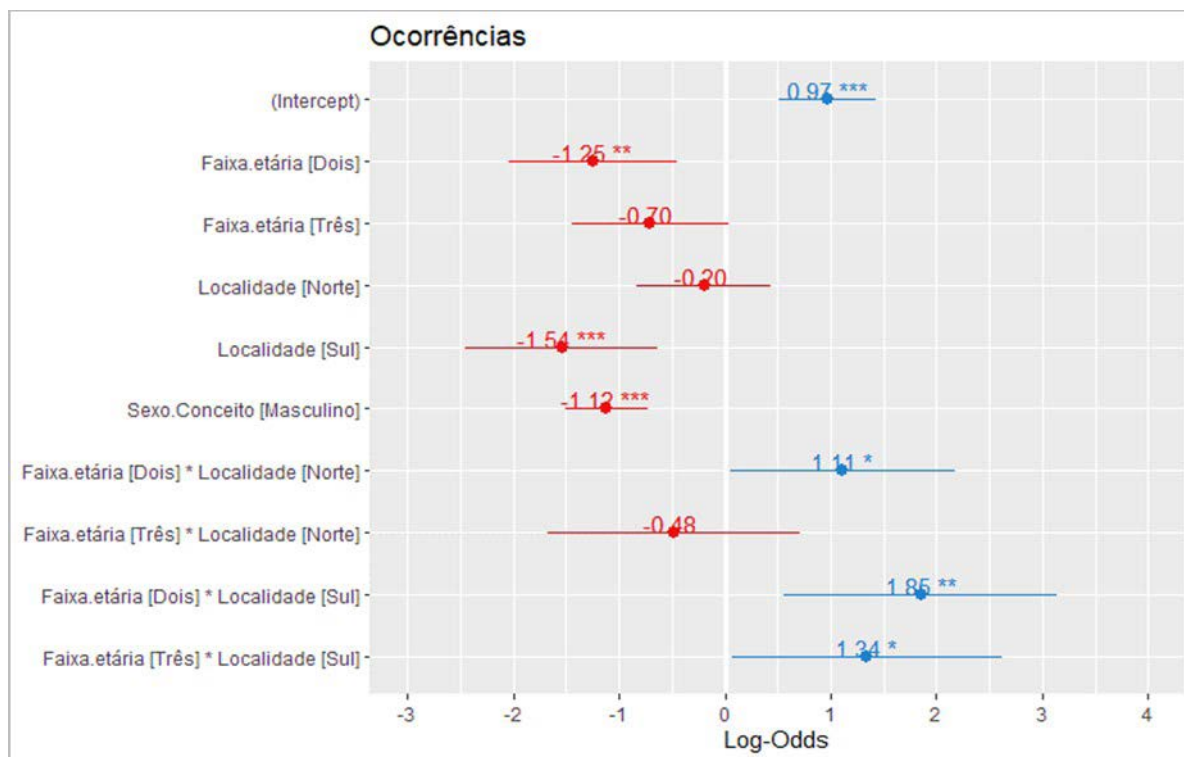
Uma vez aplicadas estas etapas à amostra de análise, neste estudo, incorporaram-se os efeitos aleatórios, observando que o modelo com esses efeitos não convergiu, não sendo, portanto, aqui exportado, selecionando o modelo com interações como sendo o mais explicativo. Na tabela 3, pode-se visualizar o aprimoramento do modelo, desde sua forma básica, com as variáveis preditoras e sem interações, até o modelo mais explicativo da variação onomasiológica conceptual de itens léxicos referentes aos órgãos genitais-sexuais feminino e masculino.

Tabela 3: Comparação entre modelos de regressão logística com efeitos principais, com interações e com efeitos mistos (conforme AIC, C, Dxy e R²)

Modelo	AIC	C	Dxy	R ²
Efeitos principais	629.03	0.679	0.358	0.121
Efeitos principais mais interações	620.93	0.700	0.400	0.156
Efeitos mistos	--	--	--	Não convergiu

Fonte: Elaboração do autor

O modelo final inclui todas as variáveis preditoras, exceto ‘Sexo.Gênero’ que não apresentou significância estatística. Isso indica que não há diferenças entre homens e mulheres quanto aos usos variáveis do fenômeno em foco. Ademais, as outras variáveis preservadas no modelo demonstraram significâncias como efeitos principais e em interações. As importâncias relativas das ditas variáveis, incluindo as interações, podem ser vistas na figura 1, em seguida.

Figura 1: Importância dos preditores fixos inclusos no modelo final

Fonte: Elaboração do autor

Como pode ser visto, a figura 1 apresenta todos os efeitos com o intervalo de confiança, com os asteriscos representando o grau de significância estatística de cada efeito (efeito negativo em vermelho e efeito positivo em azul). Logo, o uso de itens léxicos ‘menos prototípicos’ (variante observada) foi favorecido pela interação entre faixa etária e localidade e desfavorecido pela localidade, faixa etária e sexo a que se refere o conceito. O modelo não encontrou um efeito para o sexo do falante, como dito anteriormente. Na figura, o Intercepto (Intercept) representa a Faixa.etária[Um], a Localidade[Central], o Sexo.Conceito[Feminino], a Faixa.etária[Um]*Localidade[Norte] e a Faixa.etária[Um]*Localidade[Sul]. Uma vez sumarizado os detalhes para seleção do modelo, explicam-se, na sequência, os resultados da regressão logística.

A aplicação das funções *glm* e *lrm* no R (R Core Team, 2020) proporcionou a tabela 4 de resultados, logo mais apresentada, cujos valores em *Log-Odds* foram exibidos na figura 1. Os valores proporcionados pela regressão ilustram quais preditores influenciam significativamente sobre a variável resposta: itens léxicos mais prototípicos e itens léxicos menos prototípicos. Concretamente, tomou-se o primeiro nível como base (mais prototípico), a fim de calcular o efeito dos níveis (ou mesmo regressores) de cada preditor sobre o resultado ‘menos prototípico’.

Tabela 4: Resultado da análise de regressão logística com interações

	Estimativa	Erro Padrão	Valor Z	Significância (p)
(Intercepto)	0.9708	0.2365	4.104	4.06e-05 ***
Faixa.etáriaDois	-1.2464	0.4033	-3.090	0.00200 **
Faixa.etáriaTrês	-0.7043	0.3811	-1.848	0.06461 .
LocalidadeNorte	-0.1980	0.3210	-0.617	0.53728
LocalidadeSul	-1.5441	0.4672	-3.305	0.00095 ***
Sexo.Refere.ConceitoMasculino	-1.1194	0.1976	-5.666	1.46e-08 ***
Faixa.etáriaDois:LocalidadeNorte	1.1144	0.5399	2.064	0.03903 *
Faixa.etáriaTrês:LocalidadeNorte	-0.4803	0.6075	-0.791	0.42921
Faixa.etáriaDois:LocalidadeSul	1.8526	0.6588	2.812	0.00492 **
Faixa.etáriaTrês:LocalidadeSul	1.3383	0.6504	2.057	0.03964 *
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Fonte: Elaboração do autor

Ainda que a figura 1 espelhe parte dos resultados da tabela 4, fez-se o exercício de descrevê-la, com o objetivo de compartilhar maneiras de como exportar os resultados de uma regressão logística. Na coluna Estimativa é exibido o logaritmo da probabilidade de que uma ocorrência seja ‘menos prototípica’ quando os valores listados na coluna da esquerda são dados frente aos que estão omitidos (que são a base do Intercepto), ao mesmo tempo em que se controla as demais variáveis. Nesta análise, os valores positivos, na coluna Estimativa, sinalizam que o preditor favorece o uso da variante ‘menos prototípica’, o que significa que os resultados do modelo são lidos em termos do que aumenta ou diminui a probabilidade de ocorrência dessa variante.

Vale lembrar que o modelo se configura como uma equação do 1º grau, cuja notação matemática é $y = a + bx$, em que o Intercepto está relacionado com o coeficiente linear do modelo, ou seja, ao valor de y no momento em que x é igual a zero. 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 preditoras do modelo, em caso contrário, ter-se-ia uma hipótese alternativa, resultado da negação da hipótese nula.

Voltando aos resultados do modelo apresentado na tabela 4, nota-se que a variante ‘menos prototípica’ correlaciona-se com a faixa etária, a localidade e o sexo a que se refere o conceito. A primeira linha da tabela, para o regressor ‘Faixa.etáriaDois’, indica um efeito significativo. O modelo prevê que o logit seja -1.2464 maior para a ‘Faixa.etáriaDois’ do que para a ‘Faixa.etáriaUm’ e ‘Faixa.etáriaTrês’. Esse resultado, no que lhe toca, não confirma a hipótese inicial, a de que a faixa etária mais jovem apresentaria uma maior proporção de uso de itens léxicos ‘menos prototípicos’ do que os falantes da faixa etária intermediária e da faixa etária mais velha, demonstrando uma mudança em progresso.

A segunda variável atingiu significância estatística para a ‘LocalidadeSul’. Os residentes da localidade sul e norte, ainda que este segundo regressor não tenha atestado significância, preferem fazer uso de itens léxicos ‘mais prototípicos’ do que aqueles falantes habitantes da localidade central. Para este preditor a hipótese de que os falantes tanto da Chapada Norte quanto da Chapada Sul tenderiam a utilizar itens léxicos mais prototípicos, dado as suas conjunturas históricas e sociais de formação, é confirmada.

O preditor ‘SexoRefereConceito’ também atestou significância estatística. Frente ao sexo a que se refere o conceito, o masculino, nos usos pelos falantes, é mais proeminente em estruturas ‘mais prototípicas’ do que para o conceito referente ao sexo feminino para estas mesmas estruturas. A hipótese aqui também se confirma, visto que se esperava haver uma tendência para uma menor prototipicidade na expressão de itens léxicos referentes ao órgão genital-sexual feminino e, por outro lado, a inversão deste padrão para a expressão de itens léxicos referentes ao órgão genital-sexual masculino.

Finalmente, a faixa etária em interação com a localidade inverte a tendência desses últimos fatores. Ao todo, os itens léxicos referentes aos órgãos genitais-sexuais feminino e masculino são frequentemente usados como estruturas léxicas ‘menos prototípicas’ por falantes da faixa etária dois, residentes das localidades norte e sul e falantes da faixa etária três da localidade sul.

Em síntese, o modelo de regressão logística confirmou algumas hipóteses iniciais sobre quais fatores favoreciam o uso de itens léxicos ‘menos prototípicos’ e ‘mais prototípicos’ na conceptualização de órgãos genitais-sexuais feminino e masculino, a saber: a de que falantes das localidades norte e sul tenderiam a usar itens léxicos mais prototípicos e a de que haveria uma tendência para uma menor prototipicidade na expressão de itens léxicos referentes ao órgão genital-sexual feminino em oposição ao órgão genital-sexual masculino. Ademais, não foi possível observar tendências pessoais para cada falante, visto que o modelo com variáveis aleatórias, a exemplo do informante, não convergiu, no entanto, foi possível verificar a emergência de tendências de variação social que sobreatuam pela variação individual, a exemplo da faixa etária e da localidade em interação, as quais favorecem os usos de estruturas ‘menos prototípicas’.

3. Algumas nótulas finais e uma para além delas

Ao longo das seções precedentes, buscou-se contribuir com os estudos da variação semântica de itens léxicos que denominam partes do corpo humano associadas ao sexo, mais especificamente, da variação onomasiológica conceptual de expressões referentes aos órgãos genitais-sexuais feminino e masculino, cujos usos apresentam graus de saliência “mais prototípicos” ou “menos prototípicos”, concebendo uma perspectiva sociolinguístico-cognitiva. A fim de verificar se fatores sociais e conceptuais condicionavam a variação dos ditos itens léxicos, levantou-se alguns objetivos para a proposta de investigação, cujos direcionamentos tendiam para uma melhor compreensão da variação interna da linguagem na comunidade de fala analisada neste estudo.

Levou-se a cabo uma análise para investigar o significado potencial de conceitos relacionados aos órgãos genitais-sexuais, o qual permite estudar a natureza experiencial do significado em sua

esfera corporal, mas também social, visto que a experiência do corpo é culturalmente construída. Esse exame se dedicou a apreender as preferências dos falantes por estruturas “mais prototípicas” e “menos prototípicas” em relação com fatores sociais, como o sexo e a faixa etária, espacial, como a localidade e conceptual, como o sexo a que se refere o conceito. Mediante modelo de regressão logística, os resultados da análise permitiram responder algumas das questões levantadas para o estudo. De início, observou-se que, para além da variação individual, existem padrões sociais que condicionam a expressão de itens léxicos “mais prototípicos” e “menos prototípicos, a exemplo da faixa etária, da localidade e do sexo a que se refere o conceito.

Em síntese, por meio do estudo empreendido, pôde-se chegar a uma conclusão importante: no que diz respeito à variação dos itens léxicos analisados, pôde-se afirmar que nem todos os fatores sociais e conceptuais são igualmente produtivos desde um ponto de vista sociolinguístico, visto que foi possível constatar tal assertiva mediante análise estatística das ocorrências, conquanto, não se pode duvidar de que a variação semântica é flexível em termos de informações sociais e conceptuais. Ademais, notou-se que as escolhas entre estruturas “mais prototípicas” e “menos prototípicas”, feitas pelos falantes, refletem escolhas discursivas distintas, nos seus atos de fala, que estão relacionadas com suas ideologias e identidades, nos processos de interação, na comunidade linguística a que pertencem.

A relevância deste estudo, centrado no nível onomasiológico conceptual, pode ser estendida, a princípio, a qualquer fenômeno de variação semântica. Trata-se de uma proposta original, estabelecida no seio da Geossociolinguística Cognitiva, cujos aportes estão sendo constituídos. É um modelo bastante lógico, para seu estágio inicial, que, porventura, ainda permitirá inúmeros avanços, dado a sua predisposição para a interdisciplinaridade. Vale reiterar que não é uma proposta radicalmente nova, pois já existem algumas abordagens que convergem para o seu desenvolvimento, no entanto, o que distingue este trabalho de tantos outros é a reflexão que se fez de como o estudo da variação semântica pode contribuir para a apreensão do significado da variação e a variação do significado, pois o modelo em foco, não essencialista e interdisciplinar, combinado a um desenho metodológico quantitativo, consegue abordar com sucesso fenômenos de significado que não se identificam bem com estruturas menos flexíveis.

Algumas são as possibilidades que se colocam para estudos futuros, que se veja:

- 1) Contrastar os resultados deste estudo, por exemplo, com outros já realizados em *corpus* e corpora não elicitados (material escrito), a fim de observar as convergências e divergências entre eles;
- 2) Verificar a criatividade linguística na linguagem cotidiana em sua comparação com a criatividade em textos literários, em particular sobre a temática da sexualidade;
- 3) Observar se existe correspondência entre os usos léxicos e estratégias no nível semântico, em distintas comunidades linguísticas, em relação com fatores sociais, conceptuais e contextuais;
- 4) Ampliar o estudo a outros níveis que não apenas o conceptual, tais como o semasiológico e onomasiológico formal, com o propósito de se obter uma perspectiva global sobre os fatores que condicionam a seleção onomasiológica;

- 5) Examinar conceitos que estejam relacionados com a esfera da sexualidade, a exemplo da homossexualidade, da prostituição, das práticas sexuais, do erotismo, da pornografia, do desejo etc.;
- 6) Constituir bancos de dados elicitados sobre sexualidade, inexistentes e complexos para sua composição, dada a interdição que recai sobre essa temática; e
- 7) Tratar de outros fenômenos, em vários níveis da língua, usando os fundamentos teóricos e metodológicos da Geossociolinguística Cognitiva.

Postos os resultados obtidos ao longo deste texto seria até paradoxal falar de considerações finais neste momento, pois, considera-se, este, um trabalho em curso que deixa como nota, antes de seu ponto final, que a variação semântica é um espaço que goza de privilégio por permitir que se possa examinar o significado social de inúmeras conceptualizações do ambiente que circunda os sujeitos humanos, especialmente, no âmbito da sexualidade, atividade humana que está vinculada às emoções, às frustrações, os desejos, ao prazer próprio e do outro, ao gozo e às ideologias, desenvolvendo facetas que estão ligadas profundamente à história pessoal de cada indivíduo e de seu povo.

Referências

- BAAYEN, H. *Analysing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press, 2008.
- BEECHING, K. Politeness-induced semantic change: the case of *quand même*. *Language Variation and Change*, v. 17, pp. 155-180, 2005.
- BOBERG, C. Ethnic patterns in the phonetics of Montreal English. *Journal of Sociolinguistics*, v. 8, n. 4, pp. 538-568, 2004.
- CARDOSO, S. *et al.* *Atlas Linguístico do Brasil*, v. I e II. Londrina: EDUEL, 2015.
- CHAMBERS, J. K. *et al.* (ed.). *The Handbook of Language Variation and Change*. Blackwell Handbooks in Linguistics. Malden, Mass.; Oxford: Blackwell Publishers, 2002.
- CHESHIRE, J. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics*, v. 11, n. 2, pp. 155-193, 2007.
- DIVJAK, D. *Ways of intending*: delineating and structuring near synonyms. *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. S. T. Gries and A. Stefanowitsch. Berlin, Mouton de Gruyter, 2006.
- FERRARI, L. *Introdução à Linguística Cognitiva*. São Paulo: Contexto, 2016.
- GEERAERTS, D.; GRONDELAERS, S.; BAKEMA, P. *The Structure of Lexical Variation: Meaning, and Context*. Berlin/New York: Mouton de Gruyter, 1994.
- GEERAERTS, D.; CUYCKENS, H. (ed.). *The Oxford handbook of cognitive linguistics*. Oxford: University Press, 2007.

GEERAERTS, D.; SPEELMAN, D. Heterodox concept features and onomasiological heterogeneity in dialects. In: GEERAERTS, D.; KRISTIANSEN, G.; PEIRSMAN, Y. (ed.), *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 2010, pp. 23-39.

GEERAERTS, D. *Ten Lectures on Cognitive Sociolinguistics*. Leiden: Brill, 2018.

GRIES, S. T. *Corpus-based methods and cognitive semantics: The many senses of run*. Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis. S. T. Gries and A. Stefanowitsch. Berlin, Mouton de Gruyter, 2006.

GRIES, S. T.; DIVJAK, D. *Behavioural profiles: A corpus-based approach to cognitive semantic analysis* New Directions in Cognitive Linguistics. V. Evans and S. S. Pourcel. Amsterdam: Benjamins, 2009, pp. 57-75.

GRIES, S. T. *Statistics for Linguistics with R: A Practical Introduction*. Berlin/Boston: Mouton de Gruyter, 2013.

HASAN, R. Semantic variation and sociolinguistics. *Australian Journal of Linguistics* v. 9, pp. 221-275, 1989.

HASAN, R. *Meaning in sociolinguistic theory Sociolinguistics Today: International Perspectives* K. Bolton and H. Kwok. London: Routledge, 1992.

HASAN, R. *Semantic Variation: Meaning in Society and in Sociolinguistics*. The Collected Works of Ruqaiya Hasan. Collected Works of Ruqaiya Hasan. London, Equinox, 2009.

HOSMER, D. W.; LEMESHOW, S. *Applied Logistic Regression*. New York: John Wiley and Sons, 1989.

KLEINBAUM, D. *Logistic Regression: A Self-Learning Text*. New York: Springer, 1994.

LABOV, W. The social motivation of a sound change. *Word*, v. 19, pp. 273-309, 1963.

LABOV, W. The linguistic variable as a structural unit. *Washington Linguistics Review*, v. 3, pp. 4-22, 1966.

LABOV, W. *Padrões sociolinguísticos*. Tradução de Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues Cardoso. São Paulo: Parábola, 2008 [1972].

LEVSHINA, N. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015.

MACAULAY, R. K. S. *Talk that counts: age, gender, and social class differences in discourse*. New York; Oxford: Oxford University Press, 2005.

MACAULAY, R. K. S. Pure grammaticalization: The development of a teenage intensifier. *Language Variation and Change*, v. 18, n. 3, pp. 267-283, 2006.

ORTON, H. et al. (ed.). *Survey of English Dialects*. Leeds: Arnold, 1962.

OUSHIRO, L. *Introdução à Estatística para Linguistas*. (1.0.0) Zenodo, 2017. DOI: <http://doi.org/10.5281/zenodo.822070>.

POPLACK, S. et al. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, v. 26, pp. 47-104, 1988.

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

ROBINSON, J. A. Awesome insights into semantic variation. In: GEERAERTS, G.; KRISTIANSEN, G.; PIERSMAN, Y. (ed.). *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 2010, pp. 85-109.

SANTOS, E. de S. *Recortes do português diamantino: um estudo sobre a variação léxica da categoria órgãos genitais-sexuais à luz da Geossociolinguística Cognitiva*. 2021. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, 2021.

SPEELMAN, D.; GEERAERTS, D. De structuur van lexicale onzekerheid. *Taal en Tongval*, Theme number 20, pp. 47-61, 2007.

SPEELMAN, D.; GEERAERTS, D. The role of concept characteristics in lexical dialectometry. *International Journal of Humanities and Arts Computing*, v. 2, n. 1-2, pp. 221- 242, 2008.

STENSTRÖM, A. B. *It's enough funny, man: Intensifiers in teenage talk*. Corpora Galore: Analyses and Techniques in Describing English. J. M. Kirk. Amsterdam: Rodopi, 2000, pp. 177-190.

TABACHNICK, B. G.; FIDELL, L. S. *Using Multivariate Statistics*. Boston: Allyn and Bacon, 2001.

TAGLIAMONTE, S. A.; D'ARCY, A. He's like, she's like: The quotative system in Canadian youth. *Journal of Sociolinguistics*, v. 8, n. 4, pp. 493-514, 2004.

TRUDGILL, P. J. *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press, 1974.

UPTON, C.; WIDDOWSON, J. D. A. *Lexical Erosion in English Regional Dialects*. Sheffield: The University of Sheffield, 1999.

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

WONG, A. *The Semantic Derogation of Tongzhi: A Synchronic Perspective*. Language and Sexuality: Contesting Meaning in Theory and Practice. K. Campbell-Kibler, R. Podesva, S. J. Roberts and A. Wong. Stanford, California: CSLI, 2002, pp. 161-174.

WONG, A. On the actuation of semantic change: The case of tongzhi. *Language Sciences*, v. 30, n. 4, pp. 423-449, 2008.