

ESQUEMA DE ORDENAÇÃO DE *HALLIG E WARTBURG*: UMA AVALIAÇÃO DE SUA APLICAÇÃO À ANÁLISE LEXICAL DO PORTUGUÊS

RESUMO

O presente estudo teve como objetivo aplicar o esquema de ordenação de Hallig e Wartburg ao léxico do diário de Maurício Graboís redigido durante a Guerrilha do Araguaia na época da ditadura no Brasil. Testou-se a hipótese de que esse esquema não seria capaz de dar conta do léxico presente no diário de Graboís. A análise teve como fundamentação teórica a lexicologia social de Matoré e a metodologia baseou-se na lematização do diário de Graboís, com análise dos 50 itens mais frequentes. Os dados confirmaram a hipótese avaliada, pois 34% dos itens não apresentam correspondência no esquema considerado. Esse resultado sugere que cada cultura de cada época organizaria seu patrimônio lexical de uma forma própria, sendo improvável existir um esquema que seja aplicável ao léxico de qualquer cultura de qualquer época.

Palavras-chave: Lexicologia; Ditadura Militar; Guerrilha do Araguaia.

ORDINATION SCHEME OF HALLIG AND WARTBURG: AN EVALUATION OF ITS APPLICATION TO LEXICAL ANALYSIS OF PORTUGUESE

ABSTRACT

The present study aimed to apply the ordination scheme of Hallig and Wartburg to the lexicon of Maurício Graboís' diary written during the Guerrilla of the Araguaia at the time of the dictatorship in Brazil. It was hypothesized that this scheme would not be able to account for the lexicon in Graboís's diary. The analysis was based on the social lexicology of Matoré and the methodology was based on the lemmatization of the diary of Graboís, with the analysis of the 50 most frequent items. The data confirmed the evaluated hypothesis, since 34% of the items do not correspond in the considered scheme. This result suggests that each culture of each time would organize its lexical heritage in its own way, being unlikely to exist a scheme that is applicable to the lexicon of any culture of any time.

Key words: Lexicology; Military dictatorship; Guerrilla of the Araguaia

ESQUEMA DE ORDENACIÓN DE HALLIG Y WARTBURG: UNA EVALUACIÓN DE SU APLICACIÓN AL ANÁLISIS LÉXICO DEL PORTUGUÉS

RESUMEN

El presente estudio tuvo como objetivo aplicar el esquema de ordenación de Hallig y Wartburg al léxico del diario de Maurício Graboís redactado durante la Guerrilla del Araguaia en la época de la dictadura en Brasil. Se probó la hipótesis de que ese esquema no sería capaz de dar cuenta del léxico presente en el diario de Graboís. El análisis tuvo como fundamentación teórica la lexicología social de Matoré y la metodología se basó en la lematización del diario de Graboís, con análisis de los 50 ítems más frecuentes. Los datos confirmaron la hipótesis evaluada, pues el 34% de los ítems no presentan correspondencia en el esquema considerado. Este resultado sugiere que cada cultura de cada época organizaría su patrimonio lexical de una forma propia, siendo improbable existir un esquema que sea aplicable al léxico de cualquier cultura de cualquier época.

Palabras clave: Lexicología; Dictadura militar; Guerrilla del Araguaia.

1. INTRODUÇÃO

Uma das tarefas da lexicologia é propor um esquema de ordenação que seja capaz de dar conta do léxico de qualquer língua, uma vez que, como ciência, deve buscar os princípios fundamentais subjacentes às línguas naturais.

Um tal esquema de ordenação, além de sua contribuição teórica, também tem uma faceta aplicada: serve, por exemplo, como base para a elaboração de dicionários (com impacto, portanto, na lexicografia). Em função da natureza conceitual que o fundamenta, o referido esquema também tem aplicação para a apreensão da visão de mundo expressa através de textos. Essa aplicação, se realizada com uma metodologia apropriada, tem como vantagem permitir o desenvolvimento de ferramentas para sua aplicação em *corpora* mais extensos, seguindo a tendência atual de exploração de grandes volumes de dados (os chamados *big data*). A constituição da linguística de *corpus* e a criação de ferramentas computacionais para análises linguísticas tem propiciado análises mais abrangentes e refinadas do léxico, explorando de forma exaustiva os textos usados como *corpus* e trabalhando com *corpora* cada vez maiores. Especialmente em relação ao português, devem-se mencionar as experiências precursoras de Duncan (1972) e Biderman (1998). Um exemplo interessante e recente, mas aplicado ao inglês, é o estudo realizado a partir da base de dados do 5 milhões de textos do Google Books (MICHEL et al., 2011), em uma abordagem a que os autores chamam de *culturonomia*¹,

No presente estudo, aplica-se a proposta de esquema de ordenação de Hallig e Wartburg (1952) ao léxico do diário de Maurício Grabois redigido durante a Guerrilha do Araguaia na época da ditadura no Brasil.

2. FUNDAMENTAÇÃO TEÓRICA: LÉXICO COMO SISTEMA

A visão de língua como sistema também foi aplicada ao nível do léxico pelo próprio Saussure (1915 [1988, p. 145]):

Os grupos formados por associação mental não se limitam a aproximar os termos que apresentam algo em comum; o espírito capta também a natureza das relações que os unem em cada caso e cria com isso tantas séries associativas quantas relações diversas existam.

Embora o modelo linguístico de Saussure trabalhasse com as noções de eixo paradigmático (relações de substituição) e eixo sintagmático (relações de combinação), o excerto apresentado anteriormente, ao fazer referência aos “grupos formados por associação mental”, trata essencialmente da questão do eixo paradigmático.

Como é sabido, essa visão de léxico como sistema foi retomada e discutida por diferentes teóricos que o sucederam, como Trier (1931) e seus campos linguísticos, Matoré (1953 [1973]) e seus campos nocionais, Coseriu (1977) e seus campos léxicos, dentre outros.

Dentre esses teóricos, Matoré (1953 [1973]) terá sido o que mais enfatizou o aspecto social do léxico. Na visão dele o léxico consiste em uma forma de mapeamento do mundo pela comunidade de falantes:

Na realidade, as palavras não exprimem as coisas, mas a consciência que os homens têm delas. Para a lexicologia, os fatos sociais têm, com efeito, o aspecto de *coisas*, mas das coisas vistas, sentidas, compreendidas pelos homens; nossa disciplina deverá então visar às realidades sociológicas das quais o vocabulário é a “tradução”, ao mesmo tempo objetivamente, como realidades independentes do indivíduo, e subjetivamente, em função dos seres que vivem em um meio concreto, em certas condições sociais, econômicas, estéticas, etc. (MATORÉ, 1953 [1973, p. 42-43], tradução minha, *itálico do original*)

Apesar dessa sua ênfase no aspecto social (que certamente envolve também um aspecto cognitivo), Matoré (1953 [1973]) manteve-se de certa forma fiel às suas raízes estruturalistas ao contemplar a questão dos eixos paradigmáticos e sintagmáticos saussureanos:

Matoré interpreta a organização do vocabulário a partir de uma concepção estruturalista, admitindo que as palavras existem na consciência em relações recíprocas (uma visão sistêmica de vocabulário): “a palavra não está isolada na consciência: ela estabelece com suas vizinhas, no contexto, relações sintagmáticas” [...] e “independentemente do contexto, a palavra está ligada na consciência a outras palavras a que se assemelham, seja pela forma, seja pelo sentido: estas são as relações associativas” [...] (CAMBRAIA, 2013, p. 161)

Embora seu método de análise lexical tenha recebido duras críticas em função de uma série de imprecisões (Cambraia, 2013, p. 164-167), propostas metodológicas posteriores (como, por exemplo, a de Coseriu) acabaram fixando-se mais em aspectos intralinguísticos do que em aspectos extralinguísticos, o que acabou ofuscado um dos aspectos que Matoré tanto enfatizou: o social.

Uma das críticas feitas à proposta de Matoré (1953 [1973]) foi a de que haveria arbitrariedade no processo de delimitação e de hierarquização dos elementos que comporiam um campo nocional (Cambraia, 2013, p. 164-167). Justamente em relação à superação dessa limitação, destaca-se a proposta de Hallig e Wartburg (1952), cujo título, de certa forma, já evidencia uma orientação mais aplicada da proposta: *Begriffssystem als Grundlage für die Lexicographie: Versuch eines Ordnungsschemas* (“Sistema de conceitos como fundamento para a lexicografia: tentativa de um esquema de ordenação”). Em primeiro lugar, cumpre enfatizar o emprego do termo *sistema*, que certamente é repercussão da visão saussureana de língua como sistema. Em segundo lugar, é curioso que, embora, na época de publicação dessa proposta, já circulassem os termos *campo linguístico* (“sprachliches Feld”) de Trier e *campo nocional* (“champ notionel”) de Matoré, não tenham os autores adotado terminologia semelhante. Pode-se hipotetizar que, visando uma aplicação ampla e universal, tenham adotado uma orientação onomasiológica, ou seja, o foco era catalogar conceitos/significados (por isso, “sistemas de conceitos”), e não itens lexicais/significantes.

A questão sobre o critério de ordenação dos dados foi discutido por Biderman (1981), que apresentou uma proposta integrativa com *rede semântica* (contemplando o significado) e *campo léxico* (contemplando o significante): “[U]ma rede semântica é composta da integração estruturada de vários campos léxicos. Um campo léxico integra uma rede semântica juntamente com muitos outros campos léxicos” (BIDERMAN, 1981, p. 139).

3. HIPÓTESE DE TRABALHO

Tomando como ponto de partida a afirmação de Matoré (1953 [1973]) de que “as palavras não exprimem as coisas, mas a consciência que os homens têm delas”, pode-se hipotetizar que, apesar da pretensão universalista subjacente ao esquema de ordenação de Hallig e Wartburg (1952), a visão de mundo de cada cultura e de cada época deve colocar limitações para a aplicação da referida proposta a qualquer *corpus*. Mais concretamente, hipotetiza-se que o esquema de ordenação de Hallig e Wartburg (1952) não é capaz de dar conta do léxico presente no diário de Grabois, pelo fato de aquele ter sido concebido em um contexto histórico e cultural diverso do em que este o foi.

4. METODOLOGIA

4.1. O corpus

Elegeram-se como *corpus* para o presente estudo o diário de Maurício Grabois (1912-1973), escrito entre abril de 1972 e dezembro de 1973. O referido diário encontra-se nos arquivos das Forças Armadas e foi trazido a público pela revista *Carta Capital* em abril de 2011. Atualmente pode-se acessar cópia transcrita do diário (GRABOIS, 1973) no Centro de Documentação e Memória do *site* da Fundação Maurício Grabois.

No que se refere ao diário, o jornalista Lucas Figueiredo apresentou alguns dados em entrevista à referida revista:

Esse diário é o registro histórico mais aprofundado da Guerrilha do Araguaia. O

documento possui mais de 86 mil palavras. Para se ter uma ideia, o texto digitalizado completou 150 páginas de tamanho A4, que cobrem 605 dias de conflito. Além de lançar luzes sobre esse episódio nebuloso da ditadura, o documento é uma peça valiosa por incluir o relato pessoal de Grabois. Toda a sua dor, angústia, solidão, saudades da família estão contempladas no texto, que revela o lado humano do guerrilheiro. [...]

Pela primeira vez temos acesso a um relato mais profundo por parte dos guerrilheiros do período mais sangrento da Guerrilha do Araguaia. Grabois foi executado em 25 de dezembro de 1973. Foi um dos últimos insurgentes a morrer. Na prática, houve três grandes campanhas dos militares contra a guerrilha. Na última, não houve preocupação de efetuar prisões, e sim de eliminar os combatentes. Como o diário vai de abril de 1972 a dezembro de 1973, temos mais informações sobre essa fase final. Os poucos sobreviventes, não mais do que meia dúzia, não deixaram relatos consistentes. Um deles, Ângelo Arroyo, morreria em 1976 na chacina da Lapa, no Rio de Janeiro. Os demais eram desertores, não quiseram falar muito sobre o que aconteceu. Esse diário está nos arquivos sigilosos das Forças Armadas desde então. Só foi revelado agora por CartaCapital. (REDAÇÃO, 2011)

4.2. Coleta dos dados

Para realizar a análise lexical, obteve-se uma cópia do arquivo digital com o diário. O site de revista *Carta Capital* disponibilizou, no início de sua divulgação em 2011, duas versões do diário: uma como a transcrição fiel ao texto original e outra com intervenções do editor a fim de esclarecer o sentido de abreviaturas e de passagens do texto. Veja-se abaixo a comparação do trecho inicial de ambas, com marcação aqui em negrito do que foi acrescentado pelo referido editor (normalmente entre parênteses):

Texto original	Texto com intervenções
30/4 – Começou a G.P. a 12/4. O inimigo, possivelmente informado por alguma denúncia, atacou de surpresa o Peazão entre as 15 e as 16 horas daquele dia. Avisado com poucas horas (2) de antecedência, pela massa, o D.A. retirou-se organizadamente para a mata. O G3 daquele D., que estava sediado no Peazão, dada a superioridade do adversário, não ofereceu combate, mas salvou seus efetivos, se armamento e diversos materiais. No entanto, muitas coisas foram deixadas no local, principalmente roupa e a oficina. O C do DA mandou avisar o G2 e o G1, ordenando-lhes que se retirassem para um ponto previamente estabelecido, onde se encontraria o DA. A manobra realizou-se com pleno êxito. Na manhã do dia seguinte, 13 de abril. Juntavam-se o G3 e o G1. Três dias depois, chegava o G2, tendo realizado marcha difícil pelos caminhos e pela mata, enquanto no ar sobrevoavam helicópteros e nas estradas ouviam-se rajadas de metralhadoras. Os G1 e G2 salvaram todo o seu equipamento, armas e todos os objetos de uso.	30/4 – Começou a G.P. (Guerra Popular) a 12/4. O inimigo, possivelmente informado por alguma denúncia, atacou de surpresa o Peazão (na Faveira, na beira do Araguaia) entre as 15 e as 16 horas daquele dia. Avisado com poucas horas (2) de antecedência, pela massa, o D.A. (Dest. "A") retirou-se organizadamente para a mata. O G3 (grupo 3) daquele D., que estava sediado no Peazão, dada a superioridade do adversário, não ofereceu combate, mas salvou seus efetivos, se armamento e diversos materiais. No entanto, muitas coisas foram deixadas no local, principalmente roupa e a oficina. O C (comando) do DA mandou avisar o G2 e o G1, ordenando-lhes que se retirassem para um ponto previamente estabelecido (os refúgios no meio da mata), onde se encontraria o DA. A manobra realizou-se com pleno êxito. Na manhã do dia seguinte, 13 de abril. Juntavam-se o G3 e o G1. Três dias depois, chegava o G2, tendo realizado marcha difícil pelos caminhos e pela mata, enquanto no ar sobrevoavam helicópteros e nas estradas ouviam-se rajadas de metralhadoras. Os G1 e G2 salvaram todo o seu equipamento, armas e todos os objetos de uso.
Na manhã do dia 13 foi enviado um mensageiro (Land) para apanhar o com. Cid e recolher informações sobre o inimigo e o estado de espírito da massa.	Na manhã do dia 13 foi enviado um mensageiro (Land) (Landim) para apanhar o com. (comandante) Cid (João Amazonas) e recolher informações sobre o inimigo e o estado de espírito da massa.

As intervenções do editor são importantes para a compreensão do diário pelo público atual, não familiarizado com a linguagem da época e do contexto em que o autor estava inserido. Entretanto, para a análise lexical, não pode ser usado como fonte de dados, uma vez que agrega palavras que não são simplesmente desenvolvimento de abreviaturas, como no caso de “na Faveira, na beira do Araguaia”, que tem como objetivo melhor precisar geograficamente os eventos então em curso. Em função disso, empregou-se como fonte de dados o texto

original, mas fez-se o desenvolvimento das abreviaturas para o texto poder ser processado adequadamente².

Uma vez eleita a fonte de dados, fez-se conversão do arquivo original em formato *pdf* para formato *txt*, pois é este o formato compatível com o programa Concorciador *AntConC* (versão 3.5.7), também utilizado nesta análise³.

Considerando o tipo de análise aqui proposto, voltado para léxico como expressão de uma visão de mundo, em que o prioritário são as palavras de valor referencial, aplicou-se no processamento do *corpus* uma lista de exclusão de itens gramaticais (artigos, pronomes, preposições, conjunções, interjeições, numerais), mantendo-se, portanto, apenas itens lexicais (substantivos, adjetivos, verbos e advérbios em *-mente*). Também foram desconsiderados os nomes próprios (antropônimos e topônimos), porque a ênfase da análise está em um mapeamento de aspectos mais gerais, e os verbos *ser*, *estar*, *ter* e *haver*, frequentemente usados como auxiliares e, portanto, semanticamente pouco relevantes para a presente análise.

Através do recurso de *Wordlist* do *AntConC*, gerou-se uma lista de formas (*types*) ordenados por frequência⁴. Considerando o fato de que as formas flexionadas, sobretudo de verbos, acabam pulverizando o peso do item lexical (*lemma*), aplicou-se um processo de lematização à lista de formas gerada pelo *AntConC*. A análise final, portanto, tem como referência os itens lexicais lematizados (reunião de todas as formas flexionadas de um mesmo item lexical). Foram eleitos para a análise os 50 itens lexicais mais frequentes. A determinação do número de itens a serem analisados é uma decisão que envolve certa arbitrariedade, na medida em que pode-

riam ser 25 ou 100: não parece haver ainda um critério seguro para servir de valor de corte na lista de palavras a serem analisadas. Entretanto, deve-se salientar que, quanto mais frequente um item lexical, naturalmente mais relevante ele é, por isso, ainda que haja certa arbitrariedade no corte, o peso que cada um apresenta é pertinente.

Após a obtenção da lista dos 50 itens lexicais mais frequentes, procedeu-se a uma organização desses itens tomando como referência a proposta de Hallig e Warburg (1952).

Antes de passar adiante, é necessário dar ainda mais alguns esclarecimentos metodológicos.

Primeiramente, o critério de delimitação de palavra adotado foi o gráfico.

Em segundo lugar, convém mencionar a questão da homonímia, sempre difícil de ser tratada quando se lida com análises pensadas para grande volume de dados. Em alguns casos, a homonímia entre substantivo e adjetivo envolve diferença semântica relevante (como o substantivo *política* e seu adjetivo homônimo), razão pela qual esses itens devem ser tratados separadamente; em outros casos, não se trata de diferença semântica relevante (como o substantivo *combatente* e seu adjetivo homônimo), pelo que não haveria problema em tratá-los de forma unificada, como se fez aqui.

5. DESCRIÇÃO DOS DADOS⁵

Após o processamento pelo *AntConC*, obteve-se a lista da direita, a qual depois do processo de lematização, resultou na lista da esquerda:

LISTA 1 – Itens lexicais não-lematizados e lematizados

Item não-lematizado	Itens associados ⁶	Item lematizado
destacamento 622	destacamento 95 + destacamento 622	DESTACAMENTO 717
combatentes 362	combatente 47 + combatentes 362	COMBATENTE 409
inimigo 322	inimigo 322 + inimigos 14 + inimiga 9 + inimigas 1	INIMIGO 346
forças 311	força 26 + forças 311	FORÇA 337
militar 282	militar 282 + militares 69	MILITAR 351
luta 266	luta 266 + lutas 19	LUTA 285
dia 242	dia 242 + dias 128	DIA 370
guerrilheiras 228	guerrilheiros 45 + guerrilheiros 112 + guerrilheira 67 + guerrilheiras 228	GUERRILHEIRO 452
povo 205	povo 205 + povos 6	POVO 211
acampamento 199	acampamento 199 + acampamentos 5	ACAMPAMENTO 204
massas 179	massa 151 + massas 179	MASSA 330
área 178	área 178 + áreas 30	ÁREA 208
ditadura 161	ditadura 161 + ditaduras 1	DITADURA 162
grande 155	grande 155 + grandes 69	GRANDE 224
guerrilha 151	guerrilha 151 + guerrilhas 17	GUERRILHA 168
massa 151	[cf. item 11]	
comissão 142	comissão 142	COMISSÃO 142
região 141	região 141 + regiões 7	REGIÃO 148
mata 134	mata ⁷ 132 + matas 29	MATA 161
dias 128	[cf. item 7]	
soldados 128	soldado 13 + soldados 128	SOLDADO 128
trabalho 125	trabalho 125 + trabalhos 1	TRABALHO 126
ponto 123	ponto 123 + pontos 26	PONTO 149
camponeses 114	camponês 29 + camponeses 114 + camponesa 5 + camponesas 4	CAMPONÊS 152
informações 114	informação 28 + informações 114	INFORMAÇÃO 142
armada 113	armado 17 + armados 15 + armada 113 + armadas 22	ARMADO 167
guerrilheiros 112	[cf. item 8]	
partido 112	partido ⁸ 108 + partidos 2	PARTIDO 110
rádio 110	rádio 110	RÁDIO 110
exército 107	exército 107	EXÉRCITO 107
situação 105	situação 105 + situações 1	SITUAÇÃO 106
farinha 98	farinha 98	FARINHA 98
movimento 98	movimento 98 + movimentos 8	MOVIMENTO 106
local 96	local 96 + locais 19	LOCAL 115
destacamentos 95	[cf. item 1]	
atividade 93	atividade 93 + atividades 16	ATIVIDADE 109
grupo 91	grupo 91 + grupos 24	GRUPO 115
armas 82	arma 19 + armas 82	ARMA 101
bom 78	bom 78 + bons 7 + boa 61 + boas 19	BOM 165
encontro 75	encontro ⁹ 73 + encontros 2	ENCONTRO 75
comandante 74	comandante 75 + comandantes 35	COMANDANTE 110
vida 74	vida 74 + vidas 5	VIDA 79
governo 73	governo 73 + governos 1	GOVERNO 74
parece 73		PARECER 95
propaganda 72	propaganda 72	PROPAGANDA 72
anos 71	ano 66 + anos 71	ANO 137
notícias 71	notícia 36 + notícias 71	NOTÍCIA 107
vez 71	vez 71 + vezes 20	VEZ 91
política 70	político 25 + políticos 8 + política ¹⁰ 52 + políticas 5	POLÍTICO 90
grandes 69	[cf. item 14]	
militares 69	[cf. item 5]	
apoio 68	apoio 68	APOIO 68
pode 68	pode 68 + podem 26 + podemos 25 + podendo 1 + poder ¹¹ 3 + poderei 1 + poderemos 6 + poderia 15 + poderiam 5 + poderá 15 + poderão 4 + poderíamos 4 + podia 16 + podiam 4 + podre 2 + podíamos 2	PODER 211
guerrilheira 67	[cf. item 8]	
maior 67	maior 67 + maiores 16	MAIOR 83
ano 66	[cf. item 46]	
fazer 66	faz 34 + faze 1 + fazem 12 + fazemos 1 + fazendo 18 + fazer 66 + fazia 4 + faziam 3 + fazê 2 + façam 4 + faço 12 + feita 11 + feitas 5 + feito 16 + fez 35 + fiz 9 + fizemos 16 + fizer 1 + fizera 5 + fizeram 12 + fizerem 1 + fizesse 1 + fizessem 1 +	FAZER 270
meses 66	mês 58 + meses 66	MÊS 124

Embora se tenha colocado como critério a análise dos 50 itens lexicais mais frequentes, a lista não lematizada ultrapassa esse número porque há nela mais de um caso de formas flexionadas de um mesmo item lexical (cf., p. ex., *massa* e *massas*).

Após esses procedimentos metodológicos, a lista final ordenada de forma decrescente em número de ocorrências foi a seguinte:

LISTA 1 – Itens lexicais lematizados e reordenados de forma decrescente

1. Destacamento 717	18. Bom 165	35. Atividade 109
2. Guerrilheiro 452	19. Ditadura 162	36. Exército 107
3. Combatente 409	20. Mata 161	37. Notícia 107
4. Dia 370	21. Camponês 152	38. Situação 106
5. Militar 351	22. Ponto 149	39. Movimento 106
6. Inimigo 346	23. Região 148	40. Arma 101
7. Força 337	24. Comissão 142	41. Farinha 98
8. Massa 330	25. Informação 142	42. Parecer 95
9. Luta 285	26. Ano 137	43. Vez 91
10. Fazer 270	27. Soldado 128	44. Político 90
11. Grande 224	28. Trabalho 126	45. Maior 83
12. Povo 211	29. Mês 124	46. Vida 79
13. Poder 211	30. Local 115	47. Encontro 75
14. Área 208	31. Grupo 115	48. Governo 74
15. Acampamento 204	32. Partido 110	49. Propaganda 72
16. Guerrilha 168	33. Rádio 110	50. Apoio 68
17. Armado 167	34. Comandante 110	

O conjunto de dados obtidos permite discutir uma série de questões teórico- metodológicas, o que se faz a seguir.

Uma primeira questão que pode ser tratada são as *classes lexicais* que se fizeram mais presentes. É interessante verificar inicialmente a predominância da classe dos substantivos (40 itens) sobre as demais (7 adjetivos e 3 verbos). Dos adjetivos, 3 tem valor semântico genérico (GRANDE, BOM e MAIOR) e 3 bem específico (GUERRILHEIRO, MILITAR, ARMADO e POLÍTICO), estes últimos nitidamente representativos do contexto em que o diário foi escrito: uma luta *armada* de cunho *político* de forças *guerrilheiras* contra governo *militar*. No caso dos verbos, o mais frequente (FAZER) revela efetivamente a

natureza pragmática da ação guerrilheira, enquanto os dois outros (PODER e PARECER) indicam exercício de avaliação sobre a situação. Dentre os substantivos, verifica-se uma predominância nítida da classe dos concretos (31 itens¹²) frente aos poucos da classe dos abstratos (9 itens¹³): essa predominância dos concretos sobre os abstratos sugere que o principal foco estava mais na manejo efetivo dos componentes das forças guerrilheiras do que na discussão sobre o processo.

Uma segunda questão, que é o cerne da discussão, é o resultado da aplicação da proposta de Hallig e Wartburg (1952) ao diário de Grabois. Primeiramente, foi possível perceber que há uma grande lacuna: dos 50 itens em questão, 17 não têm correspondência no referido sistema¹⁴ (34%). Um primeiro tipo de ausência está relacionado à existência do item como parte do rótulo de uma categoria mas não como item no seu interior: assim, por exemplo, embora haja o rótulo *A política exterior* (*La politique exterieure*), o adjetivo *política* (*politique*) não aparece como item da categoria; o mesmo acontece com o rótulo *As armas e as armaduras* (*Les armes et les armures*), em cuja categoria não consta o item *arma* (*arme*). Parece haver uma confusão entre metalinguagem (nomes dados às categorias) e objeto de estudo (itens pertencentes às categorias). Um segundo tipo de ausência parece ter fundo ideológico: embora haja uma categoria contemplando a defesa pelo Estado (cf. *A defesa nacional* [*La defense nationale*]), não há uma categoria contemplando a defesa contra o Estado, na qual entrariam os itens GUERRILHA, GUERRILHEIRO, COMBATENTE, etc.

Tomando como referência a proposta de Hallig e Wartburg (1952), mas fazendo as inclusões necessárias em função de suas lacunas, é possível organizar os itens da lista 1 da seguinte forma (itens com asterisco não têm correspondente na proposta de Hallig e Wartburg (1952)):

Quadro 1 – Esquema de ordenação de Hallig e Wartburg (1952) aplicado ao diário de Grabois¹⁵

1. O universo: As plantas: As árvores: A floresta e as árvores da floresta: MATA.
2. O homem:
 - 2.1. O homem, ser físico:
 - 2.1.1. A vida humana em geral (nascimento, as fases da vida, a morte): VIDA.
 - 2.1.2. As necessidades do ser humano: A alimentação: Os alimentos: O pão: FARINHA.
 - 2.2. A alma e o intelecto:
 - 2.2.1. Os sentimentos: Os sentimentos relacionados aos outros: Simpatia - antipatia - indiferença: INIMIGO.
 - 2.2.2. A vontade:
 - 2.2.2.1. A ação: FAZER.
 - 2.2.2.1.1. Os princípios:
 - 2.2.2.1.1.1. As aptidões: ATIVIDADE.
 - 2.2.2.1.1.2. As modalidades da ação: PODER.
 - 2.2.2.1.2. A realização: TRABALHO.
 - 2.3. O homem, ser social:
 - 2.3.1. A vida da sociedade em geral:
 - 2.3.1.1. A linguagem: A língua: As ações da voz, a expressão e a comunicação do pensamento: NOTÍCIA, INFORMAÇÃO* e PROPAGANDA*.
 - 2.3.1.2. As relações da sociedade:
 - 2.3.1.2.1. A vida da sociedade: GRUPO.
 - 2.3.1.2.2. A ajuda mútua: APOIO*.
 - 2.3.2. O homem no trabalho:
 - 2.3.2.1. A agricultura, a pecuária, a jardinagem: A fazenda e suas dependências, o rebanho, a pecuária: A fazenda e o rebanho: CAMPONÊS.
 - 2.3.2.2. Correios, telegrafia, telefonia: Telegrafia e telefonia sem fio: RÁDIO.
 - 2.4. Organização social:
 - 2.4.1. O Estado:
 - 2.4.1.1. Os sistemas políticos: DITADURA e POLÍTICO*.
 - 2.4.1.2. As classes sociais: POVO e MASSA*.
 - 2.4.1.3. Os partidos: PARTIDO.
 - 2.4.1.4. O governo, a administração: GOVERNO.
 - 2.4.2. A defesa nacional: FORÇA, MILITAR* e ARMADO*.
 - 2.4.2.1. O exército:
 - 2.4.2.1.1. A subdivisão: EXÉRCITO.
 - 2.4.2.1.2. A hierarquia, o pessoal, os graus: SOLDADO.
 - 2.4.2.1.3. As armas e as armaduras: ARMA*.
 - 2.4.3. A guerra:
 - 2.4.3.1. Generalidades: LUTA, GUERRILHA*, GUERRILHEIRO*, COMISSÃO*, COMANDANTE*, DESTACAMENTO* e COMBATENTE*.
 - 2.4.3.2. A estratégia e a tática: ACAMPAMENTO.
3. O homem e o universo: O a priori:
 - 3.1. A existência: PARECER.
 - 3.2. As qualidades e os estados:
 - 3.2.1. A dimensão: GRANDE e MAIOR*.
 - 3.2.2. A forma: PONTO.
 - 3.3. A relação, a ordem, o valor: O valor: BOM.
 - 3.4. O espaço: ÁREA*, REGIÃO*, LOCAL e SITUAÇÃO.
 - 3.5. O tempo: DIA, MÊS, ANO e VEZ.
 - 3.6. O movimento: MOVIMENTO e ENCONTRO.

mal), o resultado não é de todo satisfatório, porque as categorias disponíveis não revelam as especificidades do léxico analisado. Como caso de afinidade semântica, pode-se citar o item PROPAGANDA (fr. *propagande*), que não aparece no esquema de Hallig e Wartburg (1952), mas certamente é compatível com o item NOTÍCIA (fr. *nouvelle*), contemplado na categoria *As ações da voz, a expressão e a comunicação do pensamento* (cf. categoria 2.3.1.1). Como caso de afinidade semântica e formal (no fundo, cognatos), tem-se COMANDANTE (fr. *commandant*) e COMBATENTE (fr. *combattant*), que, embora ausentes, apresentam afinidade, respectivamente, com os itens em francês *commander* (port. *comandar*) e *combattre* (port. *combater*), presentes na categoria *A guerra: Generalidades* (cf. categoria 2.4.3.1) do esquema de Hallig e Wartburg (1952). Parece óbvio que essa afinidade não é suficiente para a devida integração ao esquema, pois não se trata apenas de generalidades, mas sim de um sistema de organização militar bastante complexo, composto, assim como no caso da categoria *O exército*, de subdivisão (DESTACAMENTO), hierarquia (COMANDANTE), pessoal (COMBATENTE), etc,

No que se refere à questão do fator social como elemento estruturador do léxico (aspecto caro a Matoré), é possível verificar uma tendência de constituição de subsistemas lexicais reservados para cada um dos setores envolvidos no conflito em questão. Trata-se, no entanto, de uma tendência, porque existem itens compartilhados para a nomeação de componentes de ambos os setores. Assim, por um lado, os itens lexicais GUERRILHA, COMISSÃO, DESTACAMENTO, GUERRILHEIRO e COMBATENTE são usados exclusivamente para nomear as forças da resistência; por outro lado, GOVERNO, DITADURA e SOLDADO são usados exclusivamente para nomear as forças da ditadura. Entretanto, os itens FORÇA e MILITAR são empregados para nomear componentes de ambos os setores. Tal compartilhamento soa, à primeira vista, estranho, já que abriria espaço para ambiguidade. No entanto, ampliando a análise para o eixo sintagmático, percebe-se que existem combinações específicas com esses itens.

No que se refere ao item FORÇA¹⁶, a distribuição se dá da seguinte maneira:

Embora seja possível incluir os itens ausentes em função de sua afinidade (seja semântica seja também for-

Quadro 2 – Combinações com FORÇA

Ditadura	Guerrilha
forças armadas da ditadura (4 ocs.)	forças armadas da reação (3 ocs.)
forças armadas do governo (4 ocs.)	forças armadas do Araguaia (1 oc.)
forças armadas governamentais (1 oc.)	força(s) guerrilheira(s) (229 ocs.)
forças da ditadura (5 ocs.)	forças revolucionárias (2 ocs.)
forças da repressão (2 ocs.)	
força(s) repressiva(s) (8 ocs.)	
forças repressoras (2 ocs.)	
força(s) do exército (2 ocs.)	
forças reacionárias (2 ocs.)	
força inimiga (1 oc.)	
forças militares (3 ocs.)	
força militar inimiga (1 oc.)	
forças militares da ditadura (3 ocs.)	
38 ocs.	235 ocs.

Não deixa de ser curioso constatar que, embora o item FORÇA apareça muito menos vezes referindo-se ao setor da ditadura, ocorre com muito mais variantes combinatórias do que em relação ao setor da guerrilha. No que se refere ao item MILITAR, a distribuição se dá da seguinte maneira:

Quadro 3 – Combinações com MILITAR

Ditadura	Guerrilha
Polícia(s) Militar(es) (12 ocs.)	Comissão Militar (134 ocs.)
policiais militares (1 ocs.)	linha militar (12 ocs.)
campanha(s) militar(es) (5 ocs.)	Regulamento Militar (2 ocs.)
os militares (21 ocs.)	ação(ões) militar(es) (11 ocs.)
Escola(s) Militar(es) (2 ocs.)	capacidade militar (12 ocs.)
golpe militar (3 ocs.)	inexperiência militar (5 ocs.)
governos militares (1 ocs.)	tática militar (4 ocs.)
operações militares do inimigo (1 oc.)	êxito(s) militar(es) (3 ocs.)
operações militares da ditadura (1 oc.)	nossas operações militares (4 ocs.)
Comando Militar (4 ocs.)	
51 ocs.	187 ocs.

Chama a atenção, no entanto, que em alguns casos, mesmo com a combinação com um outro elemento, o item MILITAR é usado para ambos os setores, como é o caso da expressão *operação militar*, usada tanto para ações do setor da ditadura quanto da guerrilha, como nos trechos abaixo extraídos do *corpus*, respectivamente:

Depois de 5 meses de investidas inúteis para liquidar o movimento guerrilheiro que se desenvolve no Sul do Pará, as Forças Armadas da ditadura organizaram aparatosa campanha visando a cercar e a aniquilar os combatentes do Araguaia. A *operação militar*, que se iniciou a 14 de setembro e durou mais de 20

dias, mobilizou cerca de 8 mil homens de unidades pertencentes ao Comando Militar da Amazônia e ao Comando Militar do Planalto. (itálico meu)

Os guerrilheiros apreenderam 6 fuzis com 36 balas, 1 sabre, 14 calças, 15 camisas, 4 cinturões, 1 revólver 32 com coldre, 4 redes, 2 lanternas, 4 botinas, 1 relógio e 180 cruzeiros. Depois de realizada a *operação militar*, os soldados, em número de 3, foram aconselhados a abandonarem a Polícia Militar e, em seguida, postos em liberdade, vestidos apenas de calção.

Pode-se dizer, enfim, que haveria uma tendência à *isomorfia*, ou seja, uma tendência de reservar um item ou expressão composta por ele para designar uma realidade diferente da designada por outro item ou expressão composta por ele. No contexto em que o autor do diário se encontrava, a diferenciação entre COMBATENTE (da guerrilha) e SOLDADO (da ditadura) era central demais para permitir compartilhamentos (designação de realidades diferentes com um mesmo termo). Será interessante no futuro investigar se em outros *corpora* a diferenciação nesse campo lexical também se manifesta, para saber se foi o contexto histórico em que o autor se encontrou que o forçou a fazê-la.

Antes de encerrar esta análise, convém discutir um aspecto do ponto de vista metodológico: a *relevância do processo de lematização*. A questão que se coloca é se, na análise lexical, é realmente necessário realizar o processo de lematização ou se o trabalho diretamente com a lista de itens não-lematizados já seria produtivo. Dito de outra forma, a questão é se os itens selecionados na lista não-lematizada diferem dos da lista lematizada. Para essa questão, a resposta é sim. Observando a coluna da esquerda da lista 1, vê-se que, dos 8 itens que foram necessários, além dos 50 primeiros não-lematizados, para permitir uma lista lematizada como 50 (*militares*, *apoio*, *poder*, *guerrilheira*, *maior*, *ano*, *fazer e meses*), 3 já tinham sido contemplados na lista através de uma de suas formas flexionadas (*militares* através de *militar*, *guerrilheira* através de *guerrilheiras* e *ano* atra-

vés de *anos*). Assim, o processo de lematização levou à inclusão de 5 novos itens (*apoio, pode, maior, fazer e meses*), ou seja, 10% do total final¹⁷. Pode-se hipotetizar, no entanto, que, quanto maior o *corpus*, menor seja essa porcentagem de itens incluídos através da lematização. Não se pode testar esta hipótese aqui pelo fato de o *corpus* não ser ampliável, pois trabalhou-se com a totalidade do diário de Graboïs, e não com uma amostragem dele.

Embora a lematização seja bastante trabalhosa levando em conta grandes *corpora*, sua relevância evidencia a necessidade de se desenvolverem métodos automatizados de lematização eficazes a ponto de serem capazes de resolver os problemas complexos desse processo, como é o caso da homonímia.

CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo testar a hipótese de que o esquema de ordenação de Hallig e Wartburg (1952) não é capaz de dar conta do léxico presente no diário de Graboïs, pelo fato de aquele ter sido concebido em um contexto histórico e cultural diverso do em que este o foi. Essa hipótese foi confirmada pelos dados analisados, pois 34% dos itens analisados não apresentaram correspondência no esquema de Hallig e Wartburg (1952), a saber: INFORMAÇÃO, PROPAGANDA, APOIO, POLÍTICO, MASSA, MILITAR, ARMADO, DESTACAMENTO, COMISSÃO, COMANDANTE, COMBATENTE, ARMA, GUERRILHA, GUERRILHEIRO, MAIOR, ÁREA e REGIÃO. Em parte, essas ausências parecem dever-se a questões mais técnicas (presença de um item no rótulo de categoria mas não como membro da categoria), como foi o caso de POLÍTICO e ARMA. Entretanto, a maior parte delas não foi contemplada no referido esquema certamente em função da visão de mundo em que ele se baseia: embora um item GUERRILHA seja atestado em sua forma francesa (*guérilla*) desde 1812¹⁸ e o item MASSA no sentido de “conjunto de pessoas vistas como uma totalidade” o seja desde 1820¹⁹, o esquema de Hallig e Wartburg (1952) não contempla nenhum dos dois.

Não se trata apenas de inserir os itens faltantes no esquema de Hallig e Wartburg (1952), pois as categorias

previstas não apresentam plena compatibilidade com esses itens. Embora haja categorias bem definidas para as forças do Estado (como *O exército*), não há previsão no esquema para as forças de resistência a um Estado ditatorial. Esse problema parece remeter ao fato de que cada cultura de cada época organizaria seu patrimônio lexical de uma forma própria, sendo improvável existir um esquema que seja aplicável ao léxico de qualquer cultura de qualquer época²⁰. Matoré (1953 [1973]), ao valorizar o componente social na estruturação do léxico, certamente estava apontando para isso.

Não é raro encontrar estudos lexicais em que, ao invés de tentar inferir a organização subjacente ao léxico de um dado *corpus*, procurando categorias compatíveis com os dados, se projete a visão do pesquisador sobre os dados. Como exemplo dessa prática projetiva, pode-se citar o estudo de Barreiros e Barreiros (2016) sobre o léxico de um autor baiano na época de ditadura: embora o referido autor tenha usado apenas termos que revelavam que concebia o evento histórico de 1964 como uma revolução (p. ex., *Revolução Brasileira, Revolução Redentora do Brasil, Revolução com R maiúsculo*, etc.), esses itens foram colocados na classificação apresentada sob o rótulo de *Ditadura de 1964*. Não se trata aqui de discutir qual termo seja o mais adequado do ponto de vista histórico ou da ciência política, mas sim de assinalar que os estudos lexicais devem trabalhar com as categorias que expressem a visão de mundo do falante/informante, e não a visão de mundo do pesquisador (mesmo que ela seja a mais correta do ponto de vista histórico).

Como demonstrado na seção anterior, a visão de mundo do falante/informante o leva a organizar o léxico de uma forma própria, em função do contexto em que está inserido. Assim, no caso do diário de Graboïs, itens como COMBATENTE e SOLDADO, que, a princípio poderiam até ser usados como sinônimos em outros contextos, foram usados de forma discreta e excludente em seu diário, ou seja, o primeiro para membro da guerrilha e o segundo para membro da ditadura. Isso significa que um esquema de ordenação do léxico do diário de Graboïs deve refletir essa diferença. O fato de se dever organizar o léxico de um *corpus* segundo a visão

de mundo subjacente a ele não significa, porém, que não deva haver nenhum critério pré-estabelecido, mas, ao contrário, que esses critérios devem ser baseados em categorias mais gerais.

NOTAS

- 1 “Culturonomics is the application of high-throughput data collection and analysis to the study of human culture” (“Culturonomia é a aplicação de coleta e análise de dados de alta produtividade para o estudo da cultura humana” [tradução minha]).
- 2 As abreviaturas desenvolvidas, que constituem basicamente de siglas, foram: C ou com = Comandante; CC = Comandantes; CE = Comissão Executiva; CM = Comissão Militar, co = combatentes; D = Destacamento; DD = Destacamentos; DG = Destacamento de Guarda; EPL = Exército Popular de Libertação; FF AA = Forças Armadas; FF GG = Forças Guerrilheiras; G = Grupamento; GG = Grupamentos; GB = Guanabara; GP = Guerra Popular; LA = Luta Armada; MLP = Movimento de Libertação do Povo; P = Partido; PA = Ponto de Apoio; PP AA = Pontos de Apoio; PTA = Partido do Trabalho Albanês; VC = Vice-Comandante; VVC ou VCC = Vice-Comandantes. Eventualmente as siglas ocorrem seguidas de ponto ou intercaladas com espaço em branco, aspectos que foram observados para o correto desenvolvimento das abreviaturas. Diferentemente do editor do Diário de Grabois, que atribui à sigla C o valor ora de *Comando* ora de *Comandante*, optou-se aqui apenas pela segunda opção, pois é a que compatível com as formas por extenso no próprio diário.
- 3 Disponível em: <http://www.laurenceanthony.net/software/antconcl/>.
- 4 Nunca é demais salientar que se deve atentar para aspectos técnicos em análises com ferramentas computacionais, como, por exemplo, o mesmo tipo de codificação no arquivo de texto e no programa de processamento (no caso, adotou-se Western Latin-1). Além disso, deve-se também atentar para a questão dos homônimos: um caso relevante no *corpus* analisado foi a homonímia entre C que não é abreviatura (como em *Destacamento C*) e C que é abreviatura de *Comandante* (como em *C de Destacamento*), caso que, se negligenciado, levaria à inclusão das formas do primeiro na do segundo.
- 5 Uma versão preliminar desta análise foi divulgada inicialmente na disciplina “Estudos Temáticos de Filologia Românica: Lexicologia Social Aplicada à Realidade Brasileira (Período da Ditadura: 1964-1985)” ministrada na Faculdade de Letras da UFMG no 1º semestre de 2013. Essa análise preliminar foi empregada como modelo para orientar os alunos na análise lexical de textos relacionados ao tema da disciplina.
- 6 Foram relacionadas as lexias vinculadas por flexão de gênero no caso de substantivos e adjetivos e de modo-tempo e número-pessoa no caso de verbos. Formas flexionadas sem ocorrência no *corpus* não foram indicadas na lista.
- 7 Foram excluídas as 2 ocorrências de *mata* como forma flexionada do verbo *matar*, por ser homônimo do substantivo que é o item lexical mais frequente.
- 8 Foram excluídas as 4 ocorrências de *partido* como forma flexionada do verbo *partir*, por ser homônimo do substantivo que é o item lexical mais frequente.
- 9 Foram excluídas as 2 ocorrências de *encontro* como forma flexionada do verbo *encontrar*, por ser homônimo do substantivo, que é o item lexical mais frequente.
- 10 Foram excluídas as 18 ocorrências de *política* como substantivo, por

ser homônimo do adjetivo, que é o item lexical mais frequente.

- 11 Foram excluídas as 14 ocorrências de *poder* como substantivo, por ser homônimo da forma infinitiva do verbo *poder*, que é o item lexical mais frequente.
- 12 Substantivos concretos: DESTACAMENTO, COMBATENTE, DIA, INIMIGO, FORÇA, MASSA, POVO, ÁREA, ACAMPAMENTO, DITADURA, MATA, CAM-PONÊS, PONTO, REGIÃO, COMISSÃO, INFORMAÇÃO, ANO, SOLDADO, MÊS, LOCAL, GRUPO, PARTIDO, RÁDIO, COMANDANTE, EXÉRCITO, NOTÍCIA, ARMA, FARINHA, VIDA, GOVERNO e PROPAGANDA.
- 13 Substantivos abstratos: LUTA, GUERRILHA, TRABALHO, ATIVIDADE, SITUAÇÃO, MOVIMENTO, VEZ, ENCONTRO e APOIO.
- 14 Como o sistema de Hallig e Wartburg (1952) está em francês, foi necessário traduzir para o francês os 50 itens em análise. As formas traduzidas buscadas foram (na ordem da lista 1): *corps d'armée**, *guérillero**, *combattant**, *jour*, *militaire**, *ennemi*, *force*, *masse**, *combat*, *faire*, *grand*, *peuple*, *pouvoir*, *zone**, *camp*, *guerrilla**, *armé**, *bon*, *dictature*, *bois*, *paysan*, *point*, *region**, *commission**, *information**, *an*, *soldat*, *travail*, *mois*, *lieu*, *groupe*, *parti*, *radio*, *commandant**, *activité*, *armée*, *nouvelle*, *situation*, *mouvement*, *arme**, *farine*, *paraître*, *fois*, *politique**, *plus grand**, *vie*, *rencontre*, *gouvernement*, *propagande** e *soutien**. Os itens com asterisco não foram encontrados na lista de Hallig e Wartburg (1952).
- 15 Em caracteres redondos estão os rótulos das categoriais do esquema de Hallig e Wartburg (1952) e em versalete estão os itens lexicais do diário de Grabois. Fez-se eliminação e renumeração dos indexadores originais para tornar simples a visualização das categorias.
- 16 Hallig e Wartburg (1952) colocam FORÇA (*force*) na categoria das aptidões (cf. categoria 2.2.2.1.1.1), mas o seu uso no *corpus* é predominantemente no sentido de unidade de defesa, por isso foi situado na categoria A *defesa nacional*.
- 17 As demais ocorrências do item FORÇA se dão de forma genérica ou em sentido da aptidão, por isso não constam deste quadro.
- 18 A relevância da lematização foi igualmente defendida por Cambraia (2015), que demonstrou que a lematização torna mais evidente a percepção da diferença lexical entre diferentes *corpora*.
- 19 J. DE MAISTRE, *Relation pour S.M. le Roi Victor-Emmanuel ds Corresp.*, t. 4, p. 282: “Ces paysans [...] changés en véritables **guérillas** et ne sachant plus que tuer, redeviendront-ils des serfs dociles?” (TFLi, 2018).
- 20 MAINE DE BIRAN, *Journal*, 1820, p.274: “Le parti libéral ou anti-monarchique veut absolument maintenir la loi d'élection qui nous perd et repousse la loi nouvelle avec des menaces, des tentatives d'insurrection populaire (...). Cependant la **masse** est tranquille et se refuse aux révolutions.” (TFLi, 2018).
- 21 Assim, por exemplo, as categorias e os itens relacionados a religião no esquema de Hallig e Wartburg (1952) não foram suficientes para a descrição do léxico cristão medieval (CAMBRAIA; VILAÇA; MELO, 2013) nem do léxico religioso católico, evangélico e adventista moderno (WOLFF, 2016).

REFERÊNCIAS BIBLIOGRÁFICAS

- BARREIROS, P. N.; BARREIROS, L. L. S. O vocabulário da Ditadura Militar nos panfletos de Eulálio Motta. **Filologia e Linguística Portuguesa**, São Paulo, v. 17, p. 385, 2016.
- BIDERMAN, M. T. C. A estrutura mental do léxico. In: QUEIROZ, T. A. (Ed.). **Estudos de filologia e lingüística**. São Paulo: Editora da Universidade de São Paulo, 1981.

BIDERMAN, M. T. C. A face quantitativa da linguagem: um dicionário de frequências do português. *Alfa*, São Paulo, v. 42, p. 157-181, 1998.

CAMBRAIA, C. N. Da lexicologia social a uma lexicologia sócio-histórica: caminhos possíveis. *Revista de Estudos de Linguagem*, Belo Horizonte, v. 21, n. 1, p. 157-188, 2013. Disponível em: <<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/5096/4553>>.

CAMBRAIA, C. N. Lexicologia e informação: um ensaio de quantificação. *Entretextos*, Londrina, v. 15, n. 2, p. 31-52, 2015. Disponível em: <<http://www.uel.br/revistas/uel/index.php/entretextos/article/view/17370/19570>>

CAMBRAIA, C. N.; VILAÇA, C. de L.; MELO, T. C. A. de. Unidade lexical e unidade cultural: o léxico românico de religião em traduções medievais. In-*Traduções*, Florianópolis, v. 5, n. 9, p.22-39, 2013. Disponível em: <<http://incubadora.periodicos.ufsc.br/index.php/intraducoes/article/download/2557/3252>>

COSERIU, E. *Princípios de semântica estrutural*. Madrid: Gredos, 1977.

DUNCAN, J. *Frequency dictionary of portuguese*. 1972. PhD Dissertation – Stanford, Stanford University.

GRABOIS, Maurício. *Diário de Maurício Grabois (Velho Mário) – 30 de abril de 1972 a 25 de dezembro de 1973*. Disponível em: http://www.grabois.org.br/uploads/arquivos/arquivo_31_12846.pdf. Acesso em 04 ago 2018.

HALLIG, R.; WARTBURG, W. von. *Begriffssystemals Grundlage für die Lexikographie*: Versuch eines Ordnungsschemas. Berlin: AkademieVerlag, 1952.

MATORÉ, G. *La méthode en lexicologie*: domaine français. Nouv. éd. Paris: Didier, 1973.

MICHEL, J.-B. et al. Quantitative analysis of culture using millions of digitized books. *Science*, v. 331, n. 6014, p. 176-182, 2011.

REDAÇÃO. Exclusivo: O diário do Araguaia, *Carta Capital*, São Paulo, 2011. Disponível em: <https://www.cartacapital.com.br/politica/exclusivo-o-diario-do-araguaia>. Acesso em: 07 set. 2018.

SAUSSURE, F. de. *Curso de linguística geral*. 14. ed. São Paulo: Cultrix, 1988.

TLFi [Trésor de la langue Française informatisé]. ATILF - CNRS / Université de Lorraine. Disponível em: <<http://www.atilf.fr/tlfi>>. Acesso em: 07 set. 2018.

TRIER, J. *Der deutsche Wortschatz im Sinnbezirk des Verstandes*: die Geschichte eines sprachlichen Feldes. Heidelberg: C. Winter, 1931.

WOLFE, D. da C. M. *Léxico do discurso religioso*: um estudo comparado. 2016. Tese (Doutorado em Estudos Linguísticos) – Universidade Federal de Minas Gerais.

O AUTOR

César Nardelli Cambraia - Possui graduação em Letras (Português-Alemão) pela Universidade Federal de Minas Gerais (1992), mestrado em Estudos Linguísticos pela Universidade Federal de Minas Gerais (1996), doutorado em Filologia e Língua Portuguesa pela Universidade de São Paulo (2000) e pós-doutorado em Linguística Românica pela Universitat de Barcelona (2010). Atualmente é professor titular de Filologia Românica na Faculdade de Letras da Universidade Federal de Minas Gerais e tem experiência na área de Linguística, com ênfase em Linguística Românica e Crítica Textual, atuando principalmente nos seguintes temas: estudo histórico e comparado de morfossintaxe de línguas românicas em uma perspectiva tipológico-funcional, lexicologia sócio histórica e edição de textos românicos antigos.

