

LAURA CAMPOS DE BORBA

**DICIONÁRIOS DE APRENDIZES:
UMA ABORDAGEM INTERDISCIPLINAR PARA O SEU APERFEIÇOAMENTO**

**PORTO ALEGRE
2022**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ÁREA: ESTUDOS DA LINGUAGEM
ESPECIALIDADE: TEORIAS LINGUÍSTICAS DO LÉXICO
LINHA DE PESQUISA: LEXICOGRAFIA, TERMINOLOGIA E TRADUÇÃO:
RELAÇÕES TEXTUAIS

DICIONÁRIOS DE APRENDIZES:

UMA ABORDAGEM INTERDISCIPLINAR PARA O SEU APERFEIÇOAMENTO

LAURA CAMPOS DE BORBA

Tese de Doutorado em Estudos da Linguagem
(Lexicografia), apresentada como requisito
parcial para a obtenção do título de Doutora pelo
Programa de Pós-Graduação em Letras da
Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Félix Valentín Bugueño
Miranda

PORTO ALEGRE
2022

CIP - Catalogação na Publicação

Borba, Laura Campos de Borba
Dicionários de aprendizes: uma abordagem
interdisciplinar para o seu aperfeiçoamento / Laura
Campos de Borba Borba. -- 2022.
409 f.
Orientador: Félix Valentín Bugeño Miranda Bugeño
Miranda.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Letras, Programa de
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2022.

1. Lexicografia. 2. Dicionários de aprendizes. 3.
Ensino de inglês como língua estrangeira. 4. Ensino de
espanhol como língua estrangeira. 5. Didatismo. I.
Bugeño Miranda, Félix Valentín Bugeño Miranda,
orient. II. Título.

Laura Campos de Borba

DICIONÁRIOS DE APRENDIZES:

UMA ABORDAGEM INTERDISCIPLINAR PARA O SEU APERFEIÇOAMENTO

Tese de Doutorado em Estudos da Linguagem (Lexicografia), apresentada como requisito parcial para a obtenção do título de Doutora pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Porto Alegre, 24 de janeiro de 2022.

Resultado: Aprovada.

BANCA EXAMINADORA:

Prof. Dr. Félix Valentín Bugueño Miranda (orientador)

Universidade Federal do Rio Grande do Sul (UFRGS)

Profa. Dra. Ana Beatriz Arêas da Luz Fontes

Universidade Federal do Rio Grande do Sul (UFRGS)

Profa. Dra. Adja Balbino de Amorim Barbieri Durão

Universidade Federal de Santa Catarina (UFSC)

Profa. Dra. Cláudia Zavaglia

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)

AGRADECIMENTOS

Às pessoas:

– “El doctorado es un largo viaje consigo mismo”. Com essas palavras, meu orientador me deu as boas-vindas ao doutorado. Sem dúvidas, a viagem foi longa, alegre em alguns momentos e amarga em outros; mas, acima de tudo, foi um período de muita aprendizagem, como deve ser. O prof. Félix contribuiu de maneira substancial para o meu desenvolvimento profissional durante o doutorado e no restante da minha trajetória acadêmica, e eu tenho muito a lhe agradecer por isso. Obrigada, professor, pelo acompanhamento, apoio e incentivo durante todos esses anos. Agora cabe a mim fazer jus à formação de excelência que tive.

Agradeço aos meus familiares pela torcida para que o doutorado tivesse êxito.

Às instituições:

Agradeço à CAPES pela bolsa concedida, sem a qual não teria sido possível me dedicar a esta pesquisa.

Além disso, agradeço e cumpro com o dever de mencionar, em função dos Termos de uso da Cambridge University Press, que a tese utilizou as seguintes ferramentas:

a) Cambridge Learner Corpus (CLC). Trata-se de um subcorpus do Cambridge English Corpus. O acesso a esta ferramenta ocorreu mediante licença temporária concedida pela Cambridge University Press. Todos os dados extraídos do CLC estão devidamente referenciados.

b) English Vocabulary Profile e English Grammar Profile. “Estes recursos são baseados em ampla pesquisa usando o Cambridge Learner Corpus e formam parte do programa English Profile, o qual almeja fornecer evidências sobre o uso da língua sobre o uso da língua que ajude a produzir melhores materiais didáticos para ensino de língua. Veja <http://www.englishprofile.org> para maiores informações”ⁱ.

Finalmente, também agradeço e cumpro com o dever de mencionar que a tese fez uso das informações do Corpus de Aprendices de Español como Lengua Extranjera, do Instituto Cervantes, e do Fórum Português-Espanhol do *site* WordReference.

ⁱ [This thesis has made use of the English Vocabulary Profile and the English Grammar Profile. These resources are based on extensive research using the Cambridge Learner Corpus and are part of the English Profile programme, which aims to provide evidence about language use that helps to produce better language teaching materials. See <http://www.englishprofile.org> for more information].

RESUMO

Embora existam dicionários específicos para o ensino-aprendizagem de inglês e de espanhol como língua estrangeira, essas obras são compiladas sem que se tenha um perfil de usuário bem delimitado. Consequentemente, fatores importantes atinentes ao usuário, tais como o seu conhecimento prévio e as suas necessidades de aprendizagem – dentre as quais estão as combinatórias léxicas – raramente são levados em conta. Essa limitação termina distanciando o dicionário de aprendizes do contexto de ensino-aprendizagem no qual está inserido o seu consulente, prejudicando o didatismo da obra lexicográfica. O objetivo da tese foi explorar e aplicar princípios teórico-metodológicos para um aprimoramento do desenho de dicionários de aprendizes de língua inglesa e de língua espanhol destinados aos níveis A2 e B1 de aprendizagem. O aprimoramento a que se refere a tese diz respeito à determinação de um perfil de usuário e à descrição de combinatórias léxicas. Trata-se de uma pesquisa de cunho eminentemente teórico, na qual se analisam diferentes perspectivas teóricas que servem de base para a elaboração de propostas de descrição lexicográfica. Nesse sentido, foi adotada uma abordagem interdisciplinar, de forma a buscar subsídios não somente na Metalexigrafia, mas também em outras áreas: Linguística Aplicada, Ciência Cognitiva, Linguística Contrastiva e Lexicologia. A Linguística Aplicada, a Ciência Cognitiva e a Linguística Contrastiva tiveram um papel fundamental no aprofundamento das variáveis Metalexigráficas que definem o perfil de usuário de um dicionário de aprendizes. A Metalexigrafia, por sua vez, forneceu subsídios relativos à constituição de segmentos informativos de um dicionário de aprendizes. A Lexicologia, por fim, forneceu subsídios relativos às combinatórias léxicas. Como resultado, foi possível elaborar uma série de orientações e propostas a respeito da determinação do perfil de usuário, da redação de paráfrases explanatórias e notas de uso e da descrição de combinatórias léxicas. A principal conclusão da tese é a de que as limitações teóricas da Metalexigrafia podem ser superadas com o auxílio de outras áreas do conhecimento. A aplicação das propostas feitas na tese depende, no entanto, da sua testagem em massa com mais itens e de pesquisas empíricas junto ao seu público-alvo.

Palavras-chave: Dicionários de aprendizes. Ensino de Inglês como Língua Estrangeira. Ensino de Espanhol como Língua Estrangeira. Didatismo.

RESUMEN

Aunque existen diccionarios específicos para la enseñanza y aprendizaje de inglés y español como lengua extranjera, estos diccionarios se compilan sin que se establezca un perfil de usuario bien definido. En consecuencia, factores importantes relativos al usuario, como su conocimiento previo y sus necesidades de aprendizaje – entre ellas las combinatorias léxicas – rara vez se tienen en cuenta. Esta limitación acaba por alejar el diccionario de aprendices del contexto de enseñanza-aprendizaje en el que se inserta su público, perjudicando el carácter didáctico de la obra lexicográfica. El objetivo de la tesis fue explorar y aplicar principios teórico-metodológicos para mejorar el diseño de diccionarios de aprendices de inglés y español orientados a los niveles de aprendizaje A2 y B1. La mejoría a que se refiere la tesis se refiere a la determinación de un perfil de usuario y a la descripción de combinaciones léxicas. Se trata de una investigación eminentemente teórica, en la que se analizan diferentes perspectivas teóricas que sirven de base para la elaboración de propuestas para la descripción lexicográfica. En este sentido, se adoptó un enfoque interdisciplinario con el fin de buscar respuestas no solo en Metalexigrafía, sino también en otras áreas: Lingüística Aplicada, Ciencias Cognitivas, Lingüística Contrastiva y Lexicología. La Lingüística Aplicada, las Ciencias Cognitivas y la Lingüística Contrastiva desempeñaron un papel fundamental en la profundización de las variables metalexigráficas que definen el perfil de usuario de un diccionario de aprendices. La Metalexigrafía, a su vez, forneció orientaciones acerca de la constitución de los segmentos informativos de un diccionario de aprendices. Finalmente, la Lexicología aclaró cuestiones relativas a las combinatorias léxicas. Como resultado, fue posible elaborar una serie de orientaciones y propuestas en cuanto a la determinación del perfil de usuario, a la redacción de paráfrasis explanatorias y notas de uso, y la descripción de combinatorias léxicas. La principal conclusión de la tesis es que las limitaciones teóricas de la Metalexigrafía se pueden superar con la ayuda de otras áreas del conocimiento. La aplicación de las propuestas de la tesis depende, sin embargo, de su testeado masivo con más ítems y de la investigación empírica junto al público a quien se dirigen.

Palabras-clave: Diccionarios de aprendices. Enseñanza de Inglés como Lengua Extranjera. Enseñanza de Español como Lengua Extranjera. Didactismo.

ABSTRACT

Although there are specific dictionaries for teaching and learning English and Spanish as a foreign language, these dictionaries are compiled without a well-defined user profile. Consequently, important factors concerning the user, such as their prior knowledge and their learning needs – among which are lexical combinations – are rarely taken into account. This limitation ends up distancing the learner's dictionary from the teaching-learning context of its user, hindering the didactic nature of the lexicographical work. This thesis aimed to explore and apply theoretical-methodological principles to improve the design of EFL and SFL dictionaries to learning levels A2 and B1. The improvement concerns the determination of a user profile and the description of lexical combinations. This is an eminently theoretical research, in which we analyse different theoretical perspectives that serve as a basis for proposals for a lexicographical description. In this sense, an interdisciplinary approach was adopted, in order to seek support not only in Metalexigraphy, but also in other areas: Applied Linguistics, Cognitive Science, Contrastive Linguistics and Lexicology. Applied Linguistics, Cognitive Science and Contrastive Linguistics played a fundamental role in deepening the Metalexigraphical variables that define the user profile of a learner's dictionary. Metalexigraphy, in turn, provided theoretical support related to informative segments of a learner's dictionary. Finally, Lexicology provided theoretical support about lexical combinations. As a result, it was possible to draw up a series of guidelines and proposals regarding the determination of the user profile, the writing of definitions and usage notes, and the description of lexical combinations. The main conclusion of the thesis is that the theoretical limitations of Metalexigraphy can be overcome with the help of other areas of knowledge. To apply the proposals of the thesis it is necessary, however, to conduct a mass testing with more items and empirical research with the user profile.

Keywords: Learner's dictionaries. EFL. SFL. User-friendliness.

LISTA DE FIGURAS

Figura 1 – Informação discriminante em um dicionário de aprendizes	35
Figura 2 – Reprodução parcial do verbete <i>fold</i> (CALD4)	47
Figura 3 – Reprodução do verbete <i>fold</i> (LWord2)	48
Figura 4 – Tipologia de Tarp (2011)	60
Figura 5 – Classificação dos dicionários pedagógicos	61
Figura 6 – Modalidades de orientação linguística	242
Figura 7 – Resultados do CLC (2019) na busca por tipos de erros	256
Figura 8 – Posições sintáticas de <i>already</i>	275
Figura 9 – Verbetes <i>decisión</i> do Señas (2001)	300
Figura 10 – Tipologia de Hausmann (2007)	307
Figura 11 – Tipologia de Sánchez Rufat (2010)	313
Figura 12 – Tipologia de Mel'čuk (2012, 2015)	319

LISTA DE QUADROS

Quadro 1 – Indicação da frequência de uso de unidades léxicas.....	33
Quadro 2 – Concepções sobre o usuário nas informações incluídas nos dicionários.....	39
Quadro 3 – Reprodução parcial do verbete <i>trabajo</i> (Señas, 2001)	48
Quadro 4 – Falta de auxílio na função de recepção.....	66
Quadro 5 – Compreensão leitora nos níveis A1 e A2	97
Quadro 6 – Vocabulário nos níveis A1 e A2.....	98
Quadro 7 – Gramática nos níveis A1 e A2	100
Quadro 8 – Competência linguística geral nos níveis A2 e B1	101
Quadro 9 – Amplitude de vocabulário nos níveis A2 e B1	103
Quadro 10 – Domínio de vocabulário nos níveis A2 e B1	104
Quadro 11 – Gramática nos níveis A2 e B1	104
Quadro 12 – Dimensão fonológica nos níveis A2 e B1	105
Quadro 13 – Comparativo entre o PCIC (2006, B1) e o Threshold 1990 (1998)	109
Quadro 14 – Unidades léxicas analisadas nos dicionários de aprendizes de inglês	162
Quadro 15 – Quantidade de itens analisados por dicionário e por nível CEFR (2001) (inglês)	163
Quadro 16 – ULs tratadas como vocabulário passivo (inglês)	165
Quadro 17 – Análise do nível do vocabulário de definidores (inglês)	165
Quadro 18 – Padrões gramaticais tratados como vocabulário passivo (inglês)	168
Quadro 19 – Análise do nível dos padrões gramaticais (inglês)	168
Quadro 20 – Facilidade de compreensão leitora: definidores e padrões gramaticais (inglês)	170
Quadro 21 – Informação de origem etimológica no OALDe (2010)	173
Quadro 22 – Tipos de romanismos.....	175
Quadro 23 – Classificação dos tipos de romanismos (inglês).....	176
Quadro 24 – Facilidade de compreensão sob a perspectiva do consulente brasileiro (inglês)	180
Quadro 25 – Ranqueamento interno de cada categoria de facilidade de compreensão (inglês)	182
Quadro 26 – Tipos de similaridade (espanhol).....	188
Quadro 27 – Unidades léxicas analisadas no Señas (2001).....	188
Quadro 28 – Facilidade de compreensão leitora: definidores e padrões gramaticais (espanhol)	190
Quadro 29 – Definidores difíceis de compreender para os brasileiros com proficiência A1-A2	195
Quadro 30 – Facilidade de compreensão sob a perspectiva do consulente brasileiro (espanhol)	196
Quadro 31 – Propostas ao nível do lema	203
Quadro 32 – Dados do Corpes (2021) sobre <i>folclor</i> , <i>folclore</i> e <i>folklore</i>	204
Quadro 33 – Propostas ao nível da representação fonológica	205
Quadro 34 – Similaridade ortográfica entre termos gramaticais.....	207
Quadro 35 – Resumo das propostas ao comentário de forma	208
Quadro 36 – Classes de palavras, tipo de paráfrases e padrões sintáticos.....	211
Quadro 37 – Propostas de aprimoramento das paráfrases explanatórias (inglês): substantivos	216
Quadro 38 – Propostas de aprimoramento das paráfrases explanatórias (inglês): verbos	217
Quadro 39 – Propostas de aprimoramento das paráfrases explanatórias (inglês): adjetivos	218
Quadro 40 – Propostas de aprimoramento das paráfrases explanatórias (inglês): preposições e advérbios	219

Quadro 41 – Padrões sintáticos de Farias (2013) adaptados ao espanhol	222
Quadro 42 – Aprimoramento das paráfrases explanatórias de ULs A2 e B1 do Señas (2001)	224
Quadro 43 – Paráfrases explanatórias do Señas (2001) difíceis de adaptar	227
Quadro 44 – Função das notas de uso	235
Quadro 45 – Notas de uso com informações extraordinárias	237
Quadro 46 – Etiquetas das notas de uso do OxD2 (2012) e do LWord2 (2008)	240
Quadro 47 – Propostas para o aspecto formal das notas de uso	246
Quadro 48 – Notas de uso em dicionários de aprendizes de inglês	248
Quadro 49 – Propostas de aprimoramento das notas de uso (inglês)	249
Quadro 50 – Notas de uso sobre a ortografia de <i>comfortable</i> e <i>address</i> nos dicionários	264
Quadro 51 – Adaptação das notas de uso de <i>comfortable</i> e <i>address</i>	265
Quadro 52 – Proposta de nota de uso em <i>house</i> para o OxD2 (2012)	270
Quadro 53 – Proposta de nota de uso em <i>dear</i> para o CaED2 (2011)	272
Quadro 54 – Proposta de nota de uso em <i>already</i> para o LWord2e (2008)	276
Quadro 55 – Proposta de nota de uso em <i>generous</i> para o CoED2 (2014)	278
Quadro 56 – Proposta de nota de uso em <i>appointment</i> para o CaED2 (2011)	279
Quadro 57 – Erros de aprendizes brasileiros de espanhol no CAES (2020)	283
Quadro 58 – Dúvidas de aprendizes brasileiros de espanhol no WordReference	285
Quadro 59 – Excerto do NDiFAEP (2014)	286
Quadro 60 – Propostas de aprimoramento das notas de uso (espanhol)	287
Quadro 61 – Distinção de acepções em verbetes polissêmicos nos dicionários de inglês	295
Quadro 62 – Descrição de colocações nos dicionários de aprendizes de inglês	299
Quadro 63 – Exemplos de coligações em dicionários	302
Quadro 64 – Coligações em dicionários de aprendizes de inglês	303
Quadro 65 – Coligações no Señas (2001)	305
Quadro 66 – FLs qualificativas	326
Quadro 67 – FLs de localização	327
Quadro 68 – FLs de verbos suporte, copulativos e de realização	328
Quadro 69 – FLs de processo	330
Quadro 70 – FLs variadas	331
Quadro 71 – Proposta de descrição de colocações gramaticais e coligações	338
Quadro 72 – Proposta de descrição de restrições de seleção	340
Quadro 73 – Proposta de descrição secundária de fraseologias	342
Quadro 74 – Proposta de adaptação das FLs	344
Quadro 75 – Proposta de descrição de colocações	345

LISTA DE TABELAS

Tabela 1 – Perfil resumido dos participantes de Wartenburger et al. (2003)	152
Tabela 2 – Tipos de similaridade (espanhol)	194
Tabela 3 – Dados gerais e do Brasil no CLC (2019)	251
Tabela 4 – Dados de brasileiros adultos A1-C2 no CLC (2019)	252
Tabela 5 – Estatísticas de erros em sinais gráficos no CLC (2019)	253
Tabela 6 – Dez erros mais frequentes de brasileiros adultos A1-B1 no CLC (2019)	257
Tabela 7 – Dez erros ortográficos mais frequentes dos brasileiros adultos A1-B1 no CLC (2019)	260
Tabela 8 – Análise relativa dos dez erros ortográficos mais frequentes	263
Tabela 9 – Incidência de erros sobre itens A2-B1	268

APOIO DE FINANCIAMENTO CAPES

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código 001.

SUMÁRIO

INTRODUÇÃO.....	18
A sociedade da informação e a oferta de informações	19
A sociedade da informação e a demanda por informações de qualidade na internet.....	20
Relevância da aprendizagem de inglês e espanhol para os brasileiros	22
1 O ESCOPO DESTA TESE DE DOUTORADO	27
1.1 O compromisso com o fornecimento de informações de qualidade de maneira didática.....	29
1.1.1 <i>Informações discretas em dicionários de aprendizes</i>	32
1.1.2 <i>Informações discriminantes em dicionários de aprendizes.....</i>	34
1.1.3 <i>Perfil de usuário em dicionários de aprendizes</i>	36
1.1.4 <i>Apresentação de informações de maneira didática.....</i>	44
1.2 Alternativas para aprimorar os dicionários de aprendizes	50
1.3 Objetivos.....	52
1.3.1 <i>Objetivo geral</i>	53
1.3.2 <i>Objetivos específicos</i>	53
1.4 Pressupostos de pesquisa.....	53
1.5 Estrutura geral do trabalho	54
2 O PERFIL DE USUÁRIO SOB A ÓTICA DA METALEXICOGRAFIA	57
2.1 A classificação dos dicionários de aprendizes e o estabelecimento do perfil de usuário.....	57
2.2 A função dos dicionários de aprendizes e o estabelecimento do perfil de usuário.....	64
2.3 Usuário: o ilustre (e convenientemente) desconhecido.....	67
2.3.1 <i>Propostas teóricas sobre o usuário</i>	67
2.3.2 <i>Pesquisas empíricas sobre o uso de dicionários</i>	71
2.3.2.1 <i>Pesquisas empíricas diretas.....</i>	72
2.3.2.2 <i>Pesquisas empíricas indiretas</i>	76
2.3.3 <i>Comentário geral sobre as contribuições de pesquisas teóricas e empíricas.....</i>	77
2.3.4 <i>Influências mercadológicas no estabelecimento do perfil de usuário: uma variável indesejada</i>	80
2.4 O perfil de usuário proposto para esta tese.....	82
3 O PERFIL DE USUÁRIO SOB A ÓTICA DA LINGUÍSTICA APLICADA.....	85
3.1 Fronteiras conceituais entre língua estrangeira, segunda língua e língua adicional.....	87
3.2 Práticas e objetivos de ensino e a concepção do perfil de usuário.....	92
3.3 Nível de proficiência e a definição do perfil de usuário	95
3.4 Necessidades de aprendizagem nos níveis A2 e B1	101
3.5 Documentos de referência para o ensino de inglês e espanhol como língua estrangeira	106
3.5.1 <i>Descrição de níveis de referência do espanhol: Plan Curricular del Instituto Cervantes (2006)</i>	108
3.5.2 <i>Descrição de níveis de referência do inglês: English Profile (2011).....</i>	110
3.6 Corolário do capítulo 3.....	114

4	O PERFIL DE USUÁRIO SOB A ÓTICA DA CIÊNCIA COGNITIVA	120
4.1	A participação da língua materna na aprendizagem da segunda língua em adultos	123
4.2	Processamento de segunda língua: duas línguas em interação	126
4.2.1	<i>Distinções terminológicas: homógrafos e cognatos na perspectiva psicolinguística</i>	132
4.2.2	<i>Psicolinguística e Linguística Contrastiva na determinação do perfil de usuário</i>	134
4.3	Processamento de L2 em bilíngues com baixa proficiência na L2	145
4.3.1	<i>Processamento fonológico, morfossintático e semântico na L2</i>	148
4.3.1.1	Processamento fonológico da L2	149
4.3.1.2	Processamento morfossintático da L2	151
4.3.1.3	Processamento semântico da L2	154
4.4	Corolário do capítulo 4	156
5	ANÁLISE DOS DICIONÁRIOS DE APRENDIZES À LUZ DO PERFIL DE USUÁRIO	159
5.1	Dicionários de aprendizes de inglês e o público brasileiro	160
5.1.1	<i>Conhecimento prévio geral na compreensão leitora dos dicionários de inglês</i>	164
5.1.1.1	Análise do nível dos definidores	164
5.1.1.2	Análise do nível dos padrões gramaticais	167
5.1.1.3	Corolário das análises léxica e gramatical	169
5.1.2	<i>Influências interlinguísticas na compreensão leitora dos dicionários de inglês</i>	171
5.1.3	<i>Facilidade de compreensão leitora dos dicionários de aprendizes de inglês A1-B1</i>	179
5.2	Dicionários de aprendizes de espanhol e o público brasileiro	184
5.2.1	<i>Conhecimento prévio geral na compreensão leitora do Señas (2001)</i>	189
5.2.2	<i>Influências interlinguísticas na compreensão leitora do Señas (2001)</i>	193
5.2.3	<i>Facilidade de compreensão leitora do Señas (2001)</i>	195
6	O PERFIL DE USUÁRIO APLICADO AOS DICIONÁRIOS DE APRENDIZES (I)	198
6.1	Macroestrutura dos dicionários de aprendizes	199
6.2	Medioestrutura dos dicionários de aprendizes	200
6.3	Microestrutura dos dicionários de aprendizes: comentário de forma	201
6.3.1	<i>Lema</i>	201
6.3.2	<i>Representação fonológica</i>	205
6.3.3	<i>Categoria morfológica</i>	206
6.4	Microestrutura dos dicionários de aprendizes: paráfrases explanatórias	209
6.5	Aprimorando as paráfrases explanatórias dos dicionários de aprendizes de inglês	214
6.6	Aprimorando as paráfrases explanatórias dos dicionários de aprendizes de espanhol	221
7	O PERFIL DE USUÁRIO APLICADO AOS DICIONÁRIOS DE APRENDIZES (II)	233
7.1	As notas de uso sob a ótica da teoria e da práxis lexicográfica	234
7.1.1	<i>A função das notas de uso</i>	235
7.1.2	<i>O aspecto formal das notas de uso</i>	238
7.1.2.1	Recursos (ícono)gráficos	239
7.1.2.2	Etiquetas	239
7.1.2.3	Língua utilizada na nota	241
7.1.2.4	Modalidade de orientação e padrão sintático	242

7.1.2.5 Recursos tipográficos.....	245
7.2 Notas de uso para dicionários de aprendizes brasileiros de inglês.....	247
7.2.1 <i>Adaptação das notas de uso já existentes com vistas ao público brasileiro</i>	248
7.2.2 <i>O Cambridge Learner Corpus (2019)</i>	250
7.2.3 <i>Seleção de conteúdo para notas de uso: buscas por tipos de erros no CLC (2019) ..</i>	255
7.2.4 <i>Seleção de conteúdo para notas de uso: buscas por itens específicos no CLC (2019)</i>	265
7.3 Notas de uso para dicionários de aprendizes brasileiros de espanhol	281
7.3.1 <i>Seleção de conteúdo para notas de uso: CAES (2020)</i>	282
7.3.2 <i>Seleção de conteúdo para notas de uso: Fórum Português-Espanhol (WR, 2020) ...</i>	285
7.3.3 <i>Seleção de conteúdo para notas de uso: NDiFAEP (2014).....</i>	286
7.3.4 <i>Esboço de notas de uso para adaptação do Señas (2001)</i>	287
8 AS COMBINATÓRIAS LÉXICAS NO DICIONÁRIO DE APRENDIZES.....	291
8.1 Breve panorama dos estudos sobre combinatórias léxicas	292
8.1.1 <i>Viés estatístico sobre as combinatórias léxicas: Sinclair (1991, 2004) e Hoey (2005)</i>	293
8.1.2 <i>Viés lexicológico sobre as combinatórias léxicas I: Hausmann (2007 [1984]).....</i>	307
8.1.3 <i>Viés gramatical sobre as combinatórias léxicas: Bosque (2001) e Sánchez Ruffat (2010)</i>	311
8.1.4 <i>Viés lexicológico sobre as combinatórias léxicas II: Mel'čuk (2015).....</i>	316
8.1.4.1 <i>Fraseologias [idioms].....</i>	319
8.1.4.2 <i>Colocações</i>	323
8.1.4.2.1 <i>FLs sintagmáticas qualificativas.....</i>	325
8.1.4.2.2 <i>FLs sintagmáticas de localização</i>	327
8.1.4.2.3 <i>FLs sintagmáticas de verbos suporte, copulativos e de realização.....</i>	328
8.1.4.2.4 <i>FLs sintagmáticas de processo</i>	330
8.1.4.2.5 <i>FLs sintagmáticas variadas</i>	331
8.1.4.3 <i>Clichês</i>	333
8.2 Proposta para a descrição de combinatórias léxicas em dicionários de aprendizes	336
8.2.1 <i>Propostas para a descrição de ligações e colocações gramaticais (contextualismo inglês)</i>	337
8.2.2 <i>Propostas para a descrição de restrições de seleção (BOSQUE, 2001)</i>	340
8.2.3 <i>Propostas para a descrição de fraseologias (MEL'ČUK, 2015)</i>	341
8.2.4 <i>Propostas para a descrição de colocações (MEL'ČUK, 2015)</i>	343
CONCLUSÕES.....	348
REFERÊNCIAS.....	355
ANEXO 1.....	380
ANEXO 2.....	383
ANEXO 3.....	386
ANEXO 4.....	389
ANEXO 5.....	392
ANEXO 6.....	393
ANEXO 7.....	394
ANEXO 8.....	395

ANEXO 9	397
ANEXO 10	398
ANEXO 11	402
ANEXO 12	403
ANEXO 13	404
ANEXO 14	408

INTRODUÇÃO

A busca por informações de qualidade constitui uma das práticas mais elementares da humanidade. A velha sentença latina “Verba volant, scripta manent”, que significa que a palavra oral é efêmera, enquanto a palavra escrita é perene, pode até ter vigorado até poucas décadas atrás ao referir-se às informações que são transmitidas por via oral (rádio, palestras etc.) e às informações transmitidas por via escrita (livros, jornais etc.). Contudo, a sentença já não se aplica tanto à realidade da chamada “sociedade da informação” do mundo contemporâneo graças à popularização da internet como meio de circulação e troca de informações. Em pouco tempo, a palavra oral adquiriu traços de perenidade com o armazenamento de conteúdo audiovisual por tempo e acesso indeterminados em servidores de internet. Esse aspecto, somado à massificação do uso da internet como meio para propagação de dados, contribuiu para uma enorme oferta de informações. Entretanto, essa oferta nem sempre satisfaz, *a priori*, o padrão de qualidade que elas deveriam ter. Fenômenos tais como as notícias falsas (*fake news*) são uma prova de que a ingente oferta de informações não é suficiente para satisfazer a necessidade de informações de qualidade.

Além disso, se é verdade que a internet facilitou o acesso das pessoas à informação, contribuindo para a consolidação da “sociedade da informação”¹, é verdade também que não tornou mais fácil a busca por informações de qualidade. Com tantos dados circulando, é cada vez mais difícil distinguir o que é de boa ou má qualidade. Apesar dessa dificuldade, não se pode negar que, para que informações de boa qualidade atinjam um público cada vez mais amplo, é necessário que estejam disponíveis na internet.

Inserida no contexto de oferta de informações e demanda por qualidade, esta tese visa contribuir para satisfazer uma parcela da demanda por informações de qualidade no âmbito do ensino-aprendizagem de línguas no Brasil, mais precisamente o inglês e o espanhol, com foco sobre o desenho de dicionários de aprendizes dessas duas línguas. Ao longo desta introdução, serão apresentados dados que respaldam esta afirmação. Primeiramente, no entanto, é prudente tratar dos meios disponíveis para a divulgação de informações, do uso da internet para fins educacionais no Brasil e da relevância da aprendizagem de inglês e espanhol para os brasileiros.

¹ Segundo Briggs e Burke (2016, p. 296), nos anos 1970, quando a expressão “sociedade da informação” começou a ser usada de maneira mais ampla, “Ela dava forma ou modelava um conjunto [...] de aspectos relacionados à comunicação – conhecimento, notícias, literatura, entretenimento –, todos permutados entre mídias e elementos de mídia diferentes – papel, tinta, telas, pinturas, celuloide, cinema, rádio, televisão e computadores”.

A sociedade da informação e a oferta de informações

Se considerarmos que a rede mundial de computadores foi criada há apenas três décadas, é curioso pensar que, até não muitos anos atrás, o conhecimento era divulgado unicamente por meio da oralidade e via publicações impressas. Ainda é comum, aliás, que a divulgação de informações das mais diversas áreas da ciência se realize em suporte impresso. Prova disso é a importância que alguns tipos de publicações, como certos atlas de anatomia humana, manuais de cálculo e materiais didáticos de ensino de línguas, representam para quem busca informações de qualidade no âmbito das ciências da saúde, das engenharias ou do ensino-aprendizagem de línguas estrangeiras, respectivamente. Um fator que contribui para a alta credibilidade de obras como essas é, justamente, a oferta de um conteúdo confiável e de qualidade, convertendo-as em fontes de consulta para quem estuda medicina, engenharia civil ou letras, por exemplo. O acesso a poucos materiais de referência, ou mesmo uma única fonte, é um possível ônus que pode ser recompensado pela boa qualidade das suas informações.

Ao longo dos últimos anos, um terceiro meio de divulgação de informações começou a se tornar cada vez mais popular: a internet. A possibilidade de acesso universal a uma ampla gama de informações, a capacidade infinita de armazenamento de dados e a rapidez dos provedores de busca pareciam ser uma solução às limitações de acesso do âmbito impresso. Entretanto, tais benefícios acarretam também alguns ônus.

Primeiramente, da mesma forma em que se pode acessar qualquer conteúdo, é possível também divulgar qualquer tipo de informação como um dado verídico – seja realmente verdadeiro ou não². Cabe observar que as informações que constituem a rede provêm tanto de materiais antes disponíveis apenas em papel (e que foram transpostos para a rede) como de conteúdo submetido exclusivamente à mesma. Por essa razão, o risco de deparar-se com informações de baixa qualidade é dobrado, se comparado ao do suporte impresso, justamente pela soma das deficiências do conteúdo em papel às deficiências do conteúdo gerado para a internet.

Diretamente relacionado com o anterior, o segundo ônus que se pode mencionar diz respeito à disparidade entre a capacidade infinita de armazenamento de dados e o controle de qualidade daquilo que está presente na rede. A internet acaba sendo um depósito ilimitado de dados que não se situam, necessariamente, em um patamar *standard* de confiabilidade, e que

² Um exemplo desse tipo de problema é a propagação de notícias falsas (*fake news*) na internet. O fenômeno impulsionou a criação de diversos *sites* e *blogs* para desmenti-las, tais como o *Fake Check – Detector de fake news* (MONTEIRO *et al.*, 2018).

podem permanecer lá indefinidamente, sem passar por qualquer filtro³. Se, nas publicações em papel, nem sempre se pode garantir a qualidade das informações oferecidas, na internet é menos possível ainda. O conteúdo de baixa qualidade que foi publicado em papel sem passar por critérios de avaliação continuará a ser propagado na internet em maior escala, possivelmente sem qualquer intervenção.

O terceiro ônus presente na internet é a contrapartida da rapidez com que os provedores de busca dão retorno ao usuário, com páginas e páginas de resultados. Se, por um lado, nem sempre é possível ter acesso a uma diversidade de fontes de consulta em suporte impresso, por outro lado, a expressiva quantidade de opções de *sites* apontados pelos provedores torna a busca dispendiosa, pois o usuário precisa avaliar a qualidade da informação de cada resultado até encontrar o que precisa. Não se pode inferir, no entanto, que um usuário da internet esteja capacitado para fazer um julgamento acerca do nível de qualidade do conteúdo que lhe é apresentado. Na mesma proporção em que esse indivíduo pode deparar-se com uma informação confiável, também pode ser induzido a dados equivocados ou falhos sem estar ciente disso. A possibilidade de ocorrência da segunda opção é maior na internet, em comparação ao papel, pelas mesmas razões apontadas nos parágrafos anteriores.

Os problemas anteriormente descritos têm uma relevância fundamental por duas razões. Por um lado, nota-se uma sensível ampliação do acesso dos brasileiros à internet. Por outro lado, cada vez mais brasileiros acessam a internet em busca de informações para fins educacionais.

A sociedade da informação e a demanda por informações de qualidade na internet

O constante crescimento e desenvolvimento da internet no Brasil atraiu não apenas novos usuários, mas também as atenções do governo em relação aos tipos de informações mais buscadas por esses usuários. Por essa razão, foi criado o *Comitê Gestor da Internet no Brasil* (CGI.br) (www.cgi.br), um órgão que se ocupa da expansão, manutenção e monitoramento das atividades realizadas pelos brasileiros na internet. Os resultados das últimas pesquisas empreendidas pelo CGI.br indicam que, em 2020, 83% dos domicílios brasileiros tinham acesso

³ Um exemplo disso é a notícia intitulada “Planta africana conhecida como ‘milagre da natureza’, é capaz de ‘Lavar’ o excesso de açúcar do sangue e combater diversas doenças” (<https://portal.ciencia-saude.com/plantamilenar-2/>). Ao final da página, há um *link* chamado “artigos de eficácia científica” onde, no item “Validação Científica Eficácia Redução do Açúcar”, reproduz-se um artigo em inglês referente a um estudo com a planta. Curiosamente, as conclusões aludem à ausência de efeito no controle do índice glicêmico por parte dos pacientes que ingeriram medicamentos à base da suposta “planta milagrosa”.

à internet, isto é, um aumento de 12 pontos percentuais em relação a 2019 (CETIC.BR; NIC.BR, 2021a, p. 3). Além disso, o número de usuários de internet no Brasil com dez anos ou mais representa 81% da população (152 milhões), isto é, um aumento de sete pontos percentuais em relação a 2019 (74%; 133 milhões) (CETIC.BR; NIC.BR, 2021a, p. 3). O aumento do acesso à internet e do número de usuários da rede torna-a cada vez mais democrática. Por outro lado, a partir desses dados se pode inferir que cada vez mais brasileiros estão sujeitos a deparar-se com informações de baixa qualidade por meio da internet, tanto as que são publicadas diretamente na rede como aquelas que são mera projeção de publicações em papel sem qualquer tipo de alteração.

Os problemas da internet enquanto meio de divulgação de informações, por sua vez, têm uma relevância fundamental para a educação nacional. Segundo os dados do CGI.br, dentre os 152 milhões de brasileiros que usam a internet, parte significativa usa as informações disponíveis na rede para fins educacionais: a) 45% usam para realizar atividades ou pesquisas escolares (proporção equivalente a 2019); b) 44% acessam a internet para estudar por conta própria (proporção equivalente a 2019); e c) 21% usam a internet para participar de cursos à distância (aumento de nove pontos percentuais em relação a 2019) (CETIC.BR; NIC.BR, 2021b, p. 86-87). Outros dados que revelam o interesse dos brasileiros pela aprendizagem via internet dizem respeito ao uso do aplicativo Duolingo para aprendizagem de idiomas. Segundo um relatório da empresa referente ao ano de 2020 (MARTINS, 2020), houve um aumento de 23% no número de brasileiros que iniciavam seus estudos em alguma língua estrangeira por meio do aplicativo, colocando os brasileiros em segundo lugar na lista de novos alunos. O mesmo relatório ainda revela um aumento de 29% no número de alunos que retomavam a sua aprendizagem de alguma língua estrangeira pelo Duolingo após haverem interrompido os estudos em algum momento (MARTINS, 2020).

Tais estatísticas refletem o interesse que os brasileiros têm por informações de qualidade e, em parte, pela aprendizagem de línguas estrangeiras. O desempenho educacional desses indivíduos termina por ser influenciado no todo (nos casos do estudo por conta própria e do Ensino a Distância) ou em parte (no caso das atividades e pesquisas escolares) pela infinidade de informações disponíveis na rede, sejam tais informações originárias de livros ou enviadas diretamente a ela, sejam de boa ou de má qualidade. A possibilidade de não estarem preparados para distinguir o que é confiável e o que não é constitui um agravante da situação.

De acordo com o exposto até o momento, os temas da oferta e do acesso a dados de qualidade são ainda controversos. A internet, em particular, não conseguiu superar os problemas das publicações em papel, em parte porque seu conteúdo provém das mesmas. Ao

mesmo tempo, há certos tipos de publicações de boa qualidade que estão disponíveis apenas em papel e cuja disseminação via internet (seja gratuitamente, seja por venda de acesso) contribuiria para o acesso às mesmas. Nossa afirmação se aplica mais especificamente a um tipo de material didático: os dicionários de aprendizes de línguas estrangeiras. Obras dessa natureza se fazem especialmente necessárias para os brasileiros por uma razão simples: aprender línguas estrangeiras, mais precisamente inglês e espanhol, se torna cada vez mais necessário para os brasileiros.

Relevância da aprendizagem de inglês e espanhol para os brasileiros

Não há dúvidas de que a língua inglesa constitui o que se denomina “língua global” (CRYSTAL, 2003; NUNAN, 2003). Essa “globalização” da língua é um fenômeno altamente pertinente para este trabalho porque o inglês é o veículo internacionalmente aceito para a difusão do conhecimento científico, para a divulgação da atividade econômica e, não menos importante, permite um acesso abrangente à informação. Tampouco se pode ignorar que, após o inglês, o espanhol figura como o segundo idioma de comunicação internacional (INSTITUTO CERVANTES, 2018, p. 6).

Durante a pandemia de COVID-19, houve um notório crescimento na procura pela aprendizagem de línguas estrangeiras em geral e de inglês e espanhol, em particular, por parte dos brasileiros. Segundo os dados do relatório sobre o uso do Duolingo no Brasil em Martins (2020), por exemplo, 61% dos brasileiros que usaram o aplicativo em 2020 o fizeram para estudar inglês, classificando o país como o que mais estudou o idioma pelo Duolingo em 2020 na comparação com usuários de outras nacionalidades. Logo após o inglês, o espanhol é a segunda língua de interesse dos brasileiros para aprendizagem via Duolingo (MARTINS, 2020). O interesse pelo estudo de inglês seguido pelo estudo de espanhol pelo Duolingo se manteve ainda em 2021 conforme o relatório global da empresa sobre o ano-base 2021 (BLANCO, 2021). Embora esses dados se refiram a um aplicativo em particular, podem ser considerados como uma amostra do interesse dos brasileiros pela aprendizagem e inglês e espanhol.

No século XIX, o ensino da língua inglesa começou a ser incentivado no Brasil em função da demanda pelo conhecimento da língua no mercado de trabalho da época, ocasionada pelas relações comerciais com a Inglaterra (NOGUEIRA, 2007). Com o passar do tempo, as razões para se aprender inglês no Brasil sofreram poucas modificações. Hoje, a habilidade de falar inglês é considerada como um fator elementar para garantir a relação das empresas

brasileiras com o mercado global; além disso, para os trabalhadores brasileiros, o conhecimento do inglês lhes permite uma ascensão social por meio do aumento salarial⁴ (BRITISH COUNCIL, 2014, p. 14-20). Contudo, conforme os dados do British Council (2014, p. 8), apenas 3,4% da população brasileira de classe média fala inglês efetivamente; em relação à classe alta, os falantes de inglês somam apenas 9,9%. A maior parte da população brasileira (47%) afirma que possui somente um nível básico de conhecimento de inglês (BRITISH COUNCIL, 2014, p. 8). Os dados divulgados pelo British Council (2014) revelam a deficiência do ensino de inglês na formação escolar básica brasileira (tanto pública como privada). Na rede pública de ensino, por exemplo, a falta de acesso a materiais didáticos adequados (entre eles, os dicionários) é o principal problema⁵ identificado por professores de inglês (BRITISH COUNCIL, 2015, p.15). Em uma pesquisa mais recente acerca do nível de proficiência em inglês em diversos países do globo, conduzida pela empresa Education First (2019), percebe-se que os brasileiros continuam apresentando um grau de proficiência abaixo de índices medianos. De acordo com o documento, o Brasil está classificado no grupo de baixa proficiência numa escala constituída pelas faixas “muito alta” [very high], “alta” [high], “moderada” [moderate], “baixa” [low] e “muito baixa” [very low] (EDUCATION FIRST, 2019, p. 6-7). Nesse mesmo estudo, o Brasil ocupa a posição nº 59/100 no *ranking*, inferior à de vários países vizinhos, como Argentina, Uruguai, Chile, Paraguai, Bolívia e Peru (EDUCATION FIRST, 2019, p. 28). Todos os participantes eram adultos; cerca de 90% estavam abaixo dos 40 anos de idade e a média de idade dos participantes foi de 23 anos.

Em relação à língua espanhola, o fator econômico também cumpre um papel importante na sua relevância para os brasileiros. De acordo com o Instituto Cervantes (2018, p. 28), 6,9% do PIB global é gerado por países hispânicos. Esse dado, somado às fronteiras geográficas do Brasil com sete países que têm o espanhol como idioma oficial, se reflete na significativa procura dos brasileiros por aprender espanhol. Segundo o Instituto Cervantes (2018, p. 11), no Brasil, há pouco mais de seis milhões de estudantes de espanhol como língua estrangeira; trata-se do segundo país com maior número de estudantes de espanhol no mundo⁶.

⁴ Segundo os dados do British Council (2014, p. 14-20), 91% dos executivos brasileiros consideram que o inglês é o principal idioma no mundo dos negócios. Entre os trabalhadores brasileiros, 96,5% consideram que o conhecimento da língua inglesa lhes proporciona um aumento salarial de 40%.

⁵ Para 81% dos professores de inglês entrevistados, a falta de acesso a materiais didáticos adequados é um problema rotineiro. Não há dados precisos da situação na rede privada de ensino básico, mas os baixos índices de falantes de inglês das classes média e alta permitem-nos supor que o fator financeiro não é a única variável que permite explicar índices tão modestos.

⁶ No total, há 21 milhões de estudantes de espanhol como língua estrangeira no mundo (INSTITUTO CERVANTES, 2018, p. 5).

Com o intuito de melhorar os baixos índices mencionados acima, o Brasil investiu nos últimos dez anos na massificação do ensino de línguas a nível universitário para que, assim, os estudantes brasileiros pudessem ir ao exterior para estadas de até um ano em universidades e centros de pesquisa (Programa Ciência Sem Fronteiras). Recentemente, os programas governamentais de incentivo à formação acadêmica no exterior foram interrompidos por razões econômicas⁷. No entanto, para a comunidade acadêmica, está claro que a competência em línguas estrangeiras, como o inglês e o espanhol, é um “investimento intelectual” que oferecerá “rendimentos” duradouros entre os jovens acadêmicos e futuros profissionais brasileiros.

Nesse contexto, qualquer iniciativa em prol do desenho de materiais didáticos que permitam um desenvolvimento contínuo da competência linguística e que, paralelamente, propiciem a autonomia na aprendizagem, representa uma ação estratégica para a expansão do conhecimento no país. Dentre os materiais didáticos, o dicionário, quando reflete a língua em uso, é uma ferramenta poderosa para ser usada de maneira complementar às atividades dos livros didáticos. Dentre as diversas classes de dicionários existentes, uma está direcionada especificamente para o ensino de línguas estrangeiras; trata-se do dicionário de aprendizes. Diferente dos dicionários bilíngues, que descrevem uma língua estrangeira por meio de relações de equivalência com a língua materna, um dicionário de aprendizes se preocupa unicamente com a descrição da língua estrangeira de que trata a obra, sendo capaz de fornecer uma descrição mais acurada sobre como a língua-objeto é usada se comparado com um dicionário bilíngue. A necessidade de uma descrição detalhada da língua-objeto surge à medida que o aprendiz vai, progressivamente, tendo contato com estruturas cada vez mais sofisticadas no decorrer do seu processo de aprendizagem da língua estrangeira. Nesse contexto, o aprendiz precisa de dicionários que, igualmente, o acompanhem nesse processo resolvendo dúvidas cada vez mais complexas.

No caso da língua inglesa, a tradição de dicionários de aprendizes é notável não somente pela qualidade das suas obras, mas também pela variedade de dicionários de aprendizes para estudantes de inglês como língua estrangeira. É possível identificar seis linhagens principais de obras: Cambridge, Collins Cobuild, Longman, Macmillan, Merriam-Webster e Oxford. Trata-se de uma indústria acadêmica a editorial firmemente estabelecida e de excelente qualidade⁸.

⁷ O Programa Ciência Sem Fronteiras ficou restrito aos estudantes de graduação e o número de bolsas foi reduzido a 5% (<http://portal.mec.gov.br/ultimas-noticias/212-educacao-superior-1690610854/46981-mec-afirma-que-o-ciencia-sem-fronteiras-tera-5-mil-bolsistas-na-pos-graduacao>).

⁸ No momento de buscar dados factuais acerca da quantidade e qualidade dos materiais didáticos de inglês como língua estrangeira, muito pouco se pode encontrar sobre avaliação qualitativa desses materiais. O pouco que se pode encontrar está majoritariamente relacionado aos livros didáticos, especificamente - estudos de caso, propostas de elaboração de livros didáticos, propostas de critérios de avaliação e análise de aspectos pontuais

Algumas das editoras de dicionários de aprendizes de inglês, como Pearson (Longman), Harper Collins, Oxford University Press e Cambridge University Press, figuram entre as 52 maiores do mundo em 2016 (MILLIOT, 2016).

Há dicionários de aprendizes de inglês para diversos níveis (*grosso modo*, iniciantes, intermediários e avançados). Se consideramos os dados do British Council (2014, 2015), que indicam que os brasileiros se encontram em um patamar básico de aprendizagem de inglês, a subclasse de obra que melhor se enquadra às necessidades dos brasileiros são os dicionários de aprendizes de níveis iniciante a pré-intermediário. Há, entretanto, duas ressalvas em relação a esse tipo de obra: a indisponibilidade de acesso pela internet (seja por site, seja por aplicativo de *smartphone*), que pode torná-las menos conhecidas entre o público aprendiz de inglês em geral; e a falta de especificação do perfil de usuário a quem se dirigem, que é uma característica intrínseca a essa classe de obras e que termina por atender parcialmente as necessidades de aprendizagem dos falantes de português brasileiro como língua materna.

No tocante aos dicionários de aprendizes de espanhol, o cenário é mais modesto. Há três obras compiladas somente em papel e aparentemente para estudantes de níveis intermediário a avançado de aprendizagem – contudo, vale ressaltar que essas obras não apresentam uma classificação acerca do público-alvo, como a aplicação da escala iniciante-intermediário-avançado. Estudantes que se encontram na faixa de níveis entre iniciante e pré-intermediário de aprendizagem de espanhol terminam tendo que recorrer a dicionários bilíngues português-espanhol, os quais apresentam diversas limitações tanto na descrição do espanhol como na descrição do próprio português (BUGUEÑO MIRANDA; BORBA, 2019).

De fato, o panorama das obras lexicográficas monolíngues de espanhol apresentado em Borba (2017a, p. 32-70) demonstra que a grande maioria das obras está disponível em suporte impresso, parte delas está veiculada também em suporte eletrônico (CD-rom ou internet) e uma parcela minoritária corresponde a conteúdo gerado aparentemente de forma exclusiva para a rede⁹.

(pragmática, competência intercultural, polidez, entre outros), por exemplo. No tocante aos dicionários para estudantes de inglês como língua estrangeira, há uma predominância de artigos que analisam aspectos pontuais de dicionários de aprendizes de inglês, como *Structuring Polysemy in English learners' dictionaries: a prototype theory-based model* (LU; WEI, 2019), *Valency and English learners' thesauri* (GAO; LIU, 2019) e *An insight into the visual presentation of signposts in English learners' dictionaries online* (DZIEMIANKO, 2016). Nota-se uma carência de estudos que avaliem as obras de maneira integral e mais exaustiva que uma resenha, por exemplo.

⁹ Há 20 dicionários monolíngues para falantes nativos de espanhol elencados no panorama da autora. Onze estão disponíveis apenas em suporte impresso, sete estão disponíveis em suporte impresso e eletrônico (CD, site de internet ou aplicativo para *smartphone*) e dois estão disponíveis somente na internet. Das sete obras disponíveis concomitantemente em meio impresso e eletrônico, há somente um caso em que a versão eletrônica, acessível

Em suma, as contingências econômicas descritas nos parágrafos anteriores e os desafios e projeções do ensino de inglês no Brasil são poderosas razões para se preocupar com essa classe de dicionários. Devido ao fato de que parte significativa da população brasileira se encontra em estágios muito iniciais de aprendizagem de inglês, e frente à ausência de dicionários de aprendizes de espanhol para níveis iniciais, nossa proposta se concentra no desenho de dicionários de aprendizes para uma etapa de transição entre os níveis básico e intermediário de aprendizagem, isto é, os níveis A2 e B1. Este trabalho propõe estratégias de descrição lexicográfica que almejam aprimorar quatro dicionários de aprendizes de inglês A1-B1 e um dicionário de aprendizes de espanhol no que se refere a duas dimensões: o perfil de usuário, que envolve o viés didático-pedagógico de uma obra lexicográfica; e a descrição de combinatórias léxicas, a qual está relacionada à descrição da língua em uso. O objetivo último da pesquisa foi aproximar a classe dos dicionários de aprendizes de inglês e de espanhol do público brasileiro adulto que aprende inglês ou espanhol no Brasil e que tem contato com a língua ao menos duas vezes por semana em atividades de aprendizagem. A aproximação foi feita por meio da proposição de alternativas para a produção de edições adaptadas de obras já existentes que visem as necessidades desse público. Não se trata, contudo, de propostas exclusivas para os dicionários de aprendizes de inglês e espanhol considerados na tese; parte significativa das considerações e conclusões às quais chegamos com as investigações são uma contribuição para a Lexicografia de Aprendizes em geral e aplicáveis a obras desse ramo que descrevam qualquer língua.

O aspecto inovador da pesquisa apresentada aqui é a perspectiva interdisciplinar adotada para investigar questões para as quais a Metalexigrafia, sozinha, não pode fornecer respostas, tais como a participação da língua materna na aprendizagem de outras línguas, a idade com que se aprende uma língua estrangeira, o conhecimento prévio do aprendiz (língua materna e proficiência) e as suas necessidades de aprendizagem, bem como as combinatórias léxicas. Essas questões têm uma incidência direta sobre o estabelecimento de um perfil de usuário de dicionário e sobre a descrição de combinatórias léxicas, e só puderam ser investigadas sob a ótica de outras áreas, como a Psicolinguística, a Linguística Aplicada, a Linguística Contrastiva (ou *Learner Corpus Research*) e a Lexicologia.

O capítulo 1, a seguir, expõe de maneira mais detalhada o escopo desta tese, com seus temas e objetivos de investigação.

via site da internet, é atualizada independentemente da atualização da edição impressa. Trata-se do *Diccionario de la Real Academia Española* (DRAEe, 2014).

1 O ESCOPO DESTA TESE DE DOUTORADO

Há dois aspectos dos quais, na nossa opinião, deriva parte significativa das orientações e diretrizes acerca de como compilar e como avaliar um dicionário¹¹: a oferta de informações de qualidade e o viés didático-pedagógico da obra.

A oferta de informações de qualidade está diretamente relacionada à descrição da língua. Uma descrição que corresponda ao que efetivamente se realiza (ou realizava, no caso das obras com perspectiva diacrônica) deveria ser o objetivo primordial de uma obra lexicográfica. Quando um indivíduo recorre a um dicionário, seja qual for a classe de obra, o faz porque precisa resolver uma tarefa linguística que requer a compreensão do significado de uma palavra ou expressão, o uso de um sinônimo, a checagem da ortografia, a verificação de que elementos podem se combinar com uma determinada palavra, entre tantos outros exemplos. Uma informação alheia às realizações da língua, isto é, que não encontra respaldo em *corpus*, não faz jus à própria língua. Contudo, sempre é válido mencionar que não há como um dicionário acompanhar em tempo real os processos de constante mudança pelos quais uma língua passa; esse tipo de instrumento oferecerá sempre um reflexo *a posteriori* da língua (BUGUEÑO MIRANDA, 2007a, p. 97).

O reflexo de como a língua se realiza (ou já se realizou algum dia), por si só, não é suficiente para que uma obra lexicográfica seja útil; é preciso que a consulta seja vantajosa para o consulente, na medida em que se faz entender e se aproxima das suas necessidades linguísticas. O aspecto didático dos dicionários diz respeito, portanto, ao quão fáceis de compreender são as suas informações, enquanto a contraparte pedagógica contempla a relevância dessas informações para o usuário.

De acordo com Fernández de la Torre (2001), toda obra lexicográfica termina adquirindo um viés didático-pedagógico. Na verdade, um dicionário tem de sê-lo, pois as informações nele contidas precisam ser apresentadas de maneira facilitada – isto é, didática – para que se cumpra a função – de natureza pedagógica – de orientar o consulente em relação ao funcionamento da língua. A questão da relevância das informações diz respeito ao seu benefício para o consulente. Por exemplo: não faz sentido incluir em um dicionário infantil para crianças brasileiras do 1º ano do ensino fundamental unidades léxicas como *watt*, porque serão pouco

¹¹ Alguns exemplos de temas e de trabalhos correspondentes são: sobre a macroestrutura do dicionário (nominata), Bugueño Miranda (2004/2005, 2007a, 2007b); sobre a microestrutura (informações contidas nos verbetes), Bugueño Miranda (2009a, 2009b) e Farias (2013); sobre a medioestrutura (sistema de remissões), Bugueño Miranda e Zanatta (2010); sobre o *Front Matter* (orientações das páginas iniciais de um dicionário), Fornari (2008) e Borba e Bugueño Miranda (2012).

(para não dizer nada) úteis ao público-alvo da obra, embora *watt* esteja presente na norma real do português¹².

Há uma vertente da área da Lexicografia cuja responsabilidade em ser didático-pedagógica é duas vezes maior: a Lexicografia Pedagógica. A sua finalidade é ir de encontro ao âmbito do ensino-aprendizagem de línguas, seja da língua materna, seja de uma língua estrangeira (cf. capítulos 2 e 3 para questões classificatórias e terminológicas). Em vista disso, esperar-se-ia um esforço redobrado por intensificar o seu carácter didático-pedagógico. Entretanto, e conforme Borba e Bugueño Miranda (2019a), o estado da arte dos dicionários pedagógicos ainda é insatisfatório para que se possa falar com propriedade de uma lexicografia que, de fato, atenda às demandas do ensino-aprendizagem, tanto no caso dos dicionários pedagógicos de língua materna como no caso dos de língua estrangeira.

A pesquisa que apresentamos aqui se enquadra em um ramo da Lexicografia Pedagógica: a Lexicografia de Aprendizizes, que se ocupa especificamente da compilação de obras monolíngues para aprendizizes de uma língua estrangeira/adicional (cf. capítulo 3 para considerações sobre esses termos). O trabalho almeja suprir algumas das carências apresentadas por dicionários de aprendizizes de espanhol e de inglês, identificadas em trabalhos como Borba e Bugueño Miranda (2019a, 2019b, 2018).

Julgamos pertinente ressaltar que, se o dicionário, em geral, é uma ferramenta da linguagem que permite otimizar o uso desta na interação com o mundo, não é possível omitir a inserção do dicionário na sociedade. Nesse sentido, este trabalho busca contribuir para o aperfeiçoamento da relação que o indivíduo estabelece com a sociedade por meio da linguagem, particularmente quando se trata da procura por informações de qualidade, que foi o tema tratado na Introdução.

Por outro lado, a tese obedece também a uma carência detectada na Metalexigrafia e no seu produto final, que é a compilação e aprimoramento das obras lexicográficas no tocante ao seu viés didático-pedagógico. Por essa razão, a tese almeja converter-se em um aporte tanto para a teoria do dicionário como para o aperfeiçoamento das obras lexicográficas para aprendizizes de língua inglesa e de língua espanhola. Nas seções a seguir trataremos das carências e limitações dos dicionários de aprendizizes e das possibilidades de aperfeiçoamento desses instrumentos.

¹² O exemplo não foi fornecido *ad hoc*; está presente no *Meu primeiro livro de palavras* (MPLP, 2010, s.v. *watt*).

1.1 O compromisso com o fornecimento de informações de qualidade de maneira didática¹³

Conforme foi comentado, todo instrumento lexicográfico apresenta um viés pedagógico justamente por estar relacionado à função de orientar o consulente a respeito de questões relativas ao uso da língua. Cabe se questionar, nesse caso, o que diferencia um dicionário pedagógico das demais classes de dicionários. Segundo Hartmann e James (2001, s.v. *pedagogical dictionary*), um dicionário pedagógico é “uma obra de referência desenhada especificamente para as necessidades didáticas práticas de professores e estudantes de uma língua”¹⁴. Dessa definição, infere-se que haveria uma relação entre o dicionário pedagógico e um programa de ensino. Nesse sentido, o dicionário pedagógico consideraria, por um lado, o patamar de aprendizagem no qual se encontra o estudante, isto é, o conhecimento já consolidado em etapas anteriores de aprendizagem; por outro lado, o dicionário pedagógico atribuiria especial atenção às necessidades da etapa na qual se encontra o estudante e pela qual deve seguir. Ademais, e como já foi comentado no início desta Introdução, esperar-se-ia que os dicionários pedagógicos refletissem uma preocupação redobrada em relação à maneira como as informações são apresentadas. Trata-se, em outras palavras, do aspecto didático dessas informações.

Sob a perspectiva metalexiconográfica, e considerando o *status* de obra de referência para suprir necessidades didáticas, mencionado por Hartmann e James (2001, s.v. *pedagogical dictionary*), poder-se-ia inferir que um dicionário pedagógico estaria preparado para fornecer informações de qualidade de maneira didática¹⁵ e estaria atrelado a um programa de ensino. Além disso, tendo em vista o papel que a língua materna desempenha no processo de aprendizagem de uma língua estrangeira/adicional (ver mais detalhes sobre o tema a seguir e no capítulo 4), esperar-se-ia da subclasse dos dicionários de aprendizes uma certa delimitação do perfil de usuário pretendo a partir da variável “língua materna”.

Na prática, as inferências acima não se ajustam à realidade dos dicionários de aprendizes de inglês e de espanhol. Os parágrafos a seguir discorrerão brevemente acerca das perspectivas metalexiconográfica (informações de qualidade) e didático-pedagógica (perfil de usuário, programa de ensino, informações didaticamente apresentadas). Observe-se que a divisão entre perspectiva didático-pedagógica e perspectiva lexicográfica foi utilizada apenas para facilitar a

¹³ Os argumentos expostos nesta seção foram apresentados também em Borba e Bugueño Miranda (2019b).

¹⁴ [A reference work specifically designed for the practical didactic needs of teachers and learners of a language]

¹⁵ Sobre o conceito de *didática* e como ele se relaciona com os dicionários pedagógicos, cf. subseção 1.1.4.

organização da nossa argumentação inicial. Como se poderá perceber ao longo da leitura da tese, as variáveis enquadradas sob a perspectiva metalexiconográfica têm uma incidência sobre a perspectiva didático-pedagógica e vice-versa.

Sob a perspectiva metalexiconográfica, convém especificar o que seriam informações de qualidade em um dicionário de aprendizes. De acordo com o que foi mencionado no início da Introdução, uma informação de qualidade é aquela que está em conformidade com a língua em uso e que, conseqüentemente, sobrevive à contraprova empírica. Na Metalexiconografia, o conceito que se refere a essa dimensão é *informação discreta* (= conformidade com a língua em uso), cunhado por Bugueño Miranda e Farias (2006; 2013, p. 4).

Sob a perspectiva didático-pedagógica, a questão primordial é a especificação do perfil de usuário, ou seja, o público-alvo a quem se dirige uma obra lexicográfica. Os dicionários de aprendizes de inglês e dois dos dicionários de aprendizes de espanhol se dirigem a todo e qualquer estudante dessas línguas como língua estrangeira sem considerar o papel da língua materna dos consulentes durante a aprendizagem de inglês ou espanhol. No caso dos dicionários de inglês, indica-se também um nível de referência, mas sem especificar se esse nível corresponde à proficiência necessária para compreender as informações da obra ou se alude à etapa de aprendizagem à qual se destina.

No âmbito do ensino-aprendizagem de uma língua estrangeira/adicional, é impossível conceber as “necessidades didáticas práticas” mencionadas por Hartmann e James (2001, s.v. *pedagogical dictionary*) sem ter em mente um perfil de usuário a quem o dicionário se destina. As referidas necessidades contemplam, primeiramente, as exigências curriculares da(s) etapa(s) pela(s) qual(is) o aprendiz deve seguir, conforme já comentado. Com base em uma experiência anterior relatada em Borba e Bugueño Miranda (2019, p. 142), neste trabalho, partimos da hipótese de que, na prática, tal relação não ultrapassa a simples indicação que alguns dicionários fazem acerca da etapa à qual almejam servir de auxílio (inicial, intermediária ou avançada). Nos capítulos 2 e 3 se discorrerá mais extensamente sobre o mérito dessa questão.

Além das exigências de um programa de ensino, as “necessidades didáticas práticas” também dizem respeito às demandas advindas da relação entre a língua materna do aprendiz (LM) e a língua estrangeira/adicional (LE/A) que está aprendendo. Os avanços na área da Psicolinguística, tais como comentam Kroll et al. (2010) e Li (2013), indicam que ambas as línguas estão ativadas a todo momento no cérebro dos bilíngües, mesmo em etapas iniciais de aprendizagem (cf. capítulo 4). Isso significa que há uma comprovada inter-relação entre a LM e a LE na qual a primeira influencia a segunda e vice-versa. No que concerne ao ensino-aprendizagem de línguas, é possível afirmar que os resultados dessa inter-relação na

aprendizagem da LE variam conforme as línguas em questão e culminam em necessidades específicas dos aprendizes de acordo com o par de línguas envolvido. Por exemplo: as necessidades de um estudante brasileiro de inglês não serão as mesmas de um estudante germânico – justamente porque se trata de pares de línguas diferentes.

Da relação entre a LM e a LE/A derivam necessidades didático-pedagógicas que são específicas dos indivíduos que possuem uma determinada LM, como o português brasileiro, e que estão aprendendo uma outra língua, como o inglês ou o espanhol. Tais necessidades são alvo de encontros periódicos entre pesquisadores e professores que ensinam espanhol a brasileiros, tais como o *Seminario de Dificultades Específicas de la Enseñanza del Español a Lusohablantes*. Os trabalhos apresentados nesses eventos tendem a atribuir uma especial atenção aos aspectos próprios do ensino de espanhol a brasileiros. Em relação ao ensino de inglês, presta-se especial atenção às necessidades específicas de falantes de certas línguas maternas quando aprendem inglês como língua estrangeira/adicional, como, por exemplo, em Kivistö-de Souza (2017), O'Donnell Christoffersen (2016) e Zacarias (2015).

A especificação do perfil de usuário almejado em relação à sua LM constitui, portanto, uma demanda fundamental para o fornecimento de um auxílio cada vez mais efetivo ao usuário. Tal demanda ainda não foi atendida pelos dicionários de aprendizes em geral justamente porque, nas diversas tradições lexicográficas que contam com essa classe de obras, é costumeiro que os dicionários de aprendizes se dirijam a todo e qualquer aprendiz daquela língua como língua estrangeira.

O estabelecimento das demandas advindas de um programa de ensino em conjunto com as da língua materna é importante porque permite que se inclua e enfatize informações relevantes para o usuário. A esse tipo de informações Bugueño Miranda e Farias (2006; 2013, p. 4) se referem através do termo *informação discriminante*. Trata-se de informações personalizadas conforme as necessidades do público-alvo e que, por esse motivo, são capazes de fornecer um auxílio efetivo.

Uma terceira – mas não menos importante – questão atrelada à perspectiva didático-pedagógica diz respeito à natureza didática das informações fornecidas em dicionários de aprendizes. Nota-se que há uma necessidade de se adotar um mecanismo de apresentação de informações que facilite a sua compreensão em um dicionário de aprendizes, pois há casos (que são comentados ainda neste capítulo) em que o consulente precisa inferir o que o dicionário está tentando indicar realmente.

As próximas subseções estão dedicadas a algumas das variáveis comentadas nos parágrafos acima, quais sejam: informações discretas, informações discriminantes, perfil de usuário e apresentação de informações de maneira didática.

1.1.1 Informações discretas em dicionários de aprendizes

Aplicado aos dicionários de aprendizes, o termo *informação discreta* se refere ao reflexo da norma contemporânea da língua. De fato, é por meio da aprendizagem da norma contemporânea que o aprendiz poderá desempenhar-se adequadamente em situações de interação na língua estrangeira/adicional – isto é, conforme Trappes-Lomax (2004, p. 144), em processos discursivos cujos participantes se manifestam com algum propósito “socialmente reconhecido”.

Há muitos avanços em relação ao fornecimento de informações discretas em dicionários de aprendizes, os quais foram protagonizados pela tradição anglo-saxônica. A longa trajetória de inovações que foram sendo incorporadas aos seus dicionários de aprendizes são, em grande parte, manifestação de um fenômeno conhecido como *Corpus Revolution* (GRANGER, 2018), que traduz os efeitos dos investimentos maciços em Linguística de *Corpus* (KLOTZ; HERBST, 2016, p. 202-210). De acordo com Chi (2015) e Heuberger (2016), os *corpora* contribuíram para que os dicionários de aprendizes de inglês refletissem de maneira mais acurada a língua em uso por meio de iniciativas como a indicação da frequência junto aos verbetes, a ordenação de acepções conforme a frequência de uso e o fornecimento de exemplos provenientes dos próprios *corpora*. O *Macmillan English Dictionary for Advanced Learner's* (MEDAL2, 2007), por exemplo, afirma empregar um *corpus* de inglês contemporâneo (*World English Corpus*, não acessível ao público) para indicar a frequência de uso nos verbetes, conferindo-lhes o *status* de informações discretas. O dicionário adota uma escala de três níveis de frequência que se manifesta nos verbetes por meio de um sistema semiótico de estrelas: mais comum [*most common*], com três estrelas; muito comum [*very common*], com duas estrelas; e razoavelmente comum [*fairly common*], com uma estrela¹⁶. Reproduzimos a seguir exemplos de verbetes de cada nível da escala:

¹⁶ No dicionário, há também verbetes que não têm estrelas. Embora os compiladores não forneçam esclarecimentos sobre isso, deduz-se que sejam unidades léxicas que não se usam tanto como as marcadas com estrelas.

Quadro 1 – Indicação da frequência de uso de unidades léxicas

<p>child /tʃaɪld/ (plural children /tʃɪldrən/) noun [C] ★★★ 1 a young person from the time they are born until they are about 14 years old: <i>The nursery has places for 30 children.</i> ♦ <i>The film is not suitable for young children.</i> ♦ <i>He can't understand – he's just a child.</i> 1a. [only before noun] for or about children: <i>a child car seat</i> ♦ <i>new programmes to address the problem of child poverty</i> 1b. [only before noun] used for showing that the person mentioned is a child: <i>a child actor</i> ♦ <i>child workers</i> 2 someone's son or daughter of any age: <i>All of our children are grown and married.</i> ♦ <i>They're expecting their second child in May.</i> → ONLY CHILD 3 an adult who behaves in ways that are not reasonable or sensible 4 someone whose character is influenced by the main political and social attitudes of a particular period of time: <i>+of a selfish, materialistic child of the '80s</i> PHRASES: be with child <i>old-fashioned</i> to be pregnant have a child or have children to be or become a parent: <i>fertility treatment for couples who can't have children</i></p>	<p>childhood /tʃaɪld.hʊd/ noun [C/U] ★★ the time of your life when you are a child: <i>We spent our childhood in a small town in the mountains.</i> a. [only before noun] connected with childhood: <i>happy childhood memories</i> childish /tʃaɪldɪʃ/ adj ★ 1 behaving in a silly and annoying way, like a small child = IMMATURE 1a. used about behaviour that is not reasonable or sensible: <i>a childish outburst of temper</i> ♦ <i>It's simply childish to refuse to talk to her.</i> 2 typical of a child or of the things children do: <i>She wrote with large, childish letters.</i> — childishly adv — childishness noun [U]</p>
---	--

Fonte: MEDAL2 (2007, s.vs. *child*, *childhood* e *childish*)

A frequência, no entanto, é indicada em relação ao lema apenas. Isso pode levar o consulente a interpretar, erroneamente, que todas as acepções s.v. *child* e s.v. *childish* possuem o mesmo nível de frequência. Seria válido questionar-se se as estrelas foram atribuídas aos verbetes a partir da somatória da frequência de ocorrência de todas as manifestações (acepções) de *child* e *childish*, ou se corresponde à frequência de uma única manifestação (acepção) que seria a mais comum. Independentemente de qual hipótese seja a verdadeira, o exemplo acima revela que o potencial de aplicação da Linguística de *Corpus* à Lexicografia pode ser explorado ainda mais para que se possa aprimorar a qualidade da informação fornecida nos dicionários.

Há casos, entretanto, em que as informações fornecidas por alguns dicionários de aprendizes são pouco ou nada discretas. Uma informação pouco discreta é aquela que conta com um número insuficiente de registros em um *corpus* considerado exaustivo de uma dada língua. A determinação em números do que se pode considerar de baixa frequência é subjetiva; no entanto, certamente se pode duvidar de unidades léxicas que contam com apenas um registro em um *corpus*. Em Borba (2019a) essa situação pode ser constatada em dois dos dicionários de aprendizes de espanhol: o *Diccionario Señas para la enseñanza de la lengua española para brasileños* (Señas, 2001, s.v. *cacera*) e o *Diccionario Salamanca de la Lengua Española* (DiSal, 2006, s.vs. *cacera*, *cachetudo*, *cachifo*, *cachurear*, *percusor*). O fato de serem pouco discretas torna tais palavras pouco discriminantes, pois sua baixa representatividade no espanhol em uso não é relevante para o aprendiz de espanhol.

Mais críticos ainda são os casos de informações não discretas. Um exemplo disso é a lematização de *ghost words*, ou seja, “uma palavra ou frase que foi incluída em um dicionário, por erro ou desenho [*sic!*], embora não haja EVIDÊNCIA da sua existência” (HARTMANN;

JAMES, 2001, s.v. *ghost word*)¹⁷. O DiSal (2006) é um exemplo disso. Uma análise presente em Borba (2019a) confirma que, em um intervalo lematizado de 158 palavras-entrada do DiSal (2006), 15 eram *ghost words*: *acantáceo*; *cacaste*; *acasearse*; *cacaxtle*; *cachaplinco*; *cachería*; *cachetero*; *cachifollar*; *cachiporrearse*; *cachiyuyo*; *peraltado*; *peralto*; *perchel*; *perciforme*; *pepónide*. Por não serem fatos de norma, não são inadequadas somente para um dicionário de aprendizes, mas para qualquer classe de dicionário de espanhol.

No estado atual da pesquisa metalexigráfica, é impossível separar a Lexicografia de Aprendizes da Linguística de *Corpus*. Como consequência disso, não há como negar que, em termos gerais, os dicionários de aprendizes de língua inglesa refletem, de fato, a língua em uso. A evidência empírica proporcionada pela abordagem *corpus-driven*, no entanto, revela sempre novos aspectos de uma língua natural. Assim, por exemplo, em Bugueño Miranda e Borba (2019) foi estabelecida uma série de padrões colocacionais e coligacionais do espanhol para os quais, até agora, a literatura não havia oferecido descrição alguma¹⁸.

1.1.2 Informações discriminantes em dicionários de aprendizes

O conceito de informações discriminantes está construído sobre a base do conceito anterior; em outras palavras, uma informação só adquire o *status* de discriminante (= relevante para o usuário) se for também discreta (= reflexo da língua em uso). Portanto, não existem informações discriminantes que não sejam também discretas, pois uma informação que não condiz com a língua em uso nunca será relevante.

O que determina se uma informação discreta é também discriminante são as características do usuário a quem se destina o dicionário. Por exemplo: uma informação sobre a etimologia de uma unidade léxica pode até ser discreta (*i.e.*, condizente com os dados empíricos sobre o seu surgimento), mas não é discriminante para um dicionário de aprendizes porque sua utilidade é praticamente nula para quem está aprendendo uma língua estrangeira/adicional. No caso dos dicionários de aprendizes, e conforme comentado anteriormente, a adequação a uma etapa específica de um programa de ensino-aprendizagem e a consideração da relação entre a língua materna e a língua estrangeira/adicional do consulente

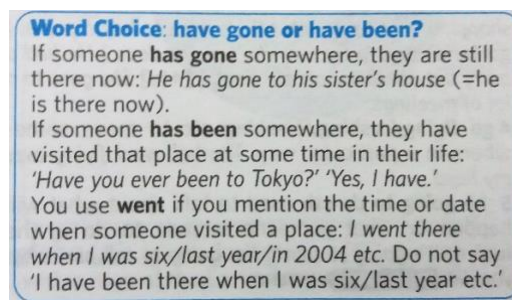
¹⁷ [A word or phrase which has been included in a dictionary, by error or design, although there is no EVIDENCE for its existence]. A marcação [*sic!*] se refere à incongruência do argumento do autor a respeito das razões pelas quais uma *ghost word* acaba sendo incluída no dicionário. Não nos parece que seja por um erro na concepção do desenho de um dicionário, mas sim por um erro na execução do projeto.

¹⁸ No âmbito das colocações, destacam-se *licencia de conducir* e *cédula de identidad*; no âmbito das coligações, destacam-se *demasiado tarde* e *jugar + preposición*, que varia conforme o tipo de complemento.

a quem a obra se dirige são duas variáveis que auxiliam a tornar uma informação discreta mais (ou menos) discriminante.

Os dicionários de aprendizes de inglês são os que mais fornecem informações discriminantes para determinados grupos de usuários, embora a falta de delimitação do perfil de usuário em relação à língua materna prejudique o fornecimento de mais informações dessa natureza. Tome-se como exemplo a nota de uso *Word Choice* presente no verbete *go*¹ do *Longman Active Study Dictionary* (LAS5, 2010):

Figura 1 – Informação discriminante em um dicionário de aprendizes



Fonte: LAS5 (2010, s.v. *go*¹)

As informações sobre a distinção entre os tempos verbais presente perfeito (*has gone, has been*) e passado simples (*went*), juntamente com as implicaturas do uso do presente perfeito com determinados verbos (*go, be*), acabam sendo sumamente úteis (ou seja, discriminantes) para um estudante brasileiro de inglês – ainda que o dicionário não tenha sido compilado visando este público-alvo especificamente. Tais informações enfatizam um aspecto que pode causar dúvidas nos estudantes brasileiros em função do fato de que o português do Brasil emprega somente um tempo verbal, o pretérito perfeito do indicativo, para realizar os mesmos conteúdos expressados pelo presente perfeito e pelo passado simples no inglês.

Contudo, também podem ser identificados casos de informações pouco discriminantes, como a presença de unidades léxicas com índice de frequência pouco significativo e que não condizem com o que o consultante pretendo necessita – apesar de constituírem feitos de norma por estarem documentadas em *corpora* e serem, portanto, discretas. É o caso (não exclusivo, naturalmente) de alguns verbetes presentes no *Collins Cobuild Advanced Dictionary* (CCAD6, 2009), um dicionário de aprendizes avançados de inglês como língua estrangeira: *saddle-bag* (9), *carbolic acid* (18), *scandalize* (18), *calve* (21), *saddler* (27), *carouse* (28), *callus* (31), *capon* (31), *satsuma* (33), *scapula* (33), *sapper* (34), *cantonment* (36), *Camembert* (38), *Samaritan* (38) e *satirize* (39). A frequência dessas unidades léxicas foi verificada no *WordBanks* (WB, 2005), *corpus* conhecido também como *Bank of English*, e que é a base de

dados utilizada para a compilação dos dicionários da linha Collins Cobuild. Em defesa da presença desses verbetes no dicionário, poder-se-ia argumentar que constituem casos de disponibilidade léxica, isto é, seriam unidades léxicas que, apesar do baixo índice de frequência em um *corpus*, foram julgadas pelo compilador do CCAD6 (2009) como elementos relevantes para suprir possíveis necessidades linguísticas do consulente em situações de interação. Isso é o que ocorre, por exemplo, com unidades léxicas referentes às partes do corpo ou aos dias da semana. Contudo, os exemplos acima não parecem ter o *status* de unidades léxicas com significativa relevância para as situações de interação de um aprendiz de inglês como língua estrangeira. A baixa representatividade dessas unidades léxicas no *corpus* da própria editora do dicionário implica que há poucas chances de o aprendiz comum de inglês como língua estrangeira deparar-se com elas em tarefas de recepção, por exemplo. Ainda que se argumente que esses elementos são relevantes para aqueles aprendizes de inglês que se especializam em áreas que requerem conhecimentos de anatomia, é preciso considerar que os dicionários de aprendizes, mesmo os avançados, não têm o objetivo de serem exaustivos a ponto de contemplar grupos específicos de aprendizes. A tarefa de descrever os casos comentados acima é delegada aos dicionários especializados e aos dicionários gerais de língua inglesa, como o Oxford English Dictionary (OED, 2021), que é exaustivo, e o Oxford Dictionary of English (ODE, 2010), que é α exaustivo (cf. distinções em Bugueño Miranda (2019b, c)). Os dicionários especializados se atêm ao léxico de campos específicos do saber, enquanto os dicionários gerais de língua almejam a exaustividade, podendo incluir palavras de baixa frequência e, conforme o nível de exaustividade do dicionário (total ou parcial), palavras de campos específicos do saber.

Em síntese, quando não há maiores especificações sobre o usuário que o modelo “aprendiz da língua X como língua estrangeira”, o investimento em informações discriminantes acaba sendo cerceado. É por isso que o perfil de usuário é um aspecto fundamental para os dicionários em geral e para o dicionário de aprendizes em particular.

1.1.3 Perfil de usuário em dicionários de aprendizes

Como já foi comentado, os dicionários de aprendizes de inglês e alguns dos dicionários de aprendizes do espanhol indicam o tipo de consulente que almejam auxiliar. No caso do inglês, o *Collins Cobuild Intermediate Learner's Dictionary* (CCID4, 2018), por exemplo, afirma, segundo o seu título, direcionar-se para aprendizes intermediários (considerando-se uma escala iniciante-intermediário-avançado). Na contracapa da obra, indica-se ainda o nível B1+ (leia-se “do nível B1 em diante”) do *Quadro Europeu Comum de Referência para as*

Linguas (2001). Outras editoras, como Oxford e Cambridge, também indicam níveis em seus dicionários de aprendizes conforme essa mesma escala, enquanto outras ainda, como Longman e Merriam-Webster, adotam outras escalas.

Nota-se, no entanto, que a indicação dos níveis, independente de qual escala se trate, não permite saber para quem os dicionários de aprendizes de inglês realmente se dirigem. Com as informações que as obras fornecem atualmente na capa e contracapa, não há como saber se os níveis indicados são relativos à proficiência que se requer do consulente ou se esses níveis são para aqueles que se encontram na etapa de aprendizagem indicada. Em outras palavras, e tomando como exemplo o CCID4 (2018), não se sabe se a obra se dirige para quem já têm proficiência de nível B1 (ou mais) ou se a obra se dirige para aqueles que estão em nível B1 (ou mais) de aprendizagem de inglês, com, ao menos, um nível A2 de proficiência já alcançada (sendo A2 um nível imediatamente anterior a B1).

Em relação aos dicionários de aprendizes de espanhol, não há qualquer indicação sobre o nível de proficiência requerido; apenas especifica-se que não são para aprendizes em etapas iniciais de aprendizagem da língua.





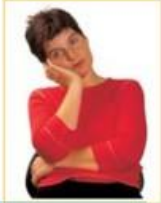
A importância de se indicar de maneira clara a quem se dirige um dicionário de aprendizes (bem como qualquer outro dicionário pedagógico) reside na incidência sobre a seleção de informações a serem incluídas, isto é, se serão discriminantes para o público-alvo, e a maneira como são apresentadas.

Ante essa lacuna na indicação do perfil de usuário dos dicionários de aprendizes, convém esclarecer desde já que a nossa proposta de tese visa auxiliar o desenho de dicionários de aprendizes de inglês e de espanhol cujo público-alvo se encontra nos níveis A2 e B1 de aprendizagem e que possui níveis A1 e A2 de proficiência já alcançada, respectivamente (cf. objetivos na seção 1.3). A lexicografia hispânica, ao que nos consta, não possui dicionários de aprendizes para tais níveis. Em contrapartida, na lexicografia de língua inglesa, já há dicionários de aprendizes que, mesmo que de maneira imprecisa, afirmam abarcar os níveis A1-B1: *Cambridge Essential English Dictionary* (CaED2, 2011), *Oxford Essential Dictionary* (OxED2, 2012) e *Cobuild Essential English Dictionary* (CoED2, 2014). A esse conjunto, cabe acrescentar uma quarta obra: o *Longman WordWise Dictionary* (LWord2, 2008), que almeja aprendizes pré-intermediários e intermediários. Embora use uma escala diferente dos três primeiros, é possível estabelecer um paralelo entre o LWord2 (2008) e o OxED2 (2012) e o CoED2 (2014), pois os três afirmam contemplar níveis pré-intermediários [pre-intermediate]. Frente a esse cenário, esta tese constitui uma proposta de aprimoramento do desenho dos dicionários de aprendizes de língua inglesa, por um lado, e uma proposta inicial de desenho de

dicionários de aprendizes de língua espanhola, por outro lado, baseada nos avanços da lexicografia de língua inglesa e nas estratégias que apresentamos neste trabalho.

Mesmo sem especificar claramente a quem se dirigem em termos de nível de aprendizagem ou nível de proficiência, as concepções dos compiladores do CaED2 (2011), do OXED2 (2012), do CoED2 (2014) e do LWord2 (2008) acerca das necessidades de aprendizagem e do conhecimento prévio do usuário almejado se reflete nos tipos de informações incluídas e na maneira como elas são descritas. Ambos os aspectos podem ser observados a seguir por meio dos excertos de verbetes das obras mencionadas:

Quadro 2 – Concepções sobre o usuário nas informações incluídas nos dicionários

CaED2	OxED2
<p style="text-align: center;">CaED2</p> <p>arrive /ə'raɪv/ verb (arriving, arrived) A to get to a place: <i>We arrived in Paris at noon.</i> ◦ <i>It was dark when we arrived at the station.</i></p> <p>Common mistake: arrive somewhere Be careful to choose the correct preposition after arrive. You arrive at a place such as a building. <i>We arrived at the hotel just after ten o'clock.</i> You arrive in a town, city, or country. <i>They arrived in Tokyo on Wednesday.</i> <i>When did David arrive in Australia?</i> You arrive home, here, or there. You do not use a preposition when arrive is used before these words. <i>We arrived home yesterday.</i> <i>I had a lot of problems when I first arrived here.</i></p> <p>arrogance /'ærəgəns/ noun [no plural] behaviour that shows you believe you are better than other people</p> <p>arrogant /'ærəgənt/ adj believing that you are better than other people</p> <p>arrow /'ærəʊ/ noun arrows</p> <p>1 a symbol used on signs to show a direction</p> <p>2 a long stick with a sharp point at one end that is shot from a bow (= curved piece of wood)</p> 	<p style="text-align: center;">OxED2</p> <p>borne form of BEAR¹</p> <p style="text-align: center;">borrow/lend</p>  <p>He's borrowing some money from his mother.</p>  <p>She's lending her son some money.</p> <p>borrow 0-π /'bɒrəʊ/ verb (borrows, borrowing, borrowed /'bɒrəʊd/) to take and use something that you will give back after a short time: <i>I borrowed some books from the library.</i> ◦ <i>Can I borrow your pen?</i></p> <p>WHICH WORD? Borrow or lend? If you borrow something, you have it for a short time and you must give it back: <i>I borrowed a CD from Alexi for the weekend.</i> If you lend something, you give it to someone for a short time: <i>Alexi lent me a CD for the weekend.</i></p> <p>boss¹ /bɒs/ noun (plural bosses) (informal) a person who controls a place where people work and tells people what they must do: <i>I asked my boss for a holiday.</i></p>
<p style="text-align: center;">CoED2</p> <p>stationary /'steɪʃənri/ adjective not moving: <i>A bus crashed into the back of a stationary vehicle.</i></p> <p>LANGUAGE HELP stationary or stationery? Stationary is an adjective; it means 'not moving'. Stationery is a noun; it means 'paper products used for writing and typing'.</p> <p>stationery /'steɪʃənri/ uncountable noun paper, envelopes and other materials or equipment used for writing and typing: <i>office stationery</i></p> <p>statistic /stə'tɪstɪk/ noun a fact that is expressed in numbers: <i>Statistics show that wages are rising.</i></p> <p>statue /'stætʃu:/ noun statue a large model of a person or an animal, made of stone or metal: <i>She gave me a stone statue of a horse.</i></p> <p>status /'stetəs/ uncountable noun the importance that people give to someone or something: <i>Older family members enjoy high status in many societies.</i></p> 	<p style="text-align: center;">LWord2e</p> <p>bored adjective UK ɹ US ɹ Go to exercises</p> <p>KEY PATTERNS bored with something</p> <p>unhappy and impatient because something is not interesting or you have nothing to do</p> <p>▶▶ <i>She soon got bored with the game.</i> ▶▶ <i>Most of the students looked bored.</i></p> <p>PHRASES be bored stiff, be bored to tears to be very bored</p> <p>WORD CHOICE</p> <p>bored or boring? Bored means "unhappy because something is not interesting": <i>You look bored!</i> Boring means "not interesting or enjoyable": <i>The film was boring.</i></p> <p>bore·dom noun (no plural) UK ɹ US ɹ the feeling you have when you are bored</p> <p>fla·vour¹ BrE, flavor AmE noun UK ɹ US ɹ</p> <p>1 the taste that a food or drink has ▶▶ <i>The ice cream comes in ten different flavours.</i></p> 

Fonte: CaED2 (2011, s.vs. *arrive-arrow*), OxED2 (2012, s.vs. *borne-boss*¹), CoED2 (2014, s.vs. *stationary-status*), LWord2 (2008, s.vs. *bored, boredom, flavour*)

Os comentários a seguir são uma breve introdução acerca dos tipos de informações incluídos nos verbetes dos quatro dicionários supracitados e a sua relevância para tarefas linguísticas de recepção e produção. Uma análise mais minuciosa requer, necessariamente, que se estabeleça uma correlação entre os dicionários e orientações sobre o ensino-aprendizagem de inglês como língua estrangeira – algo que será tratado no decorrer de toda a tese.

No estado atual da Lexicografia de Aprendizizes, a estratégia usada para que os aprendizes compreendam o funcionamento de uma língua estrangeira/segunda é o emprego da linguagem para discorrer acerca da própria linguagem, num procedimento denominado metalinguagem. Em um verbete, a metalinguagem consiste no uso da linguagem para fornecer informações semânticas (paráfrase explanatória e notas de uso, por exemplo) e gramaticais (indicação da pronúncia, classe morfológica, comportamento sintático e notas de uso, por exemplo) acerca de uma língua que é o objeto da descrição. Tecnicamente, as informações em um verbete de dicionário são denominadas *segmentos informativos* (cf. BUGUEÑO MIRANDA, 2019a, p. 15). Os dicionários de aprendizizes contêm dois tipos básicos de segmentos informativos, que seguem a dicotomia saussureana do signo linguístico: comentário de forma e comentário semântico.

O comentário de forma abriga segmentos informativos associados, por extensão, ao significante do signo linguístico. Nos dicionários em questão, trata-se da representação ortográfica de uma unidade léxica, da representação fonológica e da categoria morfológica.

A representação ortográfica é a própria palavra-entrada (também chamada de *lema*) destacada em negrito. Além de servir como mecanismo de busca, a principal função da representação ortográfica é resolver dúvidas relativas à produção linguística. Os dicionários incluem também variações ortográficas conforme a perspectiva britânica ou americana que adotam, como se pode verificar s.v. *flavour* no LWord2 (2008) em relação à variação ortográfica americana *flavor*. Os quatro dicionários analisados adotam a perspectiva britânica porque foram produzidos no Reino Unido. Isso se reflete na opção de não incluir a variante americana como uma palavra-entrada, mas sim sob o verbete da variante britânica.

Assim como a representação ortográfica, a representação fonológica também almeja auxiliar na produção linguística. Esse segmento informativo pode se apresentar de duas formas: via transcrição fonológica simplificada no caso das edições impressas ou via recurso de áudio, no caso das edições eletrônicas. Dentre os quatro dicionários analisados, dois deles possuem tanto uma edição impressa como outra eletrônica; é o caso do OxED2 (2012) e do LWord2e (2008). Via de regra, nas edições eletrônicas dos dois dicionários, os editores optaram por não mesclar as duas possibilidades de representação fonológica. Um questionamento que se pode

fazer sobre a edição eletrônica é até que ponto o aprendiz é capaz de distinguir todos os fonemas reproduzidos no áudio. Esse tema será tratado novamente no capítulo 4.

A categoria morfológica é informada por meio dos termos *noun, adjective, verb* etc. Também constituem segmentos informativos de ordem morfológica as flexões nominais, como s.vs. *arrogance* (CaED2, 2011), *boss*¹ (OxED2, 2012), *stationery* e *status* (CoED2, 2014) e *boredom* (LWord2, 2008); e as flexões verbais, como s.v. *arrive* (CaED2, 2011) e *borrow* (OxED2, 2012). Em geral, esses segmentos informativos também estão associados a tarefas de produção linguística. Apesar de não constarem nos excertos reproduzidos no Quadro 2, dentre as flexões nominais são incluídas também as formas comparativas e superlativas sufixadas dos adjetivos, tanto regulares (*taller* e *tallest* em relação a *tall*, por exemplo) como irregulares (*better* e *best* em relação a *good*, por exemplo).

No âmbito das flexões verbais, nota-se que os dicionários incluem como palavra-entrada as formas irregulares de alguns verbos, como se pode observar s.v. *borne* (OxED2, 2012). Tal estratégia contempla a função de recepção e prevê que o seu consulente é alguém que pode apresentar dificuldades em identificar e/ou associar a flexão irregular dos verbos com sua forma no infinitivo.

O comentário semântico, por sua vez, abriga segmentos informativos associados, por extensão, à dimensão do significado do signo linguístico. Trata-se das paráfrases explanatórias, marcas de uso, exemplos e sinônimos.

As paráfrases explanatórias são, por excelência, o segmento informativo da dimensão da recepção de língua e estão listadas sob a forma de acepções¹⁹. Além da sua evidente utilidade em tarefas de recepção de língua, as paráfrases explanatórias acabam sendo imprescindíveis também para a resolução de tarefas de produção, pois cada conteúdo semântico que representam está normalmente associado a uma determinada característica morfossintática. É o caso, por exemplo, do adjetivo *bored*, cujo conteúdo semântico é associado pelo LWord2 (2008) ao padrão sintático “bored with something”. Em função da sua dupla utilidade, é fundamental que a paráfrase explanatória esteja redigida de forma tal que o consulente a quem o dicionário se dirige possa compreendê-la. Sua leitura já pode ser considerada um desafio por estar redigida em inglês, isto é, a língua-objeto da descrição linguística. A única maneira de avaliar até que ponto as paráfrases explanatórias dos quatro dicionários contemplam o usuário que almejam é tomando por base orientações acerca do ensino-aprendizagem de inglês como língua estrangeira. Esse tema será tratado novamente nos capítulos 3 e 5.

¹⁹ Sob a perspectiva do signo linguístico, o verbete *arrow* do CaED2 (2011), por exemplo, abriga dois signos; os respectivos significados estão discriminados nas duas acepções listadas.

Um aspecto que pode ter chamado a atenção do leitor é o uso do termo *paráfrase explanatória*, e não *definição*, para referirmo-nos ao conteúdo semântico que esse segmento informativo comporta. De fato, por muito tempo, usou-se e ainda se usa o termo *definição*; no entanto, é cada vez mais comum que se questione a pertinência desse termo devido às limitações conceituais e práticas que apresenta. Uma das críticas de ordem conceitual que Bugueño Miranda (2009a, b) faz a esse termo é que o seu conceito abarca somente a reescrita de conteúdos proposicionais por meio da fórmula aristotélica “gênero próximo + diferença específica”²⁰ e deixa de contemplar outras formas de reescrita, tais como as *whole-sentence definitions*²¹. Nesse sentido, Farias (2009), por exemplo, prova que, em determinadas situações, as *whole-sentence definitions* são mais efetivas que a fórmula aristotélica para elucidar um conteúdo proposicional. Outra crítica, presente em Creese, McGillivray, Nesi et al. (2018) é de ordem tanto conceitual como prática. Os autores criticam o fato de que o termo *definição* almejava a exaustividade, isto é, uma definição teria de fornecer todas as características com as quais um referente se apresenta. Conforme os autores, “a palavra *definição* é, por si só, problemática: a raiz latina implica um nível de certeza e precisão que é inatingível para muitos itens do vocabulário diário”²² (CREESE et al., 2018, p. 5, grifo dos autores). Por essas razões é que Bugueño Miranda (2009b, p. 244) propõe o uso do termo *paráfrase explanatória* para referir-se à reescrita do conteúdo proposicional de uma unidade léxica em um verbete de dicionário e que abarca diferentes técnicas definitórias (ou parafrásticas) – e, nesta tese, será feito o mesmo.

Ao redor das paráfrases explanatórias, há dois segmentos informativos que normalmente aparecem nos dicionários de aprendizes: marcas de uso e exemplos. Ambos os segmentos informativos auxiliam tanto em tarefas de compreensão como de produção. As marcas de uso que se encontram nos dicionários analisados dizem respeito a usos circunscritos a um determinado espaço geográfico (Reino Unido vs. Estados Unidos) ou a uma situação específica (formal vs. informal), como se verifica s.v. *boss*^l no OxD2 (2012). Os exemplos que seguem à paráfrase explanatória são um segmento informativo muito comum, embora nem sempre estejam presentes nas acepções. É comum encontrar padrões de combinatória léxica destacados

²⁰ A fórmula aristotélica funciona por meio de uma equação, na qual a unidade está em uma relação de equivalência com a definição, podendo a primeira ser substituída pela segunda. Por exemplo, a definição “a place with a lot of books that you can read or borrow” (CaED2, 2011, s.v. *library*), equivale e pode substituir *library*.

²¹ A *whole-sentence definition* contextualiza a palavra-entrada numa oração e a define numa segunda oração (FARIAS, 2009, p. 76). Um exemplo de aplicação dessa técnica pode ser observado no LWord2 (2008, s.v. *glad*): “if you are glad, you are pleased and happy about something”.

²² [the word “definition” itself is problematic: the Latin root implies a level of certainty and precision which is in unattainable for most items of everyday vocabulary]

em negrito, como s.v. *arrive* (CaED2), *borrow* (OxED2, 2012) e *bored* (LWord2, 2008). Por essa razão, eles têm o potencial de auxiliar não somente em tarefas de recepção, almejando consolidar a compreensão do conteúdo da paráfrase explanatória, como em tarefas de produção, tendo em vista os padrões morfossintáticos ressaltados em negrito.

Um último segmento informativo do comentário semântico que aparece esporadicamente nos verbetes são os sinônimos e antônimos, que são relevantes para tarefas de produção.

Após as paráfrases, os dicionários em questão incluem alguns tipos de combinatórias léxicas, mais especificamente *phrasal verbs* e fraseologias [idioms], conforme se pode observar s.v. *bored* (LWord2, 2008). O CaED2 (2011) é o único que lista essas informações sob a forma de acepções também. Esse tipo de informação é relevante para tarefas de recepção e produção, no caso dos *phrasal verbs*, e majoritariamente de recepção, no caso das fraseologias. Duas questões que cabe levantar são até que ponto os dicionários descrevem diferentes tipos de combinatórias léxicas e que combinatórias seriam relevantes para os usuários. Mais detalhes serão comentados na próxima subseção e no capítulo 8 da tese.

Nos excertos reproduzidos no Quadro 2, buscou-se incluir verbetes que contemplassem dois elementos que são complementares aos segmentos informativos dos comentários de forma e semântico: as ilustrações e as notas de uso. Trata-se de recursos frequentemente empregados nos quatro dicionários analisados para facilitar a compreensão de determinados segmentos informativos e/ou chamar a atenção para algum aspecto, relacionando-os à dimensão didático-pedagógica das obras (cf. 1.1.4).

As ilustrações geralmente são empregadas em meio aos verbetes quando o lexicógrafo julga que a paráfrase explanatória *per se* não é suficiente para elucidar um significado. Nos dicionários analisados, as ilustrações s.v. *arrow* (CaED2, 2011), *borrow* (OxED2, 2012), *statue* (CoED2, 2014) e *bored* (LWord2, 2008) se propõem a auxiliar na compreensão da paráfrase para a execução de tarefas linguísticas de recepção.

Já as notas de uso são, segundo Farias (2011, p. 133), informações adicionais que não seguem parâmetros formais de inclusão. Nos dicionários analisados, elas podem obedecer ou a constatações sobre dados de um *corpus* de aprendizes, ou à intuição do lexicógrafo sobre dúvidas que os aprendizes podem apresentar, ou ainda a constatações de experiências prévias com o ensino-aprendizagem de inglês como língua estrangeira. Os quatro dicionários analisados não fornecem detalhes acerca da motivação para a inclusão dessas informações adicionais. As notas de uso podem complementar as informações do comentário de forma, como s.v. *arrive* (CaED2, 2011), em relação aos tipos de preposições e complementos que

acompanham o verbo, e s.v. *stationary* (CoED2, 2014), em que a nota complementa as representações ortográfica e fonológica ao comentar a homofonia e homografia; bem como o comentário semântico, como s.vs. *borrow* (OxED2, 2012) e *bored* (LWord2, 2008), cujas notas de uso se propõem a elucidar aspectos semânticos.

Em síntese, pelos tipos de segmentos informativos incluídos no CaED2 (2011), no OxED2 (2012), no CoED2 (2014) e no LWord2 (2008), percebe-se que o usuário pretense pelos editores é alguém que recorrerá a essas obras para resolver dúvidas tanto de recepção como de produção. A qualidade da descrição da língua inglesa é inquestionável; resta saber, entretanto, em que medida tais informações podem ser compreendidas por aprendizes em níveis A2 e B1 de aprendizagem e, conseqüentemente, aplicadas às tarefas de recepção e produção que motivaram a consulta às obras. Esse aspecto tem direta relação com a maneira didática de apresentar informações, tópico a ser tratado a seguir.

1.1.4 Apresentação de informações de maneira didática

Sendo o dicionário de aprendizes um instrumento didático-pedagógico, como comentamos no início deste capítulo, é impossível dissociar qualquer proposta Metalexigráfica de aprimoramento dessa classe de obras da Didática de línguas. Conforme Díaz-Corrálejo (2004, p. 243), essa área tem por finalidade fornecer respostas para “os problemas que foram surgindo e surgem no ensino-aprendizagem de línguas”²³ por meio de uma abordagem interdisciplinar. Aplicada ao âmbito da Metalexigrafia de aprendizes, a abordagem interdisciplinar advogada pela Didática ajuda a investigar e propor soluções aos problemas descritos nas subseções anteriores tomando por base subsídios não somente da Metalexigrafia, mas também da Linguística Aplicada ao ensino de línguas, da Psicolinguística e da Psicologia Cognitiva (cf. demais capítulos da tese).

O adjetivo *didático*, por sua vez, diz respeito a uma característica de materiais de ensino que associam e aplicam conhecimentos de diferentes disciplinas de maneira que o produto gerado responda às necessidades do seu público-alvo e facilitem o acesso ao conhecimento (GALISSON, 1990 apud DÍAZ-CORRALEJO, 2004, p. 247). Neste ponto, convém esclarecer que o emprego do adjetivo *didático*, e mesmo a identificação e denominação da área da Didática, não são unânimes (GALISSON; COSTE, 1983, s.v. *didática*). Segundo Díaz-Corrálejo (2004, p. 247), no termo em inglês *didactics* sobressai uma conotação negativa que,

²³ [problemas que han ido surgiendo, y surgen, en la enseñanza/aprendizaje de las lenguas]

embora presente também em outras línguas românicas, não está tão latente como no inglês²⁴. Aliás, e segundo Díaz-Corrájeo (2004, p. 247), é em função dessa conotação pejorativa latente de *didactics* que a tradição anglo-saxônica inscreve sob o rótulo de *Linguística Aplicada* tudo o que se refere à dimensão didática do ensino-aprendizagem de línguas. De maneira análoga, percebe-se que o que os trabalhos de Metalexigrafia escritos em línguas românicas e germânicas se referem por *didático* (adjetivo) e *didática* (substantivo), os trabalhos escritos em inglês chamam de *user-friendly* e *user-friendliness*.

O adjetivo *user-friendly* e o substantivo *user-friendliness* foram usados em diversos trabalhos da área ao longo dos últimos vinte anos (Por exemplo: Leech e Nesi (1999), Dziemianko (2011), Kosch (2013), Zhao (2015), entre outros). No entanto, nem sempre os estudiosos definem o conceito de *user-friendliness* e o seu respectivo adjetivo ao citarem-nos em suas publicações. Kosch (2013), por exemplo, associa vagamente o conceito à satisfação do consulente durante a busca e ao tempo demandado por ele ao recuperar da memória a informação consultada. Klotz e Herbst (2016, p. 27; 32-33) definem *user-friendliness* como um parâmetro que influencia decisivamente na compilação de um dicionário e que se refere, em linhas gerais, à facilidade de interpretação de informações gerada pela clareza do *layout* da obra e à codificação de informações de maneira clara e intuitiva. Os autores afirmam que medidas adotadas no decorrer do desenvolvimento da Lexicografia de Aprendizes de língua inglesa, tais como as melhorias nas técnicas definitórias, a inserção de informações sobre frequência e o espaço reservado aos erros comuns de aprendizes de inglês são exemplos de manifestação do conceito de *user-friendliness* nos dicionários (KLOTZ; HERBST, 2016, p. 242).

Nesta tese, os pares de termos *didático* / *user-friendly* e *didatismo* / *user-friendliness* são usados como sinônimos para se referirem a tudo o que facilita a compreensão de informações sob a perspectiva de um determinado perfil de usuário.

As iniciativas por tornar os dicionários de aprendizes mais didáticos remontam ao surgimento dessa classe de obras nos anos 1940 (para uma síntese do tema, cf. Bugueño Miranda e Borba (2019b); para explanações mais detalhadas, cf. Jackson (2002), Heuberger (2016) e Klotz e Herbst (2016, p. 242-245). Segundo assinalam Chi (2015) e Heuberger (2016), ao longo do tempo, a tradição anglo-saxônica inovou substancialmente a maneira como passou a apresentar certas informações, tais como as paráfrases explanatórias redigidas pela técnica

²⁴ Ao consultar o substantivo *didática* e o adjetivo *didático* no *Dicionário Houaiss da Língua Portuguesa* (Hou, 2009), percebe-se que o único resquício de conotação negativa está relacionado ao adjetivo *didático*, para o qual, na acepção 6, o Hou (2009, s.v., ac. 6) classifica como uso *pejorativo* e define como “que denota didatismo”, isto é, que apresenta uma “maneira afetada de ser didático” (Hou, 2009, s.v. *didatismo*, ac. 2). Esse uso pejorativo é o que parece ser predominante na língua inglesa, mais do que no português.

das *whole-sentence definitions*, o uso de vocabulários de definidores embasados pela frequência em *corpora*, o uso de distinguidores semânticos junto a verbetes polissêmicos para facilitar a consulta, a inclusão de ilustrações, entre outras iniciativas²⁵. Paralelamente, as críticas também cumpriram um papel importante no aprimoramento dessa classe de obras. As críticas ao uso de códigos para indicar padrões sintáticos no *Oxford Advanced Learner's Dictionary of Current English* (OALD3, 1974), por exemplo, deram origem a uma sucessão de modificações na maneira como essa informação era apresentada, de modo a torná-la mais elucidativa (KLOTZ; HERBST, 2016, p. 105-108).

Não há dúvidas de que as iniciativas que vêm sendo comentadas até aqui se enquadram no que chamamos de *didatismo/user-friendliness*, pois visam apresentar as informações dos dicionários de aprendizes de uma maneira cada vez mais fácil de compreender. No que diz respeito ao futuro das iniciativas em prol do *didatismo/user-friendliness* dos dicionários de aprendizes, pudemos fazer duas constatações.

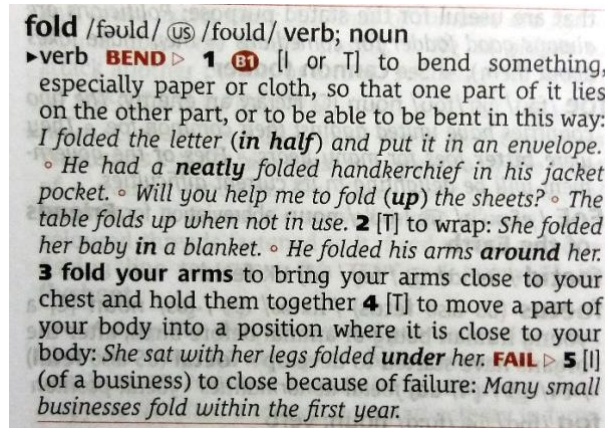
A primeira é a de que, até o momento, a Lexicografia de Aprendizes vem considerando o usuário de maneira indireta. Em outras palavras, adota-se a perspectiva da descrição da língua-alvo, da qual o usuário é beneficiário. Ainda falta analisar o dicionário de aprendizes diretamente a partir da perspectiva do usuário, tomando-o como ponto de partida das iniciativas em prol da descrição da língua-alvo. Isso implica definir quem é esse usuário, quais são as suas características e necessidades, de modo que a obra lexicográfica responda a esse perfil, tal como já comentado na seção anterior.

A segunda observação diz respeito ao futuro das iniciativas para melhorar a descrição da língua. Klotz e Herbst (2016, p. 245), após descreverem os inúmeros avanços protagonizados pelos dicionários de aprendizes de língua inglesa, observam que as obras disponíveis atualmente são muito parecidas entre si justamente por haverem adotado medidas semelhantes. Desse comentário, poder-se-ia inferir que, após o boom da segunda metade do século XX e início do século XXI, já não haveria muito o que melhorar nessa classe de obras em termos de descrição. No entanto, essa uma questão ainda não foi esgotada nos dicionários de aprendizes de inglês e de espanhol (RANALLI; NURMUKHAMEDOV, 2013, p. 4), especialmente no que diz respeito às combinatórias léxicas e às ilustrações (BORBA; BUGUEÑO MIRANDA, 2019b).

²⁵ No passado, as críticas ao uso de códigos para indicar padrões sintáticos no *Oxford Advanced Learner's Dictionary of Current English* (OALD3, 1974), por exemplo, deram origem a uma sucessão de modificações na maneira como essa informação era apresentada, de modo a torná-la mais elucidativa (Klotz; Herbst, 2016, p. 105-108).

Primeiramente, reproduz-se o verbete *fold* do *Cambridge Advanced Learner's Dictionary* (CALD4, 2013):

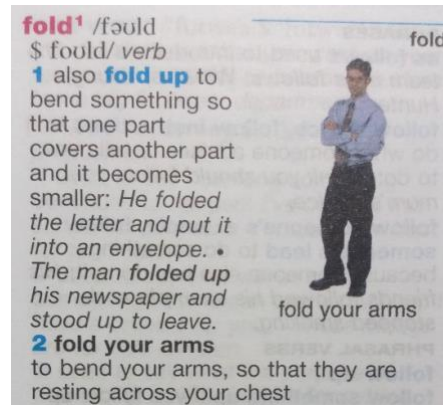
Figura 2 – Reprodução parcial do verbete *fold* (CALD4)



Fonte: CALD4 (2013, s.v. *fold*)

A leitura das quatro acepções sob o distinguidor semântico *bend* permite identificar quatro tipos de complementos: papéis, tecidos, partes do corpo e mobília. No entanto, o dicionário não apresenta essas informações de um modo elucidativo para que um consulente que tem o português brasileiro como LM, por exemplo, compreenda-as. Na primeira acepção, a especificação “especially paper or cloth” não deixa claro o que mais, além de objetos de papel e de tecido, pode ser “dobrado”. Essa falta de clareza fica mais evidente no último exemplo da acepção, que é referente a dobrar uma mesa. A segunda acepção parece não ter muita relação com o distinguidor semântico *bend* e com as acepções 1, 3 e 4. Em relação à terceira e quarta acepções, não é possível entender por que estão separadas, pois contém o mesmo significado. Por um lado, não são somente os braços que podem ser dobrados, mas as pernas também (ver exemplo na quarta acepção). Isso coloca em dúvidas o *status* de fraseologia que o CALD4 (2013) aparentemente atribui a *fold your arms* ao destacar essa expressão em negrito. Por outro lado, uma consulta ao WordBanks (WB, 2005) revelou que não é toda e qualquer parte do corpo que aparece com o verbo *to fold*, mas somente *arms*, *hands* e *legs*. Em resumo, as informações sobre os complementos de *to fold* são apresentadas no verbete de uma maneira que não proporciona ao consulente um acesso rápido às mesmas. Isso onera a busca de informações tanto para tarefas de recepção como para tarefas de produção.

Outro exemplo que ilustra a necessidade de se apresentar as combinatórias da língua em uso de maneira mais *user-friendly* é o verbete *fold* no LWord2, 2008).

Figura 3 – Reprodução do verbete *fold* (LWord2)

Fonte: LWord2 (2008, s.v. *fold*¹)

Na acepção 1, falta especificar que tipos de complementos podem acompanhar o verbo. Na acepção 2, o destaque em negrito leva a entender que *fold your arms* é uma expressão fixa quando, na verdade, admite outros complementos. A paráfrase explanatória, por sua vez, pode ser de difícil compreensão para um consulente pré-intermediário devido a palavras polissêmicas como *bend* e *rest*. Nesse sentido, a ilustração ajuda o consulente brasileiro a compreender o significado, principalmente em função de que, em português, os braços são cruzados, e não dobrados.

Um último exemplo do baixo poder elucidativo das informações sobre combinatórias léxicas em dicionários de aprendizes pode ser visualizado no verbete *trabajo* do Señas (2001):

Quadro 3 – Reprodução parcial do verbete *trabajo* (Señas, 2001)

<p>tra·ba·jo traβáxo 1 m. Actividad o dedicación física o mental; esfuerzo: <i>es un niño mimado que siempre lo ha tenido todo sin que le costara ningún ~; no sabes el ~ que me ha costado planchar esta blusa; el coche es viejo, pero hace bien su ~.</i> • trabalho [...]</p>
--

Fonte: Señas (2001, s.v. *trabajo*)

Na acepção 1, o Señas (2001) é discreto e discriminante ao incluir o padrão colocacional *costar trabajo* (ver exemplos no verbete). Entretanto, tal informação não está apresentada de maneira elucidativa, ou seja, não é didática, pois o consulente não tem como depreender dos exemplos que há um caso de combinatória entre as unidades léxicas *costar* e *trabajo*. Sob a perspectiva do consulente brasileiro que está aprendendo espanhol, é preciso considerar que, além de não poder “prever” o padrão combinatório do espanhol, o consulente brasileiro que se

ampara nas relações combinatórias que o português estabelece com a unidade léxica *trabalho* (*dar trabalho*, por exemplo) pode acabar produzindo uma combinação não preferencial no espanhol como **dar trabajo*.

Conforme já mencionado anteriormente, um mecanismo comumente encontrado nos dicionários de aprendizes e que pode contribuir para tornar a obra mais didática são as ilustrações. Nos dicionários de aprendizes de inglês, é comum encontrar ilustrações tanto junto aos verbetes como no *middle matter* (páginas que interrompem a progressão da nominata pela metade) ou no *back matter* (páginas após a nominata).

Há um evidente avanço da teoria lexicográfica principalmente na análise das relações entre elementos verbais e não verbais e entre elementos não verbais e o consulente. Trata-se da teoria da multimodalidade, que versa a respeito do papel das ilustrações, cores e recursos tipográficos no processo de construção e compreensão do texto dicionarístico sob a ótica do consulente (cf. Nascimento (2018) para uma síntese). No Brasil, a atenção dos pesquisadores recai sobre as ilustrações em dicionários escolares, como se verifica em Nascimento (2018), Pontes (2010a,b; 2018), Santos (2018) e Santos, Coriolano e Pontes (2019) – para citar alguns trabalhos.

Esses avanços constituem uma etapa posterior à identificação da função das ilustrações na obra lexicográfica e ao estabelecimento de critérios de inclusão de ilustrações. Farias (2013), por exemplo, até chega a estabelecer princípios gerais acerca da função desse segmento informativo nos dicionários; porém, os critérios de inclusão permanecem sendo uma incógnita na área, tal como apontam Farias (2013), Klotz e Herbst (2016, p. 66) e Borba (2018a). Para exemplificar a questão, tome-se o *middle matter* do *Collins Cobuild Essential English Dictionary* (CoED2, 2014) e do *Oxford Essential Dictionary* (OxED2, 2012). Ambos indicam correspondência com os níveis iniciante a pré-intermediário. O CoED2 (2014, p. P1-P8) apresenta no seu *middle matter* diversas ilustrações individuais e esquemas de ilustrações organizados conforme os seguintes temas: o corpo [the body], comida e bebida [food and drink] (com subdivisões de frutas, vegetais, outros alimentos e bebidas), casa e lar [house and home], esportes [sports], empregos [jobs] e família [family]. O OxED2 (2012, P1-P16), por sua vez, tem mais que o dobro de temas, quais sejam: transporte [transport], animais [animals], o corpo [the body], roupas [clothes], comida e bebida [food and drink], refeições [meals], frutas [fruit], vegetais [vegetables], casa [house], escola [school], empregos [jobs], lojas [shops], esportes e hobbies [sports and hobbies] e tempo atmosférico [weather].

A falta de especificação de um perfil de usuário conforme a sua língua materna também incide sobre as ilustrações. Nos mesmos dicionários citados, o verbete *piano* está acompanhado por uma ilustração correspondente. Para um aprendiz brasileiro de inglês, essa ilustração é desnecessária devido à correspondência ortográfica e fonológica com o português.

1.2 Alternativas para aprimorar os dicionários de aprendizes

A partir das considerações feitas até o momento, foi possível identificar dois temas transversais: perfil de usuário e didatismo. Foram identificados também dois temas subordinados a estes: combinatórias léxicas e ilustrações, dos quais trataremos apenas do primeiro nesta oportunidade. Uma razão para os problemas comentados na seção anterior é o fato de que não se leva em conta as perspectivas de outras áreas além da Lexicografia a respeito desses temas. Mais especificamente, carece-se de um aproveitamento ainda maior do conhecimento advindo não apenas da Linguística de *Corpus*, mas também da Lexicologia, da Linguística Aplicada, da Linguística Contrastiva e da Ciência Cognitiva para gerar mecanismos que permitam aumentar o poder de elucidação das informações fornecidas em dicionários de aprendizes.

Primeiramente, são inegáveis os benefícios que a Linguística de *Corpus* trouxe à Lexicografia, em especial aos dicionários de aprendizes. É importante frisar que o emprego de subsídios de *corpora* para a compilação de dicionários de aprendizes já é uma realidade para a tradição anglo-saxônica há muito tempo, mas o mesmo não se aplica aos dicionários de aprendizes de espanhol. A presença de *ghost words*, problema encontrado em um dos dicionários de aprendizes da tradição hispânica, constitui um entre os vários desafios a serem superados.

As contribuições da Linguística de *Corpus* aparecem de forma mais evidente na comprovação da frequência de uso das unidades léxicas incluídas nas obras²⁶, como ilustrado no exemplo sobre *child*, *childhood* e *childish* (v. Quadro 1). É importante ressaltar, porém, que o potencial de aplicação da Linguística de *Corpus* para a Lexicografia está longe de se esgotar, primeiramente porque o dicionário precisa acompanhar, com o auxílio de *corpora*, as mudanças que a língua apresenta; em segundo lugar, porque, no exemplo mencionado, é possível perceber o quão útil poderia ser a diferenciação da frequência de cada acepção dentro de um verbete; e, em terceiro lugar, porque é preciso investigar de maneira mais minuciosa como as palavras se

²⁶ Sobre este particular, Bosque (2001) já salientava que esse critério, isolado, não conduz, necessariamente, a afirmações conclusivas quando se trata de fenômenos idiomáticos.

comportam no eixo sintagmático para que se possa avaliar formas mais didáticas de apresentar esse tipo de informação.

Apesar da importância dos *corpora* para a compilação de dicionários de aprendizes (entre outras ferramentas lexicográficas, naturalmente), os casos de informações não didáticas, comentados nas páginas anteriores, demonstram que é necessário conduzir as investigações para o modo como as informações extraídas de *corpora* são apresentadas nos dicionários de aprendizes. Nesse sentido, acreditamos que a complementaridade entre a Linguística de *Corpus* e a Lexicologia é uma solução que enriquece ainda mais a descrição lexicográfica de um fenômeno e contribui para que os dicionários de aprendizes sejam mais didáticos. A Teoria Sentido-Texto de Mel'čuk (2015) é um exemplo de aporte teórico de cunho lexicológico que, ao fundamentar-se em pesquisas com *corpus*, comprova a necessária relação de complementaridade entre as duas áreas. A metodologia advinda da Teoria Sentido-Texto, aliás, pode contribuir para uma apresentação mais didática das informações em dicionários de aprendizes (BORBA, 2018a). A importância da complementaridade entre a Linguística de *Corpus* e a Lexicologia para melhorar o aspecto didático dos dicionários de aprendizes reside no fato de que, tal como assinala Vasiljevic (2014, p. 50), os casos de combinatórias léxicas, como as colocações, são “um dos maiores desafios para os estudantes de segunda língua, mesmo em estágios avançados de proficiência linguística”²⁷. Soma-se a isso o fato de que, segundo os dados arrolados por Sánchez Rufat (2018, p. 261), entre 52% e 58% do discurso de falantes nativos de uma língua está constituído por alguma classe de combinatórias léxicas. Por essas duas razões, é de suma importância que se invista no aprimoramento constante dos dicionários de aprendizes com o auxílio de distintas áreas.

Em segundo lugar, atualmente, a Linguística Aplicada conta com as propostas do *Quadro Europeu Comum de Referência para as Línguas* (QECRL, 2001) já aplicadas ao inglês e ao espanhol: o *English Profile* (EP, 2011) e o *Plan Curricular del Instituto Cervantes* (PCIC, 2006). Ambos os documentos contêm propostas acerca da organização de programas de ensino de inglês e de espanhol como língua estrangeira, respectivamente, que podem servir de base para a compilação de dicionários de aprendizes dessas línguas.

Em terceiro lugar, nos dicionários de aprendizes, a determinação de um perfil mais específico de público-alvo a partir do fator *língua materna* permitiria o fornecimento de informações cada vez mais discriminantes e didáticas ao consulente pretense. No campo da Linguística Contrastiva, os *corpora* de análise de erros podem se converter em importantes

²⁷ [one of the biggest challenges for second language learners even at very advanced stages of language proficiency]

aliados na tarefa de discriminar as informações do inglês e do espanhol que, em função das suas divergências em comparação com o português, podem oferecer dúvidas para os aprendizes brasileiros. Há *corpora*, tais como o *International Corpus of Learner English* (ICLE, 2009)²⁸ e o *Corpus de Aprendizes de Español* (CAES, 2018), nos quais se pode consultar a produção escrita de estudantes brasileiros de cada uma dessas línguas. Os dados advindos desses *corpora* possuem um claro potencial de aproveitamento no desenho de dicionários de aprendizes.

Embora os estudos de Linguística Contrastiva não sejam muitos entre os pares de línguas português-espanhol (*e.g.* Durão, 2004) e português-inglês (*e.g.* O'Donnell Christoffersen, 2016), não se pode desestimar o valor que a Linguística Contrastiva possa ter para o aprimoramento das informações a serem inseridas em um dicionário de aprendizes.

Em síntese, a necessidade de aperfeiçoamento do desenho dos dicionários de aprendizes de inglês e de espanhol poderia ser suprida não apenas pela aplicação de parâmetros metalexográficos, mas também pela aplicação de conhecimentos de outras áreas, como as mencionadas acima, que podem ser convertidos em parâmetros metalexográficos. Conforme defendem Andersen e Nielsen (2009, p. 356), “como a lexicografia é, por natureza, interdisciplinar, desenvolve-se em muitos casos a partir da contribuição de um amplo número de disciplinas dentre as quais uma é a linguística”²⁹. O que se propõe nesta tese é explorar ainda mais essa natureza interdisciplinar para aprimorar o desenho de dicionários de aprendizes por meio da criação de estratégias de descrição e disposição de informações que permitam ao usuário tirar ainda mais proveito da sua consulta, tornando o dicionário de aprendizes um instrumento mais eficiente.

1.3 Objetivos

Esta tese possui um viés interdisciplinar, pois se ampara em diferentes áreas para propor um aprimoramento do desenho de dicionários de aprendizes no que diz respeito à determinação de um perfil de usuário e à descrição de combinatórias léxicas. Nosso enfoque recai sobre o desenho de dicionários de aprendizes de inglês e de espanhol para estudantes brasileiros que se encontram em níveis A2 e B1 de aprendizagem, isto é, cujo processo de aprendizagem está passando por essas etapas.

²⁸ No início das investigações para esta tese, o *corpus* ainda podia ser acessado online; no entanto, a princípios de 2020, constatamos que, infelizmente, não está mais disponível na rede.

²⁹ [since lexicography is in nature interdisciplinary, it thrives in many cases on input from a large number of disciplines of which one is linguistics]

1.3.1 Objetivo geral

- Explorar e aplicar princípios teórico-metodológicos para um aprimoramento do desenho de dicionários de aprendizes de língua inglesa e de língua espanhola destinados aos níveis A2 e B1 de aprendizagem.

1.3.2 Objetivos específicos

- Analisar as capacidades e limitações da Lexicografia de Aprendizes no que se refere ao perfil de usuário, às combinatórias léxicas;
- Avaliar as contribuições que a Lexicologia tem a oferecer para o aprimoramento da Lexicografia de Aprendizes;
- Avaliar as contribuições que a Linguística Aplicada tem a oferecer para o aprimoramento da Lexicografia de Aprendizes;
- Avaliar as contribuições que a Linguística Contrastiva tem a oferecer para o aprimoramento da Lexicografia de Aprendizes;
- Avaliar as contribuições que a Ciência Cognitiva tem a oferecer para o aprimoramento da Lexicografia de Aprendizes;
- Propor estratégias advindas da Linguística Aplicada para o aprimoramento da Lexicografia de Aprendizes;
- Propor estratégias advindas da Lexicologia para o aprimoramento da Lexicografia de Aprendizes;
- Propor estratégias advindas da Linguística Contrastiva para o aprimoramento da Lexicografia de Aprendizes;
- Propor estratégias advindas da Ciência Cognitiva para o aprimoramento da Lexicografia de Aprendizes.

1.4 Pressupostos de pesquisa

Segundo comentado até este ponto, o desenho dos dicionários de aprendizes de espanhol e de inglês precisa ser aperfeiçoado para que as obras compiladas futuramente sejam ferramentas cada vez mais eficientes. Os exemplos mencionados demonstram que, em geral, apesar dos avanços no fornecimento de informações discretas e discriminantes, protagonizados

principalmente pela tradição anglo-saxônica, os dicionários de aprendizes nem sempre fornecem informações discretas e discriminantes e, em geral, demonstram ser pouco didáticos.

Estabelecemos os seguintes pressupostos de pesquisa:

- a) Os níveis A2 e B1 de aprendizagem estão entre os mais necessitados de materiais de apoio.
- b) Os dicionários de aprendizes de espanhol e de inglês não estão ancorados nos documentos orientadores do ensino de espanhol e inglês como língua estrangeira.
- c) Há um subaproveitamento de subsídios advindos de outras áreas, em especial da Psicolinguística e da Neurolinguística, que podem ser empregados para oferecer informações mais didáticas ao usuário.
- d) A descrição das relações sintagmáticas entre as unidades léxicas não obedece (ao menos não tanto quanto poderia) a um modelo teórico que sustente a sua análise e descrição.
- e) Há um saber especializado advindo da Linguística Contrastiva que pode ser aproveitado no desenho de um dicionário de aprendizes.

1.5 Estrutura geral do trabalho

Esta tese está composta de outros sete capítulos. Os capítulos 2 a 4 tratam do perfil de usuário sob as perspectivas da Metalexigrafia, da Linguística Aplicada e da Ciência Cognitiva. O capítulo 5 apresenta uma análise de dicionários de aprendizes de inglês e de espanhol à luz da noção de perfil de usuário desenvolvida nos capítulos precedentes. Os capítulos 6 e 7 contêm propostas de aperfeiçoamento das paráfrases explanatórias e das notas de uso de dicionários de aprendizes de inglês e de espanhol. Finalmente, o capítulo 8 trata das teorias sobre as combinatórias léxicas e apresenta propostas para a sua descrição à luz do perfil de usuário.

O capítulo 2 expõe as contribuições da Metalexigrafia no que diz respeito ao perfil de usuário. A partir da tríade de parâmetros classificação-função-usuário, discorre-se acerca dos avanços na definição do perfil de usuário tanto a partir de pesquisas teóricas como de pesquisas empíricas. Ao final, e a partir de uma série de variáveis que, conforme o referencial teórico, incidem sobre a determinação do perfil de usuário de um dicionário, são estabelecidas as características do usuário que interessa à tese, isto é, brasileiros que aprendem inglês e espanhol no Brasil.

O capítulo 3 organiza as contribuições da Linguística Aplicada na determinação do perfil de usuário. Após uma discussão terminológica inicial sobre a variedade de adjetivos usados com o termo *língua* (estrangeira, adicional, segunda), discorre-se sobre níveis de proficiência e programas de ensino. O capítulo se detém na análise do Quadro Europeu Comum de Referência para as Línguas e assinala as suas contribuições para a determinação do perfil de usuário de um dicionário conforme a proficiência já atingida e a que se almeja atingir. Também são analisados documentos específicos acerca da aprendizagem de inglês e de espanhol como língua estrangeira. Ao final, são feitas considerações sobre o papel do dicionário no ensino-aprendizagem comunicativo e sobre as contribuições da Linguística Aplicada.

O capítulo 4 trata do perfil de usuário sob a perspectiva da Ciência Cognitiva (Psicolinguística e Neurolinguística). O capítulo se detém sobre o modo como são processadas as línguas que são aprendidas após a consolidação da língua materna. São apresentados subsídios teóricos a respeito das variáveis que incidem sobre o processamento de línguas: língua materna, período da vida em que ocorre o contato com a língua estrangeira/adicional, proficiência atingida na língua materna e na língua estrangeira/adicional, entre outros. Reserva-se ainda um espaço especial no capítulo para tratar de uma área correlata à Psicolinguística, que é a Linguística Contrastiva. Como corolário, são feitas considerações acerca de como o referencial teórico empregado auxilia na especificação das características do perfil de usuário de um dicionário e, de maneira particular, daquele que interessa à tese.

O capítulo 5 emprega o referencial teórico comentado nos capítulos anteriores para avaliar o quão compreensíveis são as informações presentes em dicionários de aprendizes de inglês e de espanhol tendo em vista as características do perfil de usuário estabelecido como objeto e interesse da tese no capítulo 2. São coletadas amostras dois dos segmentos informativos de um verbete que, conforme já comentado, mais exigem proficiência em leitura na língua estrangeira: as paráfrases explanatórias e as notas de uso. Ambas são avaliadas em relação ao conhecimento prévio necessário para a sua compreensão, o qual abarca o nível de proficiência atrelado aos definidores e as influências interlinguísticas da língua materna. Como corolário das análises realizadas, são apresentados *rankings* de facilidade de compreensão das paráfrases e notas analisadas.

O capítulo 6 se ampara no referencial teórico e nos dados presentes nos capítulos anteriores para propor o aprimoramento de segmentos informativos do comentário de forma e as paráfrases explanatórias de dicionários de aprendizes de inglês e de espanhol à luz das necessidades do público brasileiro estabelecido no capítulo 2. As propostas são aplicadas ao conjunto de dados analisados no capítulo 5.

O capítulo 7 segue com as propostas de aprimoramento dos dicionários de aprendizes de inglês e de espanhol, mas com foco sobre as notas de uso. Neste capítulo, primeiramente são feitas considerações teóricas sobre o segmento informativo em questão que culminam em um conjunto de propostas acerca dos recursos iconográficos e tipográficos a serem empregados, as modalidades de orientação disponíveis, a língua empregada e as etiquetas indicativas do tema da nota. Em uma segunda parte do capítulo, as propostas feitas são aplicadas sobre um conjunto de dados relativos aos erros e dúvidas de aprendizes brasileiros de inglês e de espanhol. De especial valia para a aplicação das propostas foram os dados do Cambridge Learner Corpus, do Corpus de Aprendices de Español e de outras fontes.

O capítulo 8 trata, por fim, das combinatórias léxicas. Primeiramente, são expostas as principais teorias a respeito do tema, desde as de viés estatístico até as de viés gramatical e lexicológico. Em seguida, é elencada uma série de fenômenos combinatórios a partir dos conceitos mais bem delimitados, sem que se restrinja a uma única perspectiva teórica. Ao final, são apresentadas propostas de descrição de combinatórias léxicas em dicionários de aprendizes de inglês e espanhol.

2 O PERFIL DE USUÁRIO SOB A ÓTICA DA METALEXICOGRAFIA

O perfil de usuário constitui uma das decisões metalexigráficas que devem ser tomadas para determinar como um dicionário será compilado e avaliado. Em referência às variáveis comentadas no capítulo 1 (seções 1.1.1 a 1.1.4), a determinação de um perfil de usuário auxilia na inclusão de informações discriminantes (ou seja, relevantes para o usuário) e está diretamente relacionada também com a natureza didático-pedagógica dos dicionários de aprendizes.

Este é o primeiro de uma série de capítulos sobre o perfil de usuário. Nesta oportunidade, são comentados os atuais aportes da Metalexigrafia acerca da determinação do perfil de usuário, começando pela sua relação com a classificação e a função de dicionários de aprendizes para, então, tratar das investigações teóricas e experimentais sobre o usuário (seções 2.1, 2.2 e 2.3). A partir do anterior, na seção 2.4, determina-se o usuário almejado por esta tese a partir de uma série de variáveis, que são aprofundadas nos capítulos 3 e 4.

2.1 A classificação dos dicionários de aprendizes e o estabelecimento do perfil de usuário

Conforme Bogueño Miranda (2019d, p. 12), o perfil de usuário precisa estar inter-relacionado com a classificação (taxonomia) do dicionário e com a função para a qual a obra almeja servir de auxílio. Antes de entrar no mérito específico do perfil de usuário, é pertinente comentar como os outros dois parâmetros incidem na sua determinação.

Uma classificação de dicionários permite que se obtenha um panorama das obras disponíveis. Os dicionários pedagógicos são – ou, melhor dito, deveriam ser – obras de natureza linguística (isto é, deveriam ser um reflexo da língua em uso) e, ao mesmo tempo, pedagógica (isto é, deveriam corresponder a um programa de ensino)³⁰. Sendo assim, uma proposta classificatória para essa classe de dicionários precisaria estar fundamentada em critérios que contemplassem a natureza das mesmas (*i.e.* a sua dimensão linguístico-pedagógica).

Quando os critérios que constituem uma proposta classificatória são baseados em parâmetros de imanência linguística (v. parágrafos *ad infra*), é possível empregá-los também para a compilação e avaliação de obras lexicográficas. Tais critérios auxiliam não só no

³⁰ Quando se analisa a correspondência entre dicionários de aprendizes de inglês e de espanhol e a natureza pretensamente linguístico-pedagógica de um dicionário pedagógico, percebe-se que as obras disponíveis, por diferentes razões, deixam a desejar. No decorrer de toda a teste, forneceremos exemplos que embasam nossa afirmação.

fornecimento de um panorama claro, mas também na avaliação da pertinência das informações em relação ao perfil de usuário. Cabe mencionar também que ditos critérios auxiliam ainda no estabelecimento do perfil de usuário no ato da compilação de um dicionário pedagógico.

Pode-se identificar três propostas de classificação de dicionários pedagógicos: Welker (2008), Tarp (2011) e Borba e Bugueño Miranda (2019a). Nesta seção, os comentários detêm-se sobre as duas últimas, pois a proposta de Welker (2008) apresenta dois problemas. Por um lado, utiliza critérios impressionistas, isto é, não atinentes ao conteúdo da obra, mas sim ao seu suporte (impresso/eletrônico). O autor faz alusão à diferença na estrutura de acesso das obras, defendendo que “o conteúdo dos dois tipos pode ser igual, mas no manuseio eles são bem diferentes” (WELKER (2008, p. 25). O suporte e o modo como ocorre o manuseio de um dicionário não influenciam na sua classificação, pois não são critérios que determinam a identificação e classificação de uma obra lexicográfica como sendo de uma classe e não de outra³¹. Uma obra é identificada e classificada como dicionário de aprendizes, por exemplo, pelo seu propósito de auxiliar um usuário inserido em um contexto de ensino-aprendizagem de línguas estrangeiras/adicionais – e não em função do suporte. O fato de dita obra estar disponível em suporte impresso ou eletrônico determina como o usuário navega pela estrutura de acesso – seguindo a progressão alfabética no suporte impresso ou digitando uma palavra no campo de busca no suporte eletrônico – e pelo sistema de remissões – novamente seguindo a progressão alfabética no suporte impresso ou sendo redirecionado para o alvo da remissão mediante um simples clique se o suporte for eletrônico. O *layout* de um verbete, maiores ou menores restrições de espaço para empregar ilustrações e a possibilidade de usar recursos audiovisuais também são aspectos que dependem do suporte e sobre os quais se tomam certas decisões (mas não todas) sobre o desenho de uma obra. No entanto, tais aspectos não influenciam – insistimos – na identificação e classificação de um dicionário.

O problema de usar um critério impressionista como o do suporte para classificar dicionários é que uma mesma obra pode estar presente de maneira duplicada na proposta, de modo que o panorama fornecido resulta confuso. É o caso de muitos dicionários de aprendizes de inglês que estão veiculados tanto em suporte impresso como em suporte eletrônico, como por exemplo, CCID4 (2018), que está disponível em papel e sob a forma de aplicativo de *smartphone*, e OxColl (2009) e OxThes (2008), ambos disponíveis em suporte impresso, em

³¹ Conforme já comenta Borba (2017a, p. 23) a respeito de outra proposta classificatória do autor que também emprega a distinção impresso vs. eletrônico, o argumento defendido pelo autor acerca da estrutura de acesso “carece de uma avaliação comparativa da qualidade do conteúdo de dicionários veiculados por esse suporte [*i.e.*, suporte eletrônico]”.

CD-ROM e no site da editora (com acesso restrito a partir de um código de acesso que pode ser comprado). Muito mais interessante que discriminar o suporte seria avaliar se uma edição eletrônica de um dicionário realmente oferece um conteúdo diferenciado ao seu usuário em comparação com uma edição impressa da mesma obra, se explora os recursos que o suporte oferece e se esses recursos de fato resultam em descrição lexicográfica mais clara ou mais acurada.

Tarp (2011), por sua vez, busca uma solução de compromisso entre as orientações teóricas na área da metalexigrafia e os termos utilizados nos títulos de obras consideradas como circunscritas à Lexicografia de Aprendizes.

Primeiramente, Tarp (2011) analisa a considerável variação terminológica em distintas tradições lexicográficas no que tange aos pares de termos *dicionário pedagógico* [pedagogical dictionary] vs. *dicionário didático* [didactic dictionary] e *dicionário escolar* [school dictionary] vs. *dicionário de aprendizes* [learner's dictionary]. Segundo a análise do autor, dicionários para aprendizes de língua estrangeira ora são chamados de dicionários escolares, ora de dicionários de aprendizes. O mesmo ocorre com dicionários para aprendizes de língua materna. O termo *dicionário pedagógico* é definido pelo autor como

[...] um dicionário especialmente desenhado para auxiliar aprendizes de línguas (tanto de língua materna como de língua estrangeira) e de disciplinas científicas e práticas. Um dicionário de aprendizes é sinônimo de dicionário pedagógico, mas seu enfoque recai sobre o aprendiz e não sobre o educador (TARP, 2011, p. 223-224)³².

O termo *dicionário didático*, por sua vez, é apenas comentado, sem receber uma definição do autor. Na nossa opinião, quando se trata de obras de referência, amparar-se no fazer prático para estabelecer orientações teóricas nem sempre é a melhor solução. Nesse sentido, consideramos mais adequado circunscrever os dicionários voltados para aprendizes de disciplinas científicas e práticas ao âmbito da Terminologia e da Terminografia, pois são áreas que se preocupam justamente das obras de referência voltadas para âmbitos de especialidade. No que tange ao caráter linguístico do dicionário pedagógico, o autor prefere tratar as designações *dicionário pedagógico* e *dicionário de aprendizes* como sinônimas e referentes ao mesmo conceito. Nessa mesma esteira, Tarp (2011, p. 226) define *dicionário escolar* como

[...] um dicionário de aprendizes (ou dicionário pedagógico) especialmente desenhado para auxiliar crianças em idade escolar no aprendizado de línguas (tanto da língua

³² [a dictionary especially designed to assist learners of languages (whether a native or a foreign language) and of scientific and practical disciplines. A learner's dictionary is synonymous to a pedagogical dictionary but focuses on the learner and not the educator]

materna como de língua estrangeira) e de disciplinas científicas e práticas (TARP, 2011, p. 226)³³.

Não parece óbvio que empregar a variação designativa das obras pertencentes à Lexicografia Pedagógica seja a melhor solução para a área em termos de teoria metalexigráfica. Tal variação designativa reflete uma ausência de reflexão terminológica, tanto designativa como conceitual, ao classificar e dar nome aos dicionários. Se essa variação for empregada para gerar uma classificação de dicionários, dita classificação apenas reproduzirá a confusão terminológica, sem que necessariamente forneça um panorama consistente de uma tradição lexicográfica nem auxilie na escolha de uma obra lexicográfica.

Em segundo lugar, e baseando-se na sua proposta de terminologia, Tarp (2011, p. 228-229) propõe uma classificação tipológica de dicionários pedagógicos reproduzida a seguir:

Figura 4 – Tipologia de Tarp (2011)

Typology A for pedagogical dictionaries (learner's dictionaries)

- (1) Dictionary for native-language learners
 - (a) Preschool dictionary
 - (b) School dictionary
 - (c) College dictionary

- (2) Dictionary for non-native-langue learners
 - (a) (Preschool dictionary)
 - (b) School dictionary
 - (c) College dictionary

- (3) Dictionary for learners of scientific disciplines, etc.
 - (a) (Preschool dictionary)
 - (b) School dictionary
 - (c) College dictionary

Fonte: Tarp (2011, p. 228)

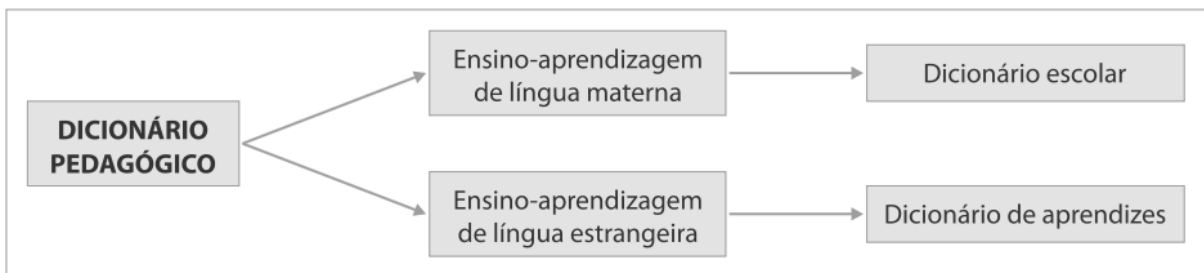
Uma tipologia tem como principal característica o fato de almejar a inclusão do maior número de obras possível sob a mesma classificação a partir do princípio da prototipicidade (BUGUEÑO MIRANDA, 2019b, p. 32). O traço prototípico que se identifica na classificação de Tarp (2011) poderia ser definido como “dicionário para fins de aprendizagem, tanto de línguas como de áreas do conhecimento”. A opção do autor por esse tipo de classificação é consonante com a sua perspectiva de refletir a prática e comportar o maior número de obras possível. Entretanto, a sua proposta de classificação não parece ser a melhor opção para a

³³ [A school dictionary is a learner's dictionary (or pedagogical dictionary) especially designed to assist school children in learning languages (whether a native or a foreign language) and scientific and practical disciplines]

Lexicografia Pedagógica a começar pela questão terminológica, já comentada anteriormente. Em segundo lugar, conforme comentam Borba e Bugueño Miranda (2019a, p. 147), há uma variação na organização do sistema educativo conforme o país de que se trate. Por essa razão, as etapas de ensino empregadas por Tarp (2011) não necessariamente corresponderão àquelas de outra tradição linguística e lexicográfica. Quando se trata de línguas estrangeiras, pode até haver uma maior consonância no sistema de ensino se se considera o ambiente europeu, em função da base comum representada pelo *Quadro Europeu Comum de Referência para as Línguas* (QECRL, 2001) – o mérito desse documento será comentado mais adiante ainda nesta seção e no capítulo 3. Em terceiro lugar, o panorama fornecido ajuda a saber quantas classes de obras existem, mas, em termos práticos, não ajuda a escolher um dicionário que realmente forneça informações de qualidade, pois não é esse o seu objetivo.

Por essas razões, somos favoráveis pela adoção de uma classificação objetiva e que, embora simples, pode ser aplicada a diversas tradições lexicográficas de dicionários de aprendizes. Trata-se da proposta de taxonomia de Borba e Bugueño Miranda (2019a). Por definição, uma taxonomia está constituída de um sistema de critérios excludentes e não almeja a inclusão de toda e qualquer obra, mas sim aquelas que estejam em conformidade com seus critérios de organização. A proposta dos autores está reproduzida a seguir:

Figura 5 – Classificação dos dicionários pedagógicos



Fonte: Borba e Bugueño Miranda (2019a, p. 141)

Os autores definem como dicionário pedagógico as obras compiladas exclusivamente para o ensino-aprendizagem de línguas. Por conter distinções que podem ser encontradas em qualquer tradição lexicográfica que possua dicionários pedagógicos, a proposta de classificação está aberta para aplicação em distintas tradições lexicográficas. O critério de classificação que constituem a proposta é genérico, pois consiste no *status* da língua-objeto descrita no dicionário em relação à comunidade linguística à qual se dirige o dicionário. A sua natureza genérica é vantajosa porque contempla a realidade de qualquer tradição lexicográfica.

Ao mesmo tempo, os autores propõem que, ao se aplicar a classificação da Figura 5 sobre uma dada tradição, sejam considerados ao menos dois critérios específicos: as diretrizes curriculares nacionais, no caso dos dicionários escolares, e as orientações sobre o ensino-aprendizagem de línguas estrangeiras, no caso dos dicionários de aprendizes. Tais critérios conferem plasticidade à proposta, pois permitem a inclusão de subclasses de obras conforme o comportamento de cada tradição. Essa plasticidade é importante porque o nível de profusão de dicionários pedagógicos varia de tradição para tradição. A tradição anglo-saxônica, por exemplo, além de empregar os níveis do QECRL (2001), também discrimina os dicionários de aprendizes de acordo com o tipo de informação linguística que o dicionário enfoca. Nesse sentido, há, por um lado, dicionários de aprendizes que poderiam ser classificados como “gerais” e diferenciados pelo nível do QECRL (2001), tais como o *Collins Cobuild Essential English Dictionary* (CoED2, 2014), para níveis A1 a B1, o *Collins Cobuild Intermediate Learner’s Dictionary* (CCID4, 2018), para níveis B1 em diante (B1+), e o *Collins Cobuild Advanced Learner’s Dictionary* (CCAD9, 2018), para níveis B2 em diante (B2+). Ao mesmo tempo, há também dicionários de aprendizes de inglês que descrevem aspectos específicos da língua, tais como os sinônimos e colocações. É o caso do *Oxford Learner’s Thesaurus* (OxThes, 2008), e do *Oxford Collocations Dictionary* (OxColl, 2009), ambos para níveis B2 em diante. Para refletir a diversidade de dicionários de aprendizes de inglês, essas duas obras poderiam ser acomodadas através da inclusão de subclasses no modelo taxonômico de Borba e Bugueño Miranda (2019a). Outras tradições, como a de língua espanhola, por exemplo, já não possuem dicionários de sinônimos para aprendizes nem dicionários de colocações para aprendizes, nem empregam os níveis do QECRL (2001) para classificar as suas obras.

No que concerne aos dicionários de aprendizes e o seu respetivo perfil de usuário, a proposta classificatória de Borba e Bugueño Miranda (2019a) pressupõe que um dicionário que se autoproclame dicionário de aprendizes precisa estar em consonância com um programa de ensino-aprendizagem, não apenas mencionando-o, mas sim que as informações descritas e a maneira como são descritas estejam amparadas nas orientações sobre o ensino-aprendizagem de línguas estrangeiras. Não é esse o caso dos dicionários de aprendizes de inglês e de espanhol. Embora os dicionários de inglês sejam produto de uma incontestável metodologia baseada no uso da língua, levada a cabo por meio da Linguística de *Corpus*, ainda se faz necessária a contrapartida pedagógica na compilação dessas obras. Já os dicionários de aprendizes de espanhol, além da questão pedagógica que se percebe no inglês, ainda não implementaram (ao menos não de maneira satisfatória) a metodologia de descrição empírica, que compromete a confiabilidade das informações apresentadas.

No que tange à contraparte pedagógica, no caso das línguas europeias, existe a base comum do QECRL (2001), cuja principal proposta consiste na escala de níveis unificada a ser empregada no ensino-aprendizagem de línguas e em testes de avaliação de proficiência. Essa escala está composta dos níveis A1, A2, B1, B2, C1 e C2. Como o QECRL (2001) contém diretrizes gerais, sem atrelar-se a uma língua específica, foram gerados documentos que, derivados da proposta, associam recursos linguísticos de línguas específicas a cada nível da escala. A seguir, serão fornecidos comentários gerais a respeito desses documentos.

Para o inglês, existe o *English Profile* (EP, 2011), que é fruto de uma pesquisa sobre os recursos linguísticos utilizados por indivíduos classificados em cada nível de proficiência da escala. Tomando-se como exemplo o nível A2, o que o EP (2011) fornece são os recursos linguísticos empregados por indivíduos que alcançaram o nível A2 de proficiência em testes dessa natureza. O inglês conta ainda com listas de vocabulário associadas a testes de proficiência para determinados níveis. Para os testes de proficiência elaborados pela Universidade de Cambridge, por exemplo, existem as listas *Key and Key for Schools Vocabulary List* (Cambridge Assessment, 2018a), referente ao nível A2, e *Preliminary and Preliminary for Schools Vocabulary List* (Cambridge Assessment, 2018b), referente ao nível B1.

No que concerne à língua espanhola, o documento derivado do QECRL (2001) que orienta o ensino-aprendizagem de espanhol como língua estrangeira é o *Plan Curricular del Instituto Cervantes* (PCIC, 2006). Esse documento fornece expoentes de recursos linguísticos que podem ser tomados como uma base mínima em termos de gramática e vocabulário.

Por serem fruto das investigações em Linguística Aplicada ao ensino de línguas, os documentos supracitados permitiriam que se estabelecesse uma correlação mais estreita entre os dicionários de aprendizes de inglês e de espanhol e o ensino-aprendizagem de ambas as línguas. Entretanto, conforme alertam Borba e Bugueño Miranda (2019a, p. 148), “não é possível afirmar que essa correlação exista realmente”. As razões para que se faça tal afirmação variam conforme a tradição lexicográfica de aprendizes. No âmbito da Lexicografia de Aprendizes de língua inglesa, os níveis indicados nas obras não permitem saber se almejam usuários cujo processo de aprendizagem está se desenvolvendo no marco dessas etapas ou se os usuários almejados são aqueles que já possuem os níveis de proficiência assinalados. No âmbito da Lexicografia de Aprendizes de língua espanhola, por sua vez, as poucas obras existentes são datadas de períodos anteriores à publicação do PCIC (2006).

Em suma, a correlação entre o perfil de usuário de dicionários de aprendizes de inglês e espanhol e a classificação de obras dessa classe ainda é fraca. Por um lado, as propostas de

classificação são relativamente recentes. Por essa razão, é precipitado esperar que os aportes desse tipo de investigação incidam em uma modificação a curto prazo na maneira como o perfil de usuário é determinado e avaliado. Por outro lado, pouco se questiona acerca da pertinência do termo “dicionário de aprendizes” a obras que não atentam para a natureza pedagógica dessa classe de obras lexicográficas. Em outras palavras, atualmente, um dicionário é pedagógico porque se diz que ele o é e/ou porque é usado em situações de ensino-aprendizagem de línguas. É por esse motivo que Borba e Bugueño Miranda (2019a), por exemplo, questionam a existência de uma lexicografia que possa se autodenominar *pedagógica* com propriedade. Nesse sentido, as obras classificadas como *dicionários pedagógicos* o seriam apenas *pro forma*, já que o fator pedagógico não está na concepção da obra, mas sim apenas na sua finalidade (uso em situações de ensino-aprendizagem).

2.2 A função dos dicionários de aprendizes e o estabelecimento do perfil de usuário

Na metalexigrafia, a função de um dicionário é um parâmetro composto de duas opções básicas: recepção e produção. Neste trabalho, opta-se por empregar o termo *recepção* ao invés de *compreensão* porque seu conceito é mais amplo que o do segundo. De acordo com Borba (2017b, p. 147), o termo *compreensão* contempla somente a dimensão do significado. Porém, há tarefas de recepção em que o significado ocupa um lugar secundário. É o caso da necessidade de confirmação de aspectos ortográficos quando se corrige um texto que possui unidades léxicas cuja ortografia causa dúvidas no leitor em relação à sua exatidão, por exemplo.

Conforme Bugueño Miranda (2019b, p. 34-36), é possível identificar duas “subfunções” atreladas a cada função: ênfase informativa no signo linguístico na sua dimensão de significante e ênfase informativa na sua dimensão de significado. Por exemplo: um dicionário de uso possui como função majoritária³⁴ a recepção linguística e sua ênfase recai sobre a dimensão do significado; um dicionário de sinônimos, por sua vez, tem como função auxiliar na produção linguística e sua ênfase também recai sobre a dimensão do significado; finalmente, um dicionário de rimas se enquadra na função de produção com ênfase no significante.

Função e perfil de usuário são dois parâmetros interdependentes. Ao se determinar o perfil de usuário antes da compilação de uma obra, considerando as suas características e possíveis necessidades, o lexicógrafo deduz qual função e subfunção seriam as mais pertinentes

³⁴ É evidente que algumas das informações fornecidas, como a transitividade verbal, são bifuncionais, ou seja, auxiliam não só na recepção, mas também na produção de língua, mesmo que o objetivo do dicionário não seja este último.

e qual classe de obra poderia responder às necessidades do usuário. Se se almeja, por exemplo, compilar um dicionário para crianças do segundo ano do ensino fundamental, tanto a função de produção como a de recepção são relevantes, pois consulentes com esse perfil precisam resolver tarefas que abarcam tanto uma como a outra. Entretanto, dificilmente um dicionário de sinônimos seria o mais adequado, embora essa classe de obras abarque um tipo de função pertinente para o perfil de usuário (função de produção com ênfase no significado). Como se trata de indivíduos que estão consolidando o sistema escrito da língua materna, a sinonímia é um fenômeno de importância secundária. Muito mais relevante seria um dicionário pela imagem, pois, embora compartilhe com o dicionário de sinônimos a mesma função e subfunção, é uma classe de obras que permite à criança corroborar ou refutar hipóteses ortográficas para resolver uma tarefa de escrita, por exemplo, sem exigir o conhecimento prévio da representação ortográfica do que ela está procurando.

Borba e Bogueño Miranda (2019a, p. 148) comentam que os dicionários de aprendizes são bifuncionais, ou seja, almejam auxiliar o seu consulente tanto na recepção como na produção de língua. Prova disso é que, paralelamente aos segmentos informativos exclusivos da função de recepção, tais como o lema, a categoria morfológica e a definição, os dicionários de aprendizes de inglês, por exemplo, incluem em seus verbetes segmentos próprios da função de produção, como é o caso dos sinônimos, colocações e notas de uso. Ainda assim, os autores alertam que a função predominante dos dicionários de aprendizes de inglês e de espanhol ainda é a de recepção. No que concerne à subfunção de um dicionário de aprendizes, os dicionários de aprendizes de inglês costumam apresentar ênfase informativa tanto em relação ao significante (ortografia e representação fonológica) como em relação ao significado (paráfrase explanatória, exemplos, notas de uso, sinônimos), sendo esta última subfunção a mais preponderante na classe de obras em questão.

A questão da função nos dicionários de aprendizes de espanhol é mais delicada. Segundo Borba (2019a), mesmo na função de recepção, as obras existentes ainda deixam muito a desejar porque essa função nem sempre pode ser identificada. Forneceremos dois exemplos que embasam a nossa afirmação. Observem-se os verbetes a seguir, reproduzidos a partir do dicionário Señas (2001) e do *Diccionario de la Lengua Española para estudiantes de español* (Espasa Calpe, 2002):

Quadro 4 – Falta de auxílio na função de recepção

<p>cham·ber·go [tʃambéryo] <i>m. fam.</i> Prenda de vestir, geralmente de abrigo: <i>salió a la calle con un ~ y una bufanda.</i> • casaca, chapéu</p> <p>(Señas, 2001)</p>	<p>fa·vor [faβór] <i>m. 1</i> Ayuda que se presta de forma gratuita: <i>Le debo muchos favores.</i> [...]</p> <p>Espasa Calpe (2002)</p>
--	---

Fonte: Señas (2001, s.v. *chambergo*) e Espasa Calpe (2002, s.v. *favor*)

No Señas (2001, s.v. *chambergo*), o exemplo fornecido não facilita a compreensão da definição; tampouco nota-se que o exemplo tenha sido incluído para auxiliar na função de produção. Embora a obra forneça equivalentes em língua portuguesa ao final do verbete, tais elementos não permitem distinguir o que, afinal, é um *chambergo*. A todos esses aspectos, soma-se ainda o fato de que, segundo dados do *Corpus del Español del Siglo XXI* (Corpes, 2018), *chambergo*, além de não apresentar uma presença significativa no *corpus* (58 registros), tem caído em desuso gradualmente (dos 58 registros, 31 são do período 2001-2005, 15 são do período 2006-2010 e 12 são do período 2011-2015). Frente a esses dados, cabe comentar que a falta de correspondência com a língua espanhola em uso é outro fator que prejudica o cumprimento da função de recepção.

O verbe *favor*, por sua vez, apresenta uma discordância entre a definição e o exemplo, tanto no que concerne à função de recepção como no que toca à função de produção. Na função de recepção, e sob a ótica da Semântica de *Frames* (FILLMORE, 1982)³⁵, a definição da acepção 1 é uma representação esquemática (*frame*) de uma das manifestações de *favor*, da qual formam os elementos “ajuda” e “ação gratuita”. O exemplo, por sua vez, contém o elemento “retribuição”, que faz alusão a uma relação de troca e que não se enquadra no *frame* expresso pela definição; trata-se, na verdade, de um *frame* diferente. Neste caso, não faz sentido exemplificar uma ação altruísta com uma oração que se refere à necessidade retribuir as ações prévias de alguém.

No tocante ao *frame* representado através do exemplo, por sua vez, e sob a perspectiva da Lexicologia e da Linguística de *Corpus*, o Espasa Calpe (2002) falha na função de produção ao não salientar que o substantivo *favor* está em uma relação de combinatória com o verbo *deber[le]*. A relação sintagmática entre *deber* e *favor* está atestada pelo *Redes - Diccionario Combinatorio del Español Contemporáneo* (Redes, 2004, s.v. *favor*) e pelo *Diccionario Combinatorio Práctico del Español Contemporáneo* (DComP, 2011, s.v. *favor*). Com o auxílio do Corpes (2018), pode-se verificar que o padrão combinatório, na verdade, inclui a duplicação

³⁵ Para uma síntese da Semântica de *Frames*, cf. Farias (2013, p. 249-250).

do dativo³⁶ (beneficiário) por meio de um pronome junto ao verbo, de forma que o padrão combinatório pode ser formulado como “deber[le/me/te/nos] + (artigo / intensificador) favor”. Do total de 9506 registros de *favor* que aparecem no Corpes (2018), 143 correspondem à combinação entre *deber* e *favor*; destas, 123 incluem a reduplicação do dativo (beneficiário). Ao não salientar esses aspectos, o dicionário acaba exigindo que o consulente interprete o comportamento do padrão combinatório em questão a partir do exemplo.

2.3 Usuário: o ilustre (e convenientemente) desconhecido

O título desta seção está baseado em uma citação de Wiegand (1977, p. 59, *apud* Bugueño Miranda, 2007b, p. 263) a respeito da dificuldade em se determinar quem é o consulente pretense no ato de compilar uma obra lexicográfica. É pertinente acrescentar que, no estado atual da questão, não faltam investigações acerca do perfil dos usuários de dicionários; porém, em contrapartida, parece haver uma falta de interesse no tema por parte dos envolvidos na compilação de obras lexicográficas, ao menos dos dicionários de aprendizes, dado que, conforme comentado no capítulo 1, o perfil de usuário dessas obras é genérico.

Há duas linhas de investigação a respeito do perfil de usuário que podem ser vistas como complementares: uma de cunho mais teórico e outra experimental. Ambas serão comentadas nas próximas subseções, seguidas por uma subseção sobre questões mercadológicas relacionadas ao perfil de usuário.

2.3.1 Propostas teóricas sobre o usuário

No que concerne a uma teoria sobre o usuário de dicionários em geral, as orientações se amparam em algumas premissas básicas. Sob uma perspectiva linguística, o usuário de qualquer dicionário deseja encontrar uma informação na obra rapidamente porque a consulta ocorre enquanto está resolvendo outra tarefa (BUGUEÑO MIRANDA, 2007b, p. 263; NESI, 2013, p. 3; NESI, 2015a, p. 69). Ou seja, a consulta ao dicionário é uma tarefa secundária para a resolução de uma tarefa primária ou principal. Sob uma perspectiva pedagógica, Bugueño Miranda (2007b) comenta que o usuário atribui ao dicionário o papel de “autoridade

³⁶ No espanhol, o elemento beneficiário da ação verbal pode aparecer: a) indicado claramente e “reiterado” por um pronome (isto é, um pleonasma), como em “[...] no **le** apetecía deberle um favor **a aquel desconocido**” (Corpes, 2018; grifo nosso); ou b) indicado apenas pelo pronome quando se pressupõe que a entidade beneficiária já é conhecida, situação ilustrada pelo exemplo do Espasa Calpe (2002, s.v. *favor* 1.).

sancionadora” em matéria de orientação no uso da língua. Em relação a essa segunda premissa, em particular, é pertinente acrescentar que Zanatta (2010, p. 19) identifica um fenômeno próprio da relação entre o usuário de uma obra lexicográfica e a obra em si: o anseio normativo, que corresponde à necessidade que o usuário tem de receber orientação em relação ao uso da língua em certas situações comunicativas nas quais, para o próprio usuário, é importante desempenhar-se adequadamente³⁷. A autora situa o anseio normativo no marco dos dicionários de língua materna; porém, o conceito pode ser aplicado a toda classe de obra lexicográfica e a qualquer perfil de usuário.

Quando se trata de dicionários pedagógicos, há mais orientações metalexigráficas acerca da determinação do perfil de usuário. Isso acontece porque o âmbito do ensino-aprendizagem de línguas permite identificar mais claramente quais são as demandas dos usuários a partir das demandas curriculares às quais estão expostos (BUGUEÑO MIRANDA, 2019b, p. 33; BORBA, 2017a, p. 24). Além das exigências do programa de ensino, Borba e Bugueño Miranda (2019a, p. 146) identificam outros dois critérios para a definição do perfil de usuário de um dicionário pedagógico: as possíveis habilidades para consultar e interpretar informações lexicográficas e as necessidades do usuário. Em relação ao caso específico dos dicionários de aprendizes, o programa de ensino é o que pode ser identificado mais facilmente (QECRL (2001) e os documentos derivados deste, por exemplo); porém, os outros dois critérios permanecem uma incógnita. No ensino-aprendizagem, as habilidades de consulta e interpretação de informações em dicionários só poderiam ser averiguadas pelo professor em sala de aula, bem como é no âmbito da sala de aula que se pode realizar intervenções para apresentar aos alunos as obras lexicográficas existentes, para que servem e como consultá-las. A esse respeito, Bogaards (2003, p. 33), ao analisar resultados de estudos empíricos acerca do uso de dicionários, comenta que “Parece que muitos aprendizes ficam relutantes em usar o dicionário e que, quando o abrem, não sabem como usá-lo. Além disso, [...] quanto mais se busca algo no dicionário, menos se encontra o que se precisa”³⁸. O comentário do autor é um alerta tanto em relação ao subaproveitamento dos dicionários no ensino-aprendizagem por parte

³⁷ A existência de um anseio normativo por parte do usuário neutraliza a distinção entre descrição e prescrição no que concerne à ênfase das obras lexicográficas em matéria de orientação no uso da língua. Tal distinção é realizada sob a perspectiva do lexicógrafo em relação à(s) norma(s) que deseja contemplar ao compilar um dicionário. No entanto, do ponto de vista do usuário, seu objetivo ao consultar um dicionário é, invariavelmente, buscar orientação sobre como usar a língua para resolver uma tarefa linguística. Assim, para o usuário, todo dicionário cumpre um papel de não apenas descrever uma ou mais normas, mas de orientar o uso da(s) mesma(s).

³⁸ [It seems that many learners are reluctant to use the dictionary and that, when they open it, they do not know how to use it. Moreover, [...] the longer one searches in the dictionary the less one finds what was needed]

de estudantes e, principalmente, de professores de língua³⁹, como também no tocante ao desencontro entre o dicionário, o programa de ensino e as necessidades dos usuários.

As necessidades do usuário, por sua vez, estão compostas por diversos fatores. Borba (2017a, p. 78), por exemplo, já questionava quais seriam as dúvidas próprias dos estudantes brasileiros no processo de aprendizagem de espanhol como língua estrangeira. Não há dúvidas de que parte das necessidades do usuário (as “dúvidas” mencionadas por Borba (2017a, p. 78) estão relacionadas ao papel que a língua materna tem no aprendizado de uma língua estrangeira. De fato, até o momento, um fator que foi bastante mencionado nos estudos metalexográficos é a interação entre a língua materna do indivíduo e a língua estrangeira que está aprendendo (v. maiores detalhes também em 2.4 e no capítulo 4). Borba e Bugueño Miranda (2019a) não são os únicos a ressaltar a importância da língua materna no momento de se estabelecer o perfil de usuário de um dicionário de aprendizes para gerar um instrumento lexicográfico o mais *tailor-made* [sob medida] possível. Há aproximadamente 20 anos, Leech e Nesi (1999, p. 300) já sugeriam que a língua materna do consulente fosse considerada no momento de se compilar um dicionário para usuários que estão em níveis iniciais de aprendizagem de uma língua estrangeira. Na mesma esteira, Tarp (2008, p. 56) também cita a língua materna como um fator a ser considerado, juntamente com outros, como a competência na língua materna (meio oral e escrito) e o nível de proficiência na língua estrangeira⁴⁰.

Dziemianko (2011, p. 54-55), por sua vez, por meio de uma revisão da literatura no que concerne ao perfil de usuário de dicionários de aprendizes de inglês, reúne críticas em relação à falta de especificação do perfil de usuário em obras dessa classe. Um dado interessante é que as críticas mais antigas datam dos anos 1970, ou seja, há mais de quarenta anos se advoga por uma especificação do perfil de usuário que vá além da premissa “estudante de inglês como língua estrangeira”. As contestações comentadas pela autora não põem em xeque a qualidade da descrição da língua inglesa nos dicionários de aprendizes; o foco recai sobre a ausência de consideração ao menos do *background* linguístico (entendido como a língua materna) dos usuários a quem se destinam os dicionários de aprendizes de inglês.

³⁹ O subaproveitamento dos dicionários por parte de professores de língua estimula trabalhos que procuram chamar a atenção para os benefícios de se consultar um dicionário para resolver tarefas linguísticas. É o caso de Borba (2017a e 2017b), nos quais, por meio da elaboração de um panorama de dicionários monolíngues de espanhol, chama-se a atenção para a profusão de obras disponíveis e também para a sua utilidade no âmbito do ensino-aprendizagem de espanhol entre brasileiros, tanto em níveis iniciais (BORBA, 2017a; BORBA, 2016) como em níveis mais avançados de aprendizagem (BORBA, 2017b).

⁴⁰ Tarp (2008, p. 54-55) acrescenta outros fatores, que envolvem o nível de conhecimento geral e em áreas de especialidade, tanto na língua materna como na língua estrangeira. O nível de conhecimento geral é um fator que dificilmente poderia ser mensurado. Já o nível de conhecimento em áreas de especialidade é um fator a ser considerado quando se deseja produzir uma obra de referência no âmbito da Terminologia.

No Brasil, as pesquisas sobre dicionários de aprendizes, embora não almejem tratar do processo de definição de um perfil de usuário de dicionário com suas características e implicações lexicográficas, reconhecem importância do português como língua materna. Essa preocupação pode ser aferida a partir de pesquisas a respeito da inclusão de paráfrases em português para definir unidades léxicas de outras línguas. A primeira proposta dessa natureza é a de Longo (2000, p. 290) a respeito de dicionários de aprendizes iniciantes de inglês. Paralelamente, Zavaglia e Zavaglia (2000a, b) e Zavaglia, A. (2000) empregam a proposta de Longo (2000) e a desenvolvem no contexto de elaboração de um dicionário trilingue temático português-francês-italiano. Mais tarde, Zavaglia, C. (2016) defende que, em dicionários bilíngues passivos (*i.e.*, obras que almejam auxiliar na compreensão de uma língua estrangeira ao descrevê-la por meio de equivalentes na língua materna), a inclusão de paráfrases na língua materna do consulente é imprescindível.

No estado atual da questão, Nesi (2015a, p. 66) mantém o mesmo ponto de vista que defendeu em 1999 no que concerne à importância de se considerar a língua materna, mas aprofunda o seu argumento, acrescentando outras variáveis que, segundo a autora, também influenciam o comportamento do usuário. Parte delas está em consonância com o que já foi comentado nesta seção, como se pode verificar a seguir:

Idade; língua materna; língua estrangeira/segunda; nível de proficiência linguística; nível educacional; nível de habilidade no uso de dicionários; atuação (como professor, aluno, tradutor [...]); local (geograficamente e em casa, no local de trabalho ou instituição educacional)⁴¹ (NESI, 2015a, p. 66).

O conjunto de variáveis proposto pela autora reúne aspectos que já vêm sendo comentados na literatura e que foram sintetizados nos parágrafos acima. Entretanto, nem Nesi (2015a) nem os demais autores considerados para a presente revisão bibliográfica fornecem detalhes a respeito do papel de cada variável no comportamento dos usuários de dicionários. Há que considerar que muitos desses autores são ou foram professores de língua estrangeira e que é muito provável que as suas afirmações estejam ancoradas, em grande medida, na sua experiência docente – o que não deixa de ser, em certa medida, a “base empírica” que motivou tais afirmações⁴². Portanto, a fim de explorar o papel e a relevância de cada variável acima no

⁴¹ [age; mother tongue; second of foreign language; language proficiency level; educational level; level of skill in dictionary use; role (as a teacher, learner, translator, [...]); location (geographically, and within the home, place of work or educational institution)]

⁴² Ao longo da história da Lexicografia Pedagógica, a experiência docente serviu de base para a compilação de dicionários de aprendizes de inglês, como o atual OALD8 (2010), tenham sido motivados pela experiência de ensino de inglês como língua estrangeira (BORBA; BUGUEÑO MIRANDA, 2019b, p. 1020-1021). Nos dias de hoje, alguns dicionários de aprendizes de inglês ainda são compilados com o auxílio de um grupo de

que concerne ao estabelecimento de um perfil de usuário, ainda neste capítulo, tais variáveis serão analisadas sob a perspectiva da Linguística Aplicada (capítulo 3) e da Ciência Cognitiva (capítulo 4). Contudo, antes de entrar no mérito dessa questão, é necessário introduzir os aportes das pesquisas empíricas sobre o perfil de usuário.

2.3.2 Pesquisas empíricas sobre o uso de dicionários

Até o momento, comentaram-se perspectivas metalexigráficas sobre a determinação do perfil de usuário de dicionários em geral e de dicionários de aprendizes, em particular. Há também uma perspectiva de investigação empírica sobre uso de dicionários, que está constituída por estudos de caso. Nesi (2015b, p. 1) afirma que esse tipo de estudo “[...] deveria ser uma parte indispensável do processo de compilação de um dicionário”⁴³. A autora salienta ainda que a importância das pesquisas sobre os usuários de dicionários reside, entre outras razões, no fato de que, “[...] sem investigação empírica, é impossível dizer como esses usuários responderiam a diferentes métodos de fornecimento de informações [...]”⁴⁴ (NESI, 2015b, p. 1), isto é, a maneira como as informações são codificadas. Domínguez Vázquez e Valcárcel Riveiro (2015, p. 166), por sua vez, são um pouco mais radicais ao defenderem que essa modalidade de pesquisa seria a forma mais adequada de estabelecer as necessidades do usuário a quem se destina um determinado dicionário. Para os autores, a carência desse tipo de investigação é contraproducente porque termina “obrigando” o lexicógrafo a intuir um perfil do usuário. Entretanto, não se pode deixar de considerar que pesquisas empíricas sobre o uso de dicionários são um entre outros recursos que podem auxiliar a compreender as necessidades do usuário. Segundo se verá nas seções 2.4 e seguintes, ainda há muitos subsídios a se aproveitar de outras áreas do conhecimento, os quais ajudam a delinear com mais precisão as características do usuário e permitem que se tomem decisões com base nesses dados⁴⁵.

De acordo com Bogaards (2003, p. 26) e Domínguez Vázquez e Valcárcel Riveiro (2015, p. 166), o interesse por investigar a relação de determinados grupos de pessoas com dicionários surgiu na década de 1980 e, desde então, têm sido alvo de estudiosos – mas ainda

profissionais responsável por contribuir com a sua experiência no ensino de inglês como língua estrangeira, tais como os consultores de didática [advisers for language pedagogy] do CALD4 (2013) e os consultores [advisory panel] do MEDAL2 (2007).

⁴³ [should be an indispensable part of the dictionary-making process]

⁴⁴ [without empirical investigation it is impossible to tell how these users might respond to different methods of information provision]

⁴⁵ Embora a proposta e aplicação de uma pesquisa empírica não tenha sido um dos objetivos desta tese, reconhecemos a sua importância caso a nossa proposta seja adotada para a compilação de um dicionário de aprendizes.

não são considerados como prioridade por parte das editoras de dicionários. Como se trata de estudos de caso, há dois aspectos gerais que permeiam esse tipo de pesquisa: as “circunstâncias locais” e as escolhas metodológicas realizadas pelos pesquisadores (NESI, 2015b, p. 1). Por “circunstâncias locais”, Nesi (2015b, p. 1) se refere aos conhecimentos prévios do usuário, tais como a sua língua materna, o nível de proficiência na língua estrangeira (quando se trata de aprendizes de língua estrangeira) e as habilidades de consulta a dicionários. Na verdade, o simples fato de considerar um determinado conjunto de variáveis para serem levadas em conta no momento de caracterizar o usuário em relação ao seu conhecimento prévio é um indício de que a pesquisa empírica, na verdade, depende da adoção de certas concepções teóricas que incidem no desenho do instrumento de coleta de dados. É pertinente questionar, primeiramente, se as habilidades de consulta não seriam um dos possíveis objetos de estudo das investigações empíricas, ao invés de uma variável “implícita”. Em segundo lugar, também cabe questionar quais seriam as premissas que os pesquisadores têm acerca da influência de cada variável no comportamento dos participantes.

As escolhas metodológicas dos pesquisadores, por sua vez, dependem do foco da pesquisa, isto é, se se almeja analisar as preferências dos usuários, as suas práticas de consulta ou a utilidade de um determinado dicionário para resolver um tipo específico de tarefa linguística; e, obviamente, dependem também dos dicionários utilizados durante a pesquisa (NESI, 2015b, p. 1).

É possível distinguir duas modalidades de pesquisa empírica sobre o usuário, uma direta e outra indireta. As pesquisas empíricas diretas têm o objetivo de coletar informações sobre o uso de dicionários. Já as pesquisas empíricas indiretas consistem no aproveitamento de informações provenientes de atividades de ensino-aprendizagem que, por alguma razão, apresentam dados considerados relevantes sobre o tema. Julgou-se conveniente estabelecer essa distinção para oferecer uma síntese mais clara dos tipos de informação empírica sobre o uso de dicionários.

2.3.2.1 Pesquisas empíricas diretas

Há seis tipos de métodos para investigar empiricamente e de maneira direta o uso de dicionários: questionários, entrevistas, observações, protocolos verbais, testes e experimentos e arquivos de registro [*log files*] (WELKER, 2010; LEW, 2011; NESI, 2015a, b). Comentaremos a seguir as principais vantagens e desvantagens de cada um.

Os questionários se destacam dentre os demais métodos em termos quantitativos, pois podem ser aplicados facilmente a um grande número de participantes; talvez, por essa razão, sejam também os mais utilizados em pesquisas empíricas sobre o uso de dicionários (NESI, 2015a, p. 63). Entretanto, os questionários também padecem de vários problemas. Nesi (2015a, p. 63), por exemplo, destaca que, muitas vezes, demandam dos participantes uma habilidade de recordação que está além das suas capacidades. Welker (2010, p. 13), por sua vez, observa que as respostas obtidas refletem aquilo que os participantes acreditam que fazem – o que, não necessariamente, corresponde ao que, de fato, acontece quando consultam um dicionário. O autor ainda comenta que os participantes podem não compreender certas perguntas dos questionários. Naturalmente, soma-se a esses problemas a questão da veracidade das respostas, pois os participantes podem conscientemente omitir certos aspectos ou inventar respostas para tentar “agradar” os pesquisadores.

Apesar dos problemas dos questionários, não se pode deixar de considerar que a maior pesquisa empírica direta já realizada sobre o uso de dicionários foi conduzida sob a forma de um questionário. Trata-se da *European Survey of Dictionary Use and Culture* (KOSEM et al., 2019), que envolveu um enorme grupo⁴⁶ de pesquisadores de diferentes países, a maioria europeus (exceto Brasil e Israel), para levantar dados acerca do uso e de concepções acerca dos dicionários gerais de língua por parte de amostras de membros de cada comunidade linguística que participou do estudo. O questionário foi traduzido para diferentes línguas, conforme o país/região participante, e foi disponibilizado para participação via site na internet durante dois meses no ano de 2017. As circunstâncias locais, isto é, as informações pessoais de cada participante, foram coletadas por meio de onze perguntas⁴⁷, enquanto as demais treze se referiram ao objetivo da pesquisa em si.

Comparadas com os questionários, as entrevistas ajudam a superar o problema da possível dificuldade de compreensão que os participantes podem ter em relação às perguntas presentes no instrumento. Segundo Nesi (2015a, p. 64), a vantagem desse método reside na

⁴⁶ Conforme Kosem et al. (2019), o núcleo central que planejou o estudo está composto de cinco pesquisadores, que elaboraram o questionário e organizaram os dados. Esse núcleo contou com a colaboração de outros 55 pesquisadores, que traduziram o questionário e foram responsáveis pela sua divulgação e aplicação em seus respectivos países.

⁴⁷ As variáveis envolvidas na coleta de informações pessoais prévias foram oito: ano de nascimento; tempo dedicado à educação formal; gênero; ocupação; se era professor; se estava obtendo ou já tinha obtido formação superior na qual a linguagem ou a linguística é um componente principal; se o trabalho do participante apresentava foco majoritário na língua; e se o participante tinha um interesse especial em línguas e/ou na sua língua materna. Na mesma esteira do que já foi comentado no início da seção 2.3.2, se os pesquisadores elencaram essas variáveis para determinar as circunstâncias locais dos participantes, está implícito que adotaram um posicionamento teórico sobre os aspectos relevantes em relação ao usuário e que, consequentemente, acreditam que essas variáveis têm alguma incidência no seu comportamento e nas suas concepções.

oportunidade que o pesquisador tem de pedir esclarecimentos ou mais detalhes aos participantes e vice-versa. Contudo, a autora salienta que muitos dos problemas verificados nos questionários se mantêm nas entrevistas, com a desvantagem de que podem ser aplicadas a um número reduzido de participantes devido à alta demanda de tempo e experiência de quem conduz a pesquisa. Lew (2011, p. 5) comenta que as entrevistas são um método pouco utilizado, o que poderia ser explicado pelas desvantagens apontadas por Nesi (2015a, p. 64).

As observações são também outro método pouco utilizado nas pesquisas sobre o uso de dicionários pelas mesmas desvantagens das entrevistas (aplicação a um número reduzido de participantes e alta demanda de tempo e experiência pelos pesquisadores). Muito embora, conforme Lew (2011, p. 5), as observações possam ser gravadas para análise e reanálise posteriores, Welker (2010, p. 15) observa que essas etapas exigem tempo de transcrição das ações dos participantes. Ademais, e visto que o método exige aplicação em um laboratório de línguas ou outro ambiente similar, o que se obtém é o uso real, mas não natural, de dicionários (WELKER, 2010, p. 15). Além disso, o autor destaca que o que se pode registrar e analisar são acontecimentos externos; não há como registrar outros aspectos relevantes relacionados ao uso de dicionários, tais como o motivo da busca, a percepção acerca do nível de auxílio dos dicionários e o grau de satisfação após a consulta.

Os protocolos verbais, por sua vez, constituem uma proposta que visa superar as limitações das observações e entrevistas. Os participantes são solicitados a expressar em voz alta (*think aloud protocols*) ou por escrito os seus pensamentos, reações e impressões que surgem durante a consulta a um dicionário. Segundo Nesi (2015a, p. 64-65), o uso de protocolos permite que se compreenda melhor o que acontece no processo de consulta a um dicionário. Contudo, para que o método funcione, é necessário treinar os candidatos a participantes a expressar seus pensamentos e selecionar os que demonstraram estar aptos a realizar esse tipo de tarefa. Esse pré-requisito reduz bastante o número de participantes. Na pesquisa por meio de protocolo verbal oral conduzida por Nesi e Boonmoh (2009), por exemplo, do número inicial de 580 participantes, somente 17 manifestaram as habilidades necessárias para seguir colaborando após o treinamento. No que concerne aos protocolos escritos, Nesi (2015, p. 65) salienta que, embora passíveis de aplicação a um número maior de participantes que os protocolos verbais orais, por exemplo, se conduzidos durante o processo de consulta a um dicionário, interferem mais que o anterior no processo de consulta. Se, por outro lado, forem conduzidos após o processo de consulta, é provável que os participantes não consigam recordar todos os detalhes da busca.

Os testes e experimentos são o(s) método(s) mais impreciso(s) em relação à sua constituição e características. Nesi (2015a) apenas os menciona, sem dar detalhes. Já Welker (2010, p. 18-20) esclarece que não estão bem definidos na metalexigrafia. Aparentemente, “[...] os testes podem ser utilizados para verificar em que medida as consultas ao dicionário [...] influenciam algum tipo de atividade linguística (leitura, aprendizagem de vocabulário, etc.)”⁴⁸ (WELKER, 2010, p. 18). Ou seja, o foco da investigação recai sobre o possível benefício da consulta para uma tarefa mais ampla ou principal. No tocante aos experimentos, apesar de Welker (2010, p. 19-20) não os definir claramente, os exemplos do autor permitem inferir que o foco recai no processo. Os experimentos envolvem procedimentos para medir uma determinada variável. É um método similar ao utilizado em determinadas áreas das ciências. Em um experimento, um grupo de participantes precisa fazer uma tarefa em um primeiro momento sem o auxílio de dicionários; em seguida, o mesmo grupo é requisitado para fazer uma tarefa similar com o uso de dicionários. Em outra modalidade, há dois grupos realizando a mesma tarefa paralelamente, um com dicionários e o outro sem esse recurso (grupo controle).

Finalmente, os arquivos de registro [*log files*] são um método que envolve dicionários eletrônicos, comumente os que estão disponíveis *on-line*. Os arquivos de registro consistem no armazenamento do que foi digitado (palavras e locuções) e dos botões que receberam cliques dos usuários (LEW, 2011, p. 5-6). Welker (2010, p. 21) acrescenta ainda como objeto de investigação as palavras e locuções que não foram digitadas, ou seja, os verbetes e partes de verbete que não foram requisitados. Esse método possui um amplo alcance em relação à quantidade de participantes (ou seja, o público que acessa o site e que, provavelmente, não está ciente do estudo). Welker (2010, p. 21) e Nesi (2015a, p. 65) citam como vantagem o fato de que os arquivos de registro não interferem no comportamento habitual dos usuários do dicionário analisado. Entretanto, e de maneira similar ao método das observações, há também ressalvas em relação ao benefício desse método. A falta de informações sobre o usuário, tais como o que Nesi (2015b, p. 1) chama de “circunstâncias locais”, a motivação da busca, as impressões do usuário acerca das informações encontradas e o eventual sucesso ou fracasso da busca são assinalados por Nesi (2015a, p. 65) e Lew (2011, p. 6) como desvantagens dos arquivos de registro no que concerne à investigação sobre o uso de dicionários.

Além das vantagens e desvantagens relacionadas a cada método, há problemas que são comuns à abordagem empírica em si. Nesi (2015b, p. 1) observa que, em geral, os resultados obtidos a partir de investigações empíricas são, com frequência, “[...] abertos a interpretações,

⁴⁸ [tests may be utilized to verify to what degree dictionary consultations [...] influence some kind of linguistic activity (reading, vocabulary learning, etc.)]

são difíceis de replicar e podem ou não ser generalizáveis a outros contextos”⁴⁹. Uma das possíveis explicações para essa condição é a ausência de estudos longitudinais, os quais incutiriam um maior rigor metodológico às investigações empíricas. Atualmente, os estudos de caso disponíveis são realizados apenas em um momento específico. Há uma carência de estudos que acompanhem os participantes em um intervalo de tempo que permita averiguar o desenvolvimento das estratégias de busca no dicionário a partir do progresso na aprendizagem da língua e das habilidades de consulta (NESI, 2015b, p. 4).

Soma-se à questão do tempo o fato de que, ainda conforme a autora, as pesquisas empíricas sobre o uso de dicionários normalmente não recebem financiamento público e, quando se trata da iniciativa privada, mais precisamente as editoras de dicionários, o excesso de envolvimento comercial pode comprometer o planejamento do estudo. Esse segundo fator permite entender por que os participantes das pesquisas sobre o uso de dicionários são, normalmente, indivíduos que frequentam o meio universitário; sem recursos financeiros, os pesquisadores, que são professores universitários, terminam tendo que restringir-se ao público da instituição à qual estão vinculados. Em função dos problemas elencados, é necessário ter cuidado ao generalizar os resultados de estudos empíricos e, ao mesmo tempo, tolerância com os avanços modestos da área, tendo em vista as limitações enfrentadas pelos pesquisadores.

Apesar das limitações comentadas *ad supra*, a partir dos estudos de caso já realizados, é possível fazer algumas afirmações gerais sobre o que já se sabe a respeito do perfil de usuário. Retornaremos a esse tema na seção 2.4 sobre o estabelecimento do perfil de usuário a ser considerado nesta tese.

Conforme comentamos no início desta subseção, além das pesquisas diretas, também há pesquisas empíricas que reaproveitam os dados gerados por atividades de ensino-aprendizagem que envolvam obras de referência. A próxima subseção está dedicada a esse tema.

2.3.2.2 Pesquisas empíricas indiretas

Como já foi mencionado no início de 2.3.2, as pesquisas empíricas indiretas consistem na análise de informações provenientes, por exemplo, de atividades de ensino-aprendizagem, ou seja, atividades cujo objetivo principal é de caráter pedagógico⁵⁰. Um primeiro exemplo disso são cursos acerca dos benefícios do uso de dicionários para fins de aprendizagem, tais

⁴⁹ [open to interpretation, are difficult to replicate, and may or may not be generalizable to other contexts]

⁵⁰ Outro exemplo de pesquisa empírica indireta é a busca pelos livros mais vendidos em livrarias, como Cultura e Saraiva, e em sites de compra e venda de livros, como Amazon e Estante Virtual.

como o relatado por Bolzan e Durão (2011). O objetivo do curso, segundo as autoras, foi a capacitação de professores de português do ensino fundamental em relação à área da Lexicografia e, conseqüentemente, ao uso de dicionários em sala de aula. Por meio do curso, Bolzan e Durão (2011, p. 186-191) tiveram acesso, entre outras informações, às percepções e hábitos de uso de dicionários em sala de aula por parte dos professores presentes.

Outro exemplo de curso de formação em lexicografia que pode fornecer dados acerca dos hábitos de uso e concepções acerca dos dicionários é o *Understanding English Dictionaries* (Coventry University), que foi disponibilizado na plataforma *on-line* Future Learn (<https://www.futurelearn.com/courses/understanding-dictionaries>) e para o qual esta autora pode contribuir como mentora. O curso teve como temas a diversidade de dicionários de língua inglesa, a estruturação dos verbetes e os tipos de informações oferecidas, a origem das informações presentes nos dicionários, o processo de compilação, técnicas definitórias e perspectiva futuras (para mais detalhes acerca da estruturação do curso, cf. Creese, McGillivray, Nesi et al. (2018)).

2.3.3 Comentário geral sobre as contribuições de pesquisas teóricas e empíricas

Esta subseção está dedicada aos aspectos das propostas teóricas e aos resultados de pesquisas empíricas que podem auxiliar na determinação de um perfil de usuário para o desenho de dicionários de aprendizes que esta tese propõe.

No tocante às propostas teóricas, as várias perspectivas comentadas em 2.3.1 permitem concluir que há variáveis diversas vezes relatadas como determinantes do comportamento do usuário, isto é, que ajudam a distinguir diferentes grupos de usuários entre si. A proposta que empregaremos nesta tese é a de Nesi (2015a, p. 66) por ser a mais detalhada e porque abrange boa parte das demais. Tal como já se havia mencionado em 2.3.1, é importante observar que o que se tem até o momento no âmbito teórico são impressões “externas” de pesquisadores envolvidos direta ou indiretamente com o ensino-aprendizagem de línguas, especialmente de línguas estrangeiras. Tais impressões levaram os pesquisadores a elencar variáveis como idade, língua materna e nível de proficiência, por exemplo. Apesar da sua relevância para a área da Lexicografia de Aprendizes e da proveitosa contribuição teórica já existente, percebe-se a necessidade de se aprofundar ainda mais as teorias sobre o perfil de usuário. Esse respaldo pode ser encontrado em uma área relativamente ampla do conhecimento, que é a Ciência Cognitiva, representada por suas diversas subáreas (Psicolinguística, Neurolinguística e Psicologia Cognitiva são algumas delas).

Em relação às pesquisas empíricas, Welker (2010, p. 3) observa que esse tipo de investigação não é comum no Brasil. De fato, do pouco que já foi feito no país, é possível contar com o trabalho de Höffling (2006), que empregou questionários e protocolos verbais junto a estudantes de licenciatura em Letras-ínglês distribuídos em diversas etapas do curso (1º ao 5º ano). O objetivo da autora foi coletar dados sobre o uso de dicionários de inglês monolíngues, bilíngues e semibilíngues. Dentre os resultados, Höffling (2006, p. 318-319) identificou uma clara preferência pelos dicionários bilíngues, dado que converge com os resultados de pesquisas internacionais da área.

Para esta tese, julgamos mais pertinente identificar alguns pontos de consenso entre os resultados de pesquisas empíricas conduzidas nos últimos trinta anos a nível nacional (mais precisamente, Höffling (2006)) e internacional. A razão para essa escolha foi a de que um conjunto maior de dados permite entender melhor o uso de dicionários no ensino-aprendizagem. As pesquisas em questão se referem ao uso de dicionários no ensino-aprendizagem de inglês como língua estrangeira, que é a classe de dicionários que conta com o maior número de dados coletados.

Dentre os pontos de convergência, Nesi (2015b, p. 2) destaca que os próprios professores de inglês preferem dicionários (monolíngues) de aprendizes para sua consulta pessoal e acabam recomendando-os a seus alunos, insistindo para que os usem em detrimento dos dicionários bilíngues. De fato, essa preferência também é comentada por Bogaards (2003, p. 27). O autor levanta a hipótese de que os professores de inglês como língua estrangeira atribuem aos dicionários (monolíngues) de aprendizes a capacidade de proporcionar aos estudantes benefícios que perdurarão a longo prazo. Bogaards (2003, p. 27) ainda comenta que essa inclinação pode estar carregando resquícios do método direto, no qual se propunha o estudo da língua estrangeira por si mesma e a exclusão da língua materna do processo de ensino.

Contudo, os estudantes de inglês como língua estrangeira tendem a preferir dicionários bilíngues de inglês (NESI, 2015b, p. 2-3). A autora enumera duas razões para essa predileção: a dificuldade de compreender a metalinguagem empregada por dicionários (monolíngues) de aprendizes e a carência de conhecimentos gramaticais suficientes pelos estudantes que os permitam compreender os códigos gramaticais usados em dicionários (monolíngues) de aprendizes. Nesi (2015b, p. 2-3) salienta também que,

“[...] embora haja evidências de que o uso do dicionário monolíngue aumente em níveis mais avançados [...], o uso do dicionário bilíngue persiste mesmo em níveis

avanzados, muitas vezes combinado com outras classes de materiais de referência [...]” (NESI, 2015b, p. 2-3)⁵¹.

Na mesma esteira, Bogaards (2003, p. 27) comenta que “[...] muitos aprendizes de L2 [segunda língua], de fato, preferem [dicionários] bilíngues, provavelmente porque eles trazem satisfação instantânea [...]”⁵². Os dados de Höffling (2006) também apontam para esta direção. Aparentemente, a preferência por dicionários bilíngues está em sentido contrário ao que esta tese propõe. No entanto, nossa proposta não pretende questionar a utilidade de um dicionário bilíngue, mas sim explorar maneiras de tornar os dicionários de aprendizes mais atrativos aos estudantes.

Ao refletir sobre as preferências de ambos os grupos (professores e estudantes de inglês), Bogaards (2003, p. 27) chama a atenção para o fato de que, apesar de existirem diversos argumentos em prol de dicionários (monolíngues) de aprendizes e diversos outros argumentos em prol de dicionários bilíngues, tais considerações estão baseadas mais em convicções pessoais que em conhecimento científico.

As observações de Nesi (2015b) e Bogaards (2003) permitem formular algumas hipóteses a respeito da predileção por uma ou outra classe de dicionários no que concerne ao contexto de ensino-aprendizagem de línguas estrangeiras/adicionais. A primeira hipótese é a de que a preferência dos professores por dicionários (monolíngues) de aprendizes de inglês pode estar ancorada na descrição mais acurada da língua em uso que essa classe de obras oferece (em comparação com os dicionários bilíngues).

A segunda hipótese que se pode levantar é a de que a preocupação dos professores em que seus alunos usem dicionários (monolíngues) para aprendizes em detrimento dos dicionários bilíngues poderia ser uma maneira de evitar a utilização da língua materna durante o aprendizado da língua estrangeira/adicional, sendo a primeira encarada como um empecilho ao desenvolvimento da aprendizagem da segunda – em outras palavras, um resquício do método direto de ensino-aprendizagem de línguas. Sob essa perspectiva, buscar-se-ia evitar o uso de materiais didáticos que considerem a língua materna, como é o caso dos dicionários bilíngues

Como terceira hipótese, a inclinação dos estudantes pelo uso de dicionários bilíngues pode estar relacionada à imediatez da solução oferecida por essa classe de obras, como assinala Bogaards (2003, p. 27). Sob uma perspectiva mais ampla, se se considera que os usuários de

⁵¹ [...] although there is some evidence that monolingual dictionary use increases at more advanced proficiency levels [...], bilingual dictionary use also persists at advanced levels, often in combination with other types of reference material [...]

⁵² [many L2 learners indeed prefer bilinguals, probably because they bring instant satisfaction [...]]

dicionários recorrem a esse tipo de instrumento enquanto estão resolvendo uma tarefa linguística mais ampla, é perfeitamente compreensível a preferência por uma obra que forneça respostas rápidas – embora nem sempre condizentes com a língua em uso ou adequadas à situação comunicativa em questão – a uma obra que demanda um maior esforço cognitivo (ver capítulo 4 para esse conceito) para procurar a informação de que se necessita.

Tais hipóteses sintetizam expectativas dos envolvidos no ensino-aprendizagem de línguas em relação ao usuário. O anseio por uma obra que descreva a língua-objeto de maneira acurada e que forneça uma solução, se não imediata, facilmente compreensível (didática), são reivindicações legítimas. Conforme já comentado, a questão da descrição da língua em uso encontra na Linguística de *Corpus* as ferramentas e a metodologia necessárias para tal. O caráter didático das informações é algo que será tratado ao longo de toda a tese, a começar pelos capítulos 3 e 4. Finalmente, a perspectiva negativa sobre a participação da língua materna no ensino-aprendizagem outras línguas é a única reivindicação que contradiz as investigações sobre como se aprende uma nova língua. Um fator tão importante quanto a qualidade da obra em termos de descrição da língua e didatismo é a noção de como os aprendizes processam uma língua não materna. Essa questão será aprofundada no capítulo 4.

2.3.4 *Influências mercadológicas no estabelecimento do perfil de usuário: uma variável indesejada*

Independentemente das duas perspectivas de investigação comentadas nas páginas anteriores, na prática, para as editoras de dicionários, é aparentemente mais vantajoso tratar o usuário como um “ilustre e convenientemente desconhecido”. Nesi (2013, p. 2) alerta para o fato de que a falta de especificação do perfil de usuário dos dicionários de aprendizes de inglês se deve a questões mercadológicas. De acordo com a autora, quanto menos específico for o perfil de usuário dessa classe de dicionários, mais ampla será a cobertura de mercado. Essa estratégia de vendas adotada pelas editoras também é comentada por Klotz e Herbst (2016, p. 33), que a justificam em função do *status* de língua global alcançado pelo inglês. Os autores ainda acrescentam que “As companhias editoriais, portanto, mais que os editores, frequentemente têm a palavra final na produção de dicionários; além disso, decisões que podem ser criticadas sob uma perspectiva metalexigráfica são, muitas vezes, bastante razoáveis em termos comerciais”⁵³ (KLOTZ; HERBST, 2016, p. 33). Na opinião de Nesi (2016, p. 585),

⁵³ [Publishers rather than editors therefore often have the final word in the production of dictionaries and decisions which can be criticised from a metalexigraphical point of view are often quite reasonable in commercial terms]

atualmente, o mercado valoriza mais os avanços técnicos na lexicografia, tais como o aprimoramento da interface (*layout*) dos dicionários e a extração automática de dados de bases lexicais (*corpora*). Tal preferência ocorre em detrimento de avanços em prol da satisfação das necessidades do usuário, a qual exige, naturalmente, que se invista em estudos sobre o estabelecimento do perfil de usuário e que os resultados dessas pesquisas sejam levados em conta pelas editoras de dicionários.

As preferências mercadológicas comentadas nesta subseção permitem entender por que os dicionários de aprendizes de inglês, por exemplo, não delimitam um perfil de usuário mais específico. Conforme comentado no capítulo 1, as editoras se limitam a assinalar que as suas obras são para todo e qualquer aprendiz de inglês que se enquadre no nível indicado. É pertinente frisar que tampouco distinguem (ou até omitem) se esse nível se refere à etapa de aprendizagem ou ao patamar de proficiência atingido.

Não negamos a importância de se produzir ferramentas que sejam aceitas no mercado e que gerem um retorno financeiro. Contudo, as preferências mercadológicas constituem uma variável indesejada quando são as únicas que detêm a palavra final sobre a compilação de um dicionário. Dessa forma, gera-se um distanciamento entre o que se pesquisa na academia e o que se produz e se oferece à população – e o maior prejudicado acaba sendo o consulente. Uma prova disso é o fato de que, dentre os quatro dicionários de aprendizes de inglês considerados na tese – CaED2 (2011), OxE2 (2012), CoED2 (2014) e LWord2 (2008) –, nenhum passou da segunda edição e o mais recente é de 2014. É importante considerar que os dicionários de aprendizes intermediários e avançados das mesmas editoras são editados até hoje com um intervalo de poucos anos entre cada edição (quatro a seis anos) e contam um número expressivo de edições (cf. capítulo 5 para mais detalhes sobre o particular). Frente a esse cenário, o fato de que os quatro dicionários supracitados não sejam mais editados é um sinal de que não foram um sucesso mercadológico, muito embora tenham sido concebidos para esse fim ao se adotar o mesmo modelo de perfil de usuário genérico dos seus pares intermediários e avançados. A conclusão a que se chega é que, ao menos quando se trata de dicionários de aprendizes de inglês de níveis iniciais, o padrão de usuário genérico não funciona. No capítulo 5, retornaremos a esse tema.

Nesta tese, defendemos que o aprimoramento do desenho de dicionários de aprendizes depende, necessariamente, da determinação do perfil de usuário, além do aprimoramento da descrição de combinatórias léxicas. Na próxima seção, os insumos teóricos e práticos comentados anteriormente serão aplicados para determinar as características do perfil de usuário desta tese.

2.4 O perfil de usuário proposto para esta tese

Segundo foi comentado em 2.3, Nesi (2015a, p. 66) enumera uma série de variáveis que determinam o comportamento do usuário. A partir dessas variáveis e das dificuldades dos consulentes de dicionários de aprendizes de inglês, comentadas em 2.3.3, é possível estabelecer com mais precisão o perfil de usuário a quem se destinam as propostas de aprimoramento que serão feitas no decorrer da tese. Trata-se de indivíduos residentes no Brasil (local), adultos (idade), que falam português brasileiro (língua materna), com Ensino Médio completo (nível educacional), que são estudantes (atuação) de inglês ou de espanhol (língua estrangeira), que atingiram os níveis de proficiência A1 ou A2 nessa língua estrangeira e que não estão habituados a usar um dicionário (nível de habilidade no uso de dicionários).

A escolha pela nacionalidade brasileira e a determinação do português brasileiro como língua materna se devem às dificuldades que o público em questão enfrenta ao tentar aperfeiçoar a sua competência linguística em língua inglesa e em língua espanhola (ver Introdução da tese). Tais dificuldades, somadas à ausência de dicionários de aprendizes de espanhol que visem a transição dos níveis básicos de proficiência para os intermediários, justificam ainda a delimitação dos níveis A1 e A2 como sendo a proficiência esperada da parte do usuário brasileiro falante de português como língua materna. Ademais, e no que concerne à nacionalidade do público-alvo considerado nesta tese, cabe mencionar o fato de que esta pesquisa foi desenvolvida no Brasil e subsidiada por recursos públicos do país visando a melhoria da educação nacional.

A faixa etária estabelecida (adultos) e o nível educacional (Ensino Médio completo, que consiste na última etapa que conforma a educação considerada de nível básico⁵⁴ no Brasil) foram determinados com base em dois aspectos. Primeiramente, optou-se por potenciais consulentes com essas características porque abrangem pessoas que têm uma necessidade mais imediata em aprender e aprimorar seus conhecimentos em inglês e espanhol, seja por estarem em período de formação técnica ou superior, seja por estarem ingressando ou já terem ingressado no mercado de trabalho. Consideramos que os adultos têm uma urgência maior no que tange à aprendizagem dessas línguas, seja para facilitar a sua inserção no mercado de trabalho, seja para ascender para cargos mais bem valorizados. Isso não significa, contudo, que, futuramente, esta proposta não possa ser aplicada para o desenho de um dicionário de inglês ou de espanhol para alguma etapa da educação básica. Outra razão que levou à delimitação do

⁵⁴ Conforme a Lei de Diretrizes e Bases da Educação Nacional (Lei 9394/96), a educação básica compreende as etapas Educação Infantil, Ensino Fundamental, Ensino Médio e Educação de Jovens e Adultos.

Ensino Médio completo como nível educacional foi a adequação aos tipos de dados aos quais tivemos acesso ao avaliar as contribuições da Ciência Cognitiva para a tese. Os indivíduos que aprendem uma nova língua já adultos e que são alvo de estudos cognitivos relevantes para esta tese são os que completaram a formação considerada básica nos sistemas educacionais em que se encontram inseridos.

A atuação, por sua vez, corresponde ao papel de estudantes cujo processo de aprendizagem está inserido no marco de um programa de ensino – mais precisamente, a escala de níveis do QECRL (2001) e seus documentos derivados. Esses estudantes podem ser tanto os que estão inseridos em um contexto institucional de ensino como os que conduzem a sua aprendizagem fora dos marcos de uma instituição de ensino (isto é, de maneira “autodidata”). Em ambos os casos, é comum que se recorra a materiais didáticos, mais especificamente livros didáticos, para auxiliar na condução da formação. Atualmente, os livros didáticos tanto de inglês como de espanhol como língua estrangeira adotam a escala de níveis do QECRL (2001), o que significa que, direta ou indiretamente, a formação desses estudantes está inserida na proposta do QECRL (2001)⁵⁵.

Finalmente, partimos de dois pressupostos para determinar o nível de habilidade no uso de dicionários. Primeiramente, devido à carência de pesquisas sobre o uso de dicionários no Brasil, optamos por partir da hipótese de que o usuário que nos interessa nesta tese não está habituado a buscar informações em um dicionário. Em segundo lugar, considerando alguns resultados gerais de estudos empíricos, que reportaram dificuldades dos consulentes em relação à compreensão da metalinguagem e dos códigos gramaticais em dicionários de aprendizes de inglês, partimos do pressuposto de que o consulente alvo desta proposta pode apresentar dificuldade em interpretar esse tipo de informação. Em termos práticos, isso significa que as informações precisam ser apresentadas de maneira clara (para que sejam encontradas facilmente) e didática (isto é, que facilite a sua interpretação).

Conforme comentado em 2.3.1 e 2.3.3, é necessário aprofundar o respaldo teórico que já existe e que, lamentavelmente, não recebeu, até agora, a devida atenção. Esse aprofundamento significa, por um lado, compreender de maneira mais detalhada qual é o papel de cada variável no comportamento do usuário e, por outro lado, que se derive dessas variáveis

⁵⁵ Segundo consulta realizada em 20 set. 2019 ao *site* da Amazon Brasil (https://www.amazon.com.br/gp/bestsellers/books/7882705011/ref=zg_bs_pg_1?ie=UTF8&pg=1), os exemplares das coleções *English for Everyone*, *Interchange*, *English File* e *Top Notch* estão entre os cem livros mais vendidos no Brasil no quesito “Importados de Educação, Referência e Didáticos”. Todos empregam a escala de níveis do QECRL (2001). Na mesma lista, no tocante aos livros didáticos de espanhol, exemplares da coleção *Gente Hoy* aparecem como os mais vendidos.

uma metodologia didática de descrição da língua em dicionários de aprendizes. Para tanto, e como consta nas seções a seguir, recorre-se à Linguística Aplicada para tratar das variáveis língua estrangeira, nível de proficiência na língua estrangeira e atuação; e à Ciência Cognitiva no que concerne às variáveis idade, língua materna, língua estrangeira, nível de proficiência linguística, nível educacional e local.

3 O PERFIL DE USUÁRIO SOB A ÓTICA DA LINGUÍSTICA APLICADA

“Nós temos dicionários para aprendizes, mas não realmente para aprender” (SCHOFIELD, 1999 *apud* BOGAARDS, 2003, p. 28)⁵⁶. Iniciamos esta seção com uma citação que parece traduzir de maneira simples a distância dos dicionários de aprendizes em relação ao usuário, mais precisamente no tocante à falta de relação com um programa de ensino e à eventual apresentação pouco didática das informações. De fato, uma das principais críticas feitas aos dicionários de aprendizes no capítulo 1 desta tese, em Borba e Bugueño Miranda (2018, 2019a, b) e em Borba (2020) diz respeito à dissociação entre o dicionário de aprendizes e o ensino-aprendizagem de línguas estrangeiras.

Conforme Hall (2016, p. 213), quando foram lançadas as primeiras edições do dicionário de aprendizes atualmente conhecido como *Oxford Advanced Learner's Dictionary* (cf. referência em OALD8 (2010)), a ênfase que os seus autores atribuíram à descrição do comportamento sintático da língua foi um reflexo do método audiolingual, em voga na época. As edições subsequentes do dicionário, bem como os demais dicionários de aprendizes que foram surgindo na tradição anglo-saxônica, afastaram-se dos sucessivos métodos e concepções de ensino de línguas que foram surgindo ao longo do século XX no âmbito da Linguística Aplicada (BORBA; BUGUEÑO MIRANDA, 2019b, p. 1020-1021; 1025). Entretanto, essa não parece uma postura acertada da parte dos compiladores de dicionários de aprendizes, pois, se almejam auxiliar no processo de ensino-aprendizagem de línguas estrangeiras/adicionais, e se esse processo está naturalmente influenciado pelos métodos e concepções de ensino vigentes, é necessário que o dicionário se adapte a esse cenário. Por outro lado, também é necessário compreender como o dicionário de aprendizes pode ser inserido no ensino-aprendizagem de línguas, isto é, como essa classe de instrumento lexicográfico pode contribuir num ensino que prega a promoção da autonomia na aprendizagem e estimula práticas como o andamento [*scaffolding*]. As respostas para essas questões estão na maneira como o estudante de línguas é concebido pelas práticas, objetivos e métodos de ensino amplamente difundidos pela Linguística Aplicada ao ensino de línguas.

Primeiramente, algumas das práticas de ensino a que nos referimos se situam no marco da Teoria Sociocultural, que promove uma docência cada vez mais centrada no uso da língua como meio de interação social. A Teoria Sociocultural se ampara em noções propostas por Vygotsky (1991) acerca do desenvolvimento infantil e as expande para o ensino-aprendizagem

⁵⁶ [We have dictionaries for learners, but not really for learning]

de línguas em geral. O conceito de zona de desenvolvimento proximal (ZDP) e a prática do andamento servem de fundamento para as orientações de como precisa ser a prática da avaliação no ensino de línguas. Nesse cenário, e segundo será comentado mais detalhadamente em 3.2, o dicionário figura como um instrumento que auxilia alunos e professores nas práticas docentes que sucedem a avaliação.

Em segundo lugar, os objetivos do ensino-aprendizagem de línguas dependem do método e da concepção de proficiência adotados. Porém, em geral, percebe-se que o desenvolvimento da autonomia na aprendizagem é um objetivo que transcende as diferentes perspectivas sobre o ensino de línguas. Conforme também se comentará em 3.2, o dicionário pode se converter em um importante aliado dos aprendizes no desenvolvimento da sua autonomia.

Finalmente, e conforme Hall (2016, p. 214-215), o que está em voga atualmente não é necessariamente um método específico, mas sim uma abordagem que, ao longo do tempo, adquiriu uma certa variação interna no seu *modus operandi*⁵⁷: a abordagem comunicativa. Essa perspectiva está centrada no conceito de competência comunicativa e se difundiu com a publicação do *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR, 2001)⁵⁸. De acordo com o *Diccionario de Términos Clave de ELE* (DiClavELE, 2008), a competência comunicativa diz respeito à

[...] capacidade de uma pessoa de se comportar de maneira eficaz e adequada em uma determinada comunidade de fala; isso implica respeitar um conjunto de regras que inclui tanto as da gramática e outros níveis da descrição linguística (léxico, fonética, semântica) como as regras de uso da língua, relacionadas com o contexto socio-histórico e cultural no qual ocorre a comunicação. [...] ou seja, trata-se da capacidade de formar enunciados que não só sejam gramaticalmente corretos, mas também socialmente apropriados. (DiClavELE, 2008, s.v. *competência comunicativa*)⁵⁹.

A competência comunicativa é um conceito amplo que, segundo o DiClavELE (2008, s.v. *proficiencia*), abarca também o conhecimento e uso tanto da língua materna como de

⁵⁷ As perspectivas *Content and Language Integrated Learning* [Aprendizagem Integrada de Conteúdos e Língua] e *Task-based Language Teaching* [Ensino de Línguas baseado em Tarefas] são exemplos dos desdobramentos da abordagem comunicativa (cf. Morton (2016) e Van den Branden (2016) para mais detalhes sobre cada uma).

⁵⁸ Segundo Borba (2017c), é curioso notar que, em nenhum momento, o CEFR (2001) define o termo *competência comunicativa*. Trata-se de uma falha grave, pois o documento representa um marco importante na história do ensino de línguas estrangeiras/adicionais e se dirige a um público bastante variado (professores, compiladores de materiais didáticos, entre outros) e que não tem por que conhecer, *a priori*, o conceito por detrás do termo.

⁵⁹ [...] capacidad de una persona para comportarse de manera eficaz y adecuada en una determinada comunidad de habla; ello implica respetar un conjunto de reglas que incluye tanto las de la gramática y los otros niveles de la descripción lingüística (léxico, fonética, semántica) como las reglas de uso de la lengua, relacionadas con el contexto socio-histórico y cultural en el que tiene lugar la comunicación. [...] es decir, se trata de la capacidad de formar enunciados que no solo sean gramaticalmente correctos sino también socialmente apropiados]

outra(s) língua(s). De 2001 até os dias de hoje, essas outras línguas foram sendo denominadas por meio de três termos: *língua estrangeira*, *segunda língua* e *língua adicional*. Entretanto, os termos em questão não são definidos nem no CEFR (2001) nem no CEFR-Co (2018). Portanto, a primeira seção do capítulo se dedica a uma distinção terminológica. As demais seções discorrem sobre as contribuições da área na definição do perfil de usuário a partir da aproximação entre o instrumento dicionário e o ensino comunicativo.

3.1 Fronteiras conceituais entre língua estrangeira, segunda língua e língua adicional

Quando foi publicado, o CEFR (2001) empregou somente o termo *língua estrangeira* para referir-se a cada nova língua que um indivíduo se propunha a aprender. Concomitantemente, no mesmo ano, um órgão vinculado à Unesco publicou um outro documento, que propunha a adoção do termo (designação e conceito) *língua adicional* em detrimento de *língua estrangeira*. Recentemente, foi publicado um documento complementar ao CEFR (2001), o *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion volume with new descriptors* (CEFR-Co, 2018), no qual se usa majoritariamente o termo *língua estrangeira* [foreign language] e, em menor escala, *segunda língua* [second language] e *língua adicional* [additional language]. É notória, no entanto, a falta de uniformidade terminológica do CEFR (2001) e do CEFR-Co (2018) no que diz respeito às designações e, principalmente, aos conceitos.

Determinar a que conceitos aludem os termos mencionados acima é importante para que se possa empregar a designação mais adequada nesta tese ao se referir às línguas inglesa e espanhola; além disso, ajuda a compreender em que consiste a variável *língua estrangeira*, chamada assim por Nesi (2015a) para referir-se a um dos aspectos que determina o perfil de usuário de um dicionário (v. seções 2.3.1, 2.3.3 e 2.4). A seguir, reproduzimos algumas definições encontradas em dicionários de Linguística e de Linguística Aplicada, e em manuais de Linguística Aplicada.

Littlewood (2004, p. 502) comenta que

Alguns escritores fazem uma distinção entre uma “segunda” língua, a qual possui funções sociais na comunidade onde é aprendida (p.ex. o inglês na Índia ou o inglês aprendido por um imigrante nos Estados Unidos), e uma língua “estrangeira”, a qual

é aprendida para contato fora da comunidade (p.ex. o francês aprendido no Reino Unido) (LITTLEWOOD, 2004, p. 502)⁶⁰.

Segundo se infere da definição do autor, o termo *segunda língua* se refere a: a) uma língua que é utilizada em uma comunidade linguística em menor escala em comparação a outra língua, mais utilizada (exemplo da Índia, que tem o hindi, o inglês e outras línguas como línguas oficiais); ou b) uma língua que não é a língua materna de um indivíduo, mas que é a mais utilizada na comunidade linguística na qual esse indivíduo está inserido e que precisa ser usada por ele no seu cotidiano (exemplo dos Estados Unidos). Em ambos os casos, há uma necessidade de uso da segunda língua no seio de uma comunidade linguística. Do termo *língua estrangeira*, por sua vez, infere-se que se trata de uma língua que não é necessária para a inserção e convivência em uma comunidade linguística.

Já Crystal (2008, s.v. *language*) faz a seguinte distinção:

[...] **primeira língua** (sc. língua materna) se distingue de **segunda língua** (uma língua outra que não a língua materna, usada para um propósito especial, p. ex., para educação, governo), que, por sua vez, se distingue de **língua estrangeira** (que não implica nenhum *status* especial) – embora a distinção entre as últimas duas não está universalmente reconhecida (especialmente nos Estados Unidos) (CRYSTAL, 2008, s.v. *language*. Grifos do autor)⁶¹.

A definição de Crystal (2008) pode até permitir que se entenda o que é *segunda língua* a partir da restrição em relação às situações de uso; porém, o termo *língua estrangeira* ainda permanece sem definição.

O DiClavELE (2008, s.v. *lengua meta*), por sua vez, oferece uma definição mais clara:

Atendendo às situações e circunstâncias em que se aprende a LM [*i.e.* língua meta], costuma-se distinguir entre língua estrangeira (LE) e segunda língua (L2). Quando a LM é aprendida em um país onde não é oficial nem autóctone, é considerada uma LE; p.ex., o espanhol no Brasil. Quando a LM é aprendida em um país onde coexiste como oficial e/ou autóctone com outra(s) língua(s), é considerada uma L2; p.ex., o guarani para as crianças paraguaias cuja L1 é o espanhol, ou também o espanhol para as crianças paraguaias cuja L1 é o guarani. Outra situação cada vez mais frequente de L2 é a que se produz quando os emigrantes com uma L1 em comum chegam a constituir uma comunidade de fala relevante no país de acolhida, como é o caso dos hispânicos nos Estados Unidos. Em situações como esta, de um ponto de vista estritamente linguístico, parece pertinente começar a considerar que a língua dessa comunidade de fala (o espanhol nesse exemplo) pode deixar de ser uma LE para se

⁶⁰ [Some writers make a distinction between a “second” language, which has societal functions in the community where it is learnt (e.g., English in India or English as learnt by an immigrant to the USA), and a “foreign” language, which is learnt for contact outside the community (e.g., French as learnt in the UK)]

⁶¹ [...] **first language** (sc. mother tongue) is distinguishable from **second language** (a language other than one’s mother-tongue used for a special purpose, e.g., for education, government), distinguishable in turn from **foreign language** (where no such special status is implied) – though the distinction between the latter two is not universally recognized (especially not in the USA)]

converter em uma L2, independentemente de sua possível co-oficialidade (DiClavELE, 2008, s.v. *lengua meta*)⁶².

Conforme o DiClavELE (2008, s.v. *lengua meta*), os requisitos determinantes para que se possa diferenciar os termos que vêm sendo comentados aqui têm a ver com as circunstâncias de uso da língua “alvo” em questão em uma dada comunidade. Tais fatores são, mais especificamente, o caráter oficial, a natureza autóctone e a formação de comunidades linguísticas de emigração. Se uma língua preenche um desses três requisitos, pode ser considerada uma segunda língua; se não, será considerada uma língua estrangeira. Nesses termos, o exemplo do espanhol em relação ao Brasil já esclarece que se trata de uma língua estrangeira. O mesmo se aplica ao caso do inglês em relação ao Brasil.

Na mesma esteira, em seu *Longman Dictionary of Language Teaching & Applied Linguistics*, Richards e Schmidt (2010, s.v. *foreign language*) definem *língua estrangeira* da seguinte forma:

Uma língua que não é a língua nativa de um grande número de pessoas em um determinado país ou região, não é usada como um meio de instrução em escolas e não é amplamente usada como um meio de comunicação no governo, mídias, etc. Línguas estrangeiras são tipicamente ensinadas como matérias escolares com o propósito de comunicar-se com estrangeiros ou para a leitura de materiais impressos nessa língua (RICHARDS; SCHMIDT, 2010, s.v. *foreign language*)⁶³.

A definição reforça a falta de preenchimento dos requisitos enumerados *ad supra*, pois depreende-se que uma língua estrangeira é de natureza não autóctone (não é língua nativa) e de caráter não oficial (não é meio de instrução nem de comunicação oficial e midiática). O *status* do inglês e do espanhol (neste último caso, com exceção das zonas de fronteira com países hispânicos) no Brasil condiz com o termo *língua estrangeira*. Os autores também definem *segunda língua* de uma forma similar ao DiClavELE (2008), amparando-se nas circunstâncias

⁶² [Atendiendo a las situaciones y circunstancias en que se aprende la LM, se suele distinguir entre lengua extranjera (LE) y lengua segunda (L2). Cuando la LM se aprende en un país donde no es ni oficial ni autóctona, se considera una LE; p. ej., el español en Brasil. Cuando la LM se aprende en un país donde coexiste como oficial y/o autóctona con otra(s) lengua(s), se considera una L2; p. ej., el guaraní para aquellos niños paraguayos cuya L1 es el español, o bien el español para aquellos niños paraguayos cuya L1 es el guaraní. Otra situación cada vez más frecuente de L2 es la que se produce cuando los emigrantes con una L1 común llegan a constituir una comunidad de habla relevante en el país de acogida, como es el caso de los hispanos en EE.UU. En situaciones como ésta, desde un punto de vista estrictamente lingüístico, parece pertinente empezar a considerar que la lengua de esa comunidad de habla (el español en este ejemplo) puede dejar de ser una LE para convertirse en una L2, al margen de su posible cooficialidad]

⁶³ [a language which is not the native language of large numbers of people in a particular country or region, is not used as a medium of instruction in schools, and is not widely used as a medium of communication in government, media, etc. Foreign languages are typically taught as school subjects for the purpose of communicating with foreigners or for reading printed materials in the language]

de uso da língua considerada segunda em uma comunidade linguística e na necessidade de conhecimento dessa língua por questões de sobrevivência na sociedade (cf. Richards e Schmidt, 2010, s.v. *second language*).

O termo língua adicional, por sua vez, foi proposto por meio do documento *Teaching Additional Languages* (JUDD et al., 2001). Seus autores são especialistas associados ao International Bureau of Education, uma organização vinculada à Unesco que tem por objetivo difundir, a nível mundial, soluções para problemas relacionados ao ensino tanto de língua como de outras matérias específicas, bem como no tocante a práticas pedagógicas.

No documento, seus autores esclarecem que sua contribuição se enquadra na abordagem comunicativa de ensino de línguas, tanto para estudantes que aprendem línguas não maternas em seu país de origem como para aqueles que estão em situação de migração (JUDD et al., 2001, p. 3; 6). São apresentados quatro argumentos para defender o uso do termo *língua adicional*. O primeiro deles é que “os estudantes devem, na verdade, estar aprendendo não uma segunda, mas uma terceira ou quarta língua. ‘Adicional’ se aplica a todas, exceto, é claro, à primeira língua aprendida” (JUDD et al., 2011, p. 6)⁶⁴. Nessa afirmação, na qual critica-se o uso do termo *segunda língua*, não fica claro o conceito ao qual se referem. Se for o mesmo presente no DiClavELE (2008) e em Richards e Schmidt (2010), isto é, se Judd et al. (2001) se referem a uma língua oficial, e/ou usada em determinadas situações (governo, mídia) ou comunidades de fala, excluem-se contextos de aprendizagem em que há uma única língua oficial e autóctone, como é o caso da maior parte do território brasileiro em relação ao uso do espanhol e do inglês. Se os autores, por outro lado, usaram o termo *segunda língua* para abarcar os termos segunda língua e língua estrangeira comentados acima, não é possível entender por que não usar variantes do termo *segunda língua*, tais como *terceira língua* ou *quarta língua*, como uma forma de cada indivíduo expressar a sua trajetória de aprendizagem. Seja qual for o caso, falta uma maior especificação sobre o conceito de *segunda língua* que serviu de base para a sua afirmação.

O segundo argumento é o de que “uma língua adicional [...] pode não ser estrangeira porque muitas pessoas no seu país podem falá-la normalmente” (JUDD et al., 2001, p. 6)⁶⁵. Essa é uma suposição demasiado genérica para que se possa defender o uso do termo *língua adicional*. Os dados do British Council (2014) a respeito da percepção dos brasileiros sobre o seu conhecimento de inglês (cf. introdução da tese) é uma prova de que, sob esse viés, o inglês

⁶⁴ [Students may actually be learning not a second but a third or fourth language. ‘Additional’ applies to all, except, of course, the first language learned]

⁶⁵ [An additional language [...] may not be foreign since many people in their country may ordinarily speak it]

e o espanhol não poderiam ser considerados línguas adicionais por uma esmagadora maioria de brasileiros.

O terceiro e o quarto argumentos se referem a possíveis preconceitos que o adjetivo *estrangeira* pode suscitar:

O termo ‘estrangeiro’ pode [...] sugerir estranho, exótico ou, talvez, alheio – todas conotações indesejáveis. Nossa escolha pelo termo ‘adicional’ assinala nossa crença de que línguas adicionais não são necessariamente inferiores nem superiores, nem uma substituição da primeira língua do estudante (JUDD et al., 2001, p. 6)⁶⁶.

Tais argumentos supõem, primeiramente, que a outra língua em questão está necessariamente inserida e bem estabelecida entre as atividades rotineiras de uma comunidade linguística e/ou de um indivíduo – pressuposto consoante ao conceito de *segunda língua*. Na verdade, isso termina sendo uma espécie de imposição de um conceito à realidade de uma comunidade linguística em detrimento da descrição do que realmente acontece linguisticamente nessa comunidade. Os baixos índices de conhecimento e capacidade de falar em inglês, comentados na introdução desta tese, são uma prova de que os brasileiros em geral não a usam em suas atividades rotineiras. Em segundo lugar, e no caso do Brasil, não há como depreender que o termo *língua estrangeira* se encontra em uma posição superior ou inferior em relação ao português ou a outras línguas como o espanhol. A última afirmação pode também levar a supor que todas as línguas não maternas possuem igual grau de importância e relevância em qualquer comunidade linguística. Contudo, isso novamente desconsidera a percepção que os membros de uma comunidade linguística têm em relação à importância das línguas não maternas nas diversas instâncias de uma sociedade (economia, política, educacional etc.). Segundo o British Council (2014, p. 18-20), entre os brasileiros, existe a crença de que o conhecimento em língua inglesa facilita a inserção e a ascensão no mercado de trabalho. Isso significa que, na concepção dos brasileiros, a língua inglesa está em um patamar de destaque. Não se trata de algo bom ou ruim, mas, sim, da realidade em si.

Em suma, percebe-se que Judd et al. (2001) tentam cunhar um termo para adequar-se à trajetória individual de cada estudante. Não nos parece ser um problema que um indivíduo decida referir-se à língua não materna que conhece ou que está aprendendo como uma língua adicional, pois essa decisão é de caráter privado. No entanto, não são apresentados argumentos (sócio)linguísticos para defender o uso do termo *língua adicional* para referir-se à possível

⁶⁶ [The term ‘foreign’ can [...] suggest strange, exotic or, perhaps, alien—all undesirable connotations. Our choice of the term ‘additional’ underscores our belief that additional languages are not necessarily inferior nor superior nor a replacement for a student’s first language]

função que uma língua não materna desempenha em uma comunidade linguística. Não se considera, sob uma perspectiva sociolinguística, que essa função varia conforme a língua e a comunidade. Os exemplos presentes em Littlewood (2004), Richards e Schmidt (2010, s.v.s. *foreign language* e *second language*) e no DiClavELE (2008, s.v. *lengua meta*) ilustram uma diversidade de tipos de relação entre comunidades linguísticas e línguas não maternas que não se vê refletida, mas ofuscada, pelo termo genérico *língua adicional*. Esse termo termina não sendo inclusivo, pois as definições em Judd et al. (2001) não abarcam todas as situações, mas somente algumas. Essa diversidade, aliás, dificilmente seria contemplada por um único termo. Nesse sentido, parece mais sensato adotar a distinção *segunda língua* e *língua estrangeira*, desde que se esteja ciente do conceito abarcado por cada um. Nesta tese, o termo *língua estrangeira* será empregado para referir-se ao tipo de relação que as línguas inglesa e espanhola têm com a maior parte da comunidade linguística brasileira.

Após a necessária distinção terminológica inicial, as seções que seguem discorrem sobre os subsídios da Linguística Aplicada ao ensino de línguas no tocante à determinação do perfil de usuário do dicionário. As próximas seções discorrem sobre como o estudante de línguas é concebido nas diversas perspectivas da Linguística Aplicada ao ensino de línguas.

3.2 Práticas e objetivos de ensino e a concepção do perfil de usuário

As práticas docentes baseadas na Teoria Sociocultural, bem como o estabelecimento da autonomia como objetivo de aprendizagem, estão em voga atualmente (THORNE; TASKER, 2011; PALFREYMAN; BENSON, 2019) e encontram lugar na abordagem comunicativa materializada pelo QECRL (2001). Embora a Teoria Sociocultural não esteja mencionada no documento, de acordo com Negueruela-Azarola e García (2016, p. 298), está intimamente associada à abordagem comunicativa em função da valorização da interação como elemento central da aprendizagem. Além disso, seus pressupostos são convergentes com alguns dos tipos de avaliação mencionados pelo QECRL (2001, p. 251-264), como é o caso da avaliação de desempenho. A autonomia, por sua vez, é citada no QERCL (2001, p. 199; 208) como uma capacidade que precisa ser promovida por ser parte integral da aprendizagem. Nesse cenário, o usuário do dicionário é definido como alguém que precisa de mediação tanto para aprender a língua estrangeira como para alcançar uma autonomia na sua aprendizagem. Os parágrafos que seguem buscam discorrer sobre o lugar do dicionário no ensino comunicativo de língua estrangeira e, conseqüentemente, no processo de aprendizagem dos alunos, na sua condição de usuários almejados por um dicionário de aprendizes. Nosso objetivo é demonstrar a necessidade

que o ensino comunicativo e, particularmente, os estudantes de língua estrangeira, têm de um instrumento como o dicionário de aprendizes.

A Linguística Aplicada adaptou os pressupostos da Teoria Sociocultural para o ensino-aprendizagem de línguas, pregando a difusão de certas práticas docentes centradas nos alunos, tais como a avaliação do potencial de desenvolvimento de cada indivíduo e a mediação da aprendizagem. A Teoria Sociocultural se fundamenta em Vygotsky (1991, p. 97), que propõe que as crianças possuem um determinado potencial de aprendizagem, chamado de zona de desenvolvimento proximal (ZDP), que é único para cada indivíduo. Conforme Vygotsky (1991, p. 101-102), aprender significa avançar por essa ZDP. Para que o aprendizado se efetive, é preciso interação (aspecto social) e mediação (aspecto cultural) (VYGOTSKY, 1991, p. 59-63; 97).

A mediação é o emprego de elementos culturalmente adquiridos para lidar com uma tarefa, sejam eles objetos, como o dicionário, ou signos, como a língua. A mediação por objetos é o primeiro de uma escala de três níveis de mediação (LANTOLF; THORNE, 2007, p. 199-200). Conforme Thorne e Tasker (2011, p. 490), trata-se de um estágio em que uma tarefa somente poderá ser desenvolvida por intermédio de um objeto, ou seja, há uma maior dependência da mediação. Um exemplo de mediação por objetos é o uso de um dicionário bilíngue para executar uma tarefa de compreensão leitora em estágios iniciais de aprendizagem de língua estrangeira.

O segundo nível de mediação é a regulação pelo outro, isto é, a interação com o professor, um colega etc., no qual a língua é o elemento intermediador para terminar uma tarefa, revisá-la ou aprimorá-la (LANTOLF; THORNE, 2007, p. 200). Nesse nível, embora a necessidade de mediação seja menor, ela ainda se faz necessária. Mais conhecida como andaime [scaffolding], esse tipo de mediação visa o auxílio na execução de uma tarefa visando um avanço pela ZDP (GIBBONS, 2015, p. 16). Em outras palavras, o andaime não é um auxílio que tem na finalização da tarefa o seu fim último, mas sim uma oportunidade de avanço pela ZDP por meio do auxílio na execução da tarefa que tem a aprendizagem como fim.

O andaime é uma prática que parece ser estimulada especialmente na avaliação de tarefas comunicativas devido à associação entre aprendizagem e interação. Nesse contexto, o andaime ocorre no momento de expor a avaliação ao aluno, na qual o professor esclarece o que está satisfatório e o que precisa de ajustes (avaliação de desempenho). Quando os ajustes envolvem recursos linguísticos nos níveis ortográfico, fonológico, morfológico, sintático, léxico-semântico ou pragmático, o professor tem duas opções: ou indicar qual é o uso correto, fornecendo uma “resposta pronta”; ou salientar o que precisa de ajustes, mas sem dar respostas,

e encaminhar o aluno para uma mediação pelo objeto dicionário. Conforme Zanatta (2010, p. 66), dicionários como os de aprendizes têm uma estrutura que permite dar conta de informações ortográficas (variação ortográfica, acentuação), fonológicas (pronúncia, sílaba tônica), morfológicas (classe de palavras, flexão de nomes e verbos), sintáticas (posicionamento sintático e padrões combinatórios com preposições e advérbios), pragmáticas (marcas de uso socioculturais) e léxico-semânticas (significado, padrões combinatórios e sinônimos). Entretanto, um dicionário de aprendizes não é capaz de fornecer orientações a nível textual/discursivo (gênero textual e discursivo, estruturação do texto e coerência, por exemplo); essa incumbência cabe ao professor.

O terceiro nível de mediação é a autorregulação, que equivale a um

[...] ganho maior de controle voluntário sobre a própria capacidade de pensar e agir, seja tornando-se mais proficiente no uso de recursos de mediação, seja por uma diminuição da dependência em relação a meios de mediação externos⁶⁷ (THORNE; TASKER, 2011, p. 490).

Portanto, a autorregulação não implica necessariamente a independência da mediação, mas sim a consciência sobre o que precisa ser feito, que tipo de mediação é necessária e onde encontrá-la. Esse nível de mediação se assemelha muito à noção de autonomia na aprendizagem, que é um objetivo muitas vezes citado, mas poucas vezes “traduzido” em orientações pedagógicas claras. Segundo sintetiza Borba (2017b) a partir de Benson (2001), a autonomia na aprendizagem é “[...] a capacidade de tomar iniciativas em relação ao próprio ensino-aprendizagem e de decidir quais conteúdos, métodos de ensino e materiais didáticos são os mais adequados para si, além de autoavaliar-se” (BORBA, 2017b, p. 33). De acordo com a autora (p. 44), cabe ao professor tomar uma série de medidas de andamento para que os alunos desenvolvam autonomia na aprendizagem por intermédio dos dicionários.

Em resumo, segundo as atuais perspectivas sobre o ensino-aprendizagem de línguas, o aprendiz, usuário do dicionário, é alguém que precisa de dois tipos de mediação, uma em relação ao dicionário e outra em relação à sua aprendizagem em meio à resolução de tarefas linguísticas. A mediação em relação ao instrumento dicionário envolve fornecimento de andamento para a localização e interpretação de informações nos dicionários. Quanto mais didática for a obra, menos esforço demandará a consulta.

⁶⁷ [gaining greater voluntary control over one’s capacity to think and act either by becoming more proficient in the use of meditational resources, or through a lessening of reliance on external meditational means]

A mediação em relação à aprendizagem diz respeito ao uso do dicionário para resolver tarefas linguísticas visando o avanço pela ZDP. Nessa categoria, enquadra-se o andamento no qual o professor fornece um retorno ao aluno sobre alguma tarefa linguística e o orienta a refazê-la sob a mediação do dicionário. À medida que o aluno se acostuma com o manuseio da obra e tem ciência do tipo de informação que o dicionário fornece, poderá tomar a iniciativa de consultar essa ferramenta sem necessariamente depender da recomendação do professor, apresentando, assim, uma autonomia na sua aprendizagem.

Após estabelecer o perfil de usuário do dicionário a partir da aproximação entre o ensino comunicativo e a obra lexicográfica, as seções que seguem invertem a equação, tratando de como o dicionário poderia se aproximar do ensino comunicativo. Tal aproximação toma como base a noção de competência comunicativa, materializada sob a forma de uma escala de seis níveis de referência (A1, A2, B1, B2, C1, C2). Esses níveis ajudam a definir o perfil de usuário de um dicionário de aprendizes tanto em relação à sua proficiência já adquirida como em relação à etapa de aprendizagem na qual se encontra.

3.3 Nível de proficiência e a definição do perfil de usuário

Na determinação do perfil de usuário de um dicionário de aprendizes, a proficiência constitui uma base⁶⁸ mínima de conhecimentos que se pode esperar do consulente de modo que ele entenda as informações incluídas nos verbetes de um dicionário. Na nossa proposta, essa base mínima corresponde aos níveis A1 e A2, que chamaremos de níveis de proficiência propriamente tal.

Proficiência é outro termo que o documento não define; a partir da escala de níveis, contudo, depreende-se que a proficiência é entendida como algo gradual; em outras palavras, não se trata de uma classificação dicotômica para avaliar se um indivíduo é ou não proficiente, mas sim o quão proficiente é. Essa perspectiva está presente em diversas definições de estudiosos de Linguística Aplicada, como Richards e Schmidt (2010, s.v. *language proficiency*), McNamara (2004, p. 771), Dilli, Schoffen e Schlatter (2012, p. 182) e ainda no DiClavELE (2008, s.v. *proficiencia*).

Além dos seis níveis já mencionados, o CEFR (2001) descreve os graus de proficiência conforme diferentes tipos de atividades comunicativas (leitura, escrita, audição e fala) e de

⁶⁸ É pertinente comentar que partir de uma base mínima para apresentar informações novas, isto é, o conhecimento prévio do consulente, está em consonância com as atuais orientações que vêm sendo investigadas há algumas décadas no âmbito da Educação e que derivam dos estudos de Piaget.

competências (linguísticas, sociolinguísticas e pragmáticas). Para esta subseção, importa verificar em que consiste a base mínima de proficiência A1 e A2, necessária para compreender as informações presentes em dicionários de aprendizes. Como o dicionário é um instrumento cujo manuseio requer somente a atividade comunicativa de leitura e contém informações majoritariamente de caráter linguístico, dentre os vários quadros de descritores do CEFR (2001), foram analisados aqueles que se referiam a essa atividade e a essa competência específicas para a determinação do perfil de usuário a ser considerado nesta tese. Conforme ilustrado no capítulo 1, os segmentos informativos que requerem maior habilidade de leitura em um verbete são a paráfrase explanatória, as notas de uso e os exemplos em função do volume de texto que contêm. Nesta oportunidade, optou-se por investigar e avaliar o nível de compreensão leitora necessário para compreender paráfrases explanatórias e notas de uso, muito embora os exemplos também possam ser alvo de análises.

No que concerne à atividade comunicativa de leitura, o CEFR (2001) fornece descritores acerca da compreensão leitora geral (CEFR, 2001, p. 69) e da compreensão leitora de tipos específicos de texto, como correspondência (p. 69), textos informativos (p. 70) (matérias de jornal, por exemplo) e instruções (p. 71). Não há, no entanto, descritores para o tipo de texto *dicionário*, muito embora o próprio CEFR (2001) reconheça que o dicionário é um tipo de texto (CEFR, 2001, p. 95; QECRL, 2001, p. 139) que figura dentre aqueles próprios do âmbito educacional (CEFR, 2001, p. 49; QECRL, 2001, p. 78), classificando-o como ferramenta mediadora do processo de ensino-aprendizagem (CEFR, 2001, p. 88; QECRL, 2001, p. 130) e da atividade de leitura, em particular (CEFR, 2001, p. 92; QECRL, 2001, p. 135) e estimule a consulta a dicionários ao associar o uso destes à competência de aprendizagem⁶⁹ (CEFR, 2001, p. 12; QECRL, 2001, p. 33). A razão para isso é que, segundo o CEFR (2001, p. 88), não há escalas exemplificativas do uso de dicionários como texto.

É importante esclarecer que, segundo Mattes e Bugueño Miranda (2015, p. 101-102), quando analisado sob a ótica da Linguística de Texto, conclui-se que o dicionário é, sim, um tipo de texto. A essa mesma conclusão também chegam Pereira e Nadin (2019). Bugueño Miranda e Mattes (2015) esclarecem que o dicionário é um texto porque permite que se compreenda o conteúdo da sequência de signos que está presente nos verbetes, seja porque esse conteúdo está explicitamente construído por meio de paráfrases explanatórias, por exemplo, seja porque a disposição entre os signos linguísticos permite que se identifique uma relação

⁶⁹ “[...] a competência de aprendizagem é a capacidade para observar e participar em novas experiências e incorporar o conhecimento novo nos conhecimentos anteriores, modificando estes últimos onde for necessário” (QECRL, 2001, p. 154).

entre eles (ou seja, o seu conteúdo), como é o caso dos sinônimos em um verbete de um dicionário de sinônimos. De acordo com os autores, outros parâmetros que permitem classificar o dicionário como um texto são a presença de intencionalidade (a maneira como as informações estão dispostas permite o cumprimento de uma tarefa, isto é, recepção ou produção) e a situacionalidade (contextos de uso de unidades léxicas, tais como indicação de país/região, situação comunicativa e extrato sociocultural). Pereira e Nadin (2019), por sua vez, argumentam que o dicionário é um texto porque os seus componentes canônicos se encontram em uma relação dialógica, isto é, se comunicam entre si – ou, como os autores assinalam também, deveriam se comunicar entre si.

Falta, portanto, conceber o dicionário como uma obra de referência que é tão textual quanto outros documentos classificados como texto para que, assim, sejam estabelecidas escalas de níveis de referência relativas ao seu uso no ensino-aprendizagem de línguas estrangeiras. Enquanto tais instrumentos ainda não são formulados, e ante a necessidade de se analisar os dicionários de aprendizes como textos que precisam ser lidos e compreendidos, a solução encontrada foi recorrer aos descritores relativos à compreensão leitora geral nos níveis A1 e A2. O Quadro 5 reproduz os descritores da versão traduzida do CEFR (2001) ao português.

Quadro 5 – Compreensão leitora nos níveis A1 e A2

Atividade comunicativa: compreensão leitora geral	
A1	“É capaz de entender textos muito curtos e muito simples, uma expressão de cada vez, retirando nomes familiares, palavras e expressões básicas e relendo-as se necessário” (QECRL, 2001, p. 107).
A2	“É capaz de entender textos simples e curtos que contenham vocabulário muito frequente, incluindo uma certa proporção de vocábulos internacionais” (QECRL, 2001, p. 107). “É capaz de entender textos simples e curtos acerca de assuntos que lhe são familiares de um tipo concreto, compostos numa linguagem muito frequente, quotidiana ou relacionada com o trabalho” (QECRL, 2001, p. 107).

Fonte: QECRL (2001)

Dos descritores do Quadro 5, depreende-se somente que a leitura é pouco fluida em indivíduos de nível A1 de proficiência em função da baixa velocidade (“uma expressão de cada vez”) e da eventual necessidade de releitura, mesmo em se tratando de “nomes familiares, palavras e expressões básicas”. No tocante ao nível A2, a fluidez da leitura parece estar relacionada a um “vocabulário muito frequente”.

Nota-se que o QECRL (2001) é pouco específico acerca do que considera como “textos muito curtos e muito simples” e “textos simples e curtos”. Os descritores reproduzidos no Quadro 5 não fornecem exemplos de textos produzidos nesses moldes nem informam que propriedades possui um texto (muito) curto e (muito) simples. Tais expressões são empregadas também nos descritores de tipos específicos de textos, tais como os já mencionados textos informativos e de instrução (QECRL, 2001, p. 109), o que confirma a falta de informações mais específicas no documento.

De fato, a metalinguagem que se usa na redação de um verbete e a própria maneira como este comumente está organizado em um dicionário de aprendizes (informações sobre a unidade léxica enquanto significante, seguidas de informações sobre o seu significado e, eventualmente, de orientações normativas sob a forma de notas de uso) não isentam o consulente do exercício da sua capacidade de interpretação na língua estrangeira. Uma maneira de tornar mais concretos os adjetivos *simples* e *curto* do QECRL (2001), acompanhados ou não do intensificador *muito*, é por meio do uso de algoritmos que calculam o nível de sofisticação de um texto, como o *Crowdsourced Algorithm of Reading Comprehension* (CROSSLEY et al., 2019) (cf. detalhes no capítulo 5).

Os descritores da competência linguística esperada nos níveis A1 e A2 em relação ao vocabulário, por sua vez, estão reproduzidos no Quadro 6.

Quadro 6 – Vocabulário nos níveis A1 e A2

Competência linguística: amplitude e domínio de vocabulário	
A1	“Tem um repertório vocabular elementar, constituído por palavras isoladas e expressões relacionadas com certas situações concretas” (QECRL, 2001, p. 160).
A2	<p>“Possui vocabulário suficiente para satisfazer as necessidades comunicativas elementares. Tem vocabulário suficiente para satisfazer necessidades simples de sobrevivência” (QECRL, 2001, p. 160).</p> <p>“Tem vocabulário suficiente para conduzir transacções do dia-a-dia que envolvam situações e assuntos que lhe são familiares” (QECRL, 2001, p. 160).</p> <p>“É capaz de dominar um repertório limitado relacionado com necessidades quotidianas concretas” (QECRL, 2001, p. 161).</p>

Fonte: QECRL (2001, p. 160-161)

Novamente, os descritores são abstratos, não permitindo que se possa inferir de que se tratam as “certas situações concretas” do nível A1 nas quais se necessita um vocabulário

“elementar” também desconhecido. No nível A2, tampouco se exemplificam quais seriam as “necessidades comunicativas elementares”, as “necessidades simples de sobrevivência” ou “as situações quotidianas concretas”, muito menos o “vocabulário suficiente” ou o “repertório limitado” necessário para agir nestas situações ou satisfazer as necessidades mencionadas.

No que tange a um dicionário de aprendizes, a falta de especificidade dos descritores acaba afetando a determinação de um vocabulário de definidores, isto é, um conjunto limitado de unidades léxicas passíveis de serem usadas na redação de paráfrases explanatórias e notas de uso e que incluem tanto as palavras ditas lexicais como as palavras ditas gramaticais. Um expediente que, em princípio, poderia servir de auxílio para a determinação de um vocabulário de definidores são as pesquisas sobre vocabulário básico [core vocabulary], que é um conjunto de “itens léxicos que são aceitos como centrais e indispensáveis para o uso da língua” [lexical items which are accepted as being central and indispensable to language use] (BELL, 2013, p. 1).

No entanto, não é fácil determinar em que consiste um vocabulário “básico”, “elementar” ou “suficiente”. Geralmente, associa-se a noção de *vocabulário básico* ao índice de frequência em um *corpus*. Nesse sentido, seriam classificados como vocabulário básico as representações ortográficas mais frequentes. Essa perspectiva deu origem a diversos compilados sobre as palavras (representações ortográficas) mais frequentes em certas línguas, que constituiriam o seu vocabulário básico e que, em certos casos, também serviriam como vocabulário de definidores, especialmente de dicionários de aprendizes. Alguns exemplos são a Oxford 3000 (OXFORD, 2019a), com as três mil representações ortográficas mais frequentes no inglês; as listas de representações ortográficas (*tokens*) e lemas (*types*) disponibilizadas pela Real Academia Española (RAE, 2021) a partir dos dados do Corpes (2021); e o *Frequency dictionary of Italian words* (FredIt, 1973). Para o português, o trabalho mais recente é o de Davies e Preto-Bay (2008), que é um dicionário com as 5.000 palavras mais frequentes do português. Não é nosso objetivo mencionar todas as propostas de dicionários de frequência, mas sim citar alguns exemplos, pois esta tese não versa sobre o tema da frequência.

Na Lexicologia, contudo, já se chegou à conclusão de que a variável estatística, sozinha, não permite que se chegue a um vocabulário básico. As pesquisas na área ainda não conseguiram determinar critérios para definir que tipos de unidades léxicas podem ser classificadas como básicas de uma língua, e isso vale também para a determinação de temas de vocabulário. Bell (2013) até propõe uma série de critérios para tal, mas uma análise de Duarte Ferreira (2017) comprova que os critérios do autor não são suficientes e podem até apresentar alguma contradição. Na verdade, conforme a autora, pouco se sabe sobre como determinar um

vocabulário básico (DUARTE FERREIRA, 2017, p. 172). Portanto, e sob a perspectiva da Lexicologia, a frequência *per se* não é critério suficiente para se estabelecer um vocabulário que se possa chamar de básico.

Uma maneira de se determinar o vocabulário atinente a cada nível de proficiência é recorrer a dados de um *corpus* de aprendizes que classifique as unidades léxicas incluídas conforme o nível em que são empregadas pelos aprendizes de língua estrangeira. A língua inglesa já conta com uma ferramenta assim, o English Vocabulary Profile, que será comentado na subseção 3.4. Nessa oportunidade, além de avaliar a facilidade de leitura das paráfrases e notas de uso dos dicionários CaED2 (2011), OxED2 (2012), CoED2 (2014) e LWord2 (2008), os definidores empregados nesses dois segmentos informativos serão classificados conforme o nível de proficiência a que correspondem.

No que concerne ao âmbito gramatical, os descritores desse aspecto da competência linguística estão reproduzidos no Quadro 7.

Quadro 7 – Gramática nos níveis A1 e A2

Competência linguística: gramática	
A1	“Mostra apenas um controlo limitado de algumas estruturas e formas gramaticais simples, que pertencem a um repertório memorizado” (QECRL, 2001, p. 163).
A2	“Usa, com correção, estruturas simples, mas ainda comete erros elementares de forma sistemática – p. ex.: tem tendência a misturar tempos e a esquecer-se de fazer concordâncias; no entanto, aquilo que quer dizer é geralmente claro” (QECRL, 2001, p. 163).

Fonte: QECRL (2001, p. 163)

Uma vez mais, o adjetivo *simples* é demasiado vago para que se tenha uma noção do que seriam estruturas gramaticais simples no que tange à compreensão. Assim como nos quesitos compreensão leitora e vocabulário, há uma alternativa para a determinação das estruturas morfológicas e sintáticas que indivíduos de níveis A1 e A2 de proficiência já conhecem. Trata-se da ferramenta English Grammar Profile, que também será comentada na subseção 3.4 e será usada para avaliar as estruturas morfológicas e sintáticas de paráfrases e notas de uso de verbetes do CaED2 (2011), do OxED2 (2012), do CoED2 (2014) e do LWord2 (2008) no capítulo 5.

Em suma, os descritores do QECRL (2001) praticamente não auxiliam a inferir a habilidade de compreensão leitora e as propriedades léxicas e gramaticais necessárias para que indivíduos de proficiência A1 e A2 entendam as informações presentes em um dicionário. A

vagueza dos descritores do QECRL (2001), aliás, já foi muito criticada na literatura, como em Alderson et al. (2005, p. 9-13) e Weir (2005), por exemplo. A inferência acerca das habilidades de compreensão e propriedades léxicas e gramaticais só pode ser realizada com o auxílio de documentos que aplicam os descritores do CEFR (2001) a línguas específicas, como o inglês e o espanhol. Antes de entrar no mérito desses documentos em particular, é necessário analisar os descritores do CEFR (2001) no que concerne aos níveis A2 e B1 enquanto etapas de aprendizagem.

3.4 Necessidades de aprendizagem nos níveis A2 e B1

Para definir o perfil do usuário de um dicionário de aprendizes, os níveis de referência do CEFR (2001) são considerados de maneira prospectiva, isto é, como indicativos das necessidades de aprendizagem de indivíduos que se encontram em etapas A2 e B1 de aprendizagem e desejam alcançar tais patamares de proficiência. Nesse sentido, um dicionário precisa incluir unidades e combinatórias léxicas que estejam, de alguma maneira, atreladas à proficiência em níveis A2 e B1.

Na presente seção, os descritores de proficiência desses níveis no QECRL (2001) e no CEFR-Co (2018) serão considerados como meta no que concerne ao processo de aprendizagem. Como os dicionários de aprendizes são instrumentos que, como qualquer dicionário de língua, almejam fornecer informações linguísticas, foram analisados os descritores relativos à competência linguística em geral, ao vocabulário, à gramática e a aspectos fonológicos.

Em relação à competência linguística de maneira geral, os descritores são os seguintes:

Quadro 8 – Competência linguística geral nos níveis A2 e B1

Competência linguística geral	
A2	<p>“Tem um repertório limitado de expressões memorizadas curtas que cobrem situações de sobrevivência previsíveis; rupturas e incompreensões frequentes ocorrem em situações não habituais” (QECRL, 2001, p. 158).</p> <p>“É capaz de usar padrões frásicos elementares e de comunicar com expressões memorizadas, grupos de poucas palavras e de expressões feitas sobre si e sobre outras pessoas, sobre aquilo que fazem, sobre lugares, bens, etc.” (QECRL, 2001, p. 158).</p> <p>“É capaz de produzir expressões quotidianas breves de modo a satisfazer necessidades simples de tipo concreto: pormenores pessoais, rotinas quotidianas, desejos e necessidades, pedidos de informação” (QECRL, 2001, p. 158).</p>

	<p>“Tem um repertório linguístico elementar que lhe permite lidar com as situações quotidianas de conteúdo previsível, ainda que, geralmente, necessite de estabelecer um compromisso entre a mensagem e a procura de palavras” (QECRL, 2001, p. 158).</p>
B1	<p>“Possui meios linguísticos suficientes para sobreviver; tem o vocabulário suficiente para se exprimir com algumas hesitações e circunloquções sobre assuntos como família, passatempos, interesses, trabalho, viagens e actualidades, mas as limitações lexicais provocam repetições e mesmo, às vezes, dificuldades com a formulação” (QECRL, 2001, p. 158).</p> <p>“Tem um repertório linguístico suficientemente lato para descrever situações imprevistas, explicar a questão principal de uma ideia ou de um problema com bastante precisão e exprimir o seu pensamento sobre assuntos abstractos ou culturais, tais como a música ou o cinema” (QECRL, 2001, p. 158).</p>

Fonte: QECRL (2001, p. 158)

Diferente dos descritores comentados em 3.2, no Quadro 8 nota-se que o QECRL (2001) fornece exemplos de funções comunicativas de nível A2 (“pormenores pessoais, rotinas quotidianas, desejos e necessidades, pedidos de informação”) e exemplos de temas de nível B1 (“família, passatempos, interesses, trabalho, viagens e actualidades”). Esses exemplos permitem que se tenha uma ideia um pouco mais clara do que se pode esperar nesses patamares de proficiência. No entanto, para fins de definição das necessidades de aprendizagem, as informações ainda não são suficientes para que se possa estabelecer com mais precisão que temas e funções comunicativas estão atrelados a esses níveis. Tal como se comentará em 3.4, mesmo que os temas e as funções comunicativas estivessem expressos integralmente no CEFR (2001), deprender itens gramaticais e de vocabulário dos temas e funções é uma tarefa complicada. O mesmo ocorre quando os descritores *ad supra* mencionam “repertório limitado de expressões memorizadas”, “padrões frásicos elementares”, “expressões memorizadas”, “grupos de poucas palavras e de expressões feitas” etc. Aparentemente, trata-se do uso de expressões, que podem ser compreendidas como combinatórias léxicas que expressam atos de fala (saudações, por exemplo); porém, para fins de determinação das necessidades de aprendizagem do perfil de usuário que se almeja nesta tese, tais informações tampouco são suficientes.

Uma possível razão para a vagueza com que a competência linguística geral foi elaborada pode estar na natureza orientativa, e não normativa, do QECRL (2001). O documento não almeja ser específico e muito menos cercear a liberdade dos professores de língua no que diz respeito ao que devem ensinar nem como fazê-lo (QECRL, 2001, p. 11-14); ao contrário,

propõe reflexões acerca de questões atreladas ao ensino-aprendizagem de línguas e estabelece objetivos gerais e uma escala de níveis de proficiência uniforme com vistas a facilitar a troca de informações (QECRL, 2001, p. 11). Embora o QECRL (2001) faculte aos professores a tarefa de adaptar as orientações à sua realidade e não determine o que deve ou não ser ensinado e aprendido em cada nível em termos de vocabulário e gramática, acreditamos que tanta vagueza não ajuda a conduzir as reflexões que o próprio documento incentiva. Na nossa opinião, para que se pudesse compreender melhor as orientações do documento e decidir entre concordar ou discordar delas, os descritores teriam de ser mais específicos. Se o QECRL (2001) fosse mais específico, poderia cumprir melhor o seu papel de fomentador de reflexões e tomadas de decisão sobre o que ensinar e como ensinar, enquanto os professores teriam um ponto de partida sobre o qual poderiam formular livremente os seus próprios objetivos específicos de aprendizagem, currículos etc.

No tocante à dimensão do vocabulário, esta se subdivide entre o vocabulário como um todo (extensão de vocabulário) e a capacidade de selecionar itens léxicos apropriados (domínio de vocabulário). Os respectivos descritores estão reproduzidos nos quadros 9 e 10.

Quadro 9 – Amplitude de vocabulário nos níveis A2 e B1

Competência linguística: amplitude de vocabulário	
A2	<p>“Possui vocabulário suficiente para satisfazer as necessidades comunicativas elementares. Tem vocabulário suficiente para satisfazer necessidades simples de sobrevivência” (QECRL, 2001, p. 160).</p> <p>“Tem vocabulário suficiente para conduzir transacções do dia-a-dia que envolvam situações e assuntos que lhe são familiares” (QECRL, 2001, p. 160).</p>
B1	<p>“Tem vocabulário suficiente para se exprimir com a ajuda de circunloquções sobre a maioria dos assuntos pertinentes para o seu quotidiano, tais como a família, os passatempos, os interesses, o trabalho, as viagens e a actualidade” (QECRL, 2001, p. 160).</p> <p>“Tem uma boa extensão de vocabulário relacionado a tópicos familiares e situações quotidianas”⁷⁰ (CEFR-Co, 2018, p. 132).</p>

Fonte: QECRL (2001, p. 160) e CEFR-Co (2018, p. 132)

⁷⁰ [Has a good range of vocabulary related to familiar topics and everyday situations]

Quadro 10 – Domínio de vocabulário nos níveis A2 e B1

Competência linguística: domínio de vocabulário	
A2	“É capaz de dominar um repertório limitado relacionado com necessidades quotidianas concretas” (QECRL, 2001, p. 161).
B1	<p>“Usa uma ampla gama de vocabulário simples de maneira apropriada quando fala sobre tópicos familiares”⁷¹ (CEFR-Co, 2018, p. 134).</p> <p>“Mostra bom domínio do vocabulário elementar, mas ainda ocorrem erros graves quando exprime um pensamento mais complexo ou quando lida com assuntos ou situações que não lhe são familiares” (QECRL, 2001, p. 161).</p>

Fonte: QECRL (2001, p. 161) e CEFR-Co (2018, p. 134)

Quanto à amplitude de vocabulário, e em relação ao nível A2, é difícil compreender o que seriam as “necessidades comunicativas elementares”, por exemplo. O descritor do nível B1 para esse aspecto é o que permite que se tenha uma noção um pouco mais clara sobre a que se refere ao mencionar os assuntos do quotidiano, embora apenas reproduza os mesmos exemplos presentes nos descritores da competência linguística geral (cf. Quadro 8), ao invés de fornecer um panorama mais extenso. Os descritores acerca do domínio de vocabulário tampouco permitem entender a que temas, assuntos ou tipos de elementos léxicos se referem, conforme se pode verificar no Quadro 10. Sem elementos mínimos de referência, ou mesmo exemplos, as expressões “repertório limitado”, “necessidades quotidianas concretas”, “vocabulário simples” e “vocabulário elementar” resultam imprecisas.

Os descritores acerca do conhecimento gramatical (que se entende como morfologia e sintaxe), por sua vez, estão reproduzidos no Quadro 11:

Quadro 11 – Gramática nos níveis A2 e B1

Competência linguística: gramática	
A2	“Usa, com correção, estruturas simples, mas ainda comete erros elementares de forma sistemática – p. ex.: tem tendência a misturar tempos e a esquecer-se de fazer concordâncias; no entanto, aquilo que quer dizer é geralmente claro” (QECRL, 2001, p. 163).
B1	<p>“Usa, com uma correção razoável, um repertório de ‘rotinas’ e de expressões frequentemente utilizadas e associadas a situações mais previsíveis” (QECRL, 2001, p. 163).</p> <p>“Comunica, com razoável correção, em contextos familiares; tem geralmente um bom controlo, apesar das influências óbvias da língua</p>

⁷¹ [Uses a wide range of simple vocabulary appropriately when talking about familiar topics]

	materna. Podem ocorrer erros, mas aquilo que ele está a tentar exprimir é claro” (QECRL, 2001, p. 163).
--	---

Fonte: QECRL (2001, p. 163)

No nível A2, a dificuldade com tempos verbais e concordâncias é um exemplo um pouco mais concreto. Por outro lado, não se sabe a que o QECRL (2001) se refere quando menciona “estruturas simples”. No nível B1, as “expressões frequentemente utilizadas e associadas a situações mais previsíveis” parecem relativas ao vocabulário, e não à gramática. Curiosamente, é somente nesse nível que se menciona a influência da língua materna. Parece difícil que a língua materna venha a exercer um papel na aprendizagem da língua estrangeira apenas a partir do nível B1 – e a Psicolinguística confirma essa suspeita, segundo consta no capítulo 4.

Finalmente, reproduzimos também os descritores atinentes à competência fonológica:

Quadro 12 – Dimensão fonológica nos níveis A2 e B1

Competência linguística: controle fonológico geral	
A2	“A pronúncia é, geralmente, clara o suficiente para ser entendida, mas os interlocutores precisarão pedir que repita de tempos em tempos. Uma forte influência de outra(s) língua(s) que ele/a fala na sílaba tônica, ritmo e entonação pode afetar a inteligibilidade, requerendo a colaboração dos interlocutores. No entanto, a pronúncia de palavras familiares é clara” ⁷² (CEFR-Co, 2018, p. 136).
B1	“A pronúncia é normalmente inteligível; pode aproximar a entonação e a sílaba tônica tanto ao nível da sentença como ao nível da palavra. Contudo, o sotaque é normalmente influenciado por outra(s) língua(s) que ele/a fala” ⁷³ (CEFR-Co, 2018, p. 136).

Fonte: CEFR-Co (2018, p. 136)

Embora se refiram à dimensão fonológica em geral, nota-se que os descritores reproduzidos no Quadro 12 se restringem às atividades de produção, sem incluir a recepção. Percebe-se que esse conjunto de descritores é mais preciso ao mencionar aspectos como a pronúncia, a sílaba tônica, o ritmo e a entonação. Também abarca a questão da influência de outra(s) língua(s) nesses três aspectos, o que é perfeitamente plausível e compreensível, como se comentará no capítulo 4. É pertinente questionar-se por que a experiência prévia com outra(s)

⁷² [Pronunciation is generally clear enough to be understood, but conversational partners will need to ask for repetition from time to time. A strong influence from other language(s) he/she speaks on stress, rhythm and intonation may affect intelligibility, requiring collaboration from interlocutors. Nevertheless, pronunciation of familiar words is clear]

⁷³ [Pronunciation is generally intelligible; can approximate intonation and stress at both utterance and word levels. However, accent is usually influenced by other language(s) he/she speaks]

língua(s) só é contemplada na dimensão fonológica. Custa a acreditar que somente nessa dimensão ocorram influências de outra(s) língua(s) que o estudante utiliza. *Corpora* de aprendizes e professores com experiência no ensino de línguas estrangeiras certamente poderão fornecer exemplos de influências que facilitam e outras que dificultam a aprendizagem e uso de itens léxicos, combinatórias léxicas e estruturas morfológicas e sintáticas.

Dos quatro aspectos presentes no CEFR-Co (2018), um dicionário é capaz de, ao menos, fornecer informações sobre a sílaba tônica, assinalando-a no lema, por exemplo. Em dicionários de aprendizes de inglês, é comum encontrar a transcrição fonológica das palavras-entrada com a devida indicação da sílaba tônica. Uma gama maior de possibilidades pode ser oferecida via suporte digital por meio de recursos de áudio para a pronúncia das palavras e, em alguns casos, até dos exemplos⁷⁴, abarcando, assim, o ritmo e a entonação. Cabe observar, contudo, que o real aproveitamento do recurso de áudio depende da capacidade do aprendiz de perceber as propriedades fonológicas de uma língua, isto é, quais unidades são fonemas e quais são alofones, conforme se comentará mais detalhadamente em 4.3.1.1. Tal capacidade é algo subjetivo, que varia de indivíduo para indivíduo e que não se pode calcular. O sucesso do recurso de áudio isolado de uma transcrição fonológica é, portanto, relativo.

Em suma, um aspecto que prepondera em todos os descritores encontrados é a vagueza com que são redigidos. Os termos empregados nos descritores, tais como “repertório linguístico elementar”, “necessidades comunicativas elementares” e “vocabulário elementar”, são demasiado imprecisos. Na mesma esteira, os exemplos fornecidos tampouco são suficientes para elucidar os descritores. Para fins de determinação das necessidades que surgem nas etapas de aprendizagem A2 e B1, muito pouco pode ser aproveitado do CEFR (2001) e do CEFR-Co (2018).

Conforme se comentará na próxima subseção, a falta de precisão nos descritores do CEFR (2001) e do CEFR-Co (2018) não é acidental, mas intencional. Isso prejudica o trabalho não somente de lexicógrafos e metalexicógrafos, mas de todo aquele profissional interessado em compilar materiais didáticos de ensino de línguas estrangeiras.

3.5 Documentos de referência para o ensino de inglês e espanhol como língua estrangeira

A falta de precisão do CEFR (2001) / QECRL (2001) se justifica em função de que o documento não almeja servir de orientação para o ensino de línguas:

⁷⁴ Parte dos exemplos do LWord2 (2008) está acompanhada de um recurso auditivo. O mesmo acontece na versão digital do *Longman Dictionary of Contemporary English* (LDOCE6, 2014) disponível na internet.

Não se trata DE MODO ALGUM de dizer aos que trabalham nesta área o que devem fazer e como devem fazê-lo. São levantadas questões, mas não são dadas respostas. A função do *Quadro Europeu Comum de Referência* não é nem formular os objectivos que os utilizadores devem atingir, nem os métodos que devem usar (QECRL, 2001, p. 11-12. Grifos no original).

Contudo, um documento de referência que não almeja, em alguma medida, orientar o processo de ensino-aprendizagem de línguas parece uma contradição, como observam Borba e Bugueño Miranda (2018, p. 2173). O que se esperava a partir da publicação do CEFR (2001) e de sua tradução a outras línguas europeias era que, em um primeiro momento, e a partir da leitura do documento, os responsáveis pela elaboração de programas de ensino e de materiais didáticos elencassem conceitos gerais (tempo, por exemplo) e atos de fala (pedir algo, por exemplo) para associá-los a cada nível de proficiência. Em um segundo momento, os profissionais envolvidos deveriam elencar os recursos linguísticos específicos que manifestam os conceitos gerais e atos de fala nas línguas com as quais trabalham. É pertinente recordar que a abordagem comunicativa e o CEFR (2001) estão baseados na perspectiva nocional-funcional, surgida na década de 1970. Conforme Richards e Schmidt (2010, s.vs. *notional syllabus* e *functional syllabus*), as noções são conceitos gerais, enquanto as funções são atos de fala. Sob essa perspectiva, o ensino de línguas deve tomar como ponto de partida as noções e funções, que são abstratas, enquanto os recursos linguísticos de cada língua são colocados em segundo plano. Essa concepção está refletida no CEFR (2001) e no CEFR-Co (2018) e lhes confere um caráter genérico, como se pode observar nos Quadros 8 a 12.

Entretanto, o público-alvo do CEFR (2001) encontrou dificuldades em executar esses movimentos, especialmente no que concerne à identificação de recursos linguísticos (BORBA; BUGUEÑO MIRANDA, 2018, p. 2174). Em outras palavras, o desejo de dar a maior autonomia possível na elaboração de programas de ensino terminou por gerar um documento excessivamente genérico e que não atendia à demanda por orientação em relação à abordagem comunicativa. Para supri-la, o Conselho da Europa estimulou a produção de documentos referentes a línguas específicas: as Descrições de Níveis de Referência. Para o inglês, gerou-se o *English Profile* (EP, 2011) e, para o espanhol, o *Plan Curricular del Instituto Cervantes* (PCIC, 2006). Apesar de terem o CEFR (2001) como base comum, os dois documentos se diferenciam entre si em questões de metodologia de elaboração. As próximas duas subseções contêm comentários acerca de cada um, a começar pelo PCIC (2006) para seguir uma ordem cronológica.

3.5.1 Descrição de níveis de referência do espanhol: *Plan Curricular del Instituto Cervantes* (2006)

De acordo com o que foi comentado no início de 3.2, o CEFR (2001) organiza seus descritores conforme três tipos de subcompetências atrelados à competência comunicativa: linguística, pragmática e sociolinguística. Cada capítulo do PCIC (2006), por sua vez, se inscreve em uma dessas três competências. Dessa forma, a competência linguística é tratada em *Gramática, Pronunciación, Ortografía, Nociones generales* e *Nociones específicas*; a competência pragmática é tratada em *Funciones, Tácticas y estrategias pragmáticas* e *Géneros discursivos y productos textuales*; e a competência sociolinguística está presente em *Referentes culturales, Saberes y comportamientos socioculturales* e *Habilidades y actitudes interculturales*. A estrutura do documento, no entanto, não é nova. Segundo Borba e Bugueño Miranda (2018), o PCIC (2006) é fruto do CEFR (2001), mas, aparentemente, a sua estrutura foi herdada de um programa de trabalho anterior ao CEFR (2001), chamado *The Threshold Level* (1975-2001) e conhecido como T-series. A coleção de documentos propunha sentar as bases para o ensino de inglês como língua estrangeira e consiste na materialização da perspectiva nocional-funcional, sobre a qual está fundamentada a abordagem comunicativa. Cada documento da T-series serviu de base para a proposta dos seis níveis de referência consagrados pelo CEFR (2001). Dessa forma, o primeiro deles, *Threshold 1990* (1998), corresponde ao nível B1; *Waystage 1990* (1998), a A2; e *Vantage 1990* (2001) a B2⁷⁵ (EP, 2011, p. 4).

Uma diferença significativa da T-series em relação ao CEFR (2001) é o fato de que, como seus documentos estavam dirigidos ao ensino-aprendizagem de inglês, as orientações aparecem ilustradas com exemplos de recursos linguísticos dessa língua, o que facilita a sua compreensão. É curioso notar que, no PCIC (2006), à exceção do capítulo sobre ortografia, todos os demais capítulos referentes à competência linguística encontram capítulos equivalentes nas T-series. O mesmo ocorre com o capítulo dedicado às funções. Para ilustrar nossas afirmações, o Quadro 13 a seguir contém um breve comparativo de trechos similares entre o PCIC (2006) para nível B1 e o *Threshold 1990* (1998).

⁷⁵ Citamos as edições do final dos anos 1990 por serem as mais recentes. Cabe mencionar que há ainda um quarto documento cuja publicação nunca foi oficializada (mas que pode ser encontrado na internet), e que corresponde ao nível A1: *Breakthrough* (s.d.).

Quadro 13 – Comparativo entre o PCIC (2006, B1) e o Threshold 1990 (1998)

<p>2.10 expresar desacuerdo</p> <ul style="list-style-type: none"> ▪ Yo no pienso lo mismo (que tú). ▪ A mí me parece que no / sí. [-Esta película no es nada interesante.] -A mí me parece que sí. ▪ Yo creo / pienso que no / sí. [-Podemos terminar todo antes del viernes.] -Yo creo que no. ▪ Yo no / sí lo encuentro... Yo no lo encuentro caro. Yo sí las encuentro interesantes. ▪ No, no es verdad que + pres. subj. No es verdad que tenga tanto dinero. 	<p>2.2 expressing disagreement with a statement</p> <p>2.2.1 I 'don't a'gree.</p> <p>2.2.2 That's 'not ,right.</p> <p>2.2.3 You are 'wrong(,there).</p> <p>2.2.4 (with a positive statement)</p> <p>2.2.4.1 No (+ tag)</p> <p>2.2.4.2 'Not ,so.</p> <p>2.2.4.3 'Certainly ,not.</p> <p>2.2.4.4 I 'don't 'think so.</p> <p>2.2.5 (with a negative statement)</p> <p>2.2.5.1 Yes (+ tag) (Spinach is 'horrible.) 'No it ,isn't (To-morrow 'isn't "Wednesday.) 'Yes it "is.</p> <p>2.2.5.2 I think (+ positive statement)</p>
--	--

Fonte: PCIC (2006) e Threshold 1990 (1998, p. 30)

Mais curioso ainda é saber que o PCIC (2006), enquanto descritor de níveis de referência, não foi compilado *ad hoc*, mas sob a tutela do próprio Conselho da Europa. Isso reforça que a tentativa de lançar uma base comum e genérica não foi suficiente para satisfazer a demanda por orientação, sendo necessário recorrer a documentos anteriores supostamente superados. Dado que cada documento faz referência a um documento anterior, cabe questionar ainda que base teórico-metodológica subjaz o conjunto da T-series. Não fica claro como se chegou a este (e não outro) conjunto de temas e itens léxicos (noções gerais e específicas) e por que foram distribuídos de uma forma e não de outra entre os níveis de referência. A falta de informações sobre o embasamento teórico dos documentos da T-series leva-nos a questionar seu *status* de fundamentação teórica que o CEFR (2001), o PCIC (2001) e os demais documentos derivados do CEFR (2001) lhe atribuem. Llorián (2017, p. 3, 5, 12, 24, 29), por exemplo, diversas vezes usou o verbo *intuir* e palavras derivadas desta ao se referir às T-series, ao PCIC (2006) e à maneira como os inventários desses documentos foram compilados. Na opinião da autora, para que o PCIC (2006) descreva de maneira mais fidedigna que recursos linguísticos do espanhol correspondem a cada nível de referência do CEFR (2001), são necessárias duas medidas: por um lado, “validar e ajustar as especificações previamente realizadas a partir da implementação dos NRE-PCIC [Níveis de Referência para o Espanhol-PCIC]”⁷⁶ (LLORIÁN, 2017, p. 3); por outro lado, “retomar o trabalho de especificação do

⁷⁶ [validar y ajustar las especificaciones previamente realizadas a partir de la evaluación de los programas derivados de la implementación de los NRE-PCIC]

material linguístico, desta vez sobre bases empíricas, empregando, para tanto, uma metodologia de corpus”⁷⁷ (LLORIÁN, 2017, p. 3). Essas medidas, comentadas extensamente por Llorián (2017, p. 21-30), são válidas, aliás, para qualquer documento que se proponha a ser um DNR do CEFR (2001). Até o momento, o PCIC (2006) não aplicou nenhuma dessas duas medidas.

Embora não se tenha todos os esclarecimentos desejados, e apesar das limitações do PCIC (2006) e do CEFR (2001), é preciso relevar as críticas e atentar para o fato de que este é o estado da arte dos instrumentos de referência para o ensino, aprendizagem e avaliação de línguas estrangeiras, pois são estes os recursos com os quais se pode contar atualmente. Sem dúvidas, qualquer trabalho que utilize o CEFR (2001) e seus DNR como base, como é o nosso caso, se beneficiará da revisão e aprimoramento desses documentos. Enquanto tais avanços ainda não forem levados a cabo, a solução que resta é seguir amparando-se no CEFR (2001) e nos recursos linguísticos assinalados no PCIC (2006).

Em contrapartida ao que se vê em relação ao espanhol, a situação do ensino-aprendizagem de inglês como língua estrangeira no tocante à descrição de níveis de referência e de disponibilidade de materiais complementares é bem diferente. Estes temas são tratados na próxima subseção.

3.5.2 Descrição de níveis de referência do inglês: *English Profile (2011)*

Conforme já comentado anteriormente, relacionar os recursos linguísticos empregados para realizar funções (atos de fala) e exprimir noções (conceitos, podendo ser entendidos também como massa léxica) não é uma tarefa simples (BORBA; BUGUEÑO MIRANDA, 2018, p. 2173-2174). A principal característica do EP (2011) é a sua base empírica, um sinal da busca pela adequação aos avanços em Linguística de *Corpus*. O projeto, portanto, aplica a segunda medida mencionada por Llorián (2017) para, de maneira mais confiável, aplicar o CEFR (2001) ao ensino-aprendizagem de uma língua específica. Sendo o EP (2011, p. 53) uma iniciativa da Cambridge University, as fontes que compõem a base de dados do projeto provêm de produtos desenvolvidos pela própria universidade, tais como o *Cambridge Learner Corpus*, composto pelas produções de aprendizes que fazem os testes de proficiência aplicados pelo órgão Cambridge Assessment. No entanto, esta não é a única fonte de dados que compõe o EP (2011); outros documentos, tais como as listas de vocabulário Cambridge Assessment (2018a, b), se somaram aos dados do *corpus* de aprendizes para constituir a base de dados do EP.

⁷⁷ [reemprender la labor de especificación del material lingüístico, esta vez sobre bases empíricas, empleando para ello metodología de corpus]

O objetivo do EP (2011) é produzir bases de dados a respeito dos recursos linguísticos (gramática e vocabulário) efetivamente empregados por aprendizes avaliados em cada um dos seis níveis de proficiência do CEFR (2001)⁷⁸. As bases de dados geradas não são necessariamente exaustivas, mas almejam identificar características sobressalentes em cada nível que permitam distingui-lo de seus adjacentes (EP, 2011, p. 6). Considera-se que

[...] há certas propriedades linguísticas que são características e indicativas de proficiência na L2 [segunda língua] em cada nível, a partir das quais professores e examinadores fazem suas avaliações práticas (juntamente com o quanto um aprendiz cumpre as funções comunicativas requeridas pela atividade realizada)⁷⁹ (EP, 2011, p. 9).

Percebe-se que não há como evitar um certo grau de subjetividade, na medida em que não se sabe em que se basearam os professores e examinadores para identificar certas características como sobressalentes.

A base de dados gramaticais oriunda do EP (2011) é o *English Grammar Profile* (EGP, 2015) cuja primeira versão está disponível tanto no EP (2011, p. 9-35) e a versão mais atualizada pode ser acessada gratuitamente pela internet (<https://www.englishprofile.org/english-grammar-profile/egp-online>). Segundo o EP (2011), a principal fonte de dados para a elaboração do EGP (2015) é o *Cambridge Learner Corpus*. Na ferramenta de busca do EGP (2015) na internet, geralmente fornece-se um exemplo correto e outro errôneo para cada item gramatical a partir dos dados do *corpus*. No EP (2011), por sua vez, há tabelas que informam tanto as características sobressalentes em cada nível como os erros mais frequentes e em que nível costumam ser superados (EP, 2011, p. 25-33).

A base de dados de itens léxicos, por sua vez, consiste no *English Vocabulary Profile* (EVP, 2015), de acesso gratuito na internet (<https://www.englishprofile.org/wordlists/evp>). Seu objetivo é reunir unidades léxicas e combinatórias léxicas “comumente conhecidas” [commonly known] por aprendizes de inglês ao redor do mundo (EP, 2011, p. 53). É pertinente observar que os responsáveis pela compilação do EVP (2015) advertem que sua base de dados não distingue entre vocabulário passivo e vocabulário ativo, isto é, entre a massa léxica ativada

⁷⁸ Os dados gramaticais e de vocabulário atinentes a cada nível de proficiência parecem ser automaticamente atrelados pelos compiladores do EP (2011) às funções elencadas no CEFR (2001). Entretanto, não foram encontrados indícios que confirmem que as funções que esses dados exprimem coincidem com o inventário do CEFR (2001). Um objetivo secundário do EP (2011) consiste no estabelecimento de funções atreladas aos níveis C1 e C2 de proficiência, os quais, segundo seus compiladores, são descritos de maneira vaga no CEFR (2001).

⁷⁹ [there are certain linguistic properties that are characteristic and indicative of L2 proficiency at each level, on the basis of which teachers and examiners make their practical assessments (alongside the extent to which a learner fulfils the communicative functions required by the activity undertaken)]

para resolver tarefas de recepção e a massa léxica usada para resolver tarefas de produção. As fontes usadas para compilar o EVP (2015) são, por um lado, o *Cambridge Learner Corpus*, com dados de textos escritos, e, por outro lado, “uma série de fontes usadas em sala de aula, incluindo listas de palavras de cursos de destaque, listas de palavras de leitores e o conteúdo de livros para aprendizagem de vocabulário”⁸⁰. As edições das listas de vocabulário Cambridge Assessment (2018a, b) disponíveis à época também serviram de base para a compilação do EVP (2015).

No tocante ao seu papel social, os autores do EP (2011) assinalam que não almejam fornecer uma lista definitiva do que “deve” formar parte do processo de ensino-aprendizagem de inglês, mas sim auxiliar na “tomada de decisões sobre o que incluir em um curso”⁸¹ (EP, 2011, p. 8). Ou seja, na mesma esteira do CEFR (2001), insiste-se que a perspectiva adotada é descritiva, e não prescritiva. Tal afirmação é válida para professores particulares e para os responsáveis por elaborar um curso de inglês para fins específicos, pois, nesse caso, há uma seleção de situações comunicativas e elementos de vocabulário e gramaticais atrelados a um determinado âmbito. O mesmo também é válido para exames de proficiência que medem o uso do inglês acadêmico, tais como o *International English Language Testing System* na rubrica *Academic*, ou os exames Cambridge sob a rubrica *Business*, por exemplo. No entanto, as mesmas organizações que aplicam esses exames para fins específicos possuem uma variada gama de testes “gerais” de inglês que avaliam as quatro habilidades e para os quais há diversos cursos preparatórios. Ante uma série de dados que indicam quais recursos linguísticos foram empregados por candidatos exitosos em cada nível de proficiência, é de se questionar se os coordenadores pedagógicos desses cursos selecionariam apenas alguns desses recursos linguísticos para constituir o seu programa de trabalho (*syllabus*). Em suma, os compiladores do EP (2011), bem como do CEFR (2001), parecem não querer assumir totalmente o papel social desses documentos, ou seja, a função de orientação que coordenadores pedagógicos, professores e compiladores de materiais didáticos terminam invariavelmente atribuindo ao EP (2011).

Finalmente, é pertinente comentar que os compiladores do EP (2011) estabelecem uma relação entre os dados gerados e as investigações de caráter psicolinguístico que podem ser realizadas a partir dos mesmos. Em outras palavras, além de se elaborar programas de ensino, materiais didáticos e ferramentas para avaliar a proficiência, propõem que se analisem quais

⁸⁰ [a range of classroom-based sources, including wordlists from leading courses, readers’ wordlists and the content of vocabulary skills books]

⁸¹ [in making decisions about what to include in a course]

seriam as implicações pedagógicas dos efeitos da língua materna na aprendizagem, ensino e avaliação de inglês e a relação entre idade e desenvolvimento da aprendizagem. Os autores listam um conjunto de fatores que consideram como influenciadores do processo de ensino, aprendizagem e avaliação, tais como a língua materna, a idade dos aprendizes e o nível de aprendizagem de inglês (EP, 2011, p. 8). Chama a nossa atenção a correspondência dessas variáveis com aquelas comentadas em 2.3.1 e 2.4 no tocante à determinação do perfil de usuário de dicionários – e que serão tratadas no capítulo 4. Em resumo, adota-se uma perspectiva de análise interdisciplinar, na medida em que se associa Linguística Aplicada, Linguística de *Corpus* e Psicolinguística:

Além do foco na gramática e nas propriedades léxicas, fatores psicolinguísticos também foram levados em conta, incluindo os efeitos da transferência linguística (*i.e.* o impacto de diferentes línguas maternas e contextos de aprendizagem) no aprendizado de línguas. Vale a pena notar que alguns dos resultados nem sempre poderiam ser previstos a partir de teorias de aquisição de línguas nem antecipados por pesquisadores usando somente a sua experiência e intuição, o que prova a validade e utilidade de uma abordagem empírica que possui dados reais em sua essência⁸² (EP, 2011, p. 7).

Conforme o parágrafo acima, os autores do EP (2011) tomam os dados do *corpus* como ponto de partida para fazer análises sobre o processo de aprendizagem de línguas não maternas e qual é o papel da língua materna nesse processo. Sem dúvidas, os resultados gerados são interessantes sob a perspectiva da pesquisa interdisciplinar; no entanto, esse tipo de investigação fica restrito à equipe envolvida com o EP (2011), pois as bases de dados EGP (2015) e EVP (2015) não incluem a língua materna como um critério de refinamento das buscas.

Além do EP (2011), há outras ferramentas também elaboradas para associar Linguística de *Corpus* à Linguística Aplicada ao ensino-aprendizagem de inglês como língua não materna: as listas Cambridge Assessment (2018a, b), que contêm o vocabulário usado nos níveis A2 e B1, e as listas *Oxford 3000* (Oxford, 2019a) e *Oxford Phrase List* (Oxford, 2019b), com as 3000 palavras e expressões mais frequentes do inglês elencadas com base na sua frequência de ocorrência no *Oxford English Corpus* e em um *corpus* criado a partir de livros didáticos de inglês como língua estrangeira⁸³. A lista de Oxford (2019a) contém o que seus compiladores

⁸² [In addition to a focus on traditional grammatical and lexical features, psycholinguistic factors have also been taken into account, including the effects of language transfer (*i.e.* the impact of different first languages and learning contexts) on language learning. It is worth noting that some of the results would not always be predicted from language acquisition theories or anticipated by researchers using experience and intuition alone, proving the validity and usefulness of an empirical approach that has real data at its core]

⁸³ O acesso às listas Oxford (2019a, b) é gratuito. Contudo, o acesso aos *corpora* que deram origem às listas mencionadas é restrito à equipe da Oxford University Press.

consideram como itens léxicos essenciais para quem almeja atingir os níveis A1 a B2 do QECRL (2001). A essa lista, soma-se outra, de 5000 palavras, que acrescenta dados acerca dos níveis B2 a C1 de proficiência (informações em *The Oxford 3000 and the Oxford 5000*, em <https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000>). Segundo seus compiladores (THE OXFORD 3000, 2019), as listas foram elaboradas considerando-se um critério de frequência de aparição no *British National Corpus* (BNC, 2007) e em um *corpus* de colocações privado da Oxford University Press. Além do critério de frequência em textos de distintos gêneros, os compiladores também acrescentaram unidades léxicas que, após passarem pelo crivo de especialistas em ensino e aprendizagem de línguas, foram consideradas “muito familiares para a maioria dos usuários da língua inglesa. Estas incluem, por exemplo, palavras para partes do corpo, palavras usadas em viagens e palavras que são úteis para explicar o que se quer dizer quando não se sabe a palavra exata”⁸⁴.

A lista de fraseologias de Oxford (2019b), por sua vez, não fornece detalhes sobre o seu processo de compilação; assinala-se que está composta de fraseologias, *phrasal verbs*, colocações e locuções prepositivas.

No que concerne a esta tese, as informações do EP (2011) sobre os itens léxicos e estruturas gramaticais empregados por aprendizes de níveis A1, A2 e B1 de proficiência são, sem dúvidas, uma importante fonte de dados para averiguar aquilo que indivíduos de proficiência A1 e A2 comumente já conhecem em inglês e o que se pode estabelecer como meta de aprendizagem para quem está tentando atingir os níveis A2 e B1 de proficiência. Embora questionemos o posicionamento dos responsáveis pelo EP (2011) em relação ao papel social da sua ferramenta e não tenhamos acesso a buscas mais refinadas conforme a língua materna, trata-se de uma valiosa contribuição para que se possa estabelecer, por um lado, o alcance da capacidade de compreensão de itens léxicos e gramaticais e, por outro lado, as necessidades de aprendizagem do perfil de usuário desta tese.

3.6 Corolário do capítulo 3

Para este e o próximo capítulo, optou-se por reservar uma seção dedicada à elaboração de axiomas e constatações gerais, que constituem um corolário da argumentação. De acordo com Blackburn (1997, s.v. *axioma*), um *axioma* é uma

⁸⁴ [are very familiar to most users of English. These include, for example, words for parts of the body, words used in travel, and words which are useful for explaining what you mean when you do not know the exact word for something]

Proposição apresentada como uma a partir da qual podemos começar; uma asserção tida como fundamental, pelo menos para os fins de investigação em causa. O método axiomático consiste em definir um conjunto de proposições desse tipo, assim como os processos de *demonstração ou as *regras de inferência que são permitidas, para derivar então os *teoremas que daí resultam (BLACKBURN, 1997, s.v. *axioma*).

Tais axiomas e constatações constituem, portanto, orientações gerais sobre como elaborar o desenho de um dicionário de aprendizes. Apesar do foco no desenho de obras para aprendizes brasileiros de inglês e espanhol em níveis A2 e B1 de aprendizagem, alguns dos axiomas podem ser estendidos ao desenho de obras para aprendizes de outras línguas e inseridos em outros *backgrounds* linguísticos.

Neste capítulo, o perfil do usuário do dicionário de aprendizes foi definido a partir de duas perspectivas. A primeira delas diz respeito à necessidade do ensino-aprendizagem comunicativo de contar com um instrumento mediador como o dicionário. A outra perspectiva se refere à necessidade do dicionário se aproximar do ensino comunicativo. Os axiomas que seguem estão divididos em dois grupos, um para cada uma das perspectivas supracitadas.

O primeiro conjunto de axiomas diz respeito ao papel do dicionário de aprendizes no ensino comunicativo.

- (1) O dicionário de aprendizes é um instrumento auxiliar do ensino-aprendizagem de línguas estrangeiras e de segunda língua.
- (2) O dicionário enquanto material didático possibilita a promoção e desenvolvimento da autonomia na aprendizagem.
- (3) O dicionário de aprendizes é um instrumento para resolver determinados tipos de dúvidas linguísticas.
- (4) O dicionário de aprendizes pode ser usado em conjunto com outros instrumentos lexicográficos.

O primeiro axioma alude ao *status* de instrumento auxiliar, em oposição a condutor, do processo de ensino-aprendizagem. Em primeiro lugar, e conforme já mencionado em outras oportunidades, a consulta a um dicionário consiste em uma tarefa secundária, necessária para a execução de uma tarefa primária de leitura, audição, escrita ou fala (BUGUEÑO MIRANDA, 2007b, p. 263; NESI, 2013, p. 3; NESI, 2015a, p. 69). Não se recorre a um dicionário como se este fosse um livro didático, isto é, um instrumento que propõe atividades para guiar a aprendizagem, pois não faz parte das competências de um dicionário orientar o ensino-aprendizagem de línguas. Em um contexto de ensino-aprendizagem subsidiado por dicionários,

a aprendizagem é uma consequência derivada da interpretação das informações do dicionário e da subsequente aplicação dessas informações para seguir adiante na execução de uma tarefa linguística, seja esta tarefa uma proposta do livro didático, seja uma tarefa encaminhada pelo professor; seja de recepção, seja de produção. Em outras palavras, a aprendizagem continua sendo uma consequência do uso da língua.

Em segundo lugar, na condição de instrumento auxiliar, o dicionário de aprendizes é um importante aliado do professor de língua no momento de fornecer *feedback* e andamento aos alunos. Uma das características do ensino comunicativo de língua consiste em propor tarefas que consistam em (ou ao menos simulem) situações de interação na língua estrangeira, levando os alunos a produzir textos escritos e orais. O professor, por sua vez, precisa fornecer um andamento ao aluno que o estimule a avançar pela sua ZDP por meio da reescrita ou reelaboração desses textos, conforme o andamento proposto em Dilli, Schoffen e Schlatter (2012), por exemplo. Dependendo de como essa tarefa é conduzida pelo docente, pode acabar sobrecarregando-o, especialmente nos casos em que, além de ter que apontar quais unidades léxicas e combinatórias léxicas não estão em conformidade com os padrões de uso da língua estrangeira, o professor tiver que explicitar a pronúncia, ortografia, flexão ou combinatória correta. Frente a esse cenário, o dicionário de aprendizes se apresenta como um instrumento que, por um lado, auxilia os professores na tarefa de fornecer andamento aos alunos, e, por outro, auxilia-os a progredir na sua aprendizagem por meio da descrição de como a língua é usada. Trata-se, portanto, de um instrumento que pode ser empregado como mediador da aprendizagem, segundo os pressupostos da Teoria Sociocultural.

O segundo axioma diz respeito ao potencial do dicionário em relação à promoção e desenvolvimento da autonomia na aprendizagem. O dicionário *per se* não promove a autonomia; é preciso que o professor o faça. Ao orientar seus alunos a usarem um dicionário para refazer uma tarefa, conforme descrito no parágrafo anterior, o professor está promovendo a autonomia na aprendizagem, pois incentiva o aluno a tomar a iniciativa de buscar um dicionário para que o oriente em relação a aspectos do uso da língua que precisam ser melhorados. No que concerne à nossa proposta de aprimoramento do desenho de dicionários de aprendizes, por mais que se invista em estratégias para apresentar informações sobre a língua em uso da maneira mais didática possível, a eficácia de um dicionário produzido nos moldes que propomos aqui continua dependente da atuação do professor; sua intervenção é essencial para que os alunos conheçam o dicionário e para que aprendam a localizar e interpretar informações na obra.

O bom aproveitamento de um dicionário de aprendizes para mediar a aprendizagem e promover a autonomia requer que se tenha noção de que tipos de dúvidas essa classe de obras é capaz de resolver, tal como postulado no terceiro axioma. Nesse sentido, é importante conhecer as possibilidades e limites do dicionário (cf. seção 3.2 e Zanatta (2010)) para que não lhe seja cobrado injustamente um auxílio em relação a algo que, pela sua própria natureza, esse tipo de instrumento não é capaz de contemplar.

O quarto axioma se refere aos tipos de instrumentos lexicográficos passíveis de serem usados no ensino-aprendizagem. A proposta apresentada nesta tese visa o aprimoramento de dicionários de aprendizes para melhor satisfazer as necessidades do usuário, mas isso não significa que o produto lexicográfico gerado venha a substituir o uso de dicionários bilíngues, por exemplo. Isso seria um contrassenso em relação à realidade, pois, de acordo com Nesi (2015b, p. 2-3), o uso de dicionários bilíngues permanece mesmo entre estudantes de níveis avançados. É mais realista pensar que, de posse de uma edição aprimorada de dicionário de aprendizes, tal como o que propomos aqui, o aprendiz irá recorrer às duas classes de obras. A consulta a um dicionário bilíngue é inevitável, pois a mediação da língua materna é uma estratégia comumente usada para se compreender e produzir na língua estrangeira. Uma das desvantagens de se amparar somente em um dicionário bilíngue é que este termina passando a impressão errônea de que as línguas estão em uma relação de equivalência entre si, quando, na verdade, sempre haverá divergência em algum nível de processamento/organização da linguagem – isso quando houver pares de unidades léxicas com algum grau de correspondência (cf. capítulo 4). É necessário considerar também que a presença do equivalente na língua materna também tende a estar acompanhada da ativação de padrões de flexão e combinatórios da língua materna, que não necessariamente serão os mesmos da língua estrangeira. É o caso, por exemplo, de *information*, que não tem flexão no plural no inglês, e dos equivalentes *order* (inglês) e *pedido* (português), que selecionam palavras diferentes para se combinar quando se referem ao encaminhamento de um pedido de compra a uma empresa, isto é, *to place an order* e *fazer um pedido*.

Além da falsa noção de equivalência, Granger (2018), ao revisar os efeitos da revolução do *corpus* na Lexicografia, ainda menciona outra desvantagem dos dicionários bilíngues: a falta de embasamento em dados de *corpus*, comprometendo a confiabilidade das obras. É o que ocorre com os nove dicionários bilíngues espanhol-português avaliados por Bugueño Miranda e Borba (2019), por exemplo, na medida que os autores identificaram informações não discretas tanto na descrição do espanhol como na descrição do português. Segundo Granger (2018), embora já se tenham transcorrido mais de vinte anos desde as primeiras críticas à ausência de

empirismo, esse problema permanece latente. Uma das justificativas para tanto é notável falta de *corpora* de tradução de língua geral, isto é, que não sejam relativos a contextos específicos – como ocorre com o site *Linguee* e seus textos do Parlamento Europeu, por exemplo. Para se ter certeza de que uma unidade léxica da língua-alvo está descrita corretamente, seria necessário consultar outra obra de referência, como um dicionário de aprendizes, um dicionário de uso ou um dicionário geral de língua. Um linguista ou alguém altamente proficiente na língua-alvo pode até aceitar fazer esse movimento, mas o mesmo não se pode esperar de aprendizes de níveis iniciais – que são justamente o público que mais precisa dos dicionários bilíngues.

Após essa axiomática inicial, a seguir, são expostos axiomas e constatações relativos a como aproximar o dicionário de aprendizes do ensino comunicativo. Para adequar-se a essa abordagem, é imprescindível levar em conta as orientações do CEFR (2001), do CEFR-Co (2018) e dos descritores de níveis de referência (PCIC (2006), EVP (2015) e EGP (2015)), bem como dados advindos de outras fontes, como listas de palavras e *corpora* de aprendizes. Os axiomas a seguir sintetizam os subsídios da Linguística Aplicada no tocante à determinação do perfil de usuário de um dicionário de aprendizes:

(5) O perfil de usuário deve ser determinado a partir do conhecimento já adquirido pelo aprendiz. O conhecimento já adquirido corresponde, ao menos, aos recursos linguísticos atinentes ao patamar de proficiência já atingido.

(6) O perfil de usuário deve ser determinado a partir das suas necessidades de aprendizagem. As necessidades de aprendizagem correspondem, ao menos, aos recursos linguísticos atinentes ao patamar de proficiência que se almeja atingir.

No tocante ao axioma 5, o conhecimento já adquirido tem uma incidência direta na determinação do quão fáceis de compreender são as paráfrases explanatórias e notas de uso. Estabelecendo um paralelo com as variáveis assinaladas no capítulo 1, facilitar a compreensão desses dois segmentos informativos significa torná-los mais didáticos.

O axioma 6 está relacionado à determinação de um conjunto básico de unidades léxicas que devem estar descritas em um dicionário de aprendizes. Esse axioma também contempla a identificação das propriedades de cada unidade léxica que possam ser consideradas como relevantes para o usuário pretense. Trata-se, em termos metalexográficos, de determinar um conjunto de informações discriminantes para o usuário, isto é, relevantes para ele e das quais ele pode tirar proveito. Os descritores de níveis de referência e as listas de palavras classificadas pela escala do CEFR (2001) precisam formar parte dos critérios de inclusão de verbetes,

acepções e combinatórias léxicas, pois são o principal fator que motiva a consulta ao dicionário de aprendizes.

Não há dúvidas de que a interseção entre a Linguística Aplicada ao ensino-aprendizagem de línguas e a Linguística de *Corpus* contribui com ferramentas que auxiliam a Metalexigrafia no que concerne ao estabelecimento do perfil de usuário de dicionários de aprendizes e à compilação desses materiais. Ainda que as bases de dados disponíveis ainda sejam limitadas em extensão e/ou nas possibilidades de busca oferecidas, não fornecendo dados sobre o uso da língua inglesa por aprendizes de línguas maternas específicas, por exemplo, é possível tirar proveito do conteúdo disponibilizado. À medida que se avança nas investigações sobre que itens léxicos e gramaticais são relevantes para o ensino de inglês e espanhol e na compilação de *corpora*, mais precisos serão os resultados das pesquisas sobre as necessidades linguísticas dos aprendizes dessas línguas e maiores serão as possibilidades de compilar instrumentos lexicográficos para auxiliar no atendimento a essas necessidades.

A Linguística Aplicada permite definir o perfil de usuário de um dicionário sob uma perspectiva “externa”, isto é, a partir da abordagem de ensino-aprendizagem de línguas em que o usuário está inserido. Por outro lado, a contraparte “interna” da aprendizagem de línguas também se faz necessária para a definição do perfil de usuário, posto que, no caso dos adultos que aprendem uma nova língua, o conhecimento prévio da língua materna exerce um papel na determinação das necessidades desses indivíduos. Nesse sentido, os avanços nas pesquisas em Ciência Cognitiva, mais especificamente na Psicolinguística e na Neurolinguística, geram resultados importantes para que se possa compreender como duas ou mais línguas interagem no cérebro dos aprendizes. Este será o tema do próximo capítulo.

4 O PERFIL DE USUÁRIO SOB A ÓTICA DA CIÊNCIA COGNITIVA

A Ciência Cognitiva, por meio das linhas de pesquisa da Psicolinguística e da Neurolinguística, fornece subsídios para que se possa entender o papel de algumas das variáveis que, conforme já comentado em 2.4, influenciam o comportamento do usuário de um dicionário e têm uma relação direta com as suas necessidades de aprendizagem. Das variáveis propostas no âmbito da Lexicografia (NESI, 2015a, p. 66), nesta seção serão tratadas a língua materna, a língua estrangeira, o nível de proficiência linguística, a idade, o nível educacional e o local na determinação do perfil de usuário considerado para fins desta tese.

O que a Psicolinguística e a Neurolinguística têm em comum com a Metalexigrafia é o interesse por investigar as atividades de recepção e de produção na língua estrangeira. Sob a perspectiva metalexigráfica, recepção (leitura e audição) e produção (escrita e fala) são funções linguísticas. Conforme já comentado no capítulo 1, os dicionários de aprendizes em geral almejam auxiliar o usuário durante a resolução de tarefas tanto de recepção como de produção. A Psico- e a Neurolinguística, por outro lado, são duas áreas correlatas que investigam, entre outros aspectos, como os indivíduos executam na memória as atividades de recepção e produção na língua estrangeira (cf. Richards e Schmidt (2010, s.v. *neurolinguistics, psycholinguistics*). Esse procedimento recebe o nome de *processamento*. Segundo essas áreas, há duas variáveis que têm uma participação fundamental no processamento de uma língua estrangeira: a língua materna e o nível de proficiência na língua estrangeira. Ao investigar como ocorre o processamento da língua estrangeira (isto é, como ela é recebida e produzida na memória), a Psico- e a Neurolinguística permitem que se determine com mais clareza algumas das necessidades do perfil de usuário de um dicionário de aprendizes, necessidades essas relacionadas à sua língua materna e ao nível de proficiência na língua estrangeira. A partir dessas informações, é possível propor estratégias de como aprimorar materiais didáticos, entre eles os dicionários de aprendizes, para facilitar o processo de ensino-aprendizagem de línguas estrangeiras, como é o caso do inglês e do espanhol em relação ao Brasil.

Para melhor compreender as informações sobre o processamento de língua estrangeira que são relevantes para esta tese, é pertinente esclarecer alguns aspectos terminológico-conceituais (bilinguismo, aprendizagem, segunda língua) e metodológicos atinentes à pesquisa em Psico- e Neurolinguística.

Em primeiro lugar, para a Psicolinguística e a Neurolinguística, os indivíduos que processam duas ou mais línguas na memória são considerados bilíngues. Embora, para a teoria, os bilíngues sejam aqueles que usam duas ou mais línguas na sua vida diária (GROSJEAN,

2013, p. 5), na prática, diversos estudos classificam indivíduos com baixa proficiência na segunda língua e uso não cotidiano da segunda língua como bilíngues também (WARTENBURGER ET AL. (2003); GOLESTANI ET AL. (2006); NAIR ET AL. (2015)).

Naturalmente, é preciso considerar que brasileiros que possuem níveis ainda muito elementares de proficiência (A1 e A2) em inglês e espanhol e que estão aprendendo essas línguas no seu país de origem, onde ambas não são oficiais, apresentam pouca probabilidade de usar o inglês e o espanhol cotidianamente. Para este trabalho, parece-nos mais apropriado empregar um meio-termo e considerar que o usuário sobre o qual recai nosso interesse corresponde a alguém que, transitoriamente, denominaremos de bilíngue em formação. Por bilíngue em formação, referimo-nos ao estudante que entra em contato com a segunda língua ao menos duas vezes por semana em função de situações formais de aprendizagem (seja por aulas conduzidas por instituições de ensino e atividades extraclasse, seja pelo uso de materiais didáticos para aprendizagem autodirigida).

Em segundo lugar, em Psico- e Neurolinguística, usa-se o termo *segunda língua* para referir-se à língua que foi (ou está sendo) assimilada após a consolidação da língua materna, comumente denominada *primeira língua* nessas áreas⁸⁵. A distinção entre primeira e segunda línguas leva em conta a progressão temporal do contato e assimilação de cada língua e é útil para fins de pesquisa sobre o processo de aprendizagem da segunda língua em níveis iniciais a moderados e para a análise e classificação de participantes de pesquisas psico- e neurolinguísticas conforme o seu histórico de aprendizagem. Dado que essa é a terminologia usada na bibliografia consultada para redigir esta subseção, preferimos empregá-la temporariamente para comentar as contribuições relevantes para a nossa proposta.

Em terceiro lugar, é necessário comentar o que significa aprender uma nova língua sob a concepção da Psico- e da Neurolinguística. Em indivíduos adultos, cuja língua materna já está consolidada, a aprendizagem de uma nova língua é, conforme Bogulski, Bice e Kroll (2019, p. 1053), “um processo difícil que envolve a incorporação de novos sons, palavras e estruturas em

⁸⁵ Os psicolinguistas são cautelosos ao empregar a distinção entre primeira e segunda línguas em pesquisas sobre o desempenho linguístico de indivíduos altamente proficientes na segunda língua e que a usam corriqueiramente e em maiores proporções que a primeira língua. Isso ocorre pelo fato de que comumente associa-se a essa terminologia a concepção de que a primeira língua seria a mais usada e gozaria de maior proficiência, enquanto a segunda seria a menos usada e gozaria de menor proficiência – algo que não se aplica a quem apresenta alta proficiência na língua não materna. Nesse e em diversos outros casos em que não se analisa necessariamente o processo de aprendizagem, prefere-se o uso dos termos *língua dominante* (mais usada e, geralmente, de maior proficiência) e *língua menos dominante* (menos usada e com menor proficiência). A título de curiosidade, Grosjean (2013, p. 13), por exemplo, salienta que a língua dominante nem sempre será a língua materna e que a língua considerada dominante pode variar ao longo da vida, dando como exemplo o caso de um/a bilíngue cuja língua dominante mudou quatro vezes em cinquenta anos.

um sistema linguístico estabelecido”⁸⁶. Em outras palavras, não se cria um novo sistema linguístico, mas sim, incorpora-se novas informações ortográficas, fonológicas, morfológicas, sintáticas e léxico-semânticas a um sistema pré-estabelecido e moldado pela língua materna. Esse tópico será retomado na seção 4.1.

Em quarto lugar, para entender como se chegam a determinadas constatações por meio das pesquisas psico- e neurolinguísticas, faz-se necessário comentar brevemente que tipo de metodologia se emprega nessas áreas e, conseqüentemente, o que as diferencia. A Psicolinguística analisa o comportamento visível dos participantes de seus experimentos, tais como: 1) a acurácia das respostas; 2) o tempo de resposta empregado para executar uma tarefa, indicando nível de esforço cognitivo; e 3) o desempenho na língua, aferido por meio da fluência⁸⁷. Trata-se de um método indireto de aferição do processamento de línguas cujos resultados são empregados para a proposição de modelos teóricos a respeito de como ocorre o processamento de línguas (DÖRNYEI, 2009, p. 43-48).

A Neurolinguística, por sua vez, associa o comportamento visível dos participantes à análise do comportamento do cérebro, empregando, para tanto, escaneamento via ressonância magnética funcional (RMF), tomografia por emissão de pósitrons (PET) e eletroencefalograma – ou seja, os mesmos métodos empregados para detectar tumores e outras anomalias no corpo humano. Esta é uma metodologia direta de aferição para verificar quais são as áreas que se ativam ou não no cérebro durante a execução de uma tarefa e qual é a função atrelada a essa área. Analisa-se também a quantidade dessa ativação associada aos dados acerca do tempo e do nível de acurácia das respostas (DÖRNYEI, 2009, p. 53-66). Como os custos de usar qualquer dos métodos de escaneamento do cérebro são altos e tais metodologias são relativamente recentes em pesquisas relacionadas a bilíngues (anos 2000 em diante), há menos estudos neurolinguísticos em comparação com os estudos psicolinguísticos sobre o particular (DÖRNYEI, 2009, p. 53-54). Apesar disso, cada vez mais se associam os resultados e a metodologia de pesquisas psicolinguísticas com os de pesquisas neurolinguísticas; isso demonstra que as duas áreas são, na verdade, complementares.

No que concerne a esta tese, as contribuições que serão levadas em consideração para detalhar o perfil de usuário de um dicionário de aprendizes de níveis A2 e B1 de aprendizagem são aquelas que gozam de significativa aceitação na literatura. Devido à existência de pontos

⁸⁶ [a difficult process that involves incorporating novel sounds, words, and structures into an established language system]

⁸⁷ Fluência consiste na “habilidade de falar a L2 [segunda língua] em frases extensas com poucas oscilações iniciais, hesitações e pausas longas” [the ability to speak the L2 in larger chunks with fewer false-starts, hesitations, and filled pauses] (TRAXLER, 2012, p. 437).

em comum entre as variáveis que ajudam a definir o perfil de usuário de um dicionário e as variáveis investigadas em pesquisas cognitivas, é possível tirar proveito desse tipo de pesquisa de maneira que os seus dados sejam empregados nesta tese para precisar as características do usuário que investigamos na tese.

Como o perfil de usuário esta tese está em processo de aprendizagem da segunda língua, na subseção 4.1, comenta-se a perspectiva psicolinguística sobre esse tema. Estabelece-se um paralelo com as variáveis metalexigráficas *língua materna, idade, nível educacional e local* (NESI, 2015a, p. 66).

A seção 4.2, por sua vez, está dedicada ao tema do processamento de segunda língua, isto é, como os bilíngues usam o seu conhecimento da segunda língua armazenado na memória para resolver tarefas de recepção e produção e como a língua materna participa desse processamento. Serão comentadas descobertas que refletem o estado da arte sobre o processamento de segunda língua e como elas auxiliam a entender o papel das variáveis metalexigráficas *língua materna e nível de proficiência linguística*.

Na seção 4.3, discorrer-se-á sobre detalhes acerca do processamento fonológico, semântico e morfossintático com especial atenção aos indivíduos que possuem baixa proficiência na segunda língua – ou seja, com perfil similar ao usuário estabelecido nesta tese. Como os verbetes de dicionário têm campos para esses três tipos de informações, as descobertas específicas advindas da Psicolinguística e da Neurolinguística acerca de cada nível de processamento permitem elaborar propostas de descrição metalexigráfica e reforçar propostas lançadas nas seções prévias.

4.1 A participação da língua materna na aprendizagem da segunda língua em adultos

Nesta subseção, comentaremos as particularidades da aprendizagem de uma segunda língua (L2) em indivíduos adultos, isto é, cuja língua materna já está consolidada (idade) e que passaram por todas as etapas da educação básica (nível educacional). A argumentação também está condicionada ao contexto em que ocorre a aprendizagem da L2 no Brasil (local), isto é, um país cuja população, em sua maioria, não precisa do inglês nem do espanhol para executar a maior parte das tarefas comunicativas da vida cotidiana.

A primeira observação a ser feita é que não é possível tratar da aprendizagem de L2 pelo perfil acima descrito sem antes atentar para as circunstâncias nas quais a língua materna (LM) é adquirida. Conforme Dörnyei (2009, p. 23), a aquisição da LM está atrelada ao desenvolvimento cognitivo, individual e social do ser humano. Isso significa que a descoberta

de mundo, a formação e expressão da individualidade e as interações sociais não apenas acontecem juntamente com a aquisição da LM, mas também por seu intermédio. Em outras palavras, e estabelecendo um paralelo com a noção de *funções comunicativas* da Linguística Aplicada (QECRL, 2001, p. 178-179), por meio da sua LM, os falantes aprendem a executar funções comunicativas básicas, como expressar atitudes, “estabelecer relações sociais” e “dar e pedir informações factuais”.

A aprendizagem de uma L2 na fase adulta, por outro lado, não está associada à construção do conhecimento de mundo e à expressão individual e interação social, mas sim à sua ampliação. Em outras palavras, o falante precisa aprender a executar, por meio da segunda língua, as funções comunicativas que já executa na sua LM e que estão intrinsecamente associadas a ela.

No que tange ao desenvolvimento cognitivo, Dörnyei (2009, p. 23) assinala que a LM auxilia na formação e rotulação do sistema de representações semânticas; na aprendizagem da L2, por outro lado, aprendem-se novos “rótulos”, que são associados a um sistema de representações semânticas prévio. Tais “rótulos” nada mais são que as unidades léxicas e suas propriedades fonológicas, ortográficas, morfológicas e sintáticas (combinatórias). Já o sistema de representações semânticas abarca informações sobre o “conteúdo”. Sem dúvidas, pode-se adicionar novos conceitos próprios da L2, mas não se adquire um sistema inteiramente novo de conceitos (TRAXLER, 2012, p. 416). Portanto, parte dos objetivos da aprendizagem de línguas são diferentes quando se trata da LM e da L2. Nos dois casos, o fator *comunicação* está presente; porém, no primeiro, adquire-se uma LM para ajudar a construir uma imagem conceitual do mundo, enquanto no caso da L2, aprende-se uma nova língua para ampliar um sistema formal e conceitual pré-existente. Sendo o nosso usuário hipotético um indivíduo já adulto, naturalmente o sistema de representações semânticas já foi modalizado por intermédio de uma única língua, o português.

Na mesma esteira do desenvolvimento cognitivo, Dörnyei (2009, p. 22) lembra que é com o subsídio da LM que o sujeito constrói a sua individualidade e participa das suas primeiras experiências de inserção social. Por meio das experiências proporcionadas pelo uso da L2, pode-se complementar a identidade individual e social, mas não “construir” uma nova. Trata-se de vínculos fortes que se estabelecem com a LM e que, como são pré-existentes ao contato com a L2, fazem com que esta tenha uma participação complementar, mas não basilar, na sua formação como indivíduo.

No que concerne ao perfil de usuário desta tese, cabe ressaltar que o nosso usuário conta com poucas oportunidades de aprendizagem e interação social na L2 para aprimorar os seus

conhecimentos. Tal escassez de oportunidades se deve tanto às limitações em função do nível ainda elementar de proficiência na L2 (A1 e A2) como também ao contexto linguístico em que está inserido, isto é, um país onde o inglês e o espanhol não são usados cotidianamente para comunicação e no qual o português é a língua proporcionalmente mais requisitada para participar de interações sociais. Ou seja, para aprender línguas como o inglês e o espanhol, o brasileiro precisa enfrentar obstáculos externos (contexto linguístico-social) e internos (proficiência). Acrescenta-se a esse cenário o fato de que, em geral, o contato prévio do brasileiro com o inglês e o espanhol na educação básica ainda está muito aquém do necessário para que se possa atingir um nível pré-intermediário ou intermediário de proficiência. De acordo com o comentado na Introdução da tese, por um lado, metade da população considera que tem níveis básicos de conhecimento em inglês (BRITISH COUNCIL, 2014, p. 8) – e, de fato, há comprovação da sua baixa proficiência em inglês (EDUCATION FIRST, 2019); por outro lado, há também um paradoxo entre o significativo interesse dos brasileiros em aprender espanhol e a carência de materiais de referência para estudantes dessa língua. Ainda há muito o que melhorar no tocante à oferta de oportunidades de ensino-aprendizagem de línguas estrangeiras na educação básica, que é, aliás, o *background* educacional que possuem os usuários almejados por esta tese. Em vista disso, as poucas oportunidades de contato com a L2 que os adultos têm – por meio de algumas horas de aprendizagem formal em cursos de língua, por exemplo – devem ser bem aproveitadas e subsidiadas por materiais didáticos de qualidade, tais como os dicionários de aprendizes.

Em resumo, não há como tratar da aprendizagem de uma L2 na fase adulta sem levar em consideração a LM dos aprendizes, pois ela é um elemento integrante desse processo. Uma vez que o perfil de usuário hipotético desta tese tem o seu processo de aprendizagem de novas línguas ancorado nas experiências com a LM, é preciso tirar o máximo proveito desta para facilitar a aprender a língua estrangeira/segunda. Em termos de estratégias de descrição metalexográfica, isso significa que, para tornar os dicionários de aprendizes de níveis iniciais ferramentas mais didáticas, faz-se necessário aproveitar o conhecimento prévio da LM, de forma a se estabelecer uma “ponte” com a língua estrangeira/segunda. Essa proposta é um aspecto central nesta tese e será desenvolvida ao longo das demais subseções.

Além de participar do processo de aprendizagem de uma L2, a LM também participa do seu processamento, isto é, da execução de tarefas linguísticas de recepção e produção na L2. Segundo Bogulski, Bice e Kroll (2019, p. 1053) por exemplo, a LM está presente no processamento da L2 mesmo quando não é requisitada pela tarefa em questão. Essa é mais uma razão pela qual os dicionários de aprendizes precisam dar importância à LM do seu usuário,

especialmente em níveis iniciais de proficiência na L2, como A1 e A2. A próxima subseção entrará no mérito dessas questões.

4.2 Processamento de segunda língua: duas línguas em interação

Conforme já comentado anteriormente, o processamento de uma língua corresponde ao que ocorre na memória durante a execução de tarefas de compreensão e produção tanto orais como escritas. Em tarefas de recepção, e sob a perspectiva da Psicolinguística, o processamento consiste no acesso ao sistema de representações semânticas (ou nível semântico) a partir de estímulos provenientes da leitura ou da escrita; em tarefas de produção, o processamento consiste no acesso ao sistema formal de uma língua a partir do sistema de representações semânticas.

O objetivo desta seção é definir o perfil de usuário alvo das pesquisas da tese a partir da perspectiva psicolinguística sobre o processamento da segunda língua (L2). Há características do processamento da L2 que são de natureza geral, isto é, se aplicam a todo indivíduo bilíngue, e que são tratadas aqui; há também características do processamento de L2 que são próprias do processamento em bilíngues com baixos níveis de proficiência na L2 e que serão tratadas na próxima seção. Em ambos os casos, atribui-se especial atenção aos fatores que facilitam e aos que dificultam o uso da L2 em indivíduos que, tal como o usuário considerado nesta tese, possui baixos níveis de proficiência na L2. Com base nesse conjunto de informações, no decorrer das seções, propõem-se algumas estratégias metalexográficas de descrição linguística em dicionários de aprendizes (níveis A2 e B1 de aprendizagem) para melhor corresponder ao perfil de usuário estabelecido na tese.

O processamento durante a execução de qualquer tarefa linguística envolve dois eventos básicos: ativação na memória e acesso a uma língua armazenada na memória. Embora não se tenha o costume de definir tais conceitos, a ativação pode ser entendida como uma parte da memória que é recrutada e disponibilizada ante a necessidade de se executar uma tarefa. O acesso, por sua vez, corresponde à recuperação de determinadas informações dentre as que estão ativadas para a resolução dessa tarefa.

Por algum tempo, acreditou-se que, se uma tarefa requisitasse uma determinada língua X, somente essa língua X estaria ativada do cérebro de um bilíngue que tem conhecimentos das línguas X e Y, por exemplo (VAN HELL; TANNER, 2012, p. 149). Uma vez ativada, essa e somente essa língua X seria acessada durante o processamento, enquanto a língua Y

permaneceria “desligada” (BRYLSBAERT; DUYCK, 2010, p. 360-361). Nesse sentido, sustentava-se que as línguas de um bilíngue seriam processadas individualmente.

Contudo, na literatura, já há um consenso sobre a ativação paralela das duas línguas (BOGULSKI et al., 2019, p. 1053; KROLL et al., 2015, p. 380; KROLL; BIALYSTOK, 2013, p. 498; KROLL et al., 2010, p. 374; TRAXLER, 2012, p. 429; VAN HELL; TANNER, 2012, p. 149; WU; THIERRY, 2012; WU et al., 2013). A ativação paralela é um processo automático, isto é, independe da vontade dos indivíduos (WU et al., 2013); ocorre independentemente das línguas em questão (KROLL et al., 2015, p. 380) e do nível de proficiência (BOGULSKI et al., 2019, p. 1053); e é não seletiva, isto é, mesmo que a tarefa que está sendo executada envolva apenas uma das línguas e exija, portanto o processamento somente dessa língua, não há como “desligar” a língua não requerida (KROLL; BIALYSTOK, 2013, p. 504; VAN HELL; TANNER, 2012, p. 149). Tal como observa Traxler (2012, p. 421), “[...] a ativação não respeita a distinção entre as duas línguas”⁸⁸, o que significa que a segmentação entre o que se conhece como “língua portuguesa”, “língua inglesa” e “língua espanhola”, por exemplo, é apenas externa, pois internamente, na memória, não há tais divisões.

A ativação paralela pode acontecer em qualquer nível de processamento, a saber, ortográfico, fonológico, morfológico, sintático e semântico (KROLL et al., 2015, p. 380), e ocorre ao menos em um deles (KROLL; BIALYSTOK, 2013, p. 504). Um dos dados mais curiosos sobre a ativação paralela das duas línguas do bilíngue é que esse fenômeno ocorre independentemente do nível de proficiência do indivíduo – e não somente em bilíngues com baixo nível de proficiência na L2, como se poderia supor (KROLL; BIALYSTOK, 2013, p. 505). Nas pesquisas em laboratório, observou-se que o que diferencia bilíngues de diferentes níveis de proficiência é a quantidade de ativação da LM, que é, evidentemente, inversamente proporcional ao nível de proficiência na L2.

A coativação da LM é, portanto, um fator inerente ao processamento da L2. Para efeitos do ensino-aprendizagem de língua estrangeira, e particularmente na compilação de materiais didáticos como os dicionários de aprendizes, a ativação paralela da LM em qualquer nível de proficiência é mais um argumento em prol do aproveitamento da LM para fins de ensino-aprendizagem de uma L2 em qualquer estágio, e não somente em níveis iniciais.

⁸⁸ [activation does not respect the distinction between the two languages]

A ativação paralela das línguas de um bilíngue durante o processamento da L2 gera uma competição por seleção. A essa competição, os teóricos denominam *interação*, fenômeno para o qual, conforme comenta Li (2013, p. 141), há uma série de evidências⁸⁹ na literatura.

A interação entre as línguas durante o processamento, por sua vez, pode causar influências interlinguísticas [crosslinguistic influences], isto é, a transferência⁹⁰ de formas, estruturas, significados e categorizações entre duas línguas durante o processamento de uma delas. As influências interlinguísticas podem afetar o processamento facilitando-o, acelerando-o e contribuindo para que uma língua seja compreendida ou produzida de uma maneira que condiga com seu o padrão de uso; ou podem deixar o processamento mais dispendioso, lento e/ou fazendo com que a língua-alvo da tarefa não seja compreendida ou produzida de acordo com o modo como normalmente é usada (ODLIN, 2013, p. 1). Jarvis (2013, p. 1) acrescenta que, seja nas influências interlinguísticas facilitadoras, seja nas custosas, pode haver a transferência de formas, estruturas, significados e categorizações ocorre tanto. Por um lado, e em línguas com alto grau de similaridade ortográfica e semântica como o português e o espanhol, por exemplo, as influências interlinguísticas positivas podem contribuir para a aprendizagem. Por outro lado, e conforme Jarvis (2013), influências custosas podem acarretar o uso excessivo [*overuse*], subutilização [*underuse*] e evasão [*avoidance*] de formas, estruturas e funções. Um exemplo é a presença do pronome sujeito de primeira pessoa no inglês, no português e no espanhol e os diferentes graus de incidência desse pronome em cada língua. Por influência do português, um aprendiz brasileiro pode apresentar a tendência de subutilização desse pronome no inglês; também por influência do português, um aprendiz brasileiro pode apresentar a tendência de uso excessivo do pronome no espanhol⁹¹. No entanto, somente pela análise de um *corpus* de aprendizes se pode identificar que unidades léxicas estão sujeitas a *overuse*, *underuse* ou *avoidance*.

⁸⁹ Entre os testes mais comuns usados para verificar a ativação paralela estão os que usam elementos cognatos e homógrafos e homófonos no processamento de palavras isoladas e de sentenças. No caso dos cognatos, a rapidez com que as respostas aos elementos cognatos são proferidas constituem uma vantagem no processamento e são um indicativo diversas vezes usado na literatura para aferir a ativação paralela, pois, se não houvesse tal fenômeno, tal vantagem não ocorreria (TRAXLER, 2012, p. 420-421). Já nos testes com homógrafos e homófonos, a lentidão das respostas a esse tipo de palavras em comparação a palavras não homófonas e não homógrafas é interpretada como uma evidência da ativação paralela das duas línguas do bilíngue; mais precisamente, a ativação simultânea de conceitos diferentes retarda o processamento da tarefa linguística que precisam desempenhar (TRAXLER, 2012, p. 421). Nesse caso, considera-se que as interações interlinguísticas tornam o processamento mais dispendioso (KROLL; BIALYSTOK, 2013, p. 505).

⁹⁰ Ainda conforme Jarvis (2013, p. 1), há uma preferência pelo uso do termo politicamente correto *influências interlinguísticas*, mas, na prática, é comum que também se empregue as designações *transferência*, *transferência positiva* e *transferência negativa* para os fenômenos comentados no parágrafo ao qual esta nota se refere.

⁹¹ No inglês, os pronomes em geral são notadamente mais marcados em função da morfologia mais sucinta dessa língua, frente à morfologia flexional do português que permite a sua omissão apoiada na flexão verbal. No espanhol, o pronome de primeira pessoa é evitado, pois seu uso excessivo é avaliado negativamente.

Nesse cenário, a proficiência é, uma vez mais, um fator que intervém no processamento (VAN HELL; TANNER, 2012, p. 159). Da mesma forma que regula a quantidade de ativação de cada língua, o nível de proficiência regula também a maneira como ocorrem as interações entre as línguas de um bilíngue (KROLL; BIALYSTOK, 2013, p. 499). Dito em outros termos, a proficiência determina a direção mais comum das influências entre as línguas do bilíngue, isto é, se são da primeira língua na L2, ou se são da L2 na primeira (TRAXLER, 2012, p. 423; 425). Em bilíngues que estão aprendendo a L2 já adultos, que apresentam baixa proficiência na L2 e que usam a LM muito mais que a L2 no seu cotidiano (tal como o perfil de usuário desta tese), a direção mais comum das influências interlinguísticas é da primeira língua em relação à segunda⁹². No nosso caso, é necessário atentar para as influências que o português como LM e de maior proficiência pode exercer sobre o inglês e o espanhol, línguas nas quais o usuário estipulado por esta tese apresenta níveis de proficiência consideravelmente menores (A1 e A2).

A similaridade é, em conjunto com a proficiência, outro fator que motiva as influências interlinguísticas (JARVIS, 2013, p. 2). Trata-se de um fenômeno comumente relacionado à noção de tipologia linguística – que, por sua vez, ampara-se na descrição de línguas para estabelecer universais linguísticos e variações entre as línguas (WHALEY, 1997, p. 18) e é amplamente conhecida pela investigação de famílias linguísticas, por exemplo. Na Psicolinguística, por outro lado, a questão da similaridade é considerada no âmbito da análise linguística. Nesse sentido, atenta-se, por exemplo, aos níveis de organização da linguagem, como fonológico, ortográfico, morfológico, sintático, léxico-semântico, pragmático (JARVIS; PAVLENKO, 2008, p. 19-20)⁹³. Apesar de haver estudos sobre influências interlinguísticas nos diversos níveis mencionados, o tipo de similaridade que se utiliza mais frequentemente em pesquisas psicolinguísticas é o que concerne ao nível ortográfico, com o uso de homógrafos e cognatos para conduzir experimentos (cf. final da seção para distinção terminológica).

A interação entre as línguas decorrente da sua ativação paralela leva ao questionamento de como os bilíngues resolvem a questão da competição entre ambas. Vale recordar que essa competição sempre vai existir, pois a ativação paralela das duas línguas e a consequente competição que gera as influências interlinguísticas, são processos que ocorrem automaticamente nos bilíngues. Enquanto a coativação e as influências interlinguísticas são

⁹² Li (2013, p. 143) alerta que isso não significa que não possam ocorrer influências da segunda língua na primeira, mas sim que a ocorrência dessa classe de influências é proporcionalmente menor.

⁹³ Jarvis e Pavlenko (2008, p. 19-20) citam ainda os níveis discursivo e sociolinguístico, que não serão considerados nas nossas análises.

fenômenos aferidos em laboratório, a seleção leva em conta aspectos internos e externos ao indivíduo.

Uma proposta que vem sendo considerada na literatura para explicar a capacidade de seleção frente à competição entre línguas é a hipótese do controle inibitório de Green (1998). Ao contrário da coativação e das influências interlinguísticas, o controle inibitório é, segundo Green (1998, p. 70-71), uma ação que ocorre automaticamente (isto é, sem o controle do indivíduo e sem que ele esteja ciente desse processo). O controle inibitório consiste basicamente na identificação dos elementos da língua não requisitada e a consequente inibição da ativação desses elementos, tudo em um processo automático que tem por objetivo fornecer uma resposta o mais adequada possível à situação (GREEN, 1998, p. 70-71). Green e Abutalebi (2013) tratam mais a fundo a questão, discorrendo acerca da influência do tipo de tarefa, do entorno no qual o bilíngue desempenha diferentes tarefas linguísticas e da proficiência. A noção de tarefa que os autores empregam é muito similar à utilizada em Linguística Aplicada ao ensino de línguas, pois considera o objetivo a ser alcançado, o gênero discursivo envolvido, a língua a ser empregada e o interlocutor. Esses fatores, em conjunto, determinam a inibição da ativação da língua não requisitada para a tarefa. O entorno, por sua vez, consiste nos locais onde o bilíngue usa cada língua. Green e Abutalebi (2013, p. 518) estabelecem três tipos de entorno: a) cada língua é usada em entornos separados, com interlocutores próprios de cada entorno; b) as duas línguas são usadas corriqueiramente, podendo ser nos mesmos entornos, mas separadamente com interlocutores diferentes; e c) as duas línguas são usadas corriqueiramente nos mesmos entornos e de maneira intercalada num mesmo enunciado de maneira consciente e proposital (*code-switching*).

As postulações acerca do controle inibitório se aplicam muito mais aos casos de bilíngues com alta proficiência nas duas línguas (GREEN; ABUTALEBI, 2013, p. 525). Conforme Green e Abutalebi (2013, p. 521), à medida que a proficiência na L2 aumenta, aumentam também a sua quantidade de ativação e as influências interlinguísticas que geram custos ao processamento. É com base nesse contexto que o controle inibitório é proposto. Em indivíduos com menor proficiência em uma das línguas, o controle inibitório também será menor (GREEN; ABUTALEBI, 2013, p. 525). É nessa última situação que se encontram os usuários almejados por esta tese. Portanto, é plenamente possível supor que terão menor capacidade de inibir a LM ao desempenhar tarefas na L2, pois a proficiência gera uma desproporção na ativação das duas línguas, com maior propensão às influências interlinguísticas da LM na L2 e menor controle inibitório sobre a LM ao se processar a L2. No que concerne ao entorno, tendo em vista o que já se comentou sobre o português ser a única

língua necessária para executar tarefas cotidianas, e considerando que o contato com a L2 será restrito às aulas de língua e ao menos à resolução de tarefas didáticas em casa, nenhum dos entornos identificado pelos autores parece corresponder àquele em que o usuário pretendo pela tese está inserido⁹⁴. Trata-se de uma conjuntura ainda a ser investigada para que se possa descobrir como o controle inibitório ocorre no tipo de entorno que consideramos na tese.

Em resumo, há, portanto, três fenômenos que compõem o processamento da L2 independentemente do nível de proficiência: a ativação paralela e constante da LM e da L2, a interação entre elas, as consequentes influências interlinguísticas e, finalmente, a seleção de elementos em cada nível de organização da linguagem. Portanto, uma característica geral, comum ao processamento de L2 em diversos grupos de indivíduos, é o fato de que esse tipo de processamento é influenciado pela LM, ora facilitando-o, ora tornando-o mais dispendioso. No decorrer de todo o processamento da L2, a proficiência atingida em cada língua cumpre um papel fundamental na quantidade de ativação de cada língua, na capacidade de inibição da LM e na seleção dos elementos necessários para resolver a tarefa linguística que se esteja executando. No caso dos usuários a quem a tese se dirige – ou seja, bilíngues em formação, que se encontram em etapas iniciais de aprendizagem da L2, que possuem níveis elementares de proficiência na mesma e uma capacidade de controle inibitório reduzida –, naturalmente espera-se que haja mais ativação da LM, causando uma influência acentuada desta no processamento da L2 (*i.e.*, na sua aprendizagem e uso). Consequentemente, pode-se esperar um menor controle inibitório da LM.

Antes de seguir com a explanação, julgou-se pertinente esclarecer questões terminológicas da Psicolinguística que foram comentadas nos parágrafos anteriores e que podem causar dúvidas ou mal-entendidos nos estudiosos de Linguística geral.

⁹⁴ Para ilustrar as observações sobre o entorno, mencionaremos como exemplo parte da nossa experiência pessoal com a aprendizagem de inglês na idade adulta, no Brasil, nos níveis A1 e A2, em um curso de línguas presencial para adultos orientado pela abordagem comunicativa com 3h/aula semanais e 2h dedicada a estudos em casa e atividades de lazer em língua inglesa. O uso do inglês, além de relativamente baixo comparado ao português, era predominante na sala de aula, o que poderia se enquadrar como o entorno (a). No entanto, em sala de aula, usava-se o português e o inglês, ora separadamente (b), ora indistintamente num mesmo enunciado (c), mas devido à falta de proficiência suficiente. Em nenhum momento estamos afirmando que a nossa experiência é a que ocorre em absolutamente todos os contextos de ensino-aprendizagem; o que se deseja demonstrar é que o entorno de uso das línguas inglesa e espanhola que o nosso perfil de usuário tem disponível pode não corresponder aos assinalados por Green e Abutalebi (2013).

4.2.1 Distinções terminológicas: homógrafos e cognatos na perspectiva psicolinguística

A começar pelos homógrafos, nas pesquisas em Psicolinguística, essa designação vem sendo aplicada nos casos de unidades léxicas que possuem a mesma representação ortográfica, mas significados diferentes, independentemente de se tratar de comparações entre palavras de uma mesma língua ou de mais línguas. Para especificar os instrumentos usados em seus estudos, os psicolinguistas costumam usar expressões como *interlingual homographs* e *interlexical homographs* em seus trabalhos para esclarecerem que foram utilizadas palavras homógrafas de línguas diferentes, às vezes em concomitância com *homograph(s)*, referidos ao mesmo fenômeno (cf. por exemplo De Groot (2013, p. 72-ss)). Neste capítulo, para falar de pesquisas psicolinguísticas, adota-se o uso de *homógrafo* conforme a perspectiva dessa área de estudos.

Na Linguística geral, também costumam ser consideradas homógrafas as palavras com grafia igual ou similar que apresentam significados ou funções diferentes e a literatura tampouco especifica se se trata de um fenômeno que se aplica somente a palavras de uma mesma língua ou se também se estende a palavras de línguas diferentes (CRYSTAL, 2008, s.v. *homography*). Na linguística brasileira, entretanto, o entendimento que se costuma ter é o de que a homografia se refere à similaridade gráfica entre palavras de uma mesma língua, como se verifica a partir das explanações no Hou (2009, s.v. *homógrafo*) e no Sac (2010, s.v. *homógrafo*). Alternativamente, a designação *heterossemântico* é usada no Brasil para referir-se a palavras de grafia e/ou pronúncia similares entre duas línguas, quase que exclusivamente entre línguas neolatinas como o português, o espanhol e o italiano. O termo, no entanto, não está isento de contradições. De acordo com Bugueño Miranda (2002, p. 189), os usos que se fazem do termo *heterossemântico* partem da premissa de que a pluralidade semântica ocorre somente entre palavras similares de línguas diferentes, quando, na verdade, a heterossemia também se observa em palavras de uma mesma língua. Seria uma contradição, portanto, que esse termo se aplicasse somente nos casos em que as palavras são de línguas diferentes, pois a sua designação não apresenta essa restrição. O autor aponta também que, na Linguística geral, a heterossemia está relacionada à homografia, à homofonia e à homonímia, o que torna esse problema terminológico ainda mais complexo.

Os cognatos, por sua vez, são definidos pelos psicolinguistas como “palavras que compartilham tanto forma (ortográfica e/ou fonológica) como significado em um par de línguas (embora não necessariamente de forma completa)” (DE GROOT, 2013, p. 73). Em ambos os casos, os psicolinguistas analisam somente a dimensão da similaridade ortográfica e a respectiva contraparte semântica.

Em contrapartida, na Linguística, é comum que se considere os cognatos sob a perspectiva da origem etimológica. Nesse sentido, duas unidades léxicas são classificadas como cognatos se, além de similares na forma e no significado (HARTMANN; JAMES, 2010, s.v. *cognate*), pertencem a línguas que possuem uma mesma raiz etimológica (CRYSTAL, 2008, s.v. *cognate*). É o caso, de *correr* no português e no espanhol, que são similares na forma e no significado e, como línguas românicas, derivam da mesma raiz latina *currere* (cf. DRAEe, 2014, s.v. *correr*; Hou, 2009, s.v. *correr*). Hartmann e James (2010, s.v. *cognate*) ainda acrescentam que os empréstimos não se classificam como cognatos: “às vezes, palavras em duas línguas são similares na forma e no significado, mas são empréstimos e não formas cognatas”⁹⁵. É o caso do inglês *research*, cuja origem não é germânica, mas sim um empréstimo do francês antigo *recherche* (OALD8e, 2010, s.v. *research*), que, por sua vez, possui origem latina (TLFi: Trésor de la Langue Française informatisé, 1994, s.v.s. *recherche*, *chercher*).

Para as pesquisas em Psicolinguística, portanto, a origem etimológica não está presente no conceito de *cognato* usado na área. Os estudos com cognatos são um dos alicerces da investigação dos tipos de influências interlinguísticas entre duas ou mais línguas. Segundo sintetizam Kroll et al. (2015, p. 380), “a convergência na forma e no significado léxico dos cognatos tende a facilitar o processamento durante o reconhecimento de palavras e em tarefas de produção”⁹⁶. Frente a unidades léxicas como *agenda* (espanhol) e *agenda* (português), por exemplo, espera-se que o seu *status* de cognatos, na concepção psicolinguística do termo, confira maior rapidez ao processamento. Por outro lado, a convergência de forma, acompanhada de divergência no significado, poder torna mais custoso o processamento durante o reconhecimento de palavras, como indicam os resultados de estudos com homógrafos e homófonos comentados por Traxler (2012, p. 421).

É pertinente observar que, mesmo em se tratando de cognatos, a similaridade entre elementos de duas línguas não é total, ou seja, não se aplica a todos os níveis de organização da linguagem; em alguma medida, sempre haverá anisomorfismo entre duas línguas, pois cada língua possui uma configuração única (BUGUEÑO MIRANDA, 2010, p. 66). Em função disso, a facilitação do processamento decorrente de unidades léxicas depende do tipo de tarefa linguística que se está executando (compreensão escrita e oral; produção escrita e oral). Tome-se novamente o exemplo da unidade léxica hispânica *agenda* e da unidade léxica do português *agenda*, considerada como sua equivalente. A similaridade (isomorfia) entre os níveis

⁹⁵ [Sometimes words in two languages are similar in form and meaning but are borrowings and not cognate forms]

⁹⁶ [The convergence in lexical form and meaning for cognates tends to facilitate processing on word recognition and word production tasks]

ortográfico e semântico das duas unidades léxicas é evidente. Numa tarefa de compreensão leitora de um texto em espanhol com a palavra *agenda*, é plenamente possível esperar que um falante nativo de português tenha o processamento facilitado por uma influência interlinguística entre a sua língua materna e a língua alvo da tarefa. No entanto, a similaridade não se aplica ao nível fonológico, pois a representação fonológica de *agenda* no português (/azenda/)⁹⁷ difere da representação de *agenda* no espanhol (/axénda/)⁹⁸. O problema da similaridade é que, especialmente em etapas iniciais de aprendizagem da L2, os aprendizes tendem a estabelecer uma relação de equivalência 1:1 entre a L2 e a sua LM (RINGBOM, 2006, p. 5). Nesse sentido, é possível esperar que o mesmo falante de português se depare com influências interlinguísticas que tornam o processamento de *agenda* mais custoso em uma tarefa de produção oral em espanhol.

Uma última observação sobre a similaridade é que, segundo Rigbom (2006, p. 11), se trata de um aspecto de natureza subjetiva, não sendo possível calcular quando um indivíduo vá perceber elementos similares entre duas línguas. O melhor que se pode fazer são projeções baseadas em cálculos do nível de similaridade ortográfica das unidades léxicas, usando fórmulas como a de Van Orden (1987).

4.2.2 *Psicolinguística e Linguística Contrastiva na determinação do perfil de usuário*

Sendo as influências interlinguísticas um fenômeno intrínseco ao bilinguismo; tendo em vista que o usuário estabelecido nesta tese um bilíngue em formação; e considerando que esses indivíduos possuem baixa capacidade de inibir a sua LM, definir as características do perfil de usuário de dicionários de aprendizes brasileiros de inglês e de espanhol de níveis A1 e A2 de proficiência requer que se levem em conta as interações entre o português e o inglês e entre o português e o espanhol.

No que concerne aos dicionários de aprendizes, as influências interlinguísticas da LM no processamento da L2 precisam ser consideradas a partir de duas perspectivas. A primeira é

⁹⁷ Transcrição nossa. O instrumento de referência comumente responsável por fornecer transcrições fonológicas é o dicionário. Na lexicografia de língua inglesa, uma grande variedade de classes de dicionários, como os gerais de língua, de aprendizes, de sinônimos, de colocações, entre outros, fornecem essa informação. Os dicionários gerais e de uso do português, porém, não incluem transcrição fonológica em seus verbetes de maneira geral, conforme se pode constatar em Bugueño Miranda (2019e). O Hou (2009) reserva as transcrições fonológicas aos verbetes de estrangeirismos, como s.v.s. *rave*, *kung fu* e *kyrie*. Arriscamos dizer que não há um dicionário monolíngue de língua portuguesa que forneça tal informação de maneira sistemática a todos os seus verbetes.

⁹⁸ Transcrição fonológica conforme Señas (2001, s.v. *agenda*). Não é comum encontrar transcrição fonológica em verbetes de dicionários monolíngues de espanhol, seja para falantes nativos, seja para aprendizes. O Señas (2001) é o único que apresenta esse segmento informativo em seus verbetes.

quanto às dúvidas que motivam a consulta ao dicionário, já que a resolução de dúvidas é a razão de ser da obra lexicográfica.

Com o subsídio teórico da Psicolinguística, é possível identificar as influências interlinguísticas custosas como uma das fontes de problemas no processamento da L2. As causas dos percalços no processamento da L2, no entanto, são subjetivas e difíceis de se detectar. Com base no exposto até aqui sobre a interação as línguas de um bilíngue, é possível supor duas motivações para o fenômeno das influências interlinguísticas da LM no processamento da L2. Uma hipótese é a de que o bilíngue tem conhecimento dos elementos da L2 (unidades léxicas e/ou propriedades de unidades léxicas) necessários para se desempenhar adequadamente, mas, pelo fato de a LM ser a mais usada e gozar de maior proficiência, acaba gerando um custo maior para o processamento da L2, que pode ser ainda mais prejudicado em função da baixa capacidade de controle inibitório. Outra hipótese é a de que o bilíngue não tem conhecimento de alguns dos elementos da L2 (unidades léxicas e/ou propriedades de unidades léxicas) que são necessários para um desempenho adequado, o que leva o indivíduo a recorrer à LM para processar a L2.

Seja como for, frente a esse cenário, o papel da Metalexigrafia que se ocupa dos dicionários de aprendizes é, por um lado, descrever a língua em uso para dar conta de dúvidas que possam surgir, seja em razão de um custo maior no processamento da L2, seja por qualquer outro motivo, como já se vinha argumentando anteriormente. Nesse caso, a Linguística de *Corpus* constitui um passo fundamental para se fornecer uma informação de qualidade para o usuário. Por outro lado, a área também precisa atentar para os casos em que as influências interlinguísticas da LM, além de gerar um custo maior para o processamento da L2, terminam culminando em manifestações que fogem dos padrões de uso da L2, seja porque o controle inibitório não foi suficiente para suprimir a LM em algum dos níveis de processamento, seja por qualquer outra razão. Para esse segundo caso, é possível fornecer uma orientação mais específica para o público brasileiro por meio de investigações empíricas sobre as manifestações errôneas mais recorrentes, por exemplo. A área que melhor pode auxiliar a identificar tais manifestações é a Linguística Contrastiva, mais conhecida hoje como *Learner Corpus Research* (cf. próximos parágrafos).

Enquanto a Psicolinguística analisa as etapas do processamento da L2, a atual Linguística Contrastiva se ocupa em especial das manifestações linguísticas decorrentes desse processamento (GILQUIN; GRANGER, 2015). Por essa razão, as duas áreas estão diretamente relacionadas, tendo em comum o interesse pela análise das influências interlinguísticas da LM no processamento de L2. Nesse sentido, não se pode ignorar a semelhança de concepções e de

nomenclatura entre a Psicolinguística e a Linguística Contrastiva. O que diferencia uma área da outra é que o foco da Psicolinguística recai sobre o modo como uma língua é processada e que variáveis estão envolvidas nesse processamento; em outras palavras, seu objeto de análise é o processo, e não o produto (língua). Um dos aspectos que a área demonstra por meio de seus estudos é o fato de que a LM é uma importante variável no processamento da L2 e que, em função disso, precisa ser levada em conta quando se trabalha com o bilinguismo. A área, portanto, se ocupa da identificação de fenômenos relativos ao bilinguismo e da proposição de teorias que expliquem esses fenômenos.

A Linguística Contrastiva, por outro lado, na sua corrente concepção como *Learner Corpus Research* (LCR) tem como objeto de análise o produto das influências interlinguísticas, isto é, a língua processada e produzida pelos aprendizes de L2. Atualmente, seus objetivos são, do ponto de vista teórico e prático, “obter um melhor entendimento sobre os mecanismos de aquisição de língua estrangeira ou segunda língua e projetar ferramentas e métodos de ensino de línguas mais eficientes”⁹⁹ (GRANGER, 2015a, p. 9). A metodologia mais empregada é a descrição do uso que os aprendizes fazem da língua-alvo em situações de produção reais ou realistas (GRANGER, 2017). Nota-se que a LCR e a Psicolinguística são altamente complementares. Por um lado, a LCR é capaz de ilustrar os fenômenos identificados pela Psicolinguística e reforçar modelos teóricos dessa área; por outro lado, a LCR se beneficia da Psicolinguística na medida em que as características próprias das manifestações dos aprendizes podem ser interpretadas à luz da Psicolinguística. Além disso, a principal aplicação dos resultados dos estudos em LCR é a compilação de materiais didáticos para o ensino-aprendizagem de línguas estrangeiras, como os dicionários, gramáticas pedagógicas, livros didáticos e CALL (Computer-Assisted Language Learning [Aprendizagem de língua mediada por computador]) (GRANGER, 2015b, 2017, 2018).

A atual configuração da LCR está baseada nas diferentes abordagens adotadas pela Linguística Contrastiva ao longo dos anos, seja por meio da descrição e comparação de pares de línguas (Análise Contrastiva), seja através da análise das produções de aprendizes (Análise de Erros e Interlíngua). As antigas abordagens, por sua vez passaram por uma revisão. Em função de que muitos aspectos dos primórdios da Linguística Contrastiva estão presentes na sua concepção atual como LCR, pode ser difícil distinguir em linhas gerais o que realmente faz parte da LCR. Os próximos parágrafos apresentam um breve histórico da área para, então, esclarecer o seu viés atual e a sua relevância para a Lexicografia e para a tese.

⁹⁹ [to gain a better understanding of the mechanisms of foreign or second language acquisition and to design more efficient language teaching tools and methods]

Sob o viés da Análise Contrastiva, pregava-se que a LM deveria ser evitada no ensino-aprendizagem da L2 em função das suas influências negativas (FENN, 2014, p. 81), as quais seriam, conforme seus adeptos, as únicas fontes problemas e dificuldades na aprendizagem e uso da L2 (FERNÁNDEZ, 1997, p. 14-15). Empregava-se uma metodologia de descrição de duas línguas, seguida da seleção de aspectos anisomórficos entre ambas (*ibid*), os quais serviriam de base para a determinação do conteúdo programático (*syllabus*) do ensino de uma determinada L2 a falantes de uma determinada LM. A pretensão de prever erros e o viés predominantemente negativo atribuído LM, característicos da Análise Contrastiva, foram alvo de duras críticas, que determinaram a sua decaída entre os estudos linguísticos. Contudo, é preciso considerar que a Análise Contrastiva é fruto da ciência do seu tempo e da tecnologia disponível nos anos 1940 e 1950 (MACDONALD, 2016, p. 104). O fato de haver chamado a atenção para a relação entre LM e L2 e para as dificuldades de aprendizagem de L2 é, sem dúvidas, um mérito dessa abordagem.

Com o passar do tempo, a LM passou de tabu (COOK *apud* HALASA; AL-MANASEER, 2012, p. 72) para ser considerada como uma entre as muitas variáveis que causam problemas e dificuldades na aprendizagem e uso da L2. Essa perspectiva permeia o trabalho da segunda corrente da Linguística Contrastiva, a Análise de Erros (FERNÁNDEZ, 1997, p. 17-18), que esteve em voga ao longo dos anos 1960 e 1970 (MACDONALD, 2016, p. 205). A metodologia da Análise de Erros consiste na identificação de erros em produções de aprendizes baseada na comparação intuitiva do pesquisador ou do professor sobre o uso da língua-alvo. Nota-se que o viés comparativo da Análise Contrastiva se mantém na Análise de Erros; o que muda é o objeto de análise, que passa a ser as manifestações concretas dos aprendizes. Também são investigadas as motivações desses erros, dentre as quais se considera a influência da LM, sem se ignorar, contudo, que haveria também outras razões (DURÃO, 2004; FERNÁNDEZ, 1997, p. 18-19).

A Interlíngua, por sua vez, surgiu como uma resposta ao enfoque praticamente exclusivo da Análise de Erros em relação às produções de aprendizes. Sob o viés da Interlíngua, as manifestações linguísticas dos aprendizes precisam de uma descrição própria, ao invés de serem analisadas a partir de comparações com a língua-alvo (SELINKER, 2014, p. 237). Segundo Larsen-Freeman (2014, p. 217), a adoção da língua-alvo como norma reduz as manifestações dos aprendizes a nada mais que a uma língua deficiente. De acordo com Granger (2021), a perspectiva da Interlíngua costuma ser adotada atualmente nas pesquisas em Aquisição de Segunda Língua.

Com o surgimento da Linguística de *Corpus* nos anos 1980, as pesquisas a respeito das manifestações linguísticas dos aprendizes de L2 passaram a contar com *corpora* de aprendizes. A área de estudos recebeu, então, um novo nome – *Learner Corpus Research* (LCR). A LCR é considerada como uma área nova e independente da Linguística Contrastiva e do seu viés pré-*corpus* pelos seus adeptos, como Granger (2021) e Granger et al. (2015). Entretanto, uma análise dos postulados teórico-metodológicos da área, como a que fazemos a seguir, revela que a LCR é, na verdade, uma evolução natural da Linguística Contrastiva, fruto dos avanços nos âmbitos teórico, com Sinclair, e tecnológico, com o aperfeiçoamento de equipamentos de computação de dados, além da sua massificação.

Primeiramente, no que concerne à terminologia, as manifestações linguísticas dos aprendizes ora são chamadas de interlândia [*interlanguage*], ora de língua do aprendiz [*learner language*] pela LCR. Além disso, as três metodologias de investigação da LCR recuperam aspectos das antigas metodologias em Linguística Contrastiva. São elas: *Contrastive Interlanguage Analysis* [Análise Contrastiva da Interlândia]; *Computer-aided Error Analysis* [Análise de Erros Auxiliada por Computador]; e *Integrated Contrastive Model* [Modelo Contrastivo Integrado].

De acordo com Granger (2015a, p. 7), a *Contrastive Interlanguage Analysis* (CIA) é a metodologia mais popular das três, apesar das críticas que recebe. É possível que a sua popularidade se deva às características do seu objeto de análise, isto é, *corpora* de aprendizes sem anotação de erros e, por vezes, sem anotação de classes de palavras (POS) tampouco, os quais constituem a maior parte dos *corpora* de aprendizes disponíveis. Sob essa metodologia, são investigadas características distintivas dos aprendizes, mais precisamente casos de erros, *overuse* [uso excessivo] e *underuse* [subutilização] (GILQUIN; GRANGER, 2015) – ou seja, fenômenos que estão diretamente relacionados às influências interlinguísticas, como já mencionado no início de 4.2.

A CIA prevê duas modalidades de análise baseadas no princípio da comparação (GRANGER, 2015b) – o que, aliás, recorda a Análise de Erros. A primeira modalidade consiste na comparação entre os dados de um *corpus* de aprendizes com os dados de um *corpus* de referência, como de falantes nativos, por exemplo (GRANGER, 2015b), visando determinar que aspectos são distintivos dos aprendizes em comparação com os falantes nativos ou altamente proficientes (GILQUIN; GRANGER, 2015). A segunda modalidade de análise da CIA consiste na comparação entre *corpora* de aprendizes ou subcorpora gerados a partir de um mesmo *corpus* de aprendizes (GRANGER, 2015b). Conforme Gilquin e Granger (2015), por um lado, essa modalidade permite concentrar-se nos aspectos distintivos da interlândia de um

grupo de aprendizes falantes de uma mesma LM ou família linguística; por outro lado, é possível também averiguar as dificuldades inerentes ao desenvolvimento da L2 nos aprendizes em geral e independentemente da sua LM. Em ambas as modalidades, adota-se a perspectiva de análise *corpus-driven*, em que a descrição é feita tomando-se os dados do *corpus* como ponto de partida.

A *Computer-aided Error Analysis* (CEA), por sua vez, é quase uma variante da CIA em termos de objetivos, objeto e metodologia de análise. Porém, segundo Gilquin e Granger (2015), é pouco popular. A principal diferença em relação à anterior é o uso de *corpora* de aprendizes anotados. A baixa profusão de *corpora* de aprendizes com anotação de erros é uma possível justificativa para a impopularidade da CEA. O uso de computador parece estar ressaltado em função de que, no caso da CIA, há *corpora* de aprendizes não digitalizados e analisados manualmente, enquanto *corpora* de aprendizes com anotação de erros seriam sempre processados eletronicamente.

O *Integrated Contrastive Model* (ICM), sob o ponto de vista prático, é uma versão “*corpus-based*” da Análise Contrastiva conforme assinalam Gilquin e Granger (2015, p. 424). Trata-se de uma metodologia *corpus-driven* que estabelece comparações entre um *corpus* da LM do aprendiz e um *corpus* da língua alvo da sua aprendizagem a fim de estabelecer semelhanças e diferenças entre ambos. Além do uso de *corpora*, a diferença entre o ICM e a Análise Contrastiva reside no seu objetivo: ao invés de prever erros, o ICM se propõe a estabelecer possíveis contextos de ocorrência de influências interlinguísticas da LM na L2 e “fornecer explicações plausíveis para certas características da interlíngua” (GILQUIN; GRANGER, 2015, p. 424). Em outras palavras, a metodologia é a mesma, mas os propósitos são outros. Almeja-se, portanto, que o produto da descrição comparativa entre LM e L2 possa servir de auxílio na interpretação dos dados da descrição de *corpora* de aprendizes via CIA ou CEA.

Em vista do anterior, nota-se que a CIA e sua variante CEA são bastante próximas da Psicolinguística por considerarem a LM como uma importante variável do processo de aprendizagem da L2 e pelo seu interesse comum em entender como ela participa desse processo. Enquanto uma atesta a ocorrência de influências interlinguísticas, a outra é capaz de comprovar na prática quando o aprendiz é induzido por esse fenômeno nos diferentes níveis de proficiência. O ICM, por sua vez, também é útil à Psicolinguística na medida em que os dados gerados pela aplicação dessa metodologia podem ser úteis para o desenho de experimentos que investiguem possíveis contextos de influências interlinguísticas.

A complementaridade entre a Psicolinguística e a LCR não somente traz benefícios para ambas, mas também para qualquer área que se interesse pelo ensino-aprendizagem de línguas não maternas, como é o caso da Lexicografia de Aprendizizes. As duas primeiras permitem não somente determinar com mais precisão quem é o usuário pretense de um dicionário e quais são as suas necessidades, mas também que se elaborem estratégias para que os dicionários de aprendizizes sejam mais didáticos e se aproximem mais do seu usuário. Nesse sentido, todas as abordagens da LCR são potencialmente úteis para cumprir com esses objetivos. Primeiramente, a CIA/CEA permite que se determinem as necessidades dos aprendizizes brasileiros de inglês e espanhol. Dentre essas necessidades, estão tanto as dificuldades originárias das influências interlinguísticas da LM que acarretaram um custo ao processamento da L2 e que o aprendiz não conseguiu inibir por falta de conhecimento ou proficiência, bem como outros problemas de origem ainda desconhecida. Em resposta a tais dificuldades, o dicionário de aprendizizes precisa chamar a atenção do consulente no tocante às manifestações dos aprendizizes que divergem dos padrões de uso do inglês ou do espanhol. Sob essa perspectiva, e consoante Granger (2018, p. 23), os *corpora* de aprendizizes se apresentam como uma valiosa fonte de dados sobre o usuário do dicionário. Em segundo lugar, e apesar da sua impopularidade e das eternas e inevitáveis críticas, a abordagem contrastiva da LCR, o ICM, também possui potencial de aplicação na Lexicografia. Os dados desse tipo de pesquisa referentes aos pontos em comum (ou isomórficos) entre duas línguas podem ser usados em paráfrases explanatórias e notas de uso para a criação de contextos potencialmente oportunos para a ocorrência de influências interlinguísticas facilitadoras do processamento. **É extremamente importante esclarecer que não estamos afirmando nem prevendo, muito menos que esperando, que tal medida terá êxito junto a todo e qualquer consulente da obra.** Trata-se de uma aposta em uma estratégia que, embora possa não servir a todos – como ocorre com qualquer iniciativa de aprimoramento de materiais didáticos, aliás –, cria possibilidades de que possam ocorrer influências interlinguísticas facilitadoras do processamento em ao menos parte dos aprendizizes envolvidos.

Alguns aspectos relativos ao uso da LCR precisam ser comentados, bem como as condições que viabilizam a aplicação das metodologias da área para fins lexicográficos.

Em primeiro lugar, de acordo com a proposta desta tese, em um dicionário de aprendizizes, a resolução de dúvidas se faz por duas frentes. Uma delas está baseada na descrição das manifestações das línguas inglesa e espanhola em uso com base nos dados de *corpora* “gerais” de língua, tal como todo dicionário deveria fazer, independente da sua classe. A outra frente consiste no destaque às informações que, sendo relativas à língua em uso, comprovadamente são alvo de influências interlinguísticas custosas da LM conforme dados de

corpora de aprendizes brasileiros de inglês e espanhol. Conforme consta nos próximos capítulos, esse destaque se pode fazer, ao menos, por meio de notas de uso. Não se pode esperar, contudo, que um dicionário de aprendizes (ou mesmo um dicionário) tenha respostas para todas as dúvidas dos aprendizes. Em situações assim, é válido recordar que o professor é quem possui competência para identificar esses outros problemas e realizar intervenções conforme as necessidades dos seus alunos.

Em segundo lugar, não estamos negando que não existam dúvidas e problemas no processamento da L2 advindos de outras razões que não a LM, como, por exemplo, uma informação passada erroneamente ou de maneira incompleta por um professor, colega ou material didático, ou mesmo problemas no uso da língua decorrentes de aspectos mal compreendidos da L2. Um *corpus* de aprendizes é capaz de revelar outras questões de origem desconhecida e que são igualmente relevantes para o usuário, no caso, os brasileiros adultos. O ponto central da argumentação apresentada aqui é o fato de que a LM é a única variável que pode ser identificada mais claramente e isolada para fins de determinação do perfil de usuário. De fato, Granger (2015b), a respeito das pesquisas com *corpora* de aprendizes, chega a assinalar que a LM é o principal fator que influencia o uso da L2 – mas não se nega que existam outros, como observam Gilquin e Granger (2015).

Em terceiro lugar, nossa proposta de uso de *corpora* de aprendizes para fins (meta)lexicográficos está condicionada, naturalmente, à existência desse tipo de ferramenta para os pares de línguas em questão. Conforme os dados da *Learner Corpus Association* (www.learnercorpusassociation.org), uma organização que reúne pesquisadores interessados nesse ramo da Linguística de *Corpus*, a compilação de *corpora* de aprendizes se encontra em um estágio ainda incipiente para muitos pares de línguas, como é o caso do português em relação a outras línguas (cf. uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html).

Também é preciso levar em conta a disponibilidade de acesso aos *corpora* de aprendizes existentes. No caso do inglês, a seção de produções de aprendizes do *Oxford English Corpus*, do *World English Corpus* (Macmillan) e do *Longman Corpus Network* são de acesso restrito às respectivas equipes editoriais, sem que se disponibilize qualquer opção de acesso a interessados externos ou mesmo maiores informações sobre a constituição dessas bases de dados. Tal medida é compreensível em função do fato de que os *corpora* de aprendizes são uma importante fonte de lucros gerados a partir da sua aplicação na compilação de materiais didáticos, entre os quais se incluem não apenas os dicionários de aprendizes, mas também livros didáticos e compêndios gramaticais. Curiosamente, o *corpus* de aprendizes de inglês aparentemente mais abrangente é

também o que oferece uma opção de acesso ao público externo. Trata-se do *Cambridge Learner Corpus*, com 50 milhões de palavras oriundas dos testes de proficiência que a Cambridge Assessment aplica ao redor do mundo. Existe a possibilidade de requisitar acesso ao *corpus* a partir do preenchimento de um formulário on-line; a Cambridge University Press, por sua vez, avalia a petição. Até onde sabemos, o *Cambridge Learner Corpus* é o único que oferece dados de produções de aprendizes brasileiros de inglês desde níveis iniciais de proficiência até níveis mais avançados. Uma importante vantagem desse *corpus* é que contém um subcorpus com pouco mais de 31 milhões de palavras, que constitui uma versão anotada com indicações dos tipos de erros que os aprendizes cometem.

O espanhol, por sua vez, conta com uma iniciativa do Instituto Cervantes, o *Corpus de Aprendices de Español* (CAES, 2020), que, em sua versão 1.3, contém 438.700 palavras (excluídos nomes próprios, números e símbolos) e é de acesso gratuito (<http://galvan.usc.es/caes>). O CAES (2020) é, até onde sabemos, o único *corpus* de aprendizes de espanhol com produções de brasileiros. Não há como saber que parcela do *corpus* está constituída por produções de brasileiros; a única informação fornecida é a de que 124.985 *tokens* (28,4% do total) e 5.214 *types* (59,4% do total) provêm de produções de falantes de português, tanto brasileiros como portugueses. Diferente do *Cambridge Learner Corpus*, os dados fornecidos não recebem anotação por tipo de erro, o que torna a análise dos dados mais dispendiosa.

Nos casos em que não existe um *corpus* de aprendizes para o par de línguas em questão, ou quando não se conta com acesso a ele, há duas opções que podem, inclusive, estar aliadas. Uma delas é apoiar-se na experiência de professores que ensinam uma língua α a aprendizes cuja LM é uma língua β . Essa prática não é alheia à Lexicografia de Aprendizes de língua inglesa, por exemplo, que, em seus primórdios, foi fundamentada nas experiências docentes de professores de inglês como língua estrangeira (JACKSON, 2002, p. 129-130). As observações desses profissionais acerca das dificuldades na aprendizagem e erros comuns no uso do inglês em níveis iniciais podem ser aproveitadas para fins (meta)lexicográficos.

A segunda opção para compensar a falta de *corpora* de aprendizes é o emprego da metodologia ICM para comparar as duas línguas em questão, identificar os aspectos anisomórficos da L2 em relação à LM e chamar a atenção do consultante por meio de segmentos informativos nos verbetes, como as notas de uso, ou ainda aplicando destaques tipográficos e incluindo ilustrações, por exemplo. Uma associação entre esta e a opção anterior constitui o que, segundo Fernández (1997, p. 17-18), se denomina de abordagem “branda” da Análise Contrastiva e que antecipou a Análise de Erros. Sua metodologia consiste em reunir erros de

aprendizes relatados por professores e identificar aqueles que consistem em influências negativas da LM.

Finalmente, além das questões acerca do dicionário como instrumento para resolver dúvidas, da LM como variável identificada nas influências interlinguísticas e do acesso a *corpora* de aprendizes, está a aplicação propriamente tal das metodologias da LCR para análise de erros, em particular. Para fins lexicográficos, o apanhado teórico da LCR disponível até o momento não fornece subsídios suficientes para que se possa interpretar a frequência dos erros encontrados e estabelecer critérios de seleção. Muito embora De Cock e Granger (2005) mencionem o “par perfeito” que os *corpora* informatizados de aprendizes formam com os dicionários de aprendizes, não comentam como interpretar e estabelecer critérios mais precisos de seleção de erros que a consideração do perfil de usuário e da sua proficiência na LE. As autoras até comentam que a frequência dos erros é um fator a ser considerado, e que os erros mais frequentes são fortes candidatos a inclusão nos dicionários de aprendizes, mas não sugerem nenhum parâmetro para definir o que significa ser mais frequente. Em outras palavras, não se sabe qual seria o ponto de corte para definir e separar o que é “mais frequente” – e, portanto, relevante para um dicionário – daquilo que não o é.

Ante um cenário como esse, a alternativa mais óbvia seria analisar como os lexicógrafos responsáveis pelos dicionários de aprendizes que incluem informações sobre erros procederam em relação à sua seleção. Porém, De Cock e Granger (2005) queixam-se de que esse tipo de informação metodológica não era divulgado pelas equipes dos dicionários Longman e Cambridge (à época, a editora Macmillan ainda não havia inserido erros de aprendizes em suas obras). Muito provavelmente, tais especificações, assim como os próprios *corpora* das editoras, não sejam tornados públicos por questões comerciais. Frente a essa limitação, as autoras terminam analisando os tipos de notas de uso incluídos nas obras, questionando-se se todas advinham dos dados dos *corpora* de aprendizes das respectivas editoras.

Em trabalhos mais recentes, Granger (2018, 2015b) volta a tratar da interseção entre a LCR e a Lexicografia e defende novamente a relevância da LCR para os dicionários de aprendizes. A autora comenta que, para a compilação do dicionário MEDAL2 (2007), foi estabelecida uma parceria entre a editora Macmillan e a Univesidade Católica de Louvain, responsável pelo *International Corpus of Learner English* (ICLE, 2009) um *corpus* de aprendizes de inglês que, à época, contemplava falantes de dezesseis LMs – hoje (2021), são vinte e seis. Como o MEDAL2 (2007) se destina a qualquer aprendiz de inglês, os critérios adotados para a seleção de erros foram a frequência associada à distribuição das ocorrências entre falantes de diversas LMs (GRANGER, 2015b). Contudo, para o desenho de um dicionário

de aprendizes falantes de uma LM em particular, o critério de distribuição já não se aplica, restando apenas a frequência.

Do estado da arte comentado anteriormente, conclui-se que ainda são necessárias mais investigações no tocante ao estabelecimento de critérios de seleção de erros para fins de aplicação lexicográfica e pedagógica quando se trata de públicos específicos. Somente assim os dados obtidos por meio da LCR podem ser aproveitados de maneira adequada em um dicionário de aprendizes falantes de uma determinada LM, por exemplo. Por enquanto, além da frequência, a disponibilidade de tempo associada aos recursos financeiros são os critérios mais concretos que se pode estabelecer para definir o que é relevante e executável em termos de análise de dados para aplicação lexicográfica nos moldes que a tese propõe.

Na presente subseção, foi apresentada uma alternativa de como aprimorar os dicionários de aprendizes considerando as motivações para a consulta à obra lexicográfica. Entretanto, o fenômeno das influências interlinguísticas no processamento da L2 também pode ser explorado sob a perspectiva da participação da LM na compreensão leitora das informações da obra lexicográfica. Conforme ilustrado no capítulo 1, até o momento, os dicionários de aprendizes em geral, incluindo os de níveis A1-B1, têm usado a língua-objeto em função metalinguística; nesse sentido, é prática comum usar o inglês para descrever essa língua e o espanhol para descrevê-lo. No entanto, é preciso atentar para o fato de que haverá influências interlinguísticas da LM durante a consulta a essa classe de obras especialmente no processamento de segmentos informativos que demandam maior capacidade de compreensão leitora, como as paráfrases explanatórias e as notas de uso. Uma importante questão a ser investigada é em que medida a similaridade formal entre a LM e a L2 motiva tanto as influências interlinguísticas que facilitam o processamento como as que o tornam mais oneroso. Portanto, a estratégia lexicográfica de descrever a língua-objeto por ela mesma precisa ser analisada caso a caso em relação às suas possíveis vantagens e desvantagens. Por essa razão, o capítulo 5 está dedicado, entre outros objetivos, à avaliação da facilidade de compreensão leitora de paráfrases explanatórias e notas de uso de dicionários de aprendizes de inglês e de espanhol com base nos pressupostos da Psicolinguística sobre nível de similaridade ortográfica e influências interlinguísticas do português nas L2 inglês e espanhol.

Como foi mencionado até o momento, as influências interlinguísticas consistem em um fenômeno comum a diferentes perfis de bilíngues. Há, no entanto, aspectos do processamento de L2 que se observam de maneira especial em bilíngues que se encontram em níveis ainda muito iniciais de proficiência, como é o caso de A1 e A2. Naturalmente, a LM está envolvida e precisa ser convertida em uma importante aliada do processamento das informações de

dicionários de aprendizes de níveis iniciais de proficiência. A subseção a seguir está dedicada a esse tópico.

4.3 Processamento de L2 em bilíngues com baixa proficiência na L2

Na Psicolinguística, existe um modelo sobre o processamento de L2 que segue válido até os dias de hoje, pois, apesar dos seus mais de 25 anos de existência, ainda conta com suporte empírico: o Modelo Hierárquico Revisado [Revised Hierarchical Model] (MHR), de Kroll e Stewart (1994). O MHR trata de como indivíduos que aprendem uma L2 após a consolidação da LM processam essa L2 e é especialmente útil para descrever como dito processamento ocorre em bilíngues com baixa proficiência na L2 (BRYLSBAERT; DUYCK, 2010, p. 359-360). É pertinente observar que o MHR é um dos poucos modelos que se propõe a abarcar o perfil de bilíngues mencionado (VAN HELL; TANNER, 2012, p. 149).

Embora tenha sido proposto na primeira metade dos anos 1990, quando ainda não havia tantas investigações com bilíngues como se vê nos dias atuais, para García (2015 *apud* Muñoz Martín, 2017, p. 566), o MHR é a proposta mais influente e que obteve maior sucesso nas pesquisas sobre processamento de L2. Após o seu lançamento por Kroll e Stewart (1994), o modelo se perpetuou ao longo dos anos, sendo incluído, por exemplo, em manuais de Psicolinguística, como o fizeram de Groot e van Hell (2005, p. 20) e de Groot (2013, p. 158). O MHR passou ainda por uma revisão em Kroll et al. (2010) e segue sendo usado como pressuposto teórico em estudos psicolinguísticos, como Bogulski, Bice e Kroll (2019), Kelder (2018) e Volkovyskaya (2017). Apesar das críticas feitas, por exemplo, em Brysbaert e Duyck (2010) e de Groot (2013), as afirmações do modelo continuam válidas no que se refere ao processamento de L2 em bilíngues com níveis iniciais de proficiência¹⁰⁰.

¹⁰⁰ Um aspecto que deve ser observado com muita atenção nos estudos que criticam a mediação pela LM que o MHR prega é o que os seus autores não especificam com clareza a que nível de proficiência se referem quando afirmam que os dados empíricos obtidos por meio dos seus estudos contradizem o MHR. Brysbaert e Duyck (2010, p. 365) não empregam níveis de proficiência para fundamentar a sua crítica à funcionalidade do MHR; Ma et al. (2017, p. 53; 64-65), além de não deixarem claro como mediram a proficiência dos seus participantes (p. 53), reconhecem que os resultados comportamentais obtidos não correspondem às características de indivíduos com níveis iniciais de proficiência, isto é, não testaram os níveis de proficiência mais baixos possíveis (p. 64-65); Meade, Midgley e Holcomb (2018), por sua vez, afirmam que seus participantes eram de baixos níveis de proficiência, mas não fornecem nenhum dado a respeito nem comentam que tipo de instrumento foi utilizado para medir a proficiência dos participantes do seu estudo. A ausência de mediação da LM em “aprendizes menos proficientes” [less proficient learners], como mencionado por Ma et al. (2017, p. 50), por exemplo, não significa necessariamente que esses indivíduos se encontram nos mais baixos níveis de proficiência possíveis, como são os níveis A1 e A2 do QECL (2001), por exemplo.

Conforme o MHR, quando se trata de bilíngues com baixa proficiência na L2, a LM não apenas está ativada paralelamente à segunda e influencia o seu processamento (WU; THIERRY, 2012, p. 3468), como mencionado anteriormente, mas também – e eis o seu diferencial – atua como mediadora do processamento da L2 (KROLL; STEWART, 1994, p. 158; POARCH; VAN HELL; KROLL, 2015, p. 357). O MHR postula que as “palavras” (isto é, suas representações fonológicas e ortográficas) possuem elos com o sistema de conceitos (isto é, os “conteúdos” ou representações semânticas a que se referem, ao qual nos referimos por *sistema de representações semânticas* no decorrer deste capítulo)¹⁰¹. Tais elos variam conforme a proficiência em cada língua. O MHR assume que os elos entre as representações fonológicas e ortográficas da LM e as suas representações semânticas são fortes – posto que o modelo se refere a bilíngues cuja LM já está consolidada – enquanto os elos que unem as representações fonológicas e ortográficas da L2 às respectivas representações semânticas são fracos em bilíngues com níveis iniciais de proficiência (KROLL ET AL. 2010, p. 375). Em função disso, durante o processamento da L2 em tarefas de recepção que envolvam compreensão do significado (e não somente verificação da ortografia ou pronúncia), bilíngues com esse perfil têm muita dificuldade em relacionar uma representação ortográfica ou fonológica de uma unidade léxica da L2 com o sistema de representações semânticas, mesmo tendo dita representação armazenada na memória. Frente a essa situação, para efetuar o processamento da L2, o bilíngue lança mão de uma estratégia que chamaremos aqui de *tradução mental*. A tradução mental consiste na busca por um elemento na LM que seja equivalente ao elemento da L2 que está em processamento, de forma a aproveitar o elo forte entre a representação equivalente da LM e o sistema de representações semânticas. Em outras palavras, em tarefas de produção, bilíngues com níveis iniciais de proficiência na L2 formulam

¹⁰¹ No MHR, as suas autoras se referem a *conceitos* e *sistema de conceitos*. No entanto, não definem o que entendem por *conceito*. Este não é um problema exclusivo das autoras em questão; na verdade, a definição de termos não é uma prática usual na literatura da Psicolinguística. Sendo assim, o que afirmamos a seguir faz parte de nossas constatações pessoais: aparentemente, para os pesquisadores da Psicolinguística, o sistema de conceitos reúne uma variedade de representações de caráter semântico numa perspectiva que recorda a Semântica Estrutural e o estudo dos traços menores que o significado (semas). Quando mencionam a representação semântica de uma palavra, parecem referir-se às representações semânticas realizadas por uma determinada língua α , ou seja, linguisticamente dependentes e relevantes para essa língua α – algumas das quais podem não ser relevantes para uma língua β , por exemplo. Quando se referem ao sistema de conceitos, parecem referir-se ao conjunto de representações semânticas armazenadas na memória do falante; num bilíngue, seriam tanto aquelas exclusivas da língua α como as exclusivas de uma língua β , bem como as representações semânticas compartilhadas por ambas as línguas. Gostaríamos de deixar claro que essas constatações são deduções nossas passíveis de revisão posterior, pois, até o momento da redação desta seção, não encontramos autores que se dedicassem a esclarecer a terminologia empregada. Para não inserir um novo termo (*conceito*) que, ademais, possui uma variedade de definições conforme a escola que se adote, preferimos usar uma expressão com termos da linguística – isto é, *sistema de representações semânticas* – para referir-nos ao que as autoras denominam *sistema de conceitos*.

mensagens no sistema de representações semânticas, mas apresentam muita dificuldade quando tentam acessar a contraparte formal da L2 (representações ortográficas e fonológicas), mesmo tendo-a armazenada na memória. Esse perfil de bilíngues, portanto, aproveita o fortalecimento do elo entre o sistema de representações semânticas e as representações formais da LM, formulando a mensagem na LM para então fazer uma tradução mental para a L2. O mesmo ocorre com o processamento de tarefas de recepção de acordo com o MHR. Assim como na produção, em tarefas de recepção, o acesso ao sistema de representações semânticas a partir de um elemento da L2 não ocorre de maneira direta, mas sim por intermédio de um elemento equivalente da LM.

A tradução mental consiste, portanto, em percorrer um caminho longo e que, sem dúvida, requer maior esforço do bilíngue devido à fragilidade dos elos entre as representações formais da L2 e o sistema de representações semânticas. Em vista disso, é imperativo que a Metalexigrafia voltada para a compilação de dicionários de aprendizes para níveis iniciais de proficiência, como A1 e A2, considere esse aspecto do processamento ao descrever a L2. Trata-se, mais precisamente, de estabelecer estratégias para facilitar a compreensão de segmentos informativos necessários para desempenhar tarefas de recepção e de produção na L2, como é o caso das paráfrases explanatórias e notas de uso.

Três ponderações finais precisam ser feitas em relação ao MHR. Primeiramente, conforme Kroll et al. (2010, p. 376), as postulações do MHR não devem ser levadas ao extremo, a ponto de que se pense que as tarefas de recepção e produção serão resolvidas por meio da tradução mental. Segundo os autores, no decorrer do processo de aprendizagem e com os avanços na proficiência, os elos entre as representações formais da L2 se fortalecem, de modo que não seja necessário buscar equivalentes na LM. No entanto, o simples fato de que os aprendizes de níveis iniciais precisam recorrer à tradução mental não pode ser ignorado sob a perspectiva do ensino-aprendizagem de línguas e, mais especificamente, do estabelecimento do perfil de usuário de um dicionário de aprendizes para níveis iniciais.

Em segundo lugar, apesar da unanimidade em afirmar que aprendizes de níveis iniciais sentirão necessidade de recorrer à tradução mental para processar a L2, os psicolinguistas em geral não empregam a escala do QECRL (2001) para classificar os participantes de seus estudos em relação ao nível de proficiência. Essa limitação, no entanto, não invalida a aplicação dos postulados do MHR para o nosso objetivo de definir o perfil de usuário, pois os níveis de proficiência que tomamos por base nesta tese (A1 e A2) são os mais baixos possíveis de uma escala de seis níveis.

Em terceiro lugar, é preciso considerar até que ponto há uma tradução mental entre um determinado par de línguas e até que ponto há uma associação por similaridade entre representações ortográficas e fonológicas da L2 com as da LM. Em pares de línguas com relativa similaridade formal, como é o caso do português e do espanhol, é de se esperar que o processamento se apoie muito mais nas influências interlinguísticas que na tradução mental. Por esse motivo, os subsídios da Psicolinguística precisam ser contrabalançados com o grau de proximidade estrutural entre duas línguas.

Na próxima subseção, trataremos do processamento de aspectos específicos da L2 nos âmbitos fonológico, morfossintáticos e semântico. Tais informações reforçam nossa argumentação e têm uma incidência nas decisões acerca de segmentos informativos relativos a esses níveis nos dicionários de aprendizes.

4.3.1 *Processamento fonológico, morfossintático e semântico da L2*

Além da LM, duas variáveis condicionam o processamento da L2 e auxiliam na distinção entre subgrupos de bilíngues: a idade em que se aprende a L2 e o nível de proficiência na mesma. Conforme se discorrerá nesta seção, ambas influenciam o processamento de diferentes tipos de níveis de organização da linguagem. Esta subseção atenta para os detalhes do processamento de L2 a níveis fonológico, morfossintático e semântico e como as características desses níveis de processamento ajudam a definir o perfil de usuário alvo das nossas pesquisas.

Segundo Grosjean (2013, p. 7), uma crença comum acerca do bilinguismo é a de que só seria bilíngue quem adquiriu as duas línguas durante a infância, descartando quem aprende outra(s) língua(s) na adolescência ou na fase adulta. A idade é, sim, uma variável a ser considerada, mas que incide no processamento diferenciado¹⁰² das línguas em bilíngues que adquiriram duas línguas concomitantemente durante a infância (do nascimento até os cinco anos de idade, aproximadamente) e bilíngues que aprenderam a L2 após a consolidação da aquisição da LM (PARADIS, 2009, p. 35; YIP, 2013, p. 110). Os primeiros são denominados

¹⁰² As diferenças em função da idade e do estado de consolidação da LM levam alguns estudiosos a usar termos específicos conforme a língua de que se trate. Dessa forma, o processamento da LM é chamado de *aquisição* e o da L2, *aprendizagem*. No âmbito da Psicolinguística, não há um consenso sobre o particular; ora usa-se o termo *aquisição* para as duas situações, ora usa-se *aquisição* para a primeira língua e *aprendizagem* para a segunda. De fato, Richards e Schmidt (2010, s.v. *language acquisition*) comentam que é comum que se empreguem os termos *aquisição* e *aprendizagem* como sinônimos. Neste trabalho, frente ao fato de que há diferenças no processamento de LM e de outras línguas após a consolidação desta, e considerando que o nosso perfil de usuário almejado está aprendendo inglês ou espanhol após a consolidação do português como LM, julgou-se oportuno empregar o termo *aquisição* em referência ao português e *aprendizagem* em referência à assimilação do inglês e do espanhol.

bilíngues precoces e os segundos, *bilíngues tardios*. Os dados de pesquisas que envolvem bilíngues tardios são de especial interesse para esta tese porque esse grupo de bilíngues se assemelha ao perfil de usuário de dicionário que nos interessa: adultos que estão aprendendo uma L2, já tendo a sua LM consolidada. Portanto, o grupo que anteriormente foi intitulado de maneira transitória como *bilíngues em formação* será denominado de maneira mais específica como *bilíngues tardios*. Dentre os dados de pesquisas com bilíngues tardios, os que nos interessam para esta tese são aqueles referentes a bilíngues tardios com baixa proficiência na L2.

A idade é uma variável que incide majoritariamente no processamento “gramatical”, isto é, de propriedades fonológicas e morfossintáticas. Já a proficiência tem maior relação com as propriedades semânticas. Os três níveis de processamento são tratados nas subseções a seguir.

4.3.1.1 Processamento fonológico da L2

No tocante ao nível fonológico, Li (2013, p. 137-138) comenta que os adultos têm mais dificuldade nessa etapa do processamento que as crianças – e, muitas vezes, não conseguem alcançar um sistema fonológico semelhante ao de um falante nativo da L2 ou ao de um bilíngue precoce. O autor comenta três razões para isso. Primeiramente, os adultos não têm a mesma quantidade de tempo que as crianças têm para se concentrar no processamento fonológico (LI, 2013, p. 137). Em segundo lugar, os adultos apresentam menos recursos neurais para processar e acrescentar na memória novas categorias fonológicas da L2 (*ibid.*). Em terceiro lugar, os adultos já possuem um sistema fonológico consolidado, o que torna mais difícil o estabelecimento de novas categorias fonológicas em comparação com as crianças, cujo sistema fonológico da LM não está completamente consolidado (*ibid.*). Em outras palavras, quanto maior o grau de consolidação das propriedades fonológicas da LM, mais difícil será identificar os fonemas que são distintivos e/ou exclusivos da L2. Esse grau de consolidação está em direta relação com a idade, pois, conforme a autora, as crianças têm mais facilidade em estabelecer categorias de fonemas que os adultos que estão aprendendo uma L2.

Conforme o Modelo de Aprendizagem de Fala [Speech Learning Model] (FLEGE, 1995 *apud* LI, 2013, p. 138-139), “[...] aprendizes de L2 estabelecem categorias fonéticas de acordo com o grau com que o sistema fonológico da língua alvo se assemelha ao da língua nativa”¹⁰³

¹⁰³ [L2 learners establish phonetic categories according to the degree to which the phonological system of the target language resembles that of their native language]

(LI, 2013, p. 138). Por um lado, isso significa que, se a LM estabelece mais distinções fonológicas que a L2, o aprendiz poderá reaproveitar os fonemas da LM durante a aprendizagem L2, facilitando esse processo. É o que ocorre no par de línguas português-espanhol, no qual o falante de português não terá grandes dificuldades com as vogais do espanhol porque todas elas já estão presentes no sistema fonológico do português. Em contrapartida, se a L2 estabelece como fonemas determinadas manifestações que a LM do aprendiz trata como alofones ou simplesmente não distingue tais fonemas, esse aprendiz terá mais dificuldade em aprender tais distinções fonológicas na L2 (LI, 2013, p. 138). No par de línguas português-inglês, por exemplo, o falante de português apresentará mais dificuldade em perceber e distinguir alguns fonemas vocálicos da língua inglesa, pois essa língua realiza mais distinções fonológicas que o português nesse quesito.

Em suma, a nível fonológico, se a LM do aprendiz realiza uma distinção fonológica que a L2 trata como alofonia, esse aprendiz não terá dificuldades no uso da L2, em função da simplificação desta. Ao contrário, se a LM do aprendiz não realiza uma determinada distinção fonológica que a L2 faz, esse aprendiz terá mais dificuldades em criar essa distinção, considerando o fato de o sistema fonológico da sua LM já está consolidado, dificultando o processamento dessa distinção. As características do processamento fonológico de L2 requerem, portanto, orientações diferentes conforme o suporte do dicionário de aprendizes.

No suporte impresso, propomos que a transcrição fonológica esteja sempre presente e que marque a sílaba tônica. Na tradição anglo-saxônica, a transcrição fonológica já é amplamente fornecida pelos dicionários de aprendizes. Na lexicografia hispânica, por outro lado, a mera inclusão da transcrição fonológica em dicionários de aprendizes de espanhol (aliás, em toda a lexicografia hispânica) não é consensual; há dicionários que o fazem, enquanto outros, não. Portanto, a proposta em questão se aplica mais especificamente ao desenho de um dicionário de aprendizes brasileiros de espanhol de níveis A2 e B1 de aprendizagem em suporte impresso.

No suporte eletrônico, propomos que as informações fonológicas estejam constituídas da transcrição fonológica e recurso de áudio. Dessa forma, o consulente pode comparar os fonemas que distingue no áudio com os que efetivamente conformam a palavra-entrada que está consultando. No caso da lexicografia inglesa, a edição eletrônica dos dicionários analisados no capítulo 1 incluem somente o recurso de áudio; para esses casos, é necessário acrescentar a transcrição fonológica. No caso do dicionário Señas (2001), como se trata de uma obra disponível apenas em papel, seria necessário digitalizá-lo primeiro e elaborar os arquivos de áudio com a pronúncia.

É necessário salientar que a transcrição fonológica garante que o aprendiz seja informado acerca de todos os fonemas presentes na representação sonora e serve de respaldo à mesma, especialmente nos casos em que a L2 apresenta fonemas de difícil distinção pelo aprendiz em função da idade em que está aprendendo a L2, do sistema fonológico da LM e do seu baixo nível de proficiência na L2. Além disso, e embora a bibliografia comentada nesta subseção tenha se referido somente aos fonemas e alofones, tanto o recurso de áudio como a transcrição fonológica também auxiliam o aprendiz na identificação da sílaba tônica, que pode ser mais um alvo de dificuldades na aprendizagem quando há influências interlinguísticas da LM que geram custo ao processamento da LE.

4.3.1.2 Processamento morfossintático da L2

Para ilustrar algumas características do processamento morfossintático em bilíngues tardios com baixa proficiência, recorreremos ao estudo de Wartenburger et al. (2003). Nesta subseção, comentaremos a parcela de resultados referente ao processamento morfossintático, embora seus autores também tenham investigado o processamento semântico.

Por meio de experimentos comportamentais e de ressonância magnética funcional, Wartenburger et al. (2003) avaliaram o desempenho (comportamento) e as imagens do cérebro (locais de ativação) de três grupos de bilíngues ítalo-germânicos: precoces, tardios com alto nível de proficiência na L2 (alemão) e tardios com baixo nível de proficiência na L2 (alemão). No momento do experimento, todos tinham a mesma idade e estavam em situação de imersão na Alemanha. Embora não se possa esperar esse tipo de experiência do usuário pretense por esta tese, consideramos que os dados obtidos a partir de participantes com essas características seguem sendo válidos para nossa argumentação.

Em relação ao nível educacional, todos os participantes eram estudantes de graduação, ou seja, já haviam cumprido todas as etapas da educação considerada básica. Esse traço do seu perfil, particularmente no que se refere aos bilíngues tardios com baixa proficiência, é um dos motivos que nos levou a selecionar o Ensino Médio completo como nível educacional do perfil de usuário de dicionários proposto nesta tese (cf. 2.4). Na tabela a seguir, reproduzimos alguns dos dados sobre os participantes da pesquisa:

Tabela 1 – Perfil resumido dos participantes de Wartenburger et al. (2003)

	Bilíngues precoces	Bilíngues tardios com alta proficiência	Bilíngues tardios com baixa proficiência
Idade de aquisição da L2	Ao nascer	18 anos	20 anos
Média de duração da exposição à L2	Equivalente à idade dos sujeitos	10,3 anos	4,8 anos
Média de duração da estada no país da L2	13,2 anos	6 anos	1,6 anos
Início da exposição à L2	Ao nascer	Entre 7 e 14 anos	Entre 15 e 18 anos
Proficiência de compreensão leitora (máximo 10 pontos)	9,5	9,2	4,7
Acurácia no processamento morfosintático (média de acertos)	L2: 93,1% L1: 96,7%	L2: 93,3% L1: 98,4%	L2: 82,2% L1: 98,5%
Tempo de resposta no processamento morfosintático (média de tempo)	L2 = L1 tempo de resposta com diferença não significativa (0,1s)	L2 > L1 tempo de resposta ligeiramente maior na L2 (0,4s a mais)	L2 > L1 tempo de resposta significativamente maior na L2 (1,4s a mais)

Legenda: L2 = segunda língua; L1 = primeira língua; s = segundos

Fonte: Wartenburger et al. (2003, p. 161; 163)

Para que seja possível compreender os dados acerca do processamento morfosintático da L2 entre os bilíngues tardios com baixa proficiência, é necessário não apenas comentar esses dados em contraste com os da primeira língua (L1), mas também compará-los com os dos outros dois grupos.

Em primeiro lugar, ao desempenhar-se na L2, os bilíngues tardios com baixa proficiência manifestaram acurácia menor em comparação com o seu desempenho na L1 (WARTENBURGER ET AL., 2003, p. 163). Os seus pares com alta proficiência, embora também fossem bilíngues tardios, tiveram um desempenho muito próximo ao da L1 no nível morfosintático, com diferenças pouco significativas em relação aos bilíngues precoces conforme Wartenburger et al. (2003, p. 162-163).

Em segundo lugar, os bilíngues com baixa proficiência apresentaram tempo de resposta mais longo em comparação com a L1 e com o tempo de resposta dos seus pares altamente proficientes quando processaram a L2 a nível morfosintático (*ibid.*). Um tempo maior de resposta significa um tempo maior de processamento, o que é um indicativo de dificuldade de processamento e esforço cognitivo. *Esforço cognitivo* é um termo muito empregado na literatura dos estudos cognitivos, até certo ponto facilmente compreensível, mas geralmente

não definido. Em termos comportamentais, parece se manifestar por meio de hesitação e numa demanda de tempo maior para resolver uma tarefa em comparação com indivíduos que claramente a realizam de maneira automatizada. Em exames de neuroimagem, além do tempo de resposta, o esforço cognitivo também está geralmente associado a um grau de ativação acima do esperado em uma mesma região do cérebro e/ou em outras regiões, como se nota nos resultados de RMF de Wartenburger et al. (2003).

Em terceiro lugar, e no tocante ao comportamento do cérebro ao processar a L2, as imagens de RMF demonstraram que os bilíngues tardios com baixa proficiência apresentam maior ativação no cérebro em comparação com a L1; porém, esse mecanismo compensatório (que gera mais esforço cognitivo, aliás) não foi suficiente para garantir uma maior acurácia nem para diminuir o tempo de processamento, medido pelo tempo de resposta (WARTENBURGER ET AL., 2003, p. 166-167). Em outras palavras, os dados sugerem que, por mais que se esforcem, os bilíngues tardios com baixa proficiência ainda não possuem nem uma estrutura neural nem um patamar de proficiência suficientes para processar a totalidade de informações morfossintáticas da L2; ou seja, há informações que esses bilíngues simplesmente ainda não conseguem processar¹⁰⁴. Materiais didáticos para esse público precisam, portanto, encontrar um meio-termo entre a descrição da língua em uso por meio de metalinguagem e a capacidade de processamento morfossintático de seus consulentes, como se comentará *ad infra*.

Wartenburger et al. (2003, p. 168) concluem que o fator determinante para a discrepância entre o processamento morfossintático da L1 e da L2 é a idade em que se iniciou o contato com a L2, associada à idade de aquisição da mesma e, naturalmente, ao tempo de exposição (v. Tabela 1). Em outras palavras, a idade, que motiva a diferenciação terminológica entre grupos de bilíngues, exerce influência sobre como ocorre o processamento morfossintático.

No que concerne ao estabelecimento do perfil de usuário desta tese, ante as evidências de que bilíngues com esse perfil têm dificuldades para processar informações morfossintáticas, não há dúvidas de que esse tipo de informação será alvo recorrente de buscas um dicionário de aprendizes para níveis A2 e B1 de aprendizagem. Por informações morfossintáticas em dicionários, referimo-nos à flexão irregular de substantivos e verbos, bem como padrões de

¹⁰⁴ A modo de comparação, seus pares altamente proficientes também empregaram mecanismos compensatórios ao processar informações morfossintáticas na L2 em comparação com a L1; contudo, estes ativaram ainda mais áreas no cérebro que os seus pares com baixa proficiência, obtendo acurácia e tempo de resposta similares aos da L1 (WARTENBURGER ET AL., 2003, p. 164-166). Esse segundo grupo de bilíngues consegue, portanto, equiparar o seu desempenho na L2 ao da L1 (dados comportamentais) por meio de mecanismos compensatórios (estrutura neural). O processamento desse perfil de bilíngues difere do da L1 no âmbito neurológico, mas não comportamental.

combinatória com preposições, pronomes e advérbios. Por essa razão, a decisão mais acertada para fins de descrição metalexigráfica é a de facilitar a sua localização e identificação. Para tanto, é possível empregar determinados recursos tipográficos, como o uso de negrito, itálico, cores e fontes diferentes para chamar a atenção do consultante em relação às informações morfossintáticas. Essa segunda proposta será desenvolvida com mais detalhes nos capítulos 6 e 8.

Na subseção a seguir, finaliza-se o apanhado sobre níveis específicos de processamento com as contribuições da Psicolinguística acerca do nível semântico.

4.3.1.3 Processamento semântico da L2

O processamento semântico da L2 já está, em parte, contemplado em 4.2, quando comenta-se sobre o Modelo Hierárquico Revisado (MHR). Convém, no entanto, introduzir uma complementação a esse modelo que chama a atenção para as influências interlinguísticas da LM no processamento semântico e que reforça a nossa proposta de acesso ao sistema de representações semânticas em um dicionário de aprendizes A2-B1 via LM. Trata-se do Modelo Hierárquico Modificado (MHM), de Pavlenko (2009)¹⁰⁵. De acordo com o MHM, cada língua possui uma maneira de construir e organizar esse sistema de representações semânticas. Há, por um lado, propriedades semânticas exclusivas de cada língua (LM e L2); por outro lado, há propriedades semânticas que são compartilhadas entre as duas línguas do bilíngue (PAVLENKO, 2009, p. 133-140). Dessa forma, Pavlenko (*ibid.*) identifica casos de equivalência total, (não) equivalência parcial [partial (non)equivalence] e não-equivalência [non-equivalence] entre duas línguas. No par português-inglês, há equivalência total entre *envy* e *inveja*; (não) equivalência parcial entre *to know* [saber; conhecer] e *saber*; e não-equivalência em *berry*. No par de línguas português-espanhol, por sua vez, há equivalência total entre *envidia* e *inveja*; (não) equivalência parcial entre *enseñar* [ensinar, mostrar] e *ensinar*; e não-equivalência em *sobremesa* [momento de conversa que duas ou mais pessoas têm à mesa após o término de uma refeição].

Conforme comentado em 4.2, pelo princípio do MHR que denominamos “tradução mental”, em níveis iniciais de aprendizagem e proficiência na L2, é comum que se estabeleça uma associação de equivalência entre uma forma da L2 e uma forma da LM para acessar o sistema de representações semânticas. Naquela oportunidade, argumentamos que a prática da

¹⁰⁵ De acordo com de Groot (2013, p. 160-161), o MHM está baseado em uma longa trajetória de propostas de precisão do sistema de representações semânticas previsto no MHR.

tradução mental para compreender paráfrases explanatórias e notas de uso em dicionários de aprendizes de níveis iniciais consiste em um esforço do qual o consulente poderia ser poupado por meio do uso da sua LM. O MHM, por sua vez, reforça essa decisão ao postular que, pela tradução mental, acessa-se uma determinada representação semântica modulada pela LM e que não necessariamente corresponde à representação semântica usada na L2 (DE GROOT, 2013, p. 159; PAVLENKO, 2009, p. 133-140).

A constatação de que as línguas naturais são diferentes entre si não é exatamente nova para a Linguística; a própria Linguística Contrastiva nasceu dessa constatação. Na Lexicologia e na Lexicografia Bilíngue, a identificação do fenômeno do anisomorfismo linguístico – isto é, o fato de que “cada língua apresenta uma organização absolutamente única em todos os seus níveis de estruturação” (BUGUEÑO MIRANDA, 2010, p. 66) – é um exemplo do interesse dessas áreas pela investigação das especificidades de cada língua natural. Os níveis de equivalência identificados por Pavlenko (2009) correspondem ao que Kromann et al. (1991 *apud* TEDESCO SELISTRE, 2019, p. 138) chama as duas primeiras categorias de equivalência plena e equivalência parcial, respectivamente, enquanto a não-equivalência recebe, entre outras designações, o nome de *realia*¹⁰⁶.

O aporte que a Psicolinguística apresenta por meio do MHM consiste em chamar a atenção para o problema de se acessar o sistema de representações semânticas por meio dos equivalentes da LM, por um lado, e para a necessidade de se elaborar maneiras de mostrar aos aprendizes de L2 o que há de comum e de diferente entre as representações semânticas da LM e da L2, por outro lado (PAVLENKO, 2009, p. 154). Para um dicionário de aprendizes brasileiros de inglês, o MHM fornece um respaldo teórico-argumentativo a mais em prol do fornecimento de paráfrases explanatórias na LM e do fornecimento de notas de uso que assinalem nuances semânticas próprias do inglês. Essas estratégias ajudam tanto no acesso rápido ao sistema de representações semânticas como também na criação de novas representações, seja nos casos de equivalência parcial entre o inglês e o português (*to know*), seja nos casos de *realia* do inglês (*berry*). Para um dicionário de aprendizes brasileiros de espanhol, o MHM fornece bases teórico-argumentativas para que os definidores em língua espanhola sejam escolhidos de forma a se evitar equivalência parcial, como *enseñar* [mostrar], e *realia*, como *sobremesa*, na redação de paráfrases explanatórias e notas de uso.

¹⁰⁶ O que autores como Bugueño Miranda (2010, p. 67) chamam de *realia* (“manifestações extremas e absolutamente particulares da forma de organizar a realidade que uma dada língua possui”), outros autores, como Klotz e Herbst (2016, p. 152) usam o termo *culture specific words* (“casos nos quais um lexema na L1 [língua 1] não correspondem sequer parcialmente a qualquer lexema na L2 [língua 2]” [cases where a lexeme in the L1 does not even partially correspond to any lexeme in the L2])

Com esta subseção, encerra-se o apanhado de contribuições da Psicolinguística que permitem especificar o perfil de usuário. A seção a seguir retoma as constatações feitas no decorrer deste capítulo.

4.4 Corolário do capítulo 4

A Psicolinguística e a Neurolinguística lançam luz sobre um tema que guarda resquícios de tabu ainda nos dias de hoje quando se trata do ensino-aprendizagem de língua estrangeira: a língua materna. As contribuições das duas áreas transcendem uma aplicação meramente metalexográfica, pois se estendem a qualquer pesquisa e iniciativa no âmbito do ensino-aprendizagem de língua estrangeira. Os seguintes axiomas buscam sintetizar o aporte das duas áreas no tocante ao perfil de usuário:

- (7) A LM é elemento integrante do processo de aprendizagem de LE entre adultos.
- (8) A LM media o processamento da LE em bilíngues tardios com baixa proficiência na LE.
- (9) É possível tirar proveito da LM do aprendiz para facilitar a aprendizagem de uma LE.

No que concerne ao axioma 7, a coativação da LM durante o processamento da LE indica que a primeira atua no processamento da segunda. O produto dessa interação são as chamadas influências interlinguísticas, que afetam o processamento facilitando-o ou tornando-o mais custoso. O volume e direção dessas influências dependem, ao menos, do nível de proficiência na LE em comparação com a LM. No caso do perfil de usuário considerado nesta tese, a LM goza de maior proficiência e é a mais usada cotidianamente na comunidade linguística na qual o usuário está inserido, enquanto o patamar de proficiência na LE é ainda elementar (A1 e A2) e as oportunidades de uso da LE são demasiado reduzidas se comparadas às da LM. Portanto, a direção mais comum das influências interlinguísticas no caso do nosso perfil de usuário é da LM para a LE.

O axioma 8, por sua vez, está diretamente relacionado ao seu antecessor. Sendo a língua mais usada cotidianamente, que goza de maior proficiência e pela qual o aprendiz aprendeu a executar funções comunicativas básicas, a LM media o processamento da LE seja via tradução mental (MHR), seja pela detecção automática de similaridades entre representações formais (especialmente ortográficas e/ou fonológicas) de elementos da LE e da LM.

Segundo já comentado, a tradução mental aciona o sistema de representações semânticas sob a perspectiva da LM. Em um dicionário de aprendizes, as paráfrases explanatórias e notas de uso são dois dos segmentos informativos que, por requererem compreensão leitora na LE, são um alvo em potencial das traduções mentais. Além de serem uma via mais longa para o processamento dos segmentos informativos em questão, podem ativar representações semânticas exclusivas da LM e que não condizem com o uso da LE, ou deixar de ativar nuances da LE porque não estão presentes na LM, segundo prega o MHM. É de se questionar, portanto, até que ponto o uso da LE em função metalinguística é benéfico para o perfil de usuário de um dicionário de aprendizes. Além disso, outra questão que surge associada à tradução mental são os padrões de combinação entre as palavras da LE, que são únicos para cada língua. Em tarefas de produção, a tradução mental de itens isolados pode dar lugar à produção de padrões combinatórios conforme a LM e que podem divergir daqueles usados na LE.

As similaridades formais, por sua vez, quando detectadas como tal pelos bilíngues, dão origem a influências interlinguísticas tanto facilitadoras como custosas. No que toca ao dicionário de aprendizes, deduz-se que a identificação de duas formas como similares ocorra especialmente nos segmentos informativos que requerem maior compreensão leitora, como é o caso das já mencionadas paráfrases explanatórias e das notas de uso. Ter ciência do quão similar é a LE em relação à LM, e de que tipo de similaridade se trata (com ou sem convergência semântica), é fundamental para que se explore o potencial de aproveitamento da LM como mediadora do processamento da LE.

Tendo em vista os axiomas 7 e 8, a conclusão lógica que deriva da Psicolinguística e da Neurolinguística é a de que a LM precisa ser considerada em qualquer decisão didático-pedagógica atinente ao ensino-aprendizagem de LE entre adultos e, de maneira especial, quando o processo de aprendizagem ocorre na terra natal do indivíduo e quando a LE não é usada para executar tarefas básicas do cotidiano da comunidade linguística em questão (axioma 9). O desafio de qualquer iniciativa em benefício do aprendiz reside em determinar como se pode usar a LM a favor desse processo.

Por um lado, é necessário explorar as oportunidades de ocorrência de interações interlinguísticas facilitadoras do processamento da LE. Em outras palavras, é necessário avaliar a proporção de similaridade formal e semântica entre a LE e a LM, pois tal fenômeno cria um contexto propício para a facilitação do processamento da LE. Avaliar o quão similares são as formas da LE em relação à LM e que de tipo de similaridade se trata (cognatos ou homógrafos, segundo a perspectiva da Psicolinguística sobre esses termos) fornecerá uma estimativa da

proporção de oportunidades de ocorrência de influências interlinguísticas tanto facilitadoras como custosas. Com esse dado em mãos, será possível avaliar o quão útil é o uso da LE em função metalinguística em dicionários de aprendizes de inglês e de espanhol sob o ponto de vista do usuário brasileiro de níveis A1 e A2 de proficiência.

Por outro lado, também é necessário investigar as influências interlinguísticas da LM que geram um custo maior no processamento da LE. De acordo com o que foi comentado em 4.1, a Linguística Contrastiva/*Learner Corpus Research*, por meio da análise de *corpora* de aprendizes, é uma importante aliada na identificação das necessidades de aprendizagem que são oriundas desse tipo de interação. Em um dicionário de aprendizes brasileiros, os desvios comuns no uso do inglês e do espanhol ocasionados por influência do português são um tipo de conteúdo pertinente para notas de uso e que aproxima a obra lexicográfica do seu usuário. Segundo os exemplos do capítulo 1, aparentemente, as notas de uso de dicionários de aprendizes de inglês A1-B1 e de dicionários de aprendizes de espanhol contêm informações sobre desvios comuns cometidos por aprendizes de cada língua. Entretanto, tais informações provêm de registros de falantes de diversas LMs; em função disso, em alguns casos, podem ser irrelevantes para os brasileiros (cf. exemplos nos próximos capítulos).

Frente às questões assinaladas nos parágrafos anteriores, o capítulo que segue está dedicado a analisar dicionários de aprendizes de inglês e de espanhol no que concerne ao uso da língua-objeto em função metalinguística e ao conteúdo das notas de uso.

5 ANÁLISE DOS DICIONÁRIOS DE APRENDIZES À LUZ DO PERFIL DE USUÁRIO

As contribuições da Psicolinguística acerca das influências interlinguísticas da LM no processamento de LE levantam questões sobre como tirar o máximo proveito das influências interlinguísticas facilitadoras do processamento para facilitar a compreensão de paráfrases explanatórias e notas de uso. Tanto em um como no outro caso, a associação entre os pressupostos da Psicolinguística e os subsídios da Linguística Aplicada (documentos e ferramentas) constitui a base para a elaboração das propostas atinentes ao comentário semântico, mais precisamente às paráfrases explanatórias e notas de uso. Tomando como fundamento o referencial teórico supracitado, foram dois os objetivos das investigações que seguem no presente capítulo:

1. Avaliar o nível de facilidade de compreensão leitora a partir dos níveis de referência do CEFR (2001) aos quais estão atrelados os definidores e as estruturas gramaticais empregados nas paráfrases explanatórias e notas de uso.
2. Avaliar se o uso da língua-objeto em função metalinguística para a redação de paráfrases explanatórias e notas de uso facilita o processamento (compreensão leitora) desses segmentos informativos tomando-se como referência o perfil de usuário estabelecido na tese e um determinado conjunto de dicionários de aprendizes de inglês e espanhol.

O primeiro objetivo consistiu em averiguar em que medida o conhecimento já adquirido nos níveis A1 e A2 de proficiência é empregado como base para a redação de paráfrases explanatórias e notas de uso. A análise se amparou nos dados disponibilizados pelos descritores de níveis de referência EP (2011) e respectivas ferramentas e, em menor medida, pelo PCIC (2006). Para o inglês, é possível buscar com certa facilidade que recursos linguísticos são usados em cada nível de proficiência em inglês por meio do EVP (2015) e EGP (2015). Para o espanhol, o PCIC (2006) é, de certa forma, mais limitado, pois a obra não almeja fornecer um inventário tão amplo de itens lexicais atrelados a cada nível de referência como o EVP (2015), mas sim alguns exemplos ilustrativos de funções comunicativas e noções (vocabulário). Dado que se dispõe de ferramentas tão diferentes para cada língua, as análises dos dicionários de aprendizes de inglês e de espanhol apresentarão perspectivas e metodologias ligeiramente diferentes quanto à análise do conhecimento prévio. Maiores esclarecimentos são fornecidos nas seções específicas dedicadas ao inglês e ao espanhol.

Os pressupostos que orientaram as análises relativas ao segundo objetivo foram a probabilidade de ocorrências de influências interlinguísticas desencadeadas pela similaridade ortográfica entre a LE, empregada nas paráfrases explanatórias e notas de uso, e o português. Conforme comentado no capítulo 4, a depender da convergência semântica entre representações ortográficas da LE e da LM, poderá haver facilitação ou custo maior que o esperado no processamento da LE. Quando há similaridade considerável na forma e na representação semântica entre as duas línguas, cria-se um ambiente propício para a facilitação do processamento. Quando há baixa similaridade na forma, mesmo que a contraparte semântica entre as duas línguas seja convergente, é menos provável que ocorra um efeito facilitador, embora não se possa afirmar tampouco que haverá custos no processamento.

O nível de similaridade ortográfica entre duas palavras é geralmente calculado nos estudos psicolinguísticos por meio da fórmula adaptada por Van Orden (1987). Essa fórmula será usada nas análises de similaridade ortográfica empreendidas no decorrer do capítulo e está disponível na calculadora automática NIM (2013), disponibilizada no *site* <https://psico.fcep.urv.cat/utilitats/nim/eng/graphsim.php>. Para usá-la, é necessário explicitar que palavras se deseja comparar¹⁰⁷. A ferramenta, por sua vez, indica o grau de similaridade em uma escala de 0 a 1. Não há uma escala de referência que explicita a partir de que índice duas palavras podem ser consideradas similares ou não. Nesta oportunidade, resultados a partir de 0,51 foram considerados como indicativos de similaridade ortográfica; resultados abaixo deste índice foram considerados como indicativos de similaridade insuficiente.

O capítulo se divide em duas seções principais, uma para cada tradição lexicográfica. Os resultados e considerações realizados nas próximas páginas contribuirão diretamente para elaboração de propostas de aprimoramento de paráfrases explanatórias e notas de uso em dicionários de aprendizes brasileiros de inglês e espanhol (cf. capítulos 6 e 7).

5.1 Dicionários de aprendizes de inglês e o público brasileiro

Nesta seção, avalia-se o quão compreensíveis são as paráfrases explanatórias e notas de uso de dicionários de aprendizes de inglês A1-B1 tomando como referência o aprendiz brasileiro de níveis A1 de proficiência e A2 de aprendizagem, e de níveis A2 de proficiência e B1 de aprendizagem. A facilidade de compreensão será avaliada, primeiramente, em termos de

¹⁰⁷ Naturalmente, haverá subjetividade no cálculo na medida que a palavra do português que se sugere como parâmetro de comparação é uma inferência nossa. Ainda assim, a fórmula é o que atualmente há de mais preciso no cálculo da similaridade ortográfica.

similaridade formal e semântica entre o inglês e o português. Além disso, também se avalia os definidores e padrões gramaticais empregados nas paráfrases e notas no que concerne ao seu nível de referência CEFR (2001) e o nível de referência da unidade léxica definida. Em nossas análises, conforme comentado no capítulo 3, entendemos que os definidores são as unidades léxicas empregadas nas paráfrases explanatórias para explicitar um significado ou função. As unidades léxicas, por sua vez, contemplam tanto as palavras ditas lexicais (forma + significado) como as palavras ditas gramaticais (forma + função) (Cf. Cunha (2019) para detalhes sobre o conceito de *unidade léxica*).

Primeiramente, foram selecionados os dicionários de aprendizes de inglês A1-B1 mencionados no capítulo 1, quais sejam: CaED2 (2011), OxED2 (2012), CoED2 (2014) e LWord2 (2008). Em seguida, selecionaram-se intervalos para análise da compreensão leitora exigida. Como a prioridade foi fornecer um panorama de como se comporta essa classe específica de dicionários, as amostras de cada obra podem ser consideradas pequenas se comparadas ao que se poderia analisar em uma única obra. Os critérios de seleção foram:

- (1) Distribuição entre os níveis A2 e B1. Sendo o perfil de usuário almejado por esta tese um estudante que se encontra nos níveis A2 e B1 de aprendizagem, as unidades léxicas mais pertinentes para análise são as que o EVP (2015) atrela a estes níveis de proficiência.
- (2) 26 unidades léxicas por dicionário. Esse número foi determinado em função do tempo de que dispúnhamos para efetuar as análises. Por unidade léxica, entende-se a associação de uma forma (significante), representada ortograficamente, a um conteúdo semântico (significado), tal como na dicotomia saussureana do signo linguístico. Em um dicionário de orientação semasiológica, isto é, no qual as informações orientadas da forma ao significado, as unidades léxicas que compartilham uma correspondência na sua forma são geralmente reunidas em um mesmo verbete e os significados são apresentados como acepções¹⁰⁸. Quando havia uma nota de uso atrelada a uma das acepções consideradas para análise, esta nota também foi incluída. Para o nível A2, foram selecionadas 12 unidades léxicas, enquanto para o nível B1, 14.

¹⁰⁸ No OxED2 (2012), por exemplo, a forma *lie* é a palavra-entrada de três verbetes diferentes: *lie*¹, *lie*² e *lie*³. Há, primeiramente, uma distinção orientada pela categoria morfológica, que separa os verbos *lie*¹, relativo à posição do corpo, e *lie*², relativo ao ato de enunciar algo falso, do substantivo *lie*³, referente a um enunciado falso. Os verbos *lie*¹ e *lie*², por sua vez, são separados em função das suas diferentes origens etimológicas. Conforme o OALD8 (2010, s.vs. *lie*¹, *lie*²), *lie*¹ é de origem latina e *lie*² é de origem germânica. O CoED2 (2014, s.vs. *lie*¹, *lie*²), por outro lado, adota apenas a distinção morfológica, abrangendo sob *lie*¹ os significados dos verbos e o substantivo sob *lie*².

- (3) Distribuição entre distintas classes de palavras. Foram selecionados substantivos, adjetivos, verbos, advérbios e preposições para abarcar diferentes tipos de paráfrase explanatória. Atribuiu-se especial atenção aos substantivos pela sua presença massiva nas listas de palavras citadas.
- (4) Similaridade ortográfica com unidades léxicas do português. Dois tipos de unidades léxicas foram selecionados. Em um deles, há pouca ou nenhuma similaridade ortográfica entre o inglês e uma unidade léxica do português (*edge, remember, etc.*); no outro, a palavra do inglês é similar ortograficamente a uma palavra do português, mas difere em relação ao significado (*appointment, library, rent, flavour*)¹⁰⁹.

Quadro 14 – Unidades léxicas analisadas nos dicionários de aprendizes de inglês

Classe de palavras	A2	B1
Substantivos	appointment	edge
	landscape	flavour
	library	label [information]
	---	label [company]
Adjetivos	glad	gorgeous
	pleasant	likely [will probably happen]
Verbos	remember [remember to do sth]	rent [receive money]
	rent [pay to live in]	bury [dead body]
	rent [pay to use]	remind [remember]
	---	remind [sb of sth or sb]
Advérbios	already [before now]	already [early]
	across [from one side to the other]	pretty
Preposições	across [on the opposite side]	among [in the middle]
	among [in a group]	towards [direction]
Total	12	14

Fonte: Elaboração da autora com base nos dados do EVP (2015)

Após determinar as 26 unidades léxicas que seriam analisadas, foram coletadas todas as paráfrases explanatórias e notas de uso relacionadas a esses itens. No total, foram reunidos 113 itens entre paráfrases explanatórias e notas de uso, estando 28 deles presentes no CaED2 (2011), 27 no OxED2 (2012), 29 no CoED2 (2014) e 29 no LWord2 (2008). O Quadro 15 a seguir resume essas informações. Um quadro mais completo, com as paráfrases explanatórias e notas de uso por extenso, pode ser visualizado nos Anexos 1 a 4, juntamente com marcações acerca do nível dos definidores segundo o EVP (2015), bem como a indicação dos padrões gramaticais segundo o EGP (2015).

¹⁰⁹ Grau de semelhança ortográfica: *appointment*-apontamento (0,64); *library*-livraria (0,59); *rent*-rente (0,73); *flavour*-favor (0,74). Para as demais unidades léxicas, não foi possível sugerir um candidato a símile ortográfico em português.

Quadro 15 – Quantidade de itens analisados por dicionário e por nível CEFR (2001) (inglês)

Nível CEFR	Itens analisados				
	CaED2	OxED2	CoED2	LWord2	
A2	appointment (ac.) landscape (ac.) library (ac.)	appointment (ac.) landscape (ac.) library (ac.1) library (n.)	appointment (ac.) landscape (ac.) library (ac.)	appointment (ac.) landscape (ac.) library (ac.1) library (n.1) library (n.2)	
	glad (ac.) pleasant (ac.)	glad (ac.1) glad (n.) pleasant (ac.)	glad (ac.) pleasant (ac.1) pleasant (ac.2)	glad (ac.) pleasant (ac.1) pleasant (ac.2)	
	remember (ac.2) rent (ac.1) rent (ac.2) rent (n.)	rent (ac.1)	remember (ac.3) rent (ac.1)	remember (ac.2) rent (ac.1)	
	already (ac.1) across (ac.1)	already (ac.1) already (n.) across (ac.1)	already (ac.1) already (n.) across (ac.1)	already (ac.1) already (n.) across (ac.1)	
	across (ac.2) among (ac.2)	across (ac.2) across (n.) among (ac.2) among (n.)	across (ac.2) among (ac.2) among (ac.3) among (n.)	across (ac.2) among (ac.1)	
	Total	13	15	15	15
	B1	edge (ac.) flavour (ac.) label ¹ (ac.1) label ¹ (ac.2)	edge (ac.1) flavour (ac.) label ¹ (ac.)	edge (ac.) flavour (ac.) label (ac.)	edge (ac.) flavour (ac.) label ¹ (ac.1) label ¹ (ac.2)
		gorgeous (ac.) likely (ac.1)	gorgeous (ac.) likely (ac.)	gorgeous (ac.) likely (ac.2)	gorgeous (ac.) likely (ac.1)
		rent (ac.3) bury (ac.1) remind (ac.1) remind (n.) remind (phr. v.)	rent (ac.2) bury (ac.1) remind (ac.1) remind (ac.2)	rent (ac.2) bury (ac.2) remind (ac.1) remind (ac.2) remind (ac.3)	rent (ac.2) bury (ac.1) remind (ac.1) remind (ac.2)
		already (ac.2) pretty ¹ (ac.)	pretty ¹ (ac.)	already (ac.2) pretty (ac.2)	already (ac.2) pretty ¹ (ac.)
among (ac.1) towards (ac.1)		among (ac.1) towards (ac.1)	among (ac.1) towards (ac.1)	among (ac.2) towards (ac.1)	
Total		15	12	14	14

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014) e LWord2 (2008)

Conforme comentado anteriormente, o objetivo é analisar a compreensão leitora nos dicionários de aprendizes de inglês primeiramente a partir do nível do CEFR (2001) ao qual se atrelam os definidores e estruturas gramaticais dos itens selecionados; em um segundo momento, a compreensão leitora também é avaliada com base nas potenciais influências interlinguísticas entre o português e o inglês. Com os dados obtidos em cada análise, serão

gerados *rankings* de facilidade de compreensão leitora de paráfrases explanatórias e notas de uso sob a ótica do perfil de usuário desta tese.

5.1.1 Conhecimento prévio geral na compreensão leitora dos dicionários de inglês

Esta subseção avalia o quanto as paráfrases explanatórias e notas de uso empregam o conhecimento prévio de um aprendiz de inglês genérico, isto é, sem que se leve em consideração a sua LM. Conforme mencionado no capítulo 3, tal conhecimento prévio consiste nos níveis de proficiência já atingidos, isto é, A1 no caso das unidades léxicas A2, e A2 no caso das unidades léxicas B1. A metodologia de análise foi a seguinte:

- (1) Averiguação do nível dos definidores com base nos dados do EVP (2015);
- (2) Classificação das paráfrases e notas conforme o nível dos definidores;
- (3) Averiguação do nível dos padrões gramaticais com base nos dados do EGP (2015);
- (4) Nova classificação das paráfrases e notas conforme o nível dos padrões gramaticais.

5.1.1.1 Análise do nível dos definidores

Foi possível identificar três padrões nas paráfrases explanatórias e notas de uso: 1) presença de definidores de níveis abaixo do nível da unidade léxica (UL) definida; 2) presença de definidores em nível equivalente ao da UL definida; e 3) presença de definidores de nível superior ao da UL definida.

Um detalhe importante acerca dos dados do EVP (2015) e do EGP (2015) é o fato de que seus compiladores não distinguem entre vocabulário passivo e vocabulário ativo (EP, 2011, p. 53), isto é, entre “palavras que uma pessoa consegue usar ativamente, em comparação com palavras que ela reconhece e entende, mas não usa na sua produção” (RICHARDS; SCHMIDT, 2010, s.v. *productive/receptive vocabulary*)¹¹⁰. Durante a análise de dados, chamaram a atenção alguns casos de ULs que eram enquadradas em um determinado nível, mas que poderiam ser compreendidas por extensão de significado de ULs de níveis anteriores, conforme ilustrado a seguir.

¹¹⁰ [words that a person can actively use, compared with [...] words that they recognize and understand but do not use productively]

Quadro 16 – ULs tratadas como vocabulário passivo (inglês)

Classificação original	Ajuste de nível por extensão de vocabulário
<i>in</i> [expressed] (B1) (EVP, 2015)	<i>in</i> [part of] (A2) (EVP, 2015)
<i>only</i> [nowhere else] (B1) (EVP, 2015)	<i>only</i> [nothing else] (A2) (EVP, 2015)
<i>we</i> [all people] (B1) (EVP, 2015)	<i>we</i> [group] (A1) (EVP, 2015)

Fonte: Elaborado pela autora com base nos dados do EVP (2015)

No caso das ULs da coluna da esquerda, para fins de compreensão, é possível esperar que o aprendiz deduza o significado a partir do significado atinente a níveis anteriores de proficiência, como indicam os dados da coluna da direita. O nível aferido pelo EVP (2015) às ULs da coluna da esquerda parece estar mais relacionado à sua produção. Por essas razões, os definidores em questão tiveram seu nível ajustado durante a análise, conforme indicado acima.

Os resultados da análise estão sintetizados a seguir. Uma relação dos verbetes e acepções específicos a que se referem os dados obtidos se encontra no Anexo 5.

Quadro 17 – Análise do nível do vocabulário de definidores (inglês)

Classificação dos definidores	CaED2	OxED2	CoED2	LWord2
Un. léxicas A2 com definidores A1 somente	2/13 (15,38%)	1 /15 (6,66%)	1/15 (6,66%)	-
Un. léxicas A2 com definidores A1, A2	2/13 (15,38%)	4/15 (26,66%)	3/15 (20%)	6/15 (40%)
Un. léxicas A2 com definidores A1, A2, B1+	9/13 (69,23%)	10/15 (66,66%)	11/15 (73,33%)	9/15 (60%)
Un. léxicas B1 com definidores A1, A2	3/15 (20%)	3/12 (25%)	3/14 (21,42%)	2/14 (14,28%)
Un. léxicas B1 com definidores A1, A2, B1	9/15 (60%)	8/12 (66,66%)	8/14 (57,14%)	10/14 (71,42%)
Un. léxicas B1 com definidores A1, A2, B1, B2+	3/15 (20%)	1/12 (8,33%)	3/14 (21,42%)	2/14 (14,28%)
Dados gerais	CaED2	OxED2	CoED2	LWord2
Definidores um nível abaixo da unidade léxica definida	5/28 (17,85%)	4/27 (14,81%)	4/29 (13,79%)	2/29 (6,89%)
Definidores de mesmo nível da unidade léxica definida	11/28 (39,28%)	12/27 (44,44%)	11/29 (37,93%)	16/29 (55,17%)
Definidores de um ou mais nível(is) acima da unidade léxica definida	12/28 (42,85%)	11/27 (40,74%)	14/29 (48,27%)	11/29 (37,93%)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OXED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

A análise estatística geral demonstra que as paráfrases explanatórias e notas de uso que contêm definidores cujo nível de referência é inferior ao nível da unidade léxica que definem constituem a menor parcela do total de itens analisados em cada dicionário. Itens com definidores de mesmo nível ou de nível superior ao da unidade léxica definida constituem uma

parte significativa da amostra: no CaED2 (2011), são 82,14% (23/28); no OxED2 (2012), são 85,18% (23/27 itens); no CoED2 (2014), são 86,20% (25/29 itens); e no LWord2 (2008), são 93,10% (27/29 itens). Tais dados revelam que o conhecimento necessário para compreender as paráfrases explanatórias e notas de uso de unidades léxicas A2 e B1 é superior ao conhecimento prévio atrelado aos níveis A1 e A2 de proficiência, respectivamente. Em outras palavras, podem ser considerados itens de difícil compreensão do ponto de vista dos definidores utilizados.

A análise por níveis revela dois cenários levemente diferentes. No tocante às unidades léxicas de nível A2, aproximadamente dois terços delas contêm definidores de nível B1 ou mais nos quatro dicionários analisados. Já no caso do nível B1, a maior parte dos definidores são também de nível B1. Em vista desses resultados, percebe-se que as paráfrases e notas de uso do menor nível de referência analisado, isto é, A2, são, curiosamente, as mais difíceis de compreender por exigirem um conhecimento prévio muito além do nível A1 de proficiência.

Sob uma perspectiva qualitativa, o cenário ideal seria que as unidades léxicas de nível A2 fossem definidas somente por itens léxicos atrelados ao nível A1 de referência; e que as unidades léxicas de nível B1 fossem definidas somente por itens léxicos relativos aos níveis A1 e A2. A significativa proporção de unidades léxicas A2 com definidores de níveis B1 em diante é um contrassenso em relação ao público que supostamente terá maior necessidade de auxílio em relação às unidades léxicas A2 por estar transitando pela etapa A2 de aprendizagem. Se é plausível supor que o consulente que atingiu nível A1 de proficiência e está em etapa A2 de aprendizagem já terá dificuldades em compreender as paráfrases explanatórias e notas de uso que contenham elementos de nível A2, ainda mais dificuldade enfrentará para compreender os itens com definidores de nível B1 em diante – se é que poderia compreendê-los.

As unidades léxicas de nível B1, por sua vez, parecem não oferecer tantas dificuldades ao consulente pretense quando comparadas com a análise anterior em virtude da quantidade massiva de paráfrases e notas de uso que contêm elementos de nível B1, no máximo. A única exceção é o OxED2 (2012), cuja incidência de itens com definidores A1-A2 é similar à dos itens com definidores B1. Ainda assim, é preciso considerar que o consulente pode não conhecer os definidores B1 utilizados se o seu processo de aprendizagem ainda não contemplou tais unidades léxicas.

Em resumo, o uso de definidores de nível equivalente ou acima do nível da unidade léxica que definem prejudica a compreensão das paráfrases e notas de uso. Mesmo que esses definidores estejam incluídos no dicionário, o simples movimento de se ter que buscar um segundo verbete para compreender o conteúdo do primeiro exige um esforço que o consulente pode não estar disposto a fazer. Nesse ponto, é pertinente recordar que, conforme comentado

em 2.3.1, a consulta a um dicionário *per se* consiste em uma interrupção de uma tarefa primária e que, por essa razão, deveria facilitar a localização e interpretação de informações por parte do consulente (BUGUEÑO MIRANDA, 2007b, p. 263; NESI, 2013, p. 3; NESI, 2015a, p. 69). Se esse consulente decide ir adiante com a sua busca por um segundo verbete para compreender o primeiro, e se este novo verbete também apresentar definidores de nível igual ou superior ao lema que o consulente também não entende, a busca terá sido dispendiosa e infrutífera, atrapalhando a retomada da atividade primária que levou à consulta ao dicionário.

Mesmo que a análise de 113 itens não ofereça um panorama tão preciso quanto se teria a partir de uma análise de um número maior de itens, não se pode ignorar que há um certo padrão nos quatro dicionários analisados no que concerne ao uso de definidores que não seriam os ideais para consulentes A1 e A2. A nossa hipótese é a de que o repertório vocabular de níveis A1 e A2 talvez não seja suficiente para que se possa redigir paráfrases explanatórias e notas de uso. Se essa hipótese for verdadeira, tem-se um impasse entre o uso inevitável de definidores de nível equivalente ou superior ao do lema e o prejuízo que esse procedimento causa na compreensão quando se trata de aprendizes de níveis A1 e A2 de proficiência. Na verdade, o que está por trás desse impasse são as limitações que o uso da língua-objeto em função metalinguística oferece ao consulente com proficiência A1 ou A2, seja ele brasileiro ou não. Os resultados obtidos permitem concluir que o uso do inglês em função metalinguística não é a solução mais adequada para a redação de paráfrases explanatórias e notas de uso em dicionários de aprendizes de inglês A1-B1 sob a perspectiva dos consulentes de níveis A1 e A2 de proficiência que se encontram em etapas A2 e B1 de aprendizagem.

5.1.1.2 Análise do nível dos padrões gramaticais

De maneira similar ao conjunto de análises léxicas, as análises gramaticais também revelaram três situações distintas: 1) presença de padrões gramaticais de níveis abaixo do nível da UL definida; 2) presença de padrões gramaticais em nível equivalente ao da UL definida; e 3) presença de definidores de nível superior ao da UL definida.

Assim como na análise dos definidores, julgou-se pertinente ajustar o nível de pronomes e de posicionamento sintático de preposições, conforme indicado a seguir.

Quadro 18 – Padrões gramaticais tratados como vocabulário passivo (inglês)

Classificação original	Ajuste de nível por extensão de vocabulário
<i>they</i> [gender neutral] (B2) (EGP, 2015)	<i>they</i> [person] (B1) (EVP, 2015)
<i>them</i> [gender neutral] (B2) (EGP, 2015)	<i>them</i> [to avoid him/her] (B1) (EVP, 2015)
<i>will</i> [predictions] (B1) (EGP, 2015)	<i>will</i> [future] (A2) (EVP, 2015)
[stranded preposition]	(conforme a classificação no EVP (2015))

Fonte: Elaboração da autora com base nos dados do EGP (2015) e do EVP (2015)

Os pronomes *they* e *them* e o verbo auxiliar *will* recebiam classificações diferentes conforme a ferramenta que se consultasse. Como são também unidades léxicas, adotou-se a classificação do EVP (2015). No caso da regência verbal cuja preposição estava posicionada ao final da frase (*stranded preposition*), tal como no verbete *across* (CaED2, 2011, 2), julgou-se que esse padrão sintático não oferece dificuldades no âmbito da compreensão, mas sim da produção. Por isso, nas análises, considerou-se o nível que a preposição recebia no EVP (2015).

Uma relação dos verbetes e acepções específicos a que se referem os dados obtidos se encontra no Anexo 6.

Quadro 19 – Análise do nível dos padrões gramaticais (inglês)

Classificação dos padrões gramaticais	CaED2	OxED2	CoED2	LWord2
Un. léxicas A2 com padrões A1 somente	5/13 (38,46%)	6/15 (40%)	5/15 (33,33%)	4/15 (26,66%)
Un. léxicas A2 com padrões A1, A2	5/13 (38,46%)	6/15 (40%)	5/15 (33,33%)	9/15 (60%)
Un. léxicas A2 com padrões A1, A2, B1+	3/13 (23,07%)	3/15 (20%)	5/15 (33,33%)	2/15 (13,33%)
Un. léxicas B1 com padrões A1, A2	10/15 (66,66%)	8/12 (66,66%)	9/14 (64,28%)	7/14 (50%)
Un. léxicas B1 com padrões A1, A2, B1	3/15 (20%)	3/12 (25%)	5/14 (35,71%)	6/14 (42,85%)
Un. léxicas B1 com padrões A1, A2, B1, B2+	2/15 (13,33%)	1/12 (8,33%)	-	1/14 (7,14%)
Dados gerais	CaED2	OxED2	CoED2	LWord2
Padrões um nível abaixo da unidade léxica definida	15/28 (53,57%)	14/27 (51,85%)	14/29 (48,27%)	11/29 (37,93%)
Padrões de mesmo nível da unidade léxica definida	8/28 (28,57%)	9/27 (33,33%)	10/29 (34,48%)	15/29 (51,72%)
Padrões de um ou mais nível(is) acima da unidade léxica definida	5/28 (17,85%)	4/27 (14,81%)	5/29 (17,24%)	3/29 (10,34%)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

É importante esclarecer que não se pretendeu classificar os padrões gramaticais em sua totalidade, mas sim destacar aqueles mais sobressalentes, especialmente as relações sintáticas e, em menor medida, certas propriedades morfológicas, como a flexão verbal, por exemplo.

Os dados demonstram que, em geral, parte significativa dos itens está distribuída entre aqueles cujos padrões gramaticais estão abaixo ou são equivalentes ao nível da UL definida. No CaED2 (2011), itens dessas duas categorias correspondem a 82,14% (23/28); no OxED2 (2012), a 85,18% do total (23/27); no CoED2 (2014), a 82,75% do total (24/29); e no LWord2 (2008), a 89,65% do total (26/29). Tais resultados são mais positivos que os da análise léxica geral, embora o uso de padrões gramaticais de mesmo nível da unidade léxica não seja uma boa estratégia para facilitar a compreensão.

Dentre as ULs de nível A2 do CaED2 (2011), do OxED2 (2012) e do CoED2 (2014), percebe-se que há uma distribuição proporcional dos padrões gramaticais entre as três categorias identificadas. A única exceção é o LWord2 (2008), que apresenta uma predominância de padrões de nível equivalente ao da UL que definem. As ULs de nível B1, por sua vez, são definidas, em sua maioria, por meio de padrões gramaticais que contemplam o conhecimento prévio dos consulentes de nível A2 de proficiência.

A distribuição dos itens entre as categorias que contemplam padrões A2 e/ou A1 correspondem, respectivamente, a 71,42% (20/28) no CaED2 (2011), 74,07% (20/27) no OxED2 (2012), 65,51% (19/29) no CoED2 (2014) e 68,96% (20/29) no LWord2 (2008). Tais dados sugerem que, em geral, os níveis A1 e, principalmente, A2 oferecem um repertório amplo o bastante para definir boa parte dos itens analisados – mas não todos. Por esse motivo, e do ponto de vista gramatical, as paráfrases explanatórias e notas de uso referentes a ULs B1 oferecem maior facilidade de compreensão que as paráfrases explanatórias e notas de uso referentes a ULs A2.

Embora em menor grau se comparado à análise léxica, percebe-se novamente um conflito entre o conhecimento prévio do consulente e os padrões gramaticais empregados na redação dos verbetes. Aparentemente, não é possível empregar somente padrões gramaticais cujo nível de referência esteja abaixo do nível da UL definida.

5.1.1.3 Corolário das análises léxica e gramatical

Os dados das análises léxica e gramatical obtidos até o momento permitem que se estabeleça um panorama da facilidade de compreensão leitora a partir do nível dos elementos definidores (definidores *per se* e padrões gramaticais) de cada item da amostra. Uma versão detalhada encontra-se no Anexo 7.

Quadro 20 – Facilidade de compreensão leitora: definidores e padrões gramaticais (inglês)

Classificação dos padrões gramaticais	CaED2	OxED2	CoED2	LWord2
Un. léxicas A2 definidas por elementos A1 somente	2/13 (15,38%)	1/15 (6,66%)	1/15 (6,66%)	0
Un. léxicas A2 definidas por elementos A1, A2	2/13 (15,38%)	3/15 (20%)	2/15 (13,33%)	5/15 (33,33%)
Un. léxicas A2 definidas por elementos A1, A2, B1+	9/13 (69,23%)	11/15 (73,33%)	12/15 (80%)	10/15 (66,66%)
Un. léxicas B1 definidas por elementos A1, A2	3/15 (20%)	3/12 (25%)	3/14 (21,42%)	3/14 (21,42%)
Un. léxicas B1 definidas por elementos A1, A2, B1	8/15 (53,33%)	7/12 (58,33%)	8/14 (57,14%)	9/14 (64,28%)
Un. léxicas B1 definidas por elementos A1, A2, B1, B2+	4/15 (26,66%)	2/12 (16,66%)	3/14 (21,42%)	2/14 (14,28%)
Dados gerais	CaED2	OxED2	CoED2	LWord2
Elementos definidores um nível abaixo da unidade léxica definida	5/28 (17,85%)	4/27 (14,81%)	4/29 (13,79%)	3/29 (10,34%)
Elementos definidores de mesmo nível da unidade léxica definida	10/28 (35,71%)	10/27 (37,03%)	10/29 (34,48%)	14/29 (48,27%)
Elementos definidores de um ou mais nível(is) acima da unidade léxica definida	13/28 (46,42%)	13/27 (48,14%)	15/29 (51,72%)	12/29 (41,37%)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

Os dados acima apenas confirmam o que já se vinha argumentando nas subseções anteriores: há uma incidência muito baixa de paráfrases explanatórias e notas de uso redigidas com definidores e padrões gramaticais relativos ao conhecimento prévio do usuário a quem os quatro dicionários se dirigem. Ao correlacionar os dados léxicos e gramaticais, percebe-se que somente pequenas parcelas das paráfrases explanatórias e notas de uso referentes a ULs A2 e B1 podem ser consideradas de fácil compreensão, considerando que tal facilidade está estritamente condicionada ao uso de elementos A1 e A1-A2, respectivamente. O uso de elementos cujo nível de referência é igual ou superior ao nível da UL definida é uma constante que predomina nas quatro obras analisadas.

Conforme já argumentado anteriormente, o cenário ideal no que tange à facilidade de compreensão leitora seria o uso exclusivo de definidores e padrões gramaticais de níveis abaixo do nível da UL definida. O fato de que, segundo os critérios empregados na análise, nenhuma das obras avaliadas tenha atingido esse patamar reforça a hipótese de que o inventário de opções léxicas e gramaticais relativas aos níveis A1 e A2 de proficiência não é suficiente para redigir paráfrases explanatórias e notas de uso de ULs A2 e B1. Frente a esses resultados, fica evidente que a manutenção do inglês em função metalinguística para redigir paráfrases e notas não contribui para a facilitação da compreensão. Em última análise, esse fator termina

comprometendo o didatismo do CaED2 (2011), do OxED2 (2012), do CoED2 (2014) e do LWord2 (2008) para uma parcela significativa do seu público-alvo.

Até o momento, as análises da compreensão leitora consideraram o usuário a partir da sua condição de estudante de inglês como LE, com nível de proficiência A1 e A2 e nível de aprendizagem A2 e B1. Segundo o apanhado teórico da Metalexigrafia e da Psicolinguística, a LM também constitui uma variável que define o perfil de usuário e participa do processamento da LE durante qualquer tarefa comunicativa, como o é a compreensão leitora que avaliamos nesta oportunidade. A próxima subseção está dedicada à análise do mesmo conjunto de itens sob a ótica das influências interlinguísticas da LM na língua-objeto usada em função metalinguística.

5.1.2 Influências interlinguísticas na compreensão leitora dos dicionários de inglês

Sendo a LM um componente integrante do processamento da LE, avaliar a facilidade de compreensão leitora na LE também requer que se investigue que itens podem apresentar propensão a influências interlinguísticas, sejam elas facilitadoras ou custosas. O objetivo desta subseção é avaliar o quanto o português como LM facilita o processamento das paráfrases explanatórias e notas de uso dos dicionários de aprendizes de inglês selecionados. Para tanto, é necessário, primeiramente, identificar contextos propícios para a ocorrência de influências interlinguística; em seguida, avaliar de que tipo de influência se trata.

Conforme já comentado no capítulo 4, a percepção de uma similaridade entre a LE e a LM é o fator mais reconhecido na literatura como desencadeador de influências interlinguísticas (JARVIS; PAVLENKO, 2008, p. 176-177). Em termos de compreensão leitora, a similaridade ortográfica é a que conta com ferramentas de medição, como a fórmula de cálculo da similaridade ortográfica de Van Orden (1987) disponível pela ferramenta NIM (2013). Portanto, as análises acerca da similaridade entre o inglês e o português se atêm ao nível ortográfico.

A ferramenta NIM (2013) não sugere pares similares a uma determinada unidade léxica, mas sim avalia um par predeterminado pelo pesquisador. Além disso, o cálculo gera um resultado numérico entre 0,00 e 1, sem que haja uma escala de referência que permita atribuir um valor ao resultado obtido, isto é, que permita identificá-lo como indicativo de presença ou ausência de similaridade ortográfica. Em vista disso, estabelecemos um ponto de corte de 0,51, isto é, maior que 50%, para distinguir o que é similar e o que não é (suficientemente) similar. Sob essa perspectiva, índices entre 0,00 e 0,50 atestam falta de similaridade, enquanto índices

entre 0,51 e 1,00 indicam similaridade. Interpretações dicotômicas como a que adotamos não são o ideal; porém, a Psicolinguística ainda não fornece orientações unificadas a respeito de uma gradação em relação à escala de índices de similaridade. À medida que as pesquisas na área conseguirem determinar tal gradação, será possível aprimorar a medição e, por conseguinte, as interpretações feitas sobre os dados, bem como as decisões tomadas a partir deles.

Para fins desta tese, ao invés de tentar sugerir um cognato do português para cada definidor e, a partir disso, calcular o grau de similaridade ortográfica, pareceu-nos mais interessante aproveitar a oportunidade para testar uma hipótese levantada por Reolon (2013, p. 166-182). A autora sugere que o público brasileiro teria sua compreensão leitora facilitada por meio do uso de romanismos na redação de paráfrases explanatórias de dicionários de aprendizes de inglês. O conceito de romanismo está construído sobre a noção de origem etimológica das palavras. Tal como comentado no capítulo 4, a Linguística adota essa mesma noção para definir o conceito de *cognato*; um romanismo, portanto, seria uma espécie de cognato entre línguas provenientes do latim românico, que constituíam o latim vulgar e que vieram a se tornar o que hoje conhecemos como português, espanhol, italiano, francês e romeno.

Segundo Reolon (2013, p. 166-182), os romanismos presentes na língua inglesa seriam mais fáceis de entender sob a perspectiva do público brasileiro porque remeteriam a uma origem comum entre tais palavras e palavras do português: o latim¹¹¹. Além da origem latina comum, outro pressuposto que sustenta tal hipótese é o de que muitos romanismos são similares a unidades léxicas do português não somente na forma, mas também no conteúdo. Reolon (2013) propõe, então, que os dicionários de aprendizes de inglês para brasileiros empreguem romanismos na redação de paráfrases explanatórias para facilitar a sua compreensão por parte do público brasileiro.

Os romanismos na língua inglesa podem ser identificados pela informação de origem etimológica que determinados dicionários de inglês fornecem, como o OALD8e (2010).

¹¹¹ Trata-se de empréstimos incorporados ao inglês que se devem, em sua maioria, a episódios históricos de invasões ao território do atual Reino Unido, como a ocupação romana no século I e a ocupação normanda no século XI.

Quadro 21 – Informação de origem etimológica no OALDe (2010)

<p>↔ allow [...] WORD ORIGIN Middle English (originally in the senses ‘commend, sanction’ and ‘assign as a right’): from Old French <i>alouer</i>, from Latin <i>allaudare</i> ‘to praise’, reinforced by medieval Latin <i>allocare</i> ‘to place’, from <i>ad-</i> ‘to’ + <i>locare</i>, from <i>locus</i> ‘place’. [...]</p>	<p>↔ par·ticu·lar [...] WORD ORIGIN late Middle English: from Old French <i>particuler</i>, from Latin <i>particularis</i> ‘concerning a small part’, from <i>particula</i> ‘small part’. [...]</p>	<p>↔ enjoy [...] WORD ORIGIN late Middle English: from Old French <i>enjoier</i> ‘give joy to’ or <i>enjoïr</i> ‘enjoy’, both based on Latin <i>gaudere</i> ‘rejoice’.</p>
<p>↔ in·for·ma·tion [...] WORD ORIGIN late Middle English (also in the sense ‘formation of the mind, teaching’), via Old French from Latin <i>informatio(n)-</i>, from the verb <i>informare</i> ‘shape, fashion, describe’, from <i>in-</i> ‘into’ + <i>forma</i> ‘a form’. [...]</p>	<p>↔ event [...] WORD ORIGIN late 16th cent.: from Latin <i>eventus</i>, from <i>evenire</i> ‘result, happen’, from <i>e-</i> (variant of <i>ex-</i>) ‘out of’ + <i>venire</i> ‘come’.</p>	<p>↔ re·mem·ber [...] WORD ORIGIN Middle English: from Old French <i>remembrer</i>, from late Latin <i>rememorari</i> ‘call to mind’, from <i>re-</i> (expressing intensive force) + Latin <i>memor</i> ‘mindful’. [...]</p>

Fonte: OALD8e (2010, s.vs. *allow*, *particular*, *enjoy*, *information*, *event*, *remember*. Diferenciação tipográfica no original)

De acordo com o OALD8e (2010), as palavras-entrada acima são romanismos em função da sua origem latina por intermédio do francês. No entanto, é precipitado pressupor que o estudante brasileiro consiga associar todos esses exemplos a unidades léxicas do português pela via da representação ortográfica. Também é precipitado pressupor que, nos casos em que é possível estabelecer tal relação, haverá uma convergência de significado entre o inglês e o português pelo simples fato de serem romanismos. Os comentários que seguem partem da percepção desta autora sobre o estabelecimento de candidatos do português similares às palavras em inglês tomadas como exemplo, pois, como já mencionado, não há uma forma automatizada de detecção de candidatos a símeles entre duas línguas.

Nossa primeira constatação é a de que *allow*, embora seja um romanismo, não se assemelha a nenhuma palavra do português (o termo *palavra* é usado aqui em alusão à representação ortográfica). Não foi possível estabelecer uma relação de similaridade, o que significa que dificilmente haverá influências interlinguísticas pela via ortográfica que tornem o processamento semântico mais fácil ou mais custoso. Portanto, o *status* de romanismo da palavra *allow* não a torna automaticamente um elemento facilmente compreensível por um brasileiro. Casos como o de *allow* serão chamados de *romanismo tipo 1*. Da mesma forma, casos nos quais o índice de similaridade léxica seja 0,50 ou inferior também se inserem nessa categoria.

A segunda constatação que se pode fazer diz respeito aos romanismos que, embora similares a palavras do português, apresentam conteúdo semântico divergente ao usado nessa

língua. É o caso de *particular* e *enjoy*. *Particular* [determinado], do inglês, é similar à palavra *particular* do português. Não é necessário sequer usar a ferramenta NIM (2013) para calcular a similaridade ortográfica, pois ambas possuem exatamente a mesma representação. Porém, embora sejam romanismos, o seu conteúdo semântico é completamente diferente nas duas línguas. O mesmo efeito ocorre com *enjoy* [gostar] em relação a *enjoar* do português – palavra que, conforme a nossa percepção, é a que mais se assemelha ortograficamente ao exemplo do inglês. Segundo o NIM (2013), a similaridade ortográfica de *enjoy* em relação a *enjoar* é de 0,65; considerando o parâmetro 0,51+ para considerar uma palavra ortograficamente similar a outra, *enjoar* se enquadra como similar a *enjoy*.

É importante observar que a convergência na forma e divergência semântica influenciará no processamento conforme o aprendiz tenha ciência ou não dessa divergência. Um aprendiz que percebe a similaridade ortográfica e está ciente da divergência semântica entre a LE e a LM provavelmente processará de maneira mais custosa a unidade léxica em questão. Contudo, um aprendiz que também percebe a similaridade ortográfica, mas não está ciente da divergência semântica entre a LE e a LM, poderá processar a unidade léxica em questão de maneira aparentemente facilitada, mas que incorrerá em uma compreensão ou produção errônea. Seja como for, esse tipo de romanismo não facilita verdadeiramente o processamento, contrariando novamente os pressupostos de Reolon (2013). Casos como o de *particular* e *enjoy* serão denominados *romanismo tipo 2*.

A terceira constatação que se pode fazer a partir dos exemplos elencados anteriormente é que há, sim, casos de convergência ortográfica (similaridade) acompanhados de convergência semântica total ou parcial. A palavra *information* [informação] é não apenas similar ao seu equivalente informação segundo a percepção desta autora, mas também conforme o NIM (2013), que atesta um índice de similaridade de 0,61 em relação a *informação*. O conteúdo semântico de ambas é totalmente convergente. Por outro lado, há situações de similaridade ortográfica em que o conteúdo semântico não é exatamente o mesmo, mas facilita a compreensão. Trata-se de casos como o de *event* [acontecimento; evento] e *evento*, para as quais o NIM (2013) atesta um índice de similaridade de 0,78. Dependendo do contexto em que está inserida, a palavra *event*, embora não totalmente convergente com *evento* do ponto de vista semântico, pode ter a compreensão leitora facilitada em função da influência interlinguística do símile em português. Casos análogos a *information* e *event* serão chamados *romanismos tipo 3*.

A quarta constatação diz respeito a casos de romanismos que são de difícil resolução. A identificação de palavras similares na LM depende do quão refinado é o conhecimento que o aprendiz tem da sua LM. A palavra *attached* [anexado], por exemplo, foi associado por esta

autora à palavra *atachado* (verbo atachar) do português. Contudo, *atachar* é pouco usado no português e compete com a forma preferencial *anexar*¹¹². A associação por similaridade ortográfica entre *attached* e *atachado* e as possíveis influências interlinguísticas dependerão, portanto, da amplitude de vocabulário do aprendiz. A esse tipo de casos e outros que eventualmente não se enquadrem nas categorias anteriores denominamos *romanismos tipo 4*.

Quadro 22 – Tipos de romanismos

Tipo de romanismo	Descrição	Exemplo
1	O índice de similaridade é baixo (0,50 ou inferior) ou não há similaridade ortográfica com palavras do português. Há pouca probabilidade de ocorrerem influências no processamento.	<i>allow</i>
2	Há semelhança ortográfica, mas divergência no plano semântico. Pode haver influências custosas ou que incorram em facilitação ilusória.	<i>particular</i> <i>enjoy</i>
3	Há semelhança ortográfica e semântica, com probabilidade de gerar influências verdadeiramente facilitadoras.	<i>information</i> <i>event</i>
4	Casos de difícil resolução. A percepção da semelhança ortográfica depende do conhecimento de palavras pouco frequentes na LM.	<i>attached</i>

Fonte: Elaboração da autora

A partir da classificação acima, o tipo de romanismo que realmente tem o potencial de facilitar a compreensão por parte de consulentes brasileiros são os romanismos tipo 3. Os romanismos tipo 1 não influenciam a compreensão, enquanto os romanismos tipo 2 têm o potencial de levar o consulente a uma compreensão errônea do conteúdo semântico de paráfrases explanatórias e notas de uso. Por essas razões, considera-se que os romanismos tipos 1 e 2 não facilitam a compreensão. Pela impossibilidade de se decidir se facilitarão ou não a compreensão, os romanismos tipo 4 são contabilizados nas avaliações quantitativas, mas excluídos das avaliações qualitativas.

A metodologia de análise dos tipos de romanismos nas paráfrases explanatórias e notas de uso, com a conseqüente possibilidade de gerarem influências interlinguísticas, esteve constituída das seguintes etapas:

¹¹² Dicionários gerais, tais como o *Dicionário Houaiss da Língua Portuguesa* (Hou, 2009, s.v. *atachar*) e o *Grande Dicionário Sacconi da Língua Portuguesa* (Sac, 2010, s.v. *atachar*), por exemplo, classificam *atachar* como anglicismo e recomendam evitar o seu uso, recomendando o vernáculo *anexar*. O Corpus do Português (CP, 2019), por sua vez, não registra as formas *atachar* e *atachado*. Apesar de este não ser um indicativo de que não sejam usadas no português, pois os textos do CP (2019) contemplam apenas jornais e revistas, esses dados são um indicativo de que *atachar* e *atachado* são palavras muito pouco usadas.

- (1) Identificar os romanismos, seja os intermediados pelo francês, seja os de origem latina direta, com base nas informações etimológicas do OALD8e (2010).
- (2) Sugerir candidatos do português ortograficamente similares quando for possível. No caso de verbos no particípio e adjetivos deverbais no particípio, o símile sugerido para comparação foi a forma no infinitivo em português.
- (3) Calcular o nível de similaridade ortográfica baseado na fórmula de Van Orden (1987) e disponibilizado pela ferramenta NIM (2013). Os símiles em português foram inseridos na ferramenta NIM (2013) sem os acentos (´, ^) nem o sinal diacrítico til (~); e a letra ç foi substituída por c.
- (4) Classificar os tipos romanismos de acordo com as quatro categorias propostas.

Os resultados obtidos estão dispostos a seguir de maneira sintética. O total de romanismos diz respeito ao número de diferentes romanismos (*types*) e não abarca o número de repetições de um mesmo romanismo (*tokens*). Uma lista com os romanismos, seu símile em português, o índice de similaridade ortográfica entre ambos e o equivalente em português pode ser consultada no Anexo 8.

Quadro 23 – Classificação dos tipos de romanismos (inglês)

Nível de referência CEFR dos romanismos	Romanismos Tipo 1	Romanismos Tipo 2	Romanismos Tipo 3	Romanismos Tipo 4
	19/72 (26,38%)	8/72 (11,11%)	38/72 (52,77%)	7/72 (9,72%)
A1	7	1	7	2
A2	5	1	16	1
B1	7	3	10	4
B2+	-	2	4	-
Ausente do EVP (2015)	-	1	1	-

Fonte: Elaboração da autora com base nos dados do OALD8e (2010) e do NIM (2013)

Os romanismos tipo 1 correspondem à quarta parte dos romanismos encontrados. Para grande parte desses romanismos, não foi possível sugerir um símile em português. Ante a falta de similaridade ou de similaridade significativa com uma palavra do português, considerou-se que esses romanismos não exercem influências no processamento do inglês por parte de aprendizes brasileiros pela via da LM. O fato de que ¼ dos tipos de romanismos encontrados não facilitem o processamento do inglês pelo perfil de usuário em questão constitui uma

contraprova à hipótese de Reolon (2013). O *status* de romanismo *per se* não incorre automaticamente em uma facilitação do processamento do inglês a nível de compreensão leitora.

Aparentemente, os romanismos tipo 1 se encontram no mesmo patamar que os anglicismos; na prática, seu processamento dependerá do conhecimento prévio do consulente, que nesta tese assume-se como o vocabulário relativo a cada nível de referência do QECRL (2001) indicado pelo EVP (2015). Observe-se, por exemplo, a paráfrase explanatória do verbete *gorgeous* do LWord2 (2008): “very beautiful or pleasant” (LWord2, 2008, s.v. *gorgeous*). No exemplo, há dois romanismos tipo 1: *beautiful*, de nível A1, e *pleasant*, de nível A2. Sendo *gorgeous* uma UL de nível B1, espera-se que haja uma facilitação no processamento em decorrência do esperado conhecimento prévio dos definidores citados, e não do seu *status* de romanismo ou por uma eventual similaridade ortográfica e semântica com alguma UL do português. O mesmo efeito facilitador não se pode esperar do romanismo *pleased*, definidor de nível A2 presente na paráfrase explanatória de *glad* do mesmo dicionário. Em dita paráfrase, lê-se “if you are glad, you are pleased and happy about something” (LWord2, 2008, s.v. *glad*). Sendo *glad* uma UL também de nível A2, não se pode esperar um efeito facilitador, seja pela via da identificação de um romanismo, seja pela via da similaridade ortográfica.

Os romanismos tipo 2, por sua vez, apresentam baixa incidência dentre o total de *types* de romanismos encontrados na amostra. Ainda assim, sua incidência na amostra analisada também é indicativa de que o *status* de romanismo de um definidor não confere necessariamente uma facilitação no processamento. O tipo de influência que exercem no processamento dependerá do conhecimento prévio do consulente, isto é, do nível de referência ao qual está atrelado o definidor que é romanismo tipo 2. Observe-se a paráfrase explanatória de *among* (acepção 2) do CaED2 (2011), cujo conteúdo semântico está atrelado ao nível A2 pelo EVP (2015): “in a particular group” (CaED2, 2011, s.v. *among*, 2). O definidor *particular* é um romanismo tipo 2 de nível B2+, ou seja, não se pode esperar que o consulente em nível A2 de aprendizagem o conheça, muito menos que saiba da diferença de significado entre esta UL do inglês e seu símile em português *particular*. Nesse e em outros casos similares, há um contexto propício para uma influência interlinguística de facilitação ilusória, na qual a alta similaridade ortográfica com o português, somada à falta de ciência sobre o significado da UL do inglês, leva a uma facilitação no processamento, mas não a um processamento em conformidade com o uso da língua inglesa. A associação da representação ortográfica *particular* a um significado diferente daquele que lhe é atribuído no inglês termina gerando equívocos na compreensão e na produção da UL *among* [in a group] ao resolver tarefas linguísticas. Por outro lado, se o

consulente já conhece o verdadeiro significado de um determinado romanismo tipo 2 no inglês, e sabe que há uma diferença em relação ao significado de um símile em português, o processamento será, muito provavelmente, mais custoso para o aprendiz brasileiro.

Os romanismos tipo 3 correspondem a pouco mais da metade dos tipos de romanismos identificados. Estes são os que realmente facilitam o processamento de paráfrases explanatórias e notas de uso. Sua presença nos itens analisados é especialmente relevante para o aprendiz brasileiro quando tais romanismos são classificados pelo EVP (2015) em um nível de referência equivalente ou superior ao nível da UL definida – ou seja, quando fogem ao conhecimento prévio do consulente em questão.

Finalmente, os romanismos tipo 4 constituem a menor parcela dos tipos de romanismos encontrados. Embora apresentem índice de similaridade ortográfica maior que 0,51 em relação a palavras do português, tais palavras costumam ser UL variantes com menor respaldo no *Corpus* do Português (CP, 2019) se comparadas às suas formas preferenciais. É o caso do símile menos frequente *atachado* (nenhum registro) e a forma preferencial *anexado* (1.764 registros), ou ainda do símile menos frequente *rememorar* (355 registros) e a forma preferencial *lembrar* (65.458) (CP, 2019). Como se trata de casos de difícil resolução, que dependem da capacidade subjetiva de cada consulente de reconhecer ou não um símile não preferencial do português, julgou-se mais apropriado tratar os romanismos tipo 4 como se fossem anglicismos, isto é, não os considerar como possíveis facilitadores do processamento pela via da LM.

Em resumo, na amostra dos quatro dicionários avaliados, metade dos romanismos encontrados facilita o processamento e a outra metade ou o torna mais custoso, ou não gera nem custos nem benefícios ao processamento. Frente a esses dados, conclui-se que a hipótese de Reolon (2013) está parcialmente correta, mas precisa de mais investigação para que se possa tomar o uso de romanismos como estratégia de descrição metalexigráfica. Um aproveitamento da LM no processamento de paráfrases explanatórias e notas de uso depende do uso de um tipo específico de romanismo, o qual denominamos “tipo 3”. A aplicação da proposta da autora depende, portanto, de um levantamento prévio de romanismos na língua inglesa e da identificação e filtragem dos romanismos tipo 3, gerando um banco de definidores adequado ao perfil do consulente brasileiro. Uma investigação dessa magnitude, bem como a sua testagem, constituiria, por si só, uma pesquisa a nível de mestrado, no mínimo; foge, assim, das dimensões desta tese.

5.1.3 Facilidade de compreensão leitora dos dicionários de aprendizes de inglês A1-B1

Os resultados da análise da subseção anterior permitiram reordenar o *ranking* de paráfrases e notas de uso considerando a variável LM e os definidores mais propícios a influências interlinguísticas. Os definidores que desencadearam a reordenação foram aqueles cujo nível de referência era igual ou superior ao nível da UL definida, pois não formam parte, *a priori*, do conhecimento prévio do consulente de proficiência A1 e A2. Definidores de nível abaixo do lema não foram considerados como justificativa para a reordenação, pois, independentemente do tipo de romanismo no qual se enquadram, constituem o conhecimento prévio esperado do consulente. Dentre os definidores que não formam parte do conhecimento prévio do consulente, tomaram-se as seguintes decisões: a) se todos fossem romanismos tipo 3, a paráfrase explanatória ou nota de uso receberia uma nova classificação, enquadrando-se entre aqueles itens que são fáceis de compreender; b) se algum dos definidores em questão fosse romanismo tipos 1, 2 ou 4, a paráfrase explanatória ou nota de uso permanecem com a mesma classificação recebida nas avaliações léxica e gramatical – mesmo que haja, concomitantemente, um ou mais romanismos tipo 3. É importante esclarecer que o foco da nova classificação recai sobre o efeito facilitador da similaridade ortográfica sobre o processamento da totalidade da paráfrase explanatória ou nota de uso; quando não há esse efeito, ou quando ele cobre apenas parte dos itens que não constituem o conhecimento prévio do consulente, a paráfrase ou nota é parcialmente, e não totalmente, fácil de compreender. Observa-se também que romanismos tipo 2 que não constituam o conhecimento prévio do consulente foram considerados como fontes de problemas de compreensão. Os resultados da nova classificação estão sintetizados na próxima página por questões de espaço. Uma versão detalhada, discriminando verbetes e acepções, pode ser encontrada no Anexo 9.

Quadro 24 – Facilidade de compreensão sob a perspectiva do consultante brasileiro (inglês)

Paráfrases explanatórias e notas de uso referentes a ULs A2				
	CoED2	OxED2	CoED2	LWord2
FÁCIL DE COMPREENDER Todos os definidores e padrões gramaticais são previamente conhecidos (A1). Eventuais definidores de níveis A2+ são romanismos tipo 3. Eventuais romanismos tipo 2 geram poucos problemas de compreensão por serem previamente conhecidos (A1).	2/13 (15,38%)	1/15 (6,66%)	1/15 (6,66%)	0
LIGEIRAMENTE DIFÍCIL DE COMPREENDER Há definidores e/ou padrões gramaticais A2 potencialmente desconhecidos. Se há romanismos tipo 3, estes não correspondem à totalidade dos definidores A2. Pode haver romanismos tipos 1, 2 ou 4, que geram maiores problemas de compreensão se forem A2.	3/13 (23,07%)	5/15 (33,33%)	3/15 (20%)	7/15 (46,66%)
DIFÍCIL DE COMPREENDER Há definidores e/ou padrões gramaticais B1+ potencialmente desconhecidos. Se há romanismos tipo 3, estes não correspondem à totalidade dos definidores A2+. Pode haver romanismos tipos 1, 2 ou 4, que geram maiores problemas de compreensão se forem A2+.	8/13 (61,53%)	9/15 (60%)	11/15 (73,33%)	8/15 (53,33%)
Paráfrases explanatórias e notas de uso referentes a ULs B1				
	CoED2	OxED2	CoED2	LWord2
FÁCIL DE COMPREENDER Todos os definidores e padrões gramaticais são previamente conhecidos (A1, A2). Eventuais definidores de níveis B1+ são romanismos tipo 3. Eventuais romanismos tipo 2 geram poucos problemas de compreensão por serem previamente conhecidos (A1, A2).	4/15 (26,66%)	4/12 (33,33%)	5/14 (35,71%)	4/14 (28,57%)
LIGEIRAMENTE DIFÍCIL DE COMPREENDER Há definidores e/ou padrões gramaticais B1 potencialmente desconhecidos. Se há romanismos tipo 3, estes não correspondem à totalidade dos definidores B1. Pode haver romanismos tipos 1, 2 ou 4, que geram maiores problemas de compreensão se forem B1.	8/15 (53,33%)	6/12 (50%)	9/14 (64,28%)	8/14 (57,14%)
DIFÍCIL DE COMPREENDER Há definidores e/ou padrões gramaticais B2+ potencialmente desconhecidos. Se há romanismos tipo 3, estes não correspondem à totalidade dos definidores B1+. Pode haver romanismos tipos 1, 2 ou 4, que geram maiores problemas de compreensão se forem B1+.	3/15 (20%)	2/12 (16,66%)	0	2/14 (14,28%)
Dados gerais				
	CoED2	OxED2	CoED2	LWord2
FÁCIL DE COMPREENDER	6/28 (21,42%)	5/27 (18,51%)	6/29 (20,68%)	4/29 (13,79%)
LIGEIRAMENTE DIFÍCIL DE COMPREENDER	11/28 (39,28%)	11/27 (40,74%)	12/29 (41,37%)	15/29 (51,72%)
DIFÍCIL DE COMPREENDER	11/28 (39,28%)	11/27 (40,74%)	11/29 (37,93%)	10/29 (34,48%)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

As análises das unidades léxicas de nível A2 revelam que, mesmo com a presença dos romanismos tipo 3, o efeito facilitador dessas ULs não compensa as dificuldades geradas por elementos de nível equivalente ou superior ao nível da UL definida. Na mesma linha do que já se vinha comentando antes, o consulente brasileiro de nível A2 de aprendizagem se vê muito pouco contemplado pelos quatro dicionários, pois a maior parte dos itens analisados não é de fácil compreensão para este usuário.

Quando comparados com os resultados das ULs A2, os resultados das análises de itens referentes a ULs B1 são, sem dúvidas, mais otimistas, pois os itens de fácil compreensão giram em torno de uma média de 30%. Por outro lado, é pertinente observar que, *grosso modo*, esses valores indicam que a probabilidade de um consulente brasileiro de nível B1 de aprendizagem deparar-se com uma paráfrase explanatória ou nota de uso que pode lhe apresentar algum grau de dificuldade de compreensão é de, no mínimo, 50%. Novamente, a presença de romanismos tipo 3 não é suficiente para que o consulente brasileiro se veja beneficiado por esse tipo de UL no exercício da sua compreensão leitora.

Uma pergunta de pesquisa que surgiu durante a etapa de análise em questão foi se havia uma variação interna de facilidade de compreensão entre os itens de cada categoria proposta no Quadro 24 (fácil / ligeiramente difícil / difícil de compreender). Para responder a essa questão, recorreremos a uma proposta teórico-metodológica sobre a mensuração da sofisticação de textos: o *Crowdsourced algorithm of reading comprehension* (CAREC), de Crossley, Skalicky e Dascalu (2019)¹¹³. A fórmula emprega variáveis psicolinguísticas, léxico-estatísticas e de coesão textual¹¹⁴, gerando um valor numérico de sofisticação de um texto em língua inglesa. Não há uma escala de referência que atribua valor ao índice obtido; até o momento, é possível usar o resultado que a ferramenta gera acerca de cada texto para estabelecer um *ranking* de textos, dos mais aos menos sofisticados em termos de compreensão leitora. O cálculo da sofisticação textual é feito por meio da ferramenta *ARTE: Automatic Readability Tool for English* (CHOI; CROSSLEY, 2020), que calcula não somente o índice de sofisticação (CAREC), mas também de velocidade de leitura, bem como os índices de fórmulas tradicionais de complexidade textual e velocidade de leitura.

¹¹³ Agradecemos à profa. Ana Fontes, que sugeriu esse caminho durante a avaliação da tese na etapa de qualificação.

¹¹⁴ As variáveis psicolinguísticas são a idade de aquisição das palavras, o quão imagináveis são os referentes das palavras de conteúdo e a prosódia semântica positiva de adjetivos. As variáveis léxico-estatísticas envolvem frequência das palavras, caracteres por palavra, entropia dos caracteres, proporção de palavras de conteúdo e proporção de combinações léxicas formadas por dois a três elementos. As variáveis de coesão são a sobreposição de parágrafos e substantivos e a presença de conectores temporais.

Para fins desta tese, cada paráfrase explanatória e nota de uso analisadas foram submetidas à calculadora ARTE (CHOI; CROSSLEY, 2020) para aferição da sofisticação textual. O Quadro 25 a seguir sistematiza os resultados obtidos, enquanto o Anexo 10 discrimina os resultados fornecidos pela ferramenta ARTE (CHOI; CROSSLEY, 2020).

Quadro 25 – Ranqueamento interno de cada categoria de facilidade de compreensão (inglês)

Facilidade de compreensão	Ranqueamento de paráfrases e notas de uso em cada dicionário			
	CaED2	OxED2	CoED2	LWord2
FÁCIL DE COMPREENDER	<i>glad</i> (ac.) <i>remember</i> (ac. 2)	<i>glad</i> (ac.)	<i>remember</i> (ac. 3)	-
LIGEIRAMENTE DIFÍCIL DE COMPREENDER	<i>library</i> (ac.) <i>rent</i> (ac. 2) <i>rent</i> (n.)	<i>among</i> (n.) <i>rent</i> (ac. 1) <i>already</i> (ac.) <i>library</i> (n.) <i>already</i> (n.)	<i>glad</i> (ac.) <i>pleasant</i> (ac. 2) <i>among</i> (n.)	<i>already</i> (ac. 1) <i>among</i> (ac. 1) <i>pleasant</i> (ac. 2) <i>glad</i> (ac.) <i>rent</i> (ac. 1) <i>library</i> (ac.) <i>remember</i> (ac. 2)
DIFÍCIL DE COMPREENDER	<i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>pleasant</i> (ac.) <i>landscape</i> (ac.) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>among</i> (ac. 2) <i>appointment</i> (ac.)	<i>across</i> (ac. 1) <i>pleasant</i> (ac.) <i>landscape</i> (ac.) <i>across</i> (ac. 2) <i>across</i> (n.) <i>glad</i> (n.) <i>library</i> (ac.) <i>among</i> (ac. 2) <i>appointment</i> (ac.1)	<i>landscape</i> (ac. 1) <i>rent</i> (ac. 1) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 2) <i>among</i> (ac. 3) <i>library</i> (ac.) <i>already</i> (n.) <i>already</i> (ac. 1) <i>appointment</i> (ac.1) <i>pleasant</i> (ac. 1)	<i>landscape</i> (ac.) <i>pleasant</i> (ac. 1) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>appointment</i> (ac. 1) <i>library</i> (n. 1) <i>already</i> (n.) <i>library</i> (n. 2)
FÁCIL DE COMPREENDER	<i>gorgeous</i> (ac.) <i>towards</i> (ac. 1) <i>pretty</i> ¹ (ac.) <i>among</i> (ac. 1)	<i>among</i> (ac. 1) <i>gorgeous</i> (ac.) <i>likely</i> (ac.) <i>towards</i> (ac. 1)	<i>among</i> (ac. 1) <i>likely</i> (ac. 2) <i>gorgeous</i> (ac.) <i>towards</i> (ac. 1) <i>edge</i> (ac. 1)	<i>gorgeous</i> (ac.) <i>likely</i> (ac. 1) <i>towards</i> (ac. 1) <i>label</i> ¹ (ac. 2)
LIGEIRAMENTE DIFÍCIL DE COMPREENDER	<i>already</i> (ac. 2) <i>flavour</i> (ac.) <i>bury</i> (ac. 1) <i>label</i> ¹ (ac. 2) <i>rent</i> (ac. 3) <i>remind</i> (ph.v.) <i>edge</i> (ac.) <i>remind</i> (ac. 1)	<i>bury</i> (ac. 1) <i>flavour</i> (ac.) <i>remind</i> (ac. 1) <i>edge</i> (ac. 1) <i>remind</i> (ac. 2) <i>pretty</i> ¹ (ac.)	<i>bury</i> (ac. 2) <i>rent</i> (ac. 2) <i>flavour</i> (ac.) <i>label</i> (ac.) <i>already</i> (ac. 2) <i>pretty</i> (ac. 2) <i>remind</i> (ac. 3) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2)	<i>already</i> (ac. 2) <i>rent</i> (ac. 2) <i>remind</i> (ac. 2) <i>flavour</i> (ac.) <i>edge</i> (ac. 1) <i>remind</i> (ac. 1) <i>among</i> (ac. 2) <i>pretty</i> ¹ (ac.)
DIFÍCIL DE COMPREENDER	<i>likely</i> (ac. 1) <i>remind</i> (n.) <i>label</i> ¹ (ac. 1)	<i>label</i> ¹ (ac.) <i>rent</i> (ac. 2)	-	<i>bury</i> (ac. 1) <i>label</i> ¹ (ac. 1)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015), EGP (2015) e CAREC (CROSSLEY et al., 2019)

Se, antes do cálculo CAREC, os itens da amostra estavam inseridos nos diferentes níveis de facilidade de compreensão sem que fosse possível estabelecer uma graduação interna em cada nível, o índice CAREC permitiu uma graduação mais fina dos dados. Assim, por exemplo, os

itens A2 do CaED2 (2011) classificados como difíceis de compreender, que, inicialmente, estavam apenas catalogados sob esse rótulo, sem uma gradação interna, foram organizados conforme o nível de dificuldade que ofereciam segundo o CAREC. O mesmo procedimento se aplica às demais células do quadro.

Se comparado a outras fórmulas tradicionais, como o índice Flesch, por exemplo, o índice CAREC se apresenta como uma proposta recente e promissora no que concerne à aferição do nível de sofisticação de textos. A fórmula em questão é produto de anos de investigação no tocante à mensuração de propriedades textuais, as quais geralmente são avaliadas de maneira subjetiva. Até o momento, não tivemos conhecimento de outras propostas semelhantes à apresentada nesta tese no que diz respeito ao cálculo de sofisticação do texto dicionarístico. Trata-se de uma estratégia a mais para contribuir com o didatismo de um dicionário, em especial dos dicionários de aprendizes. Por enquanto, ferramentas de cálculo da sofisticação de textos estão disponíveis majoritariamente para a língua inglesa.

Como corolário da seção, e à luz dos critérios de análise adotados, a habilidade de compreensão que os dicionários de aprendizes de inglês A1-B1 requerem do seu consulente é incompatível com o nível de compreensão leitora postulado pelo CEFR (2001) em relação aos níveis A1 e A2 de proficiência. Em se tratando de um usuário que tem um nível ainda relativamente baixo de proficiência em inglês, a dificuldade na compreensão dos itens analisados pode terminar por desencorajá-lo em relação ao uso de qualquer um dos quatro dicionários. O cerne dessa dificuldade está no tipo de estratégia de descrição metalexográfica empregada em paráfrases explanatórias e notas de uso, pois não permite que esses segmentos informativos possam ser facilmente compreendidos por parte significativa dos seus consulentes pretensos. Em contrapartida, e conforme já sugerido no decorrer das análises, há que se questionar até que ponto é possível gerar paráfrases explanatórias e notas de uso léxica- e gramaticalmente compatíveis com os níveis A1 e A2 de proficiência, pois os resultados demonstraram que predominam itens com definidores de níveis A2 e B1 e padrões gramaticais A1 e A2.

Presume-se que a falta de correspondência entre a maneira como a língua inglesa é descrita e o perfil de usuário dos quatro dicionários analisados pode ter impactado negativamente na procura por essas obras e, conseqüentemente, nos investimentos em novas edições, pois os respectivos projetos não passaram da segunda edição e não possuem edições posteriores a 2014. A editora Oxford parece investir apenas no seu dicionário para aprendizes avançados, que é atualizado a cada cinco anos e cuja mais recente edição data de 2020. A editora Harper Collins tem investido somente nas atualizações de seus dicionários intermediário e

avanzado, ambos com a última edição datada de 2018 e com atualizações a cada quatro anos em média. As editoras Pearson/Longman e Cambridge, que atualizavam seus dicionários intermediários e avançados a cada cinco a seis anos, aparentemente não têm levado adiante nenhum de seus projetos desde 2013 e 2014, respectivamente. É difícil acreditar que a continuidade das edições do CaED2 (2011), do OxE2 (2012), do CoED2 (2014) e do LWord2 (2008) tenha sido interrompida porque os aprendizes de níveis iniciais de proficiência como A1 e A2 não precisem de dicionários especializados no reflexo da língua inglesa em uso. Essa parcela de estudantes se vê prejudicada devido à falta de investimento no aprimoramento de dicionários para aprendizes em níveis A2 e B1 de aprendizagem.

5.2 Dicionários de aprendizes de espanhol e o público brasileiro

Nesta seção, apresentam-se os resultados de uma avaliação de paráfrases explanatórias e notas de uso do dicionário Señas (2001) sob a perspectiva do consulente brasileiro. O objetivo foi medir o nível de compreensão leitora necessário para compreender as paráfrases explanatórias e notas de uso de uma amostra de verbetes do dicionário. A metodologia de análise será um pouco diferente da que se empregou nos dicionários de aprendizes de inglês porque o espanhol não dispõe da mesma variedade de recursos. O que há em termos de suporte acerca dos recursos linguísticos atrelados os níveis de referência do CEFR (2001) é o PCIC (2006). Conforme já comentado anteriormente, o documento sugere recursos linguísticos, arrolando-os em inventários léxicos e gramaticais e estabelecendo uma rede de referências cruzadas entre tais inventários e as funções comunicativas. Algumas das limitações do PCIC (2006) ficaram mais claras durante a análise dos definidores, conforme está detalhado mais adiante.

Como não há dicionários de aprendizes de espanhol direcionados a níveis iniciais, seja de proficiência, seja de aprendizagem, optou-se por escolher o dicionário Señas (2001), por sua tentativa de se aproximar do mesmo perfil de usuário com o qual esta tese trabalha – estudantes brasileiros de espanhol com LE. Algumas ponderações devem ser feitas em relação à obra. A primeira delas é a de que o período de compilação do dicionário é anterior à publicação do CEFR (2001) e do PCIC (2006). Em vista disso, não é possível esperar que o Señas (2001) reflita a escala de níveis e a concepção comunicativa de ensino de línguas. Conforme já comentado em Borba (2019b, p. 224), a anterioridade em relação a documentos de referência de ensino comunicativo de LEs não é exclusiva do Señas (2001); os demais dicionários de aprendizes de espanhol como LE, mais especificamente o *Diccionario Salamanca de la Lengua*

Española (DiSal, 2006) e o *Diccionario de la Lengua Española para Estudiantes de Español* (Espasa-Calpe, 2002) também são anteriores ao PCIC (2006)¹¹⁵.

A segunda ponderação que se faz necessária diz respeito aos aspectos que já foram alvo de avaliação no Señas (2001) e à diferença em relação à avaliação empreendida aqui. Bugueño Miranda (2006), por exemplo, analisou se as informações do Señas (2001) eram discretas e discriminantes e ainda a sua utilidade em tarefas de recepção e produção. Moreira (2017) analisou o quanto o Señas (2001) e outro dicionário de língua espanhola incluíam informações enciclopédicas, as quais, segundo o autor, seriam necessárias para que as obras pudessem ser utilizadas por estudantes de turismo. Teer (2018) analisa informações atreladas a fraseologias incluídas no Señas (2001) e no DiSal (2006), tais como marcas de uso e variantes. Borba (2019b) analisou os tipos de informações incluídas no Señas (2001), no DiSal (2006) e no Espasa-Calpe (2002) e a sua utilidade para os estudantes brasileiros de espanhol, bem como o reflexo da língua espanhola em uso. Portanto, uma correlação entre o conteúdo do Señas (2001) e os insumos do PCIC (2006), até onde sabemos, ainda não foi realizada.

Os critérios de seleção empregados foram os seguintes:

- (1) Distribuição entre os níveis A2 e B1. Analisaram-se paráfrases explanatórias e notas de uso referentes a ULs que, segundo o PCIC (2006), enquadravam-se nos referidos níveis de referência. A justificativa, assim como no caso anterior, foi a de que tais ULs poderiam ser alvo de dúvidas em estudantes de espanhol em níveis A2 e B1 de aprendizagem.
- (2) Delimitação da amostra a 30 ULs (uma única forma associada a um único conteúdo), distribuídas de maneira equivalente entre os níveis A2 e B1. Assim como no caso dos dicionários de aprendizes de inglês, o tempo de que se dispunha para analisar o Señas (2001) determinou o volume da amostra de dados analisados. Outro fator que foi determinante para a escolha
- (3) Distribuição entre distintas classes de palavras. Foi atribuída especial atenção às classes de palavras de tipo aberto, isto é, substantivos, verbos e adjetivos, pois são os que mais aparecem nos inventários léxicos do PCIC (2006) *Nociones generales* e *Nociones específicas*. Alguns poucos advérbios, que apareciam em menor escala no PCIC (2006)

¹¹⁵ Outra proposta de dicionário de aprendizes de espanhol é o *Diccionario de Aprendizaje del Español como lengua extranjera* (DAELE, 2010). A obra, no entanto, não foi finalizada, pois está constituída de 363 verbetes, todos atinentes a verbos. Segundo as informações de apresentação do dicionário, trata-se de uma obra concebida para o suporte eletrônico online e que, no momento, conta com apenas uma amostra disponível ao público. Em função dessas limitações, a obra não poderia ser utilizada nas avaliações relatadas neste capítulo.

se comparados às outras três classes de palavras mencionadas, também foram incluídos. No total, para cada nível de referência havia 16 itens relacionados (15 paráfrases explanatórias e uma nota de uso), somando 32 itens no total.

- (4) ULs não cognatas (ver detalhes nos próximos parágrafos) e que obedeciam a uma de duas condições: a) ULs cuja representação ortográfica tinha pouca ou nenhuma similaridade ortográfica com o português; b) ULs cuja representação ortográfica é similar à de ULs do português, mas seu significado é divergente entre as duas línguas.

As duas noções chave que fundamentam o quarto critério de seleção da amostra são *cognatos* e *falsos amigos*. O conceito de *cognato* a que nos referimos é aquele adotado pela Psicolinguística, isto é, convergência formal e semântica entre elementos de duas línguas, tal como já comentado no capítulo 4 (DE GROOT, 2013, p. 73) – ao invés do conceito adotado pela Linguística, que requer análise etimológica para chancelar duas palavras como cognatas. O conceito psicolinguístico do termo foi mais proveitoso para a tese porque retrata melhor o movimento que o aprendiz realiza ao processar a LE durante a leitura de um verbete, por exemplo. Nesse sentido, o aprendiz de LE julga que uma palavra da LE é similar a uma palavra da sua LM porque percebe algum grau de similaridade ortográfica entre ambas, e não porque detecta uma raiz etimológica comum. Além disso, em termos de processamento, palavras cognatas constituem um contexto mais propício para a ocorrência de influências interlinguísticas facilitadoras, isto é, que dificilmente ofereceriam dificuldades de compreensão aos aprendizes brasileiros. Para a seleção da amostra, portanto, ULs com essa característica foram evitadas.

Intrinsecamente relacionado ao conceito psicolinguístico de *cognato* está o conceito de *falsos amigos*. O termo será empregado em referência à similaridade formal, acompanhada de uma divergência semântica entre duas ULs. Conforme Bugueño Miranda (2002), há uma certa variedade de designações para o fenômeno em questão, dentre os quais constam *falsos cognatos* e *heterossemânticos*. Em relação a *falsos cognatos*, o autor argumenta que, sob a perspectiva linguística de cognato, apenas poderiam ser enquadradas como falsos cognatos as ULs que tivessem uma raiz etimológica comum, excluindo casos de similaridade ortográfica que não cumprem com esse pré-requisito (BUGUEÑO MIRANDA, 2002, p. 188). No caso de *heterossemânticos*, Bugueño Miranda (2002, p. 189) esclarece, entre outros aspectos, que a designação denota uma pluralidade de significados, levando-a a se confundir com outro fenômeno linguístico – a polissemia. Frente a esse cenário, o autor considera mais adequado e

menos sujeito a problemas designativo-conceituais o uso da designação *falsos amigos*. Como esclarece o autor,

Desta forma, a aparente ajuda que significa encontrar uma palavra conhecida, “amiga” em uma outra língua, vira uma ajuda enganosa quando as significações não coincidem plenamente. Essa palavra é, então, uma “falsa amiga”. De outra parte, ainda que de caráter tão coloquial, a designação *falsos amigos* se tem convertido em uma expressão bem aceita entre os lingüistas, sem que, porém, tenha logrado afastar as demais (BUGUEÑO MIRANDA, 2002, p. 190).

Para esta oportunidade, amparamo-nos nos argumentos de Bugueño Miranda (2002) para empregar o termo *falsos amigos* em relação à similaridade ortográfica acompanhada de divergência semântica entre ULs de línguas diferentes.

Assim como já argumentado no caso das análises de romanismos no inglês, os falsos amigos podem exercer influências diferentes no processamento conforme o aprendiz tenha conhecimento prévio ou não da divergência semântica em questão. Nesse sentido, um falso amigo entre o espanhol e o português que esteja atrelado ao nível A1 – tal como o numeral *dos* [dois] do espanhol – muito provavelmente formará parte do conhecimento prévio de um indivíduo que já apresente dito nível de proficiência. No caso em questão, espera-se que haja uma influência interlingüística custosa devido à similaridade com a contração *dos* do português, mas que dificilmente levará à seleção do item equivocado durante o processamento devido ao conhecimento prévio relativo ao nível A1. Por outro lado, um falso amigo que não forma parte do conhecimento prévio do aprendiz por estar associado a um nível de proficiência superior ao que aquele apresenta, por exemplo, gerará um efeito aparentemente facilitador no processamento, mas que incorrerá em uma compreensão ou produção que diverge em relação aos usos da LE. É o que ocorreria no processamento de *almuerzo ligero*, por exemplo, no qual o adjetivo do espanhol está atrelado ao nível B1 e é 87% similar ao adjetivo *ligeiro* do português segundo a calculadora de similaridade ortográfica NIM (2013). Enquanto no espanhol o exemplo denota a leveza do almoço, um aprendiz brasileiro de nível A2 de proficiência que não esteja ciente da diferença semântica pode acabar interpretando como um almoço rápido.

A partir das distinções feitas, foi possível chegar a três tipos de similaridade, diferentes entre si conforme o grau de similaridade e conforme se trate de cognatos ou de falsos amigos.

Quadro 26 – Tipos de similaridade (espanhol)

Tipo de similaridade	Descrição	Exemplo
Pouca ou nula (1)	O índice de similaridade é baixo (0,50 ou inferior) ou não há similaridade ortográfica com palavras do português. Há pouca probabilidade de ocorrerem influências no processamento.	<i>ayer</i> [ontem]
Falsos Amigos (2)	Há semelhança ortográfica, mas divergência no plano semântico. Pode haver influências custosas ou que incorram em facilitação ilusória.	<i>dos</i> [dois] <i>ligero</i> [leve]
Cognatos (3)	Há semelhança ortográfica e semântica, com probabilidade de gerar influências verdadeiramente facilitadoras.	<i>gobierno</i> [governo]

Fonte: Elaboração da autora

Tendo em vista os quatro critérios de seleção adotados, foi selecionada a seguinte amostra de ULs:

Quadro 27 – Unidades léxicas analisadas no Señas (2001)

Classe de palavras	Nivel A2	Similaridade ortográfica	Nivel B1	Similaridade ortográfica
Substantivos	billete [dinero]	0,86	chiste [historieta]	-
	cita [reunión]	0,73	contraseña	-
	copa [bebida]	1,00	huelga	-
	dirección [ubicación]	0,60	rebaja	-
	gimnasio	0,89	regla [sangre]	0,78
	propina	1,00	tapa [alimento]	1,00
Verbos	bajar	-	enfadar [disgusto]	-
	contestar [responder]	1,00	ligar [relación]	1,00
	invitar [celebración]	-	saber [sabor]	1,00
	reparar [examinar]	0,93	---	---
Adjetivos	aburrido	-	fatal [algo muy malo]	1,00
	ancho [extensión]	-	ligero [alimento]	0,87
	mayor [edad]	0,78	mareado [malestar]	-
	---	---	roto [inútil]	-
	---	---	vago [perezoso]	1,00
Advérbios	demasiado	-	anoche	-
	todavía [vigente]	1,00	---	---

Fonte: Elaboração da autora com base nos dados do PCIC (2006). Os distinguidores semânticos entre colchetes foram gerados de forma *ad hoc* pela autora.

Não foi possível sugerir símiles ortográficos do português para *bajar*, *invitar*, *aburrido*, *ancho*, *demasiado*, *chiste*, *contraseña*, *huelga*, *rebaja*, *enfadar*, *mareado*, *roto* e *anoche*. As demais foram identificadas como falsos amigos entre o espanhol e o português. As designações

do português usadas como parâmetro para o cálculo da similaridade ortográfica podem ser consultadas no Anexo 11.

A análise das paráfrases e notas relativas às ULs da amostra foi realizada em duas etapas. Na primeira, os definidores foram classificados conforme os inventários do PCIC (2006) (cf. 5.2.1). Na segunda etapa, os definidores foram avaliados conforme os contextos propícios para a ocorrência de influências interlinguísticas entre o português e o espanhol (cf. 5.2.2). Da mesma forma que na análise dos dicionários de inglês, ao final, são gerados *rankings* de facilidade de compreensão de paráfrases explanatórias e notas de uso sob a perspectiva do perfil de usuário estabelecido na tese.

5.2.1 *Conhecimento prévio geral na compreensão leitora do Señas (2001)*

Nesta seção, apresentam-se os resultados da primeira etapa de avaliação do Señas (2001), que corresponde à averiguação do conhecimento prévio necessário que um estudante de espanhol genérico necessita para compreender as paráfrases explanatórias e notas de uso. A metodologia de análise foi a seguinte:

- (1) Averiguação do nível dos definidores com base nos dados do PCIC (2006);
- (2) Averiguação do nível de certas palavras gramaticais conforme dados do PCIC (2006);
- (3) Classificação das paráfrases e notas conforme o conhecimento prévio necessário para sua compreensão.

Alguns aspectos do PCIC (2006) chamaram a atenção durante a consulta ao documento e influenciaram na análise dos itens da amostra. Em primeiro lugar, notou-se que os inventários léxicos não incluem certas definidores para nenhum dos seis níveis de referência. Em alguns casos, são palavras que se enquadram nos temas discriminados pelo PCIC (2006) em seus inventários léxicos e cuja ausência chama a atenção; é o caso de *persona*, *cosa*, *dueño*, *mostrar*, *reír*, *señal*, entre outros. Por mais que o PCIC (2006) não almeje ser exaustivo, a ausência dessas palavras não deixa de causar estranheza. Nos casos de *persona* e *señal*, notou-se que ambos aparecem em frases de exemplo que ilustram outros definidores a partir dos níveis A2 e B2+. Sob essa ressalva, foram considerados como relativos a esses níveis. Os demais não aparecem sequer citados em frases de exemplo e permanecem sem atribuição de nível.

Também se notou a ausência de uma seção sobre as preposições no inventário gramatical do PCIC (2006). De fato, o próprio documento esclarece que “o tratamento das

preposições, na perspectiva do enfoque geral dos Níveis de referência para o espanhol, deve ser abordado a partir de um ponto de vista mais bem léxico que estritamente gramatical”¹¹⁶ (PCIC, 2006, Gramática: Introducción). Em termos práticos, isso significa que as preposições aparecem somente quando formam parte de algum tipo de combinatória léxica. Em um caso concreto, como não havia o registro da combinatória *entre sí* (Señas, 2001, s.v. *contraseña*), mas sim somente do pronome *sí*, a preposição ficou sem classificação.

Frente às limitações encontradas, os resultados da classificação conforme o nível de referência dos elementos definidores podem não refletir o real nível de facilidade de compreensão das paráfrases explanatórias e notas de uso analisadas.

Um tópico que precisa ser esclarecido diz respeito à análise dos padrões gramaticais. Boa parte das funções gramaticais já aparece por meio das palavras gramaticais nas paráfrases explanatórias e notas de uso da amostra selecionada. Outros padrões gramaticais já estão implícitos na classificação léxica, como é o caso dos verbos com dois e três argumentos, previstos a partir do nível A2 (PCIC, 2006, §12.1 A2). A maior parte dos verbos encontrados no PCIC (2006) está atrelada ao nível A2 em diante.

Os resultados da análise estão sintetizados a seguir. A transcrição dos verbetes e a classificação dos definidores podem ser visualizados no Anexo 12.

Quadro 28 – Facilidade de compreensão leitora: definidores e padrões gramaticais (espanhol)

Classificação dos elementos definidores	Total	Verbetes
Un. léxicas A2 definidas por elementos A1 somente	-	-
Un. léxicas A2 definidas por elementos A1, A2	1/16 (6,25%)	<i>bajar</i> (ac.1)
Un. léxicas A2 definidas por elementos A1, A2, B1+	15/16 (93,75%)	<i>billete</i> (ac.1), <i>cita</i> (ac.1), <i>copa</i> (ac.2), <i>dirección</i> (ac.3), <i>gimnasio</i> , <i>propina</i> , <i>contestar</i> (ac.1), <i>invitar</i> (ac.1), <i>repasar</i> (ac.2), <i>aburrido</i> (ac.1), <i>aburrido</i> (n.), <i>ancho</i> (ac.1), <i>mayor</i> (ac.1), <i>demasiado</i> (ac.1), <i>todavía</i> (ac.1)
Un. léxicas B1 definidas por elementos A1, A2	1/16 (6,25%)	<i>anoche</i>
Un. léxicas B1 definidas por elementos A1, A2, B1	5/16 (31,25%)	<i>chiste</i> (ac.1), <i>regla</i> (ac.5), <i>saber</i> (ac.4), <i>roto</i> (ac.2), <i>roto</i> (n.)
Un. léxicas B1 definidas por elementos A1, A2, B1, B2+	10/16 (62,50%)	<i>contraseña</i> , <i>huelga</i> , <i>rebaja</i> , <i>tapa</i> (ac.2), <i>enfadar</i> (ac.1), <i>ligar</i> (ac.6), <i>fatal</i> (ac.1), <i>ligero</i> (ac.3), <i>mareado</i> , <i>vago</i> (ac.1)
Dados gerais		
Elementos definidores um nível abaixo da unidade léxica definida	1/32 (3,12%)	<i>anoche</i>

¹¹⁶ [el tratamiento de las preposiciones, en la perspectiva del enfoque general de los Niveles de referencia para el español, debe ser abordado desde un punto de vista léxico más que estrictamente gramatical]

Elementos definidores de mesmo nível da unidade léxica definida	6/32 (18,75%)	<i>bajar</i> (ac.1), <i>chiste</i> (ac.1), <i>regla</i> (ac.5), <i>saber</i> (ac.4), <i>roto</i> (ac.2), <i>roto</i> (n.)
Elementos definidores de um ou mais nível(is) acima da unidade léxica definida	25/32 (78,12%)	<i>billete</i> (ac.1), <i>cita</i> (ac.1), <i>copa</i> (ac.2), <i>dirección</i> (ac.3), <i>gimnasio</i> , <i>propina</i> , <i>contestar</i> (ac.1), <i>invitar</i> (ac.1), <i>repasar</i> (ac.2), <i>aburrido</i> (ac.1), <i>aburrido</i> (n.), <i>ancho</i> (ac.1), <i>mayor</i> (ac.1), <i>demasiado</i> (ac.1), <i>todavía</i> (ac.1), <i>contraseña</i> , <i>huelga</i> , <i>rebaja</i> , <i>tapa</i> (ac.2), <i>enfadar</i> (ac.1), <i>ligar</i> (ac.6), <i>fatal</i> (ac.1), <i>ligero</i> (ac.3), <i>mareado</i> , <i>vago</i> (ac.1)

Fonte: Elaboração da autora com base nos dados do Señas (2001) e do PCIC (2006)

A análise demonstrou algo que já se esperava, isto é, que, sob a perspectiva dos níveis de referência, o Señas (2001) não é um dicionário para aprendizes com proficiência A1 e A2, pois não apresentam conhecimento prévio suficiente para compreender as paráfrases e notas. Conforme comentado anteriormente, o espanhol não conta com dicionários para etapas iniciais de aprendizagem ou proficiência. No caso específico da escala do CEFR (2001), o período de compilação do Señas (2001) foi anterior à publicação do CEFR (2001), razão pela qual o único que se pode fazer é analisar o Señas (2001) à luz do CEFR (2001) para estabelecer um paralelo entre ambos, mas sem exigir da obra lexicográfica uma correspondência com os atuais documentos de referência do ensino comunicativo. Tal paralelo é importante pelo fato de que o Señas (2001) não somente está ao alcance dos estudantes brasileiros de ELE como também é popular entre esse público.

Os resultados obtidos revelam que, considerando a escala de níveis de referência do CEFR (2001) e a consulta por um aprendiz de espanhol como LE genérico (*i.e.* sem que se especifique a sua LM), o Señas (2001) é uma obra que pode ser mais bem aproveitada a partir do nível B2 de proficiência. Embora seja uma obra pretensamente voltada ao público brasileiro, ainda é válido analisar as paráfrases explanatórias e notas de uso do Señas (2001) sob a perspectiva de um aprendiz genérico de espanhol em função da conjuntura em que se deu a elaboração da obra. O Señas (2001) não é um projeto totalmente original, mas sim uma adaptação de um dicionário de aprendizes de espanhol voltado para um público genérico, o *Diccionario Vox para la enseñanza de la lengua española* (DiELE, 1995). Os organizadores do Señas (2001) esclarecem essa questão na seguinte nota: “Esta obra incorpora na íntegra o dicionário publicado originalmente em espanhol com o título DICCIONARIO VOX PARA LA ENSEÑANZA DE LA LENGUA ESPAÑOLA” (Señas, 2001, p. IV. Maiúsculas no original). Isso significa que as paráfrases explanatórias e notas de uso não foram reelaboradas para o público brasileiro, mas sim mantidas tal como no DiELE (1995). A adaptação do DiELE (1995) para o consulente brasileiro consiste no acréscimo de equivalentes em português ao final de cada acepção, gerando um dicionário também conhecido como semibílingue.

O potencial de auxílio do Señas (2001) em relação aos brasileiros é limitado porque as necessidades desse público são resumidas ao nível léxico e respondidas com a inclusão de equivalentes, sem que se considerem as necessidades dos brasileiros relativas a informações de outros níveis de organização da linguagem, em especial o nível sintático. Bugueño Miranda (2006, p. 221-224), por exemplo, critica o baixo potencial de auxílio do Señas (2001) no que tange à indicação da valência verbal, o que termina gerando problemas ao se usar o dicionário para fins de produção. O autor ainda observa que o auxílio do Señas (2001) ao aprendiz brasileiro se restringe às tarefas de compreensão, embora observe que há paráfrases explanatórias pouco claras ou que poderiam ser redigidas de uma maneira mais simples (BUGUEÑO MIRANDA, 2006, p. 220-221).

O Señas (2001) também falha em refletir a língua espanhola em uso, conforme Borba (2019b). Além do inchaço macroestrutural, isto é, da inclusão de ULs pouco relevantes para o consulente brasileiro devido à sua frequência baixa ou nula, a análise feita para esta tese revelou que a falta de reflexo da língua em uso se manifesta de outras maneiras também. Por um lado, os problemas na descrição da língua em uso se manifestam no verbete *saber*, no qual chama a atenção a ausência da indicação da preposição *a*, que rege o verbo quando este significa “tener sabor”. Por outro lado, verificou-se que o dicionário, às vezes, emprega estruturas linguísticas não preferenciais na redação das suas paráfrases explanatórias. É o caso do verbete *vago*, em que a paráfrase explanatória “Que no gusta del trabajo” (Señas, 2001, s.v. *vago*, 1) emprega o padrão morfossintático *gustar de*, que, segundo a versão Manual da *Nueva Gramática de la lengua española* (MGramLE, 2010, §33.2.2a), é um registro formal do verbo *gustar*; além disso, segundo o Corpes (2020), esse padrão é bem menos usado em espanhol que *gustar[le] [algo] [a alguien]*¹¹⁷. Usar tal estrutura em um dicionário de aprendizes é uma estratégia pouco didática porque pode passar uma falsa percepção de que se trata de um padrão de uso corriqueiro.

A próxima seção analisa a amostra de verbetes do Señas (2001) sob a perspectiva do aprendiz brasileiro de espanhol com proficiência A1-A2 e das influências interlinguísticas da LM no processamento da LE.

¹¹⁷ No Corpes (2020), há um total de 93.939 registros do verbo *gustar*; destes, menos de 2.518 registros são do padrão *gustar de [algo]*. Não foi possível isolar as ocorrências desse padrão e chegar a um número preciso por via automática. Uma consulta manual fugia aos objetivos de nossa argumentação.

5.2.2 *Influências interlinguísticas na compreensão leitora do Señas (2001)*

O objetivo desta subseção é averiguar que definidores são mais propensos a influências interlinguísticas do português no espanhol e qual é a natureza dessas influências em termos de facilitação verdadeira, facilitação ilusória e custo ao processamento. O fator que se leva em consideração é a similaridade ortográfica dos definidores do espanhol em relação à grafia de ULs do português, amparado na nas justificativas teóricas já comentadas. O critério de seleção dos definidores a serem analisados, no entanto, foi diferente em relação ao inglês. Nesta oportunidade, optamos por analisar o nível de similaridade ortográfica de todos os definidores ao invés de selecionar algum grupo específico. A razão para essa escolha metodológica reside na proximidade tipológica entre o espanhol e o português e na percepção subjetiva de que são línguas com uma considerável convergência formal – maior que no par inglês-português, por exemplo.

Tendo em vista os três tipos de similaridade entre o espanhol e o português (pouca ou nula, falsos amigos e cognatos), a metodologia de análise das paráfrases e notas foi a seguinte:

- (1) Sugerir, quando possível, um candidato ortograficamente similar do português a cada um dos definidores do espanhol presentes na amostra de verbetes.
- (2) Calcular o nível de similaridade ortográfica baseado na fórmula de Van Orden (1987) e disponibilizado pela ferramenta NIM (2013). A calculadora é sensível a acentos (´, ^) e diacríticos (~ do português, ¨ do espanhol), de modo que a sua inclusão resultaria em um baixo índice de similaridade ortográfica. Acreditamos que esses elementos influenciam pouco ou nada na percepção dos aprendizes brasileiros A1-A2 no que se refere ao que é similar ou não entre o espanhol e o português. Por essa razão, para os cálculos, os acentos e diacríticos foram omitidos. No que diz respeito às letras diferenciais, o ñ do espanhol foi mantido nos cálculos, pois forma parte dos conhecimentos de nível A1 em espanhol conforme o PCIC (2006). A letra ç do português, por outro lado, foi substituída por c, pois, como a grafia do português não é o ponto de partida do cálculo, a inclusão do cedilha poderia distorcer os resultados desnecessariamente.
- (3) Classificar os definidores conforme o tipo de similaridade (pouca ou nula similaridade, falsos amigos e cognatos).
- (4) Identificar definidores que causam dificuldade de compreensão no consulente brasileiro de proficiência A1-A2.

A predominância de cada categoria de similaridade e sua relação com os níveis de referência do CEFR (2001) presentes no PCIC (2006), está discriminada a seguir. Uma versão detalhada dos dados está disponível no Anexo 13.

Tabela 2 – Tipos de similaridade (espanhol)

Tipo de similaridade / Nível CEFR do definidor	Similaridade baixa ou nula	Falsos amigos	Cognatos
Todos os níveis	32/191 (16,75%)	12/191 (6,28%)	147/191 (76,96%)
A1	17	4	40
A2	4	4	25
B1	4	1	23
B2+	5	2	34
Ausente do PCIC (2006)	2	1	25

Fonte: Elaboração da autora com base nos dados do Señas (2001), do PCIC (2006) e da calculadora NIM (2013)

Os resultados apontam para um alto potencial de facilitação do processamento de diversos definidores, pois a maior parte deles são cognatos entre o espanhol e o português, isto é, favorecem a ocorrência de influências interlinguísticas facilitadoras. Ainda que um cognato esteja atrelado a um nível de referência superior à proficiência do aprendiz brasileiro (B1 em relação a A2, por exemplo), considerou-se que a sua natureza cognata “anula” o efeito negativo do nível e facilita o processamento. Como resultado, muitos dos definidores classificados de maneira negativa em função do seu nível de referência passaram a ser considerados como facilitadores do processamento. Dos definidores com pouca ou nenhuma similaridade com o português, aproximadamente a metade é de nível A1, isto é, correspondem ao conhecimento prévio do perfil de usuário considerado nesta tese; presume-se, portanto, que não gerariam custos no processamento. Finalmente, a parcela de falsos amigos foi consideravelmente baixa, o que beneficia o consultante brasileiro.

Ao se correlacionar os dados acerca do nível de referência dos definidores segundo o PCIC (2006) com os tipos de influências interlinguísticas que cada definidor pode suscitar, foi possível identificar que definidores podem causar maiores dificuldades aos brasileiros com proficiência A1-A2. Trata-se daqueles cujo nível é equivalente ou superior à da UL que definem e que, concomitantemente, são falsos amigos ou têm baixa similaridade com o português. Também estão incluídos na lista de substituição os casos omissos do PCIC (2006) que não guardam uma relação de similaridade ortográfica e semântica com ULs do português. Não se trata necessariamente de ULs difíceis de compreender. Porém, consideramos que a falta de uma

correlação com um nível de referência torna arriscado o seu uso em um dicionário de aprendizes, e, por isso, precisam ser evitados quando possível.

Quadro 29 – Definidores difíceis de compreender para os brasileiros com proficiência A1-A2

Similaridade ortográfica	Definidor	Nível do definidor	UL definida	Nível da UL definida
Baixa ou nula	cantidad	B1	billete	A2
	cantidad	B1	rebaja	B1
	población	B1	dirección	A2
	buen	B1	propina	A2
	lo (pron. átono)	A2	contestar	A2
	dice (verbo <i>decir</i>)	A2	contestar	A2
	u (variante de <i>o</i>)	A2	invitar	A2
	aburre (verbo <i>aburrir</i>)	B2+	aburrido	A2
	lo (sustantivador)	B2+	ancho, demasiado	A2
	obrero	B2+	huelga	B1
	hembra	B1	regla	B1
	enfado	B2+	enfadar	B1
	fijación	-	cita	A2
	reír	-	chiste	B1
	señal	(B2+)	contraseña	B1
Falsos amigos	esté (verbo <i>estar</i>)	B2+	invitar	A2
	mirar	A2	reparar	A2
	hasta	A2	todavía	A2
	ligero	B1	tapa	B1
	persona(s)	(A2)	cita, dirección, invitar	A2
	aparatos	-	gimnasio	A2
	nave	-	mareado	B1

Fonte: Elaboração da autora com base nos dados do Señas (2001), do PCIC (2006) e da calculadora NIM (2013)

A partir da identificação desses definidores, estabeleceram-se quais itens da amostra realmente são difíceis de compreender sob a perspectiva dos brasileiros de proficiência A1 e A2. Tais definidores são ainda o alvo das modificações propostas no capítulo 6 no que se refere à adaptação das paráfrases da amostra ao público em questão.

5.2.3 Facilidade de compreensão leitora do Señas (2001)

Após análises anteriores, revisou-se o *ranking* inicial de facilidade de compreensão conforme a propensão a influências interlinguísticas que o aprendiz brasileiro apresenta. Para que uma paráfrase explanatória ou nota de uso fosse considerada mais fácil para o aprendiz

brasileiro, os definidores de nível equivalente ou superior ao nível do lema deveriam ser todos cognatos. Se ao menos um desses definidores equivalentes ou superiores ao nível do lema fosse um falso amigo ou apresentasse nível de similaridade insuficiente, a paráfrase ou nota permaneceriam com a classificação inicial.

Quadro 30 – Facilidade de compreensão sob a perspectiva do consulente brasileiro (espanhol)

Paráfrases explanatórias e notas de uso referentes a ULs A2	
<p>FÁCIL DE COMPREENDER</p> <p>Todos os definidores e padrões gramaticais são previamente conhecidos (A1). Eventuais definidores de níveis A2+ são cognatos. Eventuais falsos amigos geram poucos problemas de compreensão por serem previamente conhecidos (A1).</p>	<p>6/16 (37,50%)</p> <p><i>copa</i> (ac.2), <i>gimnasio</i>, <i>aburrido</i> (n.), <i>cita</i> (ac.1), <i>bajar</i> (ac.1), <i>mayor</i></p>
<p>LIGEIRAMENTE DIFÍCIL DE COMPREENDER</p> <p>Há definidores e/ou padrões gramaticais A2 potencialmente desconhecidos. Se há cognatos, estes não correspondem à totalidade dos definidores A2. Pode haver falsos amigos ou definidores com pouca/nula similaridade com o português; ambos geram maiores problemas de compreensão se forem A2.</p>	<p>3/16 (18,75%)</p> <p><i>contestar</i>, <i>reparar</i>, <i>todavía</i></p>
<p>DIFÍCIL DE COMPREENDER</p> <p>Há definidores e/ou padrões gramaticais B1+ potencialmente desconhecidos. Se há cognatos, estes não correspondem à totalidade dos definidores A2+. Pode haver falsos amigos ou definidores com pouca/nula similaridade com o português; ambos geram maiores problemas de compreensão se forem A2+.</p>	<p>7/16 (43,75%)</p> <p><i>billete</i> (ac.1), <i>dirección</i> (ac.3), <i>propina</i>, <i>invitar</i> (ac.1), <i>aburrido</i> (ac.1), <i>ancho</i> (ac.1), <i>demasiado</i> (ac.1)</p>
Paráfrases explanatórias e notas de uso referentes a ULs B1	
<p>FÁCIL DE COMPREENDER</p> <p>Todos os definidores e padrões gramaticais são previamente conhecidos (A1, A2). Eventuais definidores de níveis B1+ são cognatos. Eventuais falsos amigos geram poucos problemas de compreensão por serem previamente conhecidos (A1, A2).</p>	<p>11/16 (68,75%)</p> <p><i>anoche</i>, <i>chiste</i> (ac.1), <i>contraseña</i>, <i>ligar</i> (ac.6), <i>saber</i> (ac.4), <i>fatal</i> (ac.1), <i>ligero</i> (ac.3), <i>mareado</i>, <i>roto</i> (ac.2), <i>roto</i> (n.), <i>vago</i> (ac.1)</p>
<p>LIGEIRAMENTE DIFÍCIL DE COMPREENDER</p> <p>Há definidores e/ou padrões gramaticais B1 potencialmente desconhecidos. Se há cognatos, estes não correspondem à totalidade dos definidores B1. Pode haver falsos amigos ou definidores com pouca/nula similaridade com o português; ambos geram maiores problemas de compreensão se forem B1.</p>	<p>3/16 (18,75%)</p> <p><i>regla</i> (ac.5), <i>rebaja</i>, <i>tapa</i> (ac.2)</p>
<p>DIFÍCIL DE COMPREENDER</p> <p>Há definidores e/ou padrões gramaticais B2+ potencialmente desconhecidos. Se há cognatos, estes não correspondem à totalidade dos definidores B1+. Pode haver falsos amigos ou definidores com pouca/nula similaridade com o português; ambos geram maiores problemas de compreensão se forem B1+.</p>	<p>2/16 (12,5%)</p> <p><i>huelga</i>, <i>enfadar</i> (ac.1)</p>
Dados gerais	
FÁCIL DE COMPREENDER	17/32 (53,12%)
LIGEIRAMENTE DIFÍCIL DE COMPREENDER	6/32 (18,75%)
DIFÍCIL DE COMPREENDER	9/32 (28,12%)

Fonte: Elaboração da autora com base nos dados do Señas (2001), do PCIC (2006) e do NIM (2013)

Há uma sensível diferença em relação aos resultados da análise léxica, pois boa parte dos itens recebeu uma nova classificação em virtude da massiva presença de cognatos. Tal como argumentado anteriormente, os cognatos estimulam a ocorrência de influências

interlinguísticas facilitadoras do processamento e permitem que as paráfrases explanatórias e notas de uso onde estão presentes sejam mais facilmente compreendidas por parte do consulente brasileiro.

Ainda assim, nota-se que o aprendiz brasileiro de espanhol com nível A1 de proficiência enfrentará dificuldades para compreender as paráfrases explanatórias e notas de uso, pois somente a terça parte será inteiramente fácil de compreender. Em dois casos, s.v. *ancho*, acepção 1, e s.v. *demasiado*, acepção 1, a única razão para a sua classificação como difíceis de compreender foi o baixo índice de similaridade (0,45) entre o artigo neutro *lo* em função substantivadora no espanhol e o artigo masculino *o* do português. Questionamos se, nesse caso em específico, o aprendiz brasileiro realmente não poderia associar *lo* com *o* em “[...] mayor de lo normal” (Señas, 2001, s.v. *ancho*) e em “Superando lo justo o lo necessário [...]” (Señas, 2001, s.v. *demasiado*).

No que concerne aos aprendizes de nível A2 de proficiência, o cenário já é bem diferente: enquanto dois terços dos itens analisados são de fácil compreensão, apenas um terço apresenta algum grau de dificuldade de compreensão.

Uma proposta para usar o Señas (2001) como base para compilar um dicionário de aprendizes brasileiros de espanhol em nível A2 e B1 de aprendizagem precisa explorar formas de facilitar a compreensão das paráfrases explanatórias por parte desse público. Mais precisamente, trata-se de avaliar em que medida é possível substituir falsos amigos e elementos de nível equivalente ou superior ao do lema por ULs cognatas ou classificadas como de nível A1 e/ou A2 pelo PCIC (2006) conforme o nível da UL de que se trate. Uma primeira tentativa de aplicação dessa proposta será apresentada nos próximos capítulos.

6 O PERFIL DE USUÁRIO APLICADO AOS DICIONÁRIOS DE APRENDIZES (I)

Este capítulo e o próximo correlacionam os pressupostos teóricos comentados anteriormente com os resultados das análises do capítulo 5 para propor estratégias de descrição metalexiconográfica baseadas nas características do perfil de usuário. Devido à considerável extensão que os temas tratados foram adquirindo, a divisão em dois capítulos se fez necessária para fins de organização. O presente capítulo faz considerações a respeito dos componentes canônicos do dicionário de aprendizes e propõe modificações no desenho do comentário de forma e na apresentação das paráfrases explanatórias. O capítulo seguinte, por sua vez, continua no âmbito do comentário semântico e se dedica às notas de uso.

Conforme já comentado no capítulo 2, Bugueño Miranda (2019d, p. 12) identifica três parâmetros básicos nos quais o desenho de um dicionário deveria estar fundamentado: classificação, função e perfil de usuário. A classificação diz respeito ao lugar que um dicionário ocupa em uma tradição lexicográfica. Esse tipo de expediente, em especial o de tipo taxonômico, ajuda no estabelecimento de uma série de características próprias de cada classe de obras e que ajudam a diferenciá-la das demais. No caso da nossa proposta, e na esteira de Borba e Bugueño Miranda (2019a), um dicionário de aprendizes é uma obra de referência que almeja servir de auxílio no ensino-aprendizagem de línguas estrangeiras. Em função disso, o dicionário de aprendizes precisa se adequar às características do contexto em questão e responder às necessidades que dele advêm.

A função se refere aos tipos de tarefas linguísticas para as quais o dicionário almeja servir de auxílio. O dicionário de aprendizes cujo desenho almejamos aprimorar contempla as funções de recepção e produção linguística. A ênfase informativa das duas funções recai majoritariamente sobre a dimensão do significado, isto é, a sua compreensão e produção. É por essa razão que as principais propostas de aprimoramento que fazemos no momento dizem respeito às paráfrases explanatórias e notas de uso, ou seja, segmentos informativos atrelados à dimensão do significado. Os dicionários de aprendizes também possuem ênfase informativa sobre o significante na medida em que se fornecem as representações ortográfica e fonológica das unidades léxicas. Porém, como a dimensão do significante não constitui o foco dos dicionários de aprendizes, as propostas em relação a ela são de menor envergadura.

O perfil de usuário, por sua vez, precisa ser o ponto de partida para o estabelecimento de critérios de seleção de informações. Nosso usuário, segundo comentado em 2.4, é definido como alguém residente no Brasil, adultos, que fala português brasileiro, com Ensino Médio completo, estudante de inglês ou de espanhol como língua estrangeira, que atingiu os níveis

de proficiência A1 ou A2 nessa língua e que não está habituado a usar um dicionário. Para que um dicionário de aprendizes satisfaça as necessidades desse perfil de usuário, seu desenho deve se pautar no fornecimento de informações de qualidade, que sejam pedagógicas e que estejam apresentadas de maneira didática. As informações de qualidade correspondem às informações discretas, isto é, acerca da língua em uso. O viés pedagógico corresponde à relevância das informações para o usuário e está intrinsecamente relacionado a um plano de ensino (identificação de etapas de aprendizagem) e às necessidades próprias de grupos específicos de usuários (dúvidas e erros, por exemplo). Já o didatismo ou *user-friendliness* se refere facilidade com que o usuário poderá identificar e compreender as informações apresentadas. As três variáveis devem guiar o desenho dos três grandes componentes canônicos de um dicionário de aprendizes: macroestrutura, microestrutura e medioestrutura.

6.1 Macroestrutura dos dicionários de aprendizes

Conforme Bugueño Miranda (2019a, p. 16), a macroestrutura diz respeito aos critérios de inclusão de lemas (palavras-entrada, geralmente destacadas em negrito) e ao número de palavras-entrada. Quanto à dimensão quantitativa, isto é, quantas palavras-entrada são necessárias para uma classe de dicionário, o autor assinala a Metalexigrafia carece de maiores investigações. Sendo assim, nossa proposta a nível macroestrutural não entrará no mérito da determinação de um número de palavras-entrada, mas sim da macroestrutura qualitativa, ou seja, o tipo de unidades léxicas que precisam constar em um dicionário de aprendizes para níveis A2 e B1 de aprendizagem. Propomos, tal como comentado ao final do capítulo 3, que um dicionário dessa classe para o público-alvo assinalado inclua todas as unidades léxicas atreladas aos níveis A1, A2 e B1 de proficiências segundo os descritores de níveis de referência do CEFRL (2001) – EVP (2015) e EGP (2015), no caso do inglês, e PCIC (2006), no caso do espanhol.

A proposta não se refere à inclusão única e exclusiva de unidades léxicas A1, A2 e B1 na macroestrutura, excluindo unidades léxicas que não estejam presentes nesses inventários, mas sim à determinação desse critério em particular. Cabe ressaltar que tampouco almejamos analisar e julgar as unidades léxicas incluídas em dicionários de aprendizes A1-B1 já existentes e que correspondem a nível B2 ou mais, ou mesmo que não estão discriminadas nos descritores de níveis de referência do inglês e do espanhol. A principal observação que fazemos em relação a estas outras unidades léxicas é a de que devem ser informações discretas, ou seja, com respaldo empírico, pois disso depende a qualidade da obra

lexicográfica. No caso dos dicionários de aprendizes de inglês CaED2 (2011), OxED2 (2012), CoED2 (2014) e LWord2 (2008), naturalmente, faz-se necessária uma revisão macroestrutural para garantir que seus lemas sejam informações discretas. Essa revisão macroestrutural é ainda mais necessária para o espanhol, pois, conforme comentado no capítulo 1, a tradição hispânica infelizmente ainda peca na inclusão de informações pouco ou nada discretas.

6.2 Medioestrutura dos dicionários de aprendizes

A medioestrutura corresponde ao sistema de remissões de um dicionário. Não é necessário propor orientações específicas para a medioestrutura dos dicionários de aprendizes porque as já disponíveis atualmente se aplicam a toda e qualquer classe de obras lexicográficas, perfil de usuário e função. De acordo com Bugueño Miranda (2019a, p. 27), esse componente canônico não costuma receber a devida atenção teórica e metodológica durante o desenho e compilação de dicionários em geral. De fato, Atkins e Rundel (2008), em seu *Oxford Guide to Practical Lexicography*, não tratam dos mecanismos de remissão dos dicionários, o que demonstra que a teoria ainda não se deu conta da importância da medioestrutura. Conseqüentemente, e conforme observa Bugueño Miranda (2019a),

os mecanismos medioestruturais empregados pela maioria das obras lexicográficas deixam a desejar, ora porque nem sempre o consultante entende que está perante uma operação medioestrutural, ora porque [...] o usuário não consegue compreender a necessidade de executar uma operação medioestrutural (BUGUEÑO MIRANDA, 2019a, p. 27).

Nota-se que as críticas do autor têm como fundamento a perspectiva do usuário, tal qual a perspectiva adotada nesta tese. Extremamente útil para resolver a questão das remissões são os princípios de composição das remissões propostos por Bugueño Miranda e Zanatta (2010):

1. uma referência medioestrutural deve levar o usuário rapidamente à informação que o dicionário deseja fornecer: A referência medioestrutural deve obedecer sempre a um único movimento;
2. uma referência medioestrutural deve ser sempre elucidativa: A referência deve ser compreendida pelo usuário tanto pela motivação do impulso como pela meta de remissão proposta;
3. uma referência medioestrutural deve ser sempre funcional: a referência medioestrutural deve acarretar um ganho para o usuário. (BUGUEÑO MIRANDA; ZANATTA, 2010, p. 85).

Novamente, o consulente é um componente central dos princípios medioestruturais dos autores. Sob a ótica das distinções empregadas no capítulo 1, a sua orientação visa tornar as remissões didáticas (princípios 1 e 2) e discriminantes (princípio 3) para o usuário do dicionário, seja ele/a quem for. A convergência de perspectivas, portanto, permitiu que adotássemos o conjunto de princípios acima às nossas propostas de aprimoramento que envolvessem remissões.

6.3 Microestrutura dos dicionários de aprendizes: comentário de forma

A microestrutura é o conjunto de informações relativas à palavra-entrada ou lema, constituindo os verbetes do dicionário (BUGUEÑO MIRANDA, 2019a, p. 24). No capítulo 1, por exemplo, as análises do CaED2 (2011), OxED2 (2012), CoED2 (2014) e LWord2 (2008) se referem à microestrutura desses dicionários.

A microestrutura se divide em dois conjuntos de informações: comentário de forma, com segmentos informativos relativos à dimensão do significante; e comentário semântico, cujos segmentos informativos são atinentes à dimensão do significado. Trata-se de uma divisão didática, pois, na prática, significante e significado estão diretamente relacionados. Embora o principal objeto das nossas propostas seja o comentário semântico dos verbetes de dicionários de aprendizes, a revisão bibliográfica do capítulo 4 nos motivou a fazer uma revisão do comentário de forma à luz de pesquisas que trataram dessa dimensão das unidades léxicas.

A propostas desta seção são atinentes ao lema, à representação fonológica e à categoria morfológica. Visto que há muitos pontos de convergência entre os dicionários de aprendizes e inglês e de espanhol, optou-se por tratá-los em conjunto, fazendo observações específicas de uma ou outra tradição quando necessário. As propostas se amparam em especial na didatização de informações; em menor medida, também se contempla a sua natureza discriminante.

6.3.1 Lema

O lema, como já comentado anteriormente, corresponde à representação ortográfica. Esse segmento informativo também é usado para indicar variantes ortográficas de ordem diafásico-diastrática (situações de uso e extratos socioculturais) ou diatópica (espaço geográfico).

No caso das variantes ortográficas que têm diferentes registros diafásico-diastráticos, é comum tomar como preferencial a forma mais usada ou a que seja considerada neutra.

No que se refere à variação nos usos da língua proveniente da sua distribuição diatópica, seja no território de um mesmo país, seja considerando diferentes países, os dicionários costumam adotar ou uma perspectiva sintópica, ou ainda uma perspectiva diatópica. Na perspectiva sintópica, descrevem-se os usos da língua de uma única variedade. É o caso, por exemplo, do CoED2 (2014) e do Señas (2001), cujos compiladores optaram por da preferência às variedades britânica e da Península Ibérica, respectivamente.

Na perspectiva diatópica, descreve-se ao menos uma outra variedade, adotando-se, porém, uma determinada variedade como referência. A variedade preferencial costuma ser aquela usada no local onde a obra lexicográfica foi compilada. Quando a forma considerada variante está registrada na macroestrutura, a diferença entre uma e outra é marcada pela explicitação da paráfrase explanatória; a forma preferencial será aquela acompanhada da paráfrase, enquanto a variante apresentará apenas uma remissão à primeira.

Dicionários como o CaED2 (2011), o OxED2 (2012) e o LWord2 (2008), compilados no Reino Unido, adotam a perspectiva britânica, mas descrevem também os usos americanos. No âmbito ortográfico, os três dicionários mencionados incluem como palavras-entrada somente as de grafia britânica. Quando uma palavra-entrada possui uma grafia americana, acrescenta-se uma marca de uso ao lado do lema, para indicar que este corresponde à grafia britânica, e, em seguida, indica-se a variante americana, acompanhada pela respectiva marca de uso americano. A variedade americana, no entanto, não é incluída na macroestrutura.

Em dicionários de aprendizes que optam por descrever uma determinada variedade diatópica além daquela de referência, é necessário considerar que o consulente pode acabar buscando informações a partir da grafia que não é a de referência. Se essa grafia não está registrada na macroestrutura, o consulente pode acabar não tendo a sua dúvida sanada. Essa situação está presente na edição impressa do CaED2 (2011), do OxED2 (2012) e do LWord2 (2008) e na edição eletrônica do último (LWord2e, 2008), pois são obras que não fornecem resultados a partir de buscas pela forma americana. A exceção a esse cenário é a edição eletrônica do OxED2e (2012), que conduz o consulente automaticamente ao verbete encabeçado pela representação ortográfica britânica mesmo nos casos em que a palavra digitada corresponda à ortografia americana.

É arriscado partir do pressuposto de que o consulente sabe que uma representação ortográfica não é a de referência e que há uma outra representação de referência que deve ser procurada, ainda mais em níveis iniciais de proficiência e aprendizagem. Em última instância,

a obra pode acabar não sendo tão didática quanto poderia ser. Nesse sentido, e no que diz respeito ao suporte impresso, propomos que a representação ortográfica tratada como variante seja incluída na macroestrutura acompanhada de uma marca de uso e de uma remissão ao verbete relativo à variante de referência, também acompanhada de uma marca de uso. Dessa maneira, evidencia-se o impulso da remissão, de forma que o consulente entende por que está sendo conduzido a outro verbete. Para o suporte eletrônico, em que a questão do espaço não influencia tanto o tratamento lexicográfico, há uma maior variedade de opções. Uma delas é a inclusão da paráfrase nos dois verbetes, sem remissão e com a indicação da variante e sua marca de uso. Outra opção é seguir o modelo dos dicionários impressos, incluindo a variante preferencial na macroestrutura acompanhada de sua marca de uso e, ao lado, a remissão à forma preferencial com um *hiperlink* embutido nela e a sua marca de uso ao lado, de forma que basta um clique para que o consulente seja redirecionado à meta da remissão. É possível ainda que a remissão seja automática, tal como adotado pela edição eletrônica do OxED2e (2012). Na obra, ao digitar *flavor*, o dicionário abre o verbete *flavour*, que contém a indicação da forma variante. As decisões sobre qual metodologia adotar no formato eletrônico dependerão das possibilidades oferecidas pelo *software* disponível. O Quadro a seguir ilustra a proposta.

Quadro 31 – Propostas ao nível do lema

Verbete original	Versão adaptada
flavour UK (US flavour) CaED2 (2011)	flavor US → flavour UK flavour UK (US flavour)
flavour ↔ (British) (American flavor) OxED2 (2012)	flavor (American) → flavour (British) flavour ↔ (British) (American flavor)
fla·vour ¹ BrE, flavour AmE LWord2 (2008)	fla·vour ¹ BrE, flavour AmE fla·vor AmE → fla·vour ¹ BrE, fla·vour ² BrE
fol·clor folklor <i>m.</i> fol·clo·re folklore <i>m.</i> → <i>folclor</i> . ☒ La Real Academia Española prefiere la forma <i>folclor</i> . fol·klo·re folklore <i>m.</i> → <i>folclor</i> . ☒ La Real Academia Española prefiere la forma <i>folclor</i> .	fol·clo·re folklore <i>m.</i> fol·klo·re folklore <i>m.</i> → <i>folclor</i> .

Fonte: Señas (2001) e proposta da autora baseada em dados do CaED2 (2011), OxED2 (2012) e LWord2 (2008).

Os verbetes do inglês incluem a variedade americana na macroestrutura, juntamente com uma remissão ao verbete encabeçado pela variedade britânica e ainda uma marca de uso junto à indicação a forma de destino da remissão para auxiliar o consulente a entender o motivo da remissão e diferença entre uma e outra forma.

Os verbetes *folclor*, *folclore* e *folklore* do Señas (2001), por sua vez, são um exemplo de algo relativamente comum nos dicionários que não são compilados pela Real Academia Española (RAE). Quando se trata de norma ortográfica, é comum que os dicionários que não foram compilados pela RAE – como é o caso do Señas (2001) – indiquem a forma ortográfica tida como preferencial pela instituição, seja para concordar com a orientação da RAE, seja para discordar dela. Nos verbetes reproduzidos acima, as notas de uso s.v. *folclore* e *folklore* salientam que o registro ortográfico preferencial segundo a RAE é *folclor*. O Señas (2001), por sua vez, segue a orientação da RAE, pois inclui a paráfrase explanatória somente s.v. *folclor* e remete as outras duas variantes a esse verbete sob a justificativa constante nas notas de uso de cada uma.

Essa descrição reflete mais o uso de meados dos anos 1990, já que o Señas (2001) reproduz a descrição do espanhol presente no DiELE (1995), seu antecessor. Contudo, atualmente, a orientação da RAE é outra: o DRAEe (2014-2021), que é a principal fonte de orientação a nível ortográfico da RAE, registra o significado sob a forma *folclore*, indicando que esta é a preferencial. A mesma orientação está presente no *Libro de Estilo de la Lengua Española* (LEstilo, 2018), outra obra de referência compilada pela RAE. Os dados do Corpes (2021), por sua vez, revelam um panorama mais complexo, o qual esboçamos a seguir.

Quadro 32 – Dados do Corpes (2021) sobre *folclor*, *folclore* e *folklore*

<i>folclor</i>	<i>folclore</i>	<i>folklore</i>
Registros: 683	Registros: 913	Registros: 953
Colômbia: 187 (27%)	Espanha: 245 (26%)	Argentina: 286 (30%)
México: 126 (18%)	Chile: 173 (19%)	Bolívia: 110 (11%)
Chile: 85 (12%)	Argentina: 139 (15%)	Espanha: 109 (11%)

Fonte: Elaborado a partir de dados do Corpes (2021)

A primeira constatação que se pode fazer é que as três formas ortográficas estão em franca competição no espanhol, com as variantes *folclore* e *folklore* muito próximas estatisticamente e *folclor* levemente mais distante em número de ocorrências. Em termos gerais, as três variantes não apresentam preferências no tocante à frequência, mas sim de ordem diatópica. Se, ao seguir o padrão do Señas (2001), fosse dada prioridade à variedade peninsular, teriam de ser registradas *folclore* como forma preferencial e *folklore* como variante conforme o número de registros de cada uma, enquanto que a forma *folclor* deveria ser retirada. Essa foi a perspectiva adotada na versão adaptada do verbete que propusemos anteriormente. Se, por outro lado, se adota uma perspectiva diatópica, as três formas deveriam

estar presentes na macroestrutura, juntamente com suas marcas de uso e indicações das formas variantes e respectivas marcas de uso. Além disso, e caso o lexicógrafo opte por associar a paráfrase explanatória a uma forma preferencial, teria de estabelecer critérios para isso. Dado que a competição entre as três formas aparentemente envolve apenas a localidade geográfica, em um dicionário em formato eletrônico de perspectiva diatópica, seria perfeitamente possível incluir a paráfrase explanatória junto às três variantes.

6.3.2 Representação fonológica

No tocante ao segmento informativo da pronúncia, as propostas dependerão do tipo de suporte da obra. No caso do suporte impresso, a transcrição fonológica segue sendo necessária e única forma viável de representação. Nossa proposta é que se faça uma adaptação para facilitar a identificação da sílaba tônica, destacando-a em negrito. Já no caso das obras em suporte eletrônico, propomos uma conjunção entre transcrição fonológica e recurso de áudio com base no referencial teórico do capítulo 4.

Quadro 33 – Propostas ao nível da representação fonológica

Verbetes original	Versão adaptada
flavour UK (US flavour) /'fleɪvəʳ/ (CaED2, 2011)	flavour UK (US flavour) /'fleɪvəʳ/
flavour ↔ (British) (American flavor) /'fleɪvə(r)/ (OxED2, 2012)	flavour ↔ (British) (American flavor) /'fleɪvə(r)/
flavour ↔ (British) (American flavor) ◀ ◀ /'fleɪvə(r)/ (OxED2e, 2012)	flavour ↔ (British) (American flavor) ◀ ◀ /'fleɪvə(r)/
fla·vour ¹ BrE, flavour AmE /'fleɪvə \$ - əɹ/ (LWord2, 2008)	fla·vour ¹ BrE, flavour AmE /'fleɪvə \$ - əɹ/
fla·vour ¹ BrE, flavour AmE UK ◀ US ◀ /'fleɪvə \$ - əɹ/ <i>noun</i> (LWord2e, 2008)	fla·vour ¹ BrE, flavour AmE UK ◀ US ◀ /'fleɪvə \$ - əɹ/ <i>noun</i>
flavour /'fleɪvə/ (CoED2, 2014)	flavour /'fleɪvə/
jer·sey xerséi (Señas, 2001)	jer·sey xerséi
lí·mi·te límite (Señas, 2001)	lí·mi·te límite ◀

Fonte: Proposta da autora baseada em dados do CaED2 (2011), OxED2 (2012) e LWord2 (2008)

Considerando que o usuário que delimitamos no início da tese é alguém cujo grau de instrução é o Ensino Médio completo, e que o alfabeto fonético não faz parte do conjunto de conhecimentos dessa etapa educacional, provavelmente terá dificuldades de compreender

alguns dos fonemas das transcrições. Em uma situação assim, a participação do professor é fundamental para que possa instruir os alunos acerca de como interpretar as transcrições. Por mais que se invista no didatismo da obra, dificilmente seria possível consultar um dicionário de aprendizes sem ter que aprender algo, como o próprio significado da marcação em negrito ou a dinâmica de distribuição de informações nos verbetes. Tal como comentado no conjunto de axiomas do capítulo 3, para que um dicionário de aprendizes possa ser bem aproveitado no ensino-aprendizagem de línguas em prol da autonomia na aprendizagem, o professor precisa orientar os alunos sobre como localizar e interpretar informações na obra (axioma 2).

Não se trata de ensinar todo o alfabeto fonético aos alunos, mas sim de tirar proveito das transcrições para apontar ao aluno uma palavra que contenha um fonema que ele não reconheça ou tenha dificuldades em articular, por exemplo. Sempre que dito fonema aparecer em uma transcrição, o aluno poderá usar o conhecimento aprendido para aprimorar a sua audição e oralidade. Nos dicionários eletrônicos, a escuta do áudio acompanhada da leitura da transcrição e de uma repetição oral também pode ser convertida em um exercício de sala de aula para tirar dúvidas e superar dificuldades.

6.3.3 Categoria morfológica

O segmento informativo da categoria morfológica corresponde à indicação da classe gramatical e à explicitação de flexões irregulares de nomes e verbos, além da indicação das formas comparativas e superlativas de adjetivos no caso específico do inglês. No tocante à terminologia, percebe-se, por um lado, um alto grau de similaridade ortográfica de certos termos em inglês e espanhol com a terminologia do português, juntamente com uma convergência conceitual, como se pode observar a seguir.

Quadro 34 – Similaridade ortográfica entre termos gramaticais

Inglês	Português	Similaridade	Espanhol	Português	Similaridade
adjective	adjetivo	0,67	sustantivo	substantivo	0,95
verb	verbo	0,73	adjetivo	adjetivo	1,00
adverb	advérbio	0,72	verbo	verbo	1,00
pronoun	pronome	0,68	adverbio	advérbio	1,00
preposition	preposição	0,65	preposición	preposição	0,73
conjunction	conjunção	0,66	conjunción	conjunção	0,68
article	artigo	0,60	artículo	artigo	0,76
plural	plural	1,00	interjección	interjeição	0,62
quantifier	quantificador	0,83	singular	singular	1,00
			plural	plural	1,00

Fonte: Elaboração da autora com o auxílio da calculadora NIM (2013).

Nota: Foram consideradas similares as palavras cujo resultado foi igual ou superior a 0,51. No cálculo, não foram considerados os acentos do português e do espanhol; e a letra ç do português foi substituída por c

Naturalmente, tanto o inglês como o espanhol têm particularidades gramaticais para além das classes de palavras e que culminam em conceitos que estão ausentes no português, como os *(un)countable nouns* e *phrasal verbs* do inglês e o *beneficiario* do espanhol. Há também casos em que o conceito existe, mas a denominação ou difere em termos de equivalentes, como *exclamation* para interjeição; ou ainda casos em que há convergência conceitual, mas falta de similaridade ortográfica, como *noun* para substantivo. O problema de se empregar uma terminologia gramatical em português esbarra na questão das particularidades conceituais do inglês e do espanhol, para as quais não há equivalente. Uma tentativa de propor termos adaptados em português para fenômenos próprios do inglês e do espanhol não pouparia o consulente de ter que estudar conceito de que se trata. Por esse motivo, a descrição da língua-objeto no nível morfológico se vê mais favorecida pelo uso de termos gramaticais na própria língua-objeto. Uma vantagem do consulente brasileiro é que, conforme demonstrado *ad supra*, há uma considerável similaridade ortográfica e também conceitual entre o inglês e o português e entre o espanhol e o português. A alta probabilidade de ocorrerem influências interlinguísticas facilitadoras como produto dessa similaridade tende a favorecer o consulente brasileiro.

Foi possível identificar ainda uma alternativa do inglês em que é possível adaptar a terminologia original de modo que se estimule uma facilitação do processamento sem recorrer a um termo em português. Trata-se do caso de flexão nominal dos *uncountable nouns*, usada pelo CoED2 (2014). Essa expressão pode ser substituída pela alternativa também válida no inglês *no plural*, tal como já o fazem o CaED2 (2011), o OxED2 (2012) e o LWord2 (2008).

Um aspecto bastante positivo dos dicionários de aprendizes de inglês relativo à flexão verbal é a inclusão de flexões irregulares e particípios ainda no segmento informativo da

categoria morfológica, seguida da indicação da representação fonológica, como *flew* e *flown* em relação a *fly*. A curta extensão do paradigma de flexão morfológica do inglês permite que se empregue o segmento informativo em questão para assinalar esses aspectos. Essa estratégia é extremamente acertada e deve ser mantida no desenho de dicionários de aprendizes de inglês.

No caso dos dicionários de espanhol, é preciso observar como as informações sobre flexão nominal estão distribuídas no verbete. No dicionário Señas (2001), por exemplo, tais informações costumam aparecer somente em notas de uso ao final do verbete. Essa não parece ser a melhor solução, pois, dessa forma, as informações morfológicas básicas ficam divididas no verbete. Faz mais sentido deixá-las ao início do verbete, junto da categoria morfológica, como ocorre com os dicionários de aprendizes de inglês analisados, e reservar as notas de uso para outros propósitos (cf. capítulo 7).

Ainda no âmbito morfológico, os dicionários de aprendizes do espanhol costumam incluir em suas páginas finais modelos de conjugação verbal. Para uma edição impressa, esta é a única maneira possível de mostrar o paradigma de flexão verbal, dada a sua considerável extensão. Os verbos do dicionário Señas (2001), por exemplo, costumam ser associados a um número limitado de modelos de conjugação, nos quais costuma-se incluir observações sobre a sílaba tônica que são extremamente discriminantes para o usuário. Para o desenho de uma edição eletrônica de um dicionário de aprendizes de espanhol, a solução encontrada pelo DRAEe (2014-2021), por exemplo, parece ser a mais acertada: a inclusão de um botão que conduz o consulente a uma tela com o paradigma de conjugação verbal. Nesse último caso, a marcação da sílaba tônica de algumas flexões também é discriminante para o consulente brasileiro.

A seguir, há um resumo das propostas acerca do comentário de forma.

Quadro 35 – Resumo das propostas ao comentário de forma

lema. (marca de uso diatópica do lema) (variante ortográfica e respectiva marca de uso diatópica)
 /transcrição fonológica/ (🔊 recurso de áudio)
categoria morfológica [flexões irregulares ou específicas; remissão a paradigmas de conjugação verbal ou botão de conjugação verbal]

Fonte: Elaboração da autora

Os ajustes propostos para o comentário de forma visam o aumento do didatismo dos segmentos informativos comentados, mas não esgotam o conjunto de informações relativas à contraparte formal das unidades léxicas. Os padrões sintáticos também constituem esse

conjunto de informações, mas, como se enquadram no âmbito das combinatórias léxicas, serão tratados no capítulo sobre o referido tema.

6.4 Microestrutura dos dicionários de aprendizes: paráfrases explanatórias

Esta e as demais seções tratam das consequências práticas do estabelecimento de um perfil de usuário específico para dicionários de aprendizes em relação ao aspecto formal das paráfrases explanatórias. Essa dimensão diz respeito ao didatismo ou *user-friendliness*, que é o que vem sendo tratado anteriormente, e ao rigor teórico-metodológico empregado na redação.

Sob a perspectiva do didatismo, o marco teórico empregado nos capítulos anteriores fornece subsídios para que se analise o grau de compreensão leitora mínimo na língua-objeto que se pode esperar do consulente brasileiro. Esse índice foi medido a partir do conhecimento prévio do público-alvo, definido conforme o nível de proficiência já atingido na LE, bem como a LM do usuário pretense.

No que se refere aos dicionários de aprendizes de inglês, os resultados obtidos no capítulo 5 demonstram que o uso da língua-objeto em função metalinguística não é a melhor opção para aprendizes brasileiros de inglês de níveis A1 e A2 de proficiência. Primeiramente, é preciso apresentar uma proficiência maior do que A1 e A2 para compreender as paráfrases analisadas. Além disso, são poucos os contextos propícios para a ocorrência de influências interlinguísticas do português no inglês que sejam de tipo facilitador. Em outras palavras, o conhecimento prévio do português não é um fator facilitador do processamento das paráfrases explanatórias e notas de uso redigidas em língua inglesa.

As análises do dicionário Señas (2001), por outro lado, revelaram que existe uma similaridade ortográfica e semântica (os chamados definidores cognatos) suficiente para que o aprendiz brasileiro que utiliza esse dicionário seja beneficiado em função de ter o português como LM. Nesse caso, é plenamente possível empregar o espanhol na redação das paráfrases, desde que sejam realizados alguns ajustes nos definidores que não podem ser compreendidos a partir do conhecimento prévio.

Além do didatismo, o rigor teórico-metodológico também constitui o aspecto formal das paráfrases explanatórias. Tal rigor diz respeito ao amparo nos avanços da teoria metalexiconográfica e é importante para que sejam redigidas paráfrases cada vez mais elucidativas, tal como advoga Bugueño Miranda (2009b, p. 248), por exemplo. Para esta oportunidade, julgou-se pertinente não somente determinar a língua mais adequada para

ajustar as paráfrases já existentes, mas também aproveitar parte do conhecimento já disponível na Metalexigrafia na elaboração das propostas de redação desse segmento informativo. Em mudanças mais significativas, esse amparo teórico foi a base que permitiu a aplicação das propostas.

Atkins e Rundell (2008, p. 450) assinalam que “as definições são bem-sucedidas quando acertam em dois aspectos: conteúdo e forma”¹¹⁸. Bugueño Miranda (2009b, p. 243; 248), por sua vez, vai mais além e aponta o caminho a ser percorrido para a elaboração de uma teoria geral da definição. Segundo o autor, para que sejam elucidativas, devem ser produto da integração de três variáveis: uma taxonomia, isto é, uma relação de tipos de paráfrases previstos para serem usados nos verbetes; padrões sintáticos, isto é, modelos de formulação para cada tipo de paráfrase; e um modelo semântico, para que se possa estabelecer que informações são relevantes. Posteriormente, e amparada na distinção feita pelo autor, Farias (2013) propõe uma teoria geral dos mecanismos explanatórios do dicionário na qual desenvolve as três variáveis em questão. Em linhas gerais, a autora realiza uma extensa revisão bibliográfica de diferentes teorias semânticas e avalia a utilidade de cada uma para a Lexicografia. No que concerne ao aspecto formal das paráfrases (taxonomia e padrões sintáticos), Farias (2013) estabelece que tipos de paráfrases elucidam melhor cada classe de palavras e propõe padrões sintáticos de redação. No que diz respeito à seleção de informações para as paráfrases, não obstante a profunda revisão bibliográfica empreendida, a autora reconhece que ainda são necessários maiores estudos para que se possa elaborar uma metodologia completa (FARIAS, 2013, p. 268). Por essa razão, não foi possível revisar as paráfrases em relação à seleção do seu conteúdo.

Conforme Farias (2013, p. 209-219), há basicamente dois tipos de paráfrases: as que realizam metalinguagem de conteúdo, ou seja, que descrevem o significado de uma UL, e as que realizam metalinguagem de signo, ou seja, que geralmente explicitam como se usa. Os padrões sintáticos, por sua vez, variam conforme o tipo de paráfrase de que se trate e ainda a classe e subclasse de palavras. Não é o objetivo comentar todo o apanhado teórico no qual Farias (2013) se baseou para propor soluções para a contraparte formal das paráfrases; a consideração de sua proposta tem um viés prático na tese, e não de discussão teórica. A seguir, elaborou-se um esquema reunindo de maneira sintética as contribuições da autora. Os exemplos são, em sua maioria, os mesmos fornecidos por ela.

¹¹⁸ [Definitions succeed when they get two things right: content and form]

Quadro 36 – Classes de palavras, tipos de paráfrases e padrões sintáticos

PARÁFRASES EM METALINGUAGEM DE CONTEÚDO		
Classe de palavra	Padrão sintático	Exemplo
Substantivos concretos	Hiperônimo (Nome concreto) + Especificador(es)	guitarra <i>f</i> instrumento musical eletrônico que possui um braço longo com seis cordas que vibram ao serem tocadas pelos dedos.
Substantivos abstratos de ação	Hiperônimo (“ação de” / “processo de” / “efeito de”) + Especificador(es)	partida <i>f</i> 1 ato de ir-se [alguém / algo]. crecimento <i>m</i> 1 processo de desenvolvimento de [alguém / algo]. 2 efeito do desenvolvimento de [alguém / algo].
Verbos intransitivos	Verbo transitivo / Verbo suporte + Complementos	morrer <i>vi</i> perder a vida.
Verbos transitivos	Verbo transitivo / Verbo suporte (+ Complementos) + [Valência do signo-lema]	chutar <i>vt</i> dar pontapés em [algo / alguém].
Adjetivos qualificativos	[Restrição de atribuição] + Oração subordinada adjetiva	supersônico <i>adj</i> 1 [velocidade] que é superior à do som. 2 [aeronave] cuja velocidade é superior à do som.
Advérbios	Locução / Oração adverbial	anteontem <i>adv</i> no dia anterior ao de ontem.
Advérbios modais	“De modo” + Adjetivo	anteontem <i>adv</i> de modo exato.
PARÁFRASES EM METALINGUAGEM DE SIGNO		
Classe de palavra	Padrão sintático	Exemplo
Substantivos concretos que designam entidades únicas ou diferentes referentes	“Nome” / “Designação” / “Expressão” + para (referir-se a) alguém / algo	Altíssimo <i>m</i> nome usado para referir-se a Deus. chiqueiro <i>m</i> nome usado para referir-se a um lugar muito sujo.
Substantivos concretos que designam uma categoria	“Nome dado a / com que se designa” / “Designação para” + Hiperônimo (Nome) + Especificador(es) + (Exponentes da categoria)	baga <i>f</i> nome dado aos frutos de pele fina, polpa carnosa e com várias sementes pequenas no seu interior, como o <i>mamão</i> , o <i>tomate</i> e a <i>uva</i> .
Adjetivos relacionais	[Restrição de atribuição] + “Relativo a” + Sintagma nominal	vocal <i>adj</i> [algo] relativo à voz humana.
Advérbios cujo conteúdo é difícil de explicitar	“Expressa” / “Usa-se para expressar” + Noção expressa	sim <i>adv</i> Expressa afirmação / consentimento. não <i>adv</i> Expressa negação / recusa.
Conjunções	“Expressa” / “Usa-se para expressar” + Noção expressa	porém <i>conj.</i> Usa-se para expressar oposição a algo recém mencionado.
Preposições como marcadores sintáticos (A)	“Marcador de” + Função sintática indicada	a <i>prep.</i> A 1 marcador del objeto indirecto: <i>Le dieron a Juan un premio.</i> 2 marcador del objeto directo de persona: <i>A Juan lo encontré en el ascensor. [...]</i>
Preposições com significação categorial (B)	Descritor: Significação categorial	a <i>prep.</i> [...] B 1 movimiento: con verbos de movimiento: <i>Voy a la ciudad. Viajé a la capital.</i> 2 posición: señala la posición de [algo / alguien] [...]
Expressões prepositivas (C)	Verbo / Advérbio + Preposição	a <i>prep.</i> [...] C <i>accesible a, acogerse a, a sabiendas de, a la buena de Dios</i>
Artigos e pronomes pessoais	Classificação morfológica do signo-lema	o <i>art</i> artigo masculino singular te <i>pron.</i> Pronome oblíquo de 2ª pessoa singular.

Pronomes (exceto pessoais)	“Usa-se para referir / designar” + Entidade designada	este <i>pron</i> Usa-se para designar algo que está próximo fisicamente de quem fala.
Interjeições	“Expressa / Exprime” + Noção expressa	bah <i>interj</i> (<i>Sul do Brasil</i>) Expressa surpresa / admiração / espanto.

Fonte: Organização da autora a partir dos dados de Farias (2013, p. 209-219), a exceção dos exemplos de paráfrases de conjunções, artigos e pronomes, que foram elaborados *ad hoc* nesta tese.

Os padrões sintáticos são explicitados por meio de fórmulas. Entre aspas estão os definidores que introduzem as paráfrases; os parênteses também especificam, quando necessário, que tipo de definidor se deve usar; os colchetes são usados para indicar a valência verbal¹¹⁹ (somente objetos do verbo) e as restrições de atribuição dos adjetivos. Para os casos de regência, Farias (2013, p. 212) sugere que a preposição esteja explicitada entre a indicação morfológica e a paráfrase explanatória, introduzida pelo diacrítico til (~) e destacada em negrito e sublinhado (Por exemplo: ~ **de**).

Além do forte embasamento teórico, a principal vantagem na adoção da proposta da autora como referência para a revisão de paráfrases explanatórias é a sua organização em fórmulas, ou seja, com um viés eminentemente prático e de fácil aplicação. Outra vantagem é a relação de complementaridade entre a sua perspectiva e a perspectiva desta tese. Farias (2013, p. 197, n. 379) esclarece que não considerou a variável do usuário ao propor sua metodologia, detendo-se no estudo e aplicação de teorias semânticas – o que, por si, já demonstrou ser mais que suficiente, já que foram necessárias 398 páginas de tese de doutorado para desenvolver a sua teoria geral a respeito dos mecanismos explanatórios. Por outro lado, a contraparte do usuário corresponde justamente ao enfoque da pesquisa apresentada aqui. Trata-se de temas que, devido à sua extensão, precisam ser estudados separadamente, mas cuja complementaridade só tem a contribuir para o desenvolvimento da (Meta)Lexicografia.

Embora bastante abrangente, não se trata de uma proposta definitiva sobre a redação das paráfrases. Foram identificados três pontos que precisam ser esclarecidos para que a proposta contemple as mais diversas nuances do léxico de uma língua.

O primeiro deles diz respeito aos substantivos abstratos em geral. Farias (2013) propôs uma fórmula apenas para os substantivos abstratos de ação. Aparentemente, a autora

¹¹⁹ De acordo com Richards e Schmidt (2010, s.v. *dependency grammar*), na gramática das valências, “[...] o verbo é considerado como a unidade central e mais importante. Os verbos são classificados conforme o número de sintagmas nominais requeridos para completar uma frase. Este número é chamado de **valência** do verbo” [the verb is considered to be the central and most important unit. Verbs are classified according to the number of noun phrases they require to complete a sentence. This number is called the **valency** of the verb] (grifo no original). A valência é um dos fenômenos considerados como combinatórias léxicas, pois envolvem os padrões de combinação de ULs, especialmente os verbos.

considera que esse tipo de UL oferece mais dificuldades no tocante à elaboração de paráfrases explanatórias (FARIAS, 2013, p. 125). Dado que as amostras de paráfrases do inglês e do espanhol continham substantivos abstratos, surgiu um impasse. Ante a ausência de orientação, optamos por não revisar o padrão sintático desses casos. Ainda assim é pertinente observar que, nos casos de substantivos abstratos da amostra (*landscape*, do inglês, e *cita, dirección, chiste e huelga*, do espanhol), o padrão de redação empregado pelos dicionários analisados recordava a versão fórmula dos substantivos concretos, porém em uma versão mais genérica: “Hiperônimo (Nome) + Especificador(es)”.

O segundo ponto que merece atenção futuramente é a regência nominal. Nota-se que apenas os exemplos de verbos contêm a indicação da regência. Nos exemplos relativos aos substantivos abstratos de ação *partida* e *crecimiento*, que também podem aparecer acompanhados por uma preposição (*de*), não há menção a essa combinação. Não está claro se a falta de indicação foi proposital ou se foi um esquecimento eventual.

O terceiro ponto que cabe mencionar diz respeito à aplicação da fórmula das preposições com significado categorial. Apesar de prever a indicação de um descritor do significado categorial, Farias (2013) não sugere uma fonte sobre a qual seja possível se amparar no momento de redigir paráfrases dessa natureza, as quais estavam presentes na amostra dos dicionários de inglês – como o caso de *among*, que será comentado na próxima seção. Uma busca por referências sobre o tema levou-nos ao QECRL (2001, p. 206), que trata da categorização pelo rótulo de “noções gerais”. Essas noções, por sua vez, foram propostas inicialmente nos documentos da *T-series* e podem ser observadas também no PCIC (2006), já que este documento foi baseado nos seus antecessores do inglês. As *T-series* e o PCIC (2006) foram os únicos documentos que apresentavam uma relação mais extensa das noções categoriais. Isso nos levou a escolher esses documentos como referência para auxiliar na redação das paráfrases de preposições de significado categorial.

Sem dúvidas, a solução que encontramos para essa questão é provisória. Ainda assim, usar documentos como o Vantage 1990 (2001) e o PCIC (2006) não é necessariamente um demérito. Nossa experiência com a análise de obras lexicográficas que se baseiam na identificação de categorias e no estabelecimento de ontologias, tal como os dicionários onomasiológicos *stricto sensu*, os dicionários de ideias afins e os dicionários pela imagem (cf. Borba (2017b)) nos permitiu constatar que o processo de categorização é e será sempre subjetivo, sendo igualmente subjetiva qualquer proposta de categorização. Nesse sentido, qualquer ontologia de categorias (ou noções gerais) gerará uma classificação que reflete a perspectiva de um indivíduo ou grupo, mas que, naturalmente, não abarcaria todas as

categorias existentes nem estaria isenta de críticas. É, em parte, o que se observa no PCIC (2006), por exemplo: mesmo usando as mesmas noções gerais que as *T-series*, a maneira como seus compiladores as distribuíram é diferente daquela das *T-series*. Mesmo quando a relação de noções gerais do Vantage 1990 (2001) e do PCIC (2006) não possa fornecer uma resposta satisfatória, ainda é um importante ponto de partida para que o lexicógrafo possa propor uma alternativa que julgue mais adequada.

Após as considerações iniciais sobre a revisão das paráfrases, as próximas seções apresentam as propostas específicas para o inglês e o espanhol e as aplicam sobre as amostras analisadas anteriormente.

6.5 Aprimorando as paráfrases explanatórias dos dicionários de aprendizes de inglês

Como já foi comentado no corolário das análises dos dicionários de inglês, para que suas paráfrases explanatórias sejam didáticas para o aprendiz brasileiro de níveis A2 e B1 de aprendizagem, é necessário abdicar do emprego da língua inglesa em função metalinguística. Nesse sentido, propomos, primeiramente, que se explore o conhecimento prévio da LM, de forma que as paráfrases explanatórias sejam redigidas na LM do consulente, isto é, o português, para aumentar o seu didatismo. Além das análises realizadas, nossa proposta se ampara em todo o referencial teórico da Psicolinguística no que se refere à participação da LM no processamento de segunda língua, seja durante a aprendizagem da LE, seja durante o uso desta. O embasamento no tipo de análise empreendido no capítulo 5 e na Psicolinguística constituem, ainda, o diferencial da nossa proposta em relação a outras que também advogaram pelo uso da LM nas paráfrases explanatórias de dicionários de aprendizes, como Longo (2000), Zavaglia e Zavaglia (2000a, b), Zavaglia, C. (2000) e Zavaglia, A. (2016).

Por outro lado, Longo (2000), Zavaglia e Zavaglia (2000a, b), Zavaglia, C. (2000) e Zavaglia, A. (2016) propuseram que, além da paráfrase em português, o segmento informativo da paráfrase fosse acompanhado de um equivalente também na língua materna, constituindo uma espécie de segunda paráfrase. Trata-se de um recurso interessante porque reforça o conteúdo semântico contido na paráfrase usando novamente a ponte com a LM, mas sem que se abstenha de descrever as particularidades semânticas da língua-alvo descrita no dicionário, como ocorre com os dicionários bilíngues em geral. Alguns casos para os quais esse expediente se mostrou especialmente útil foram o de preposições e advérbios, conforme se poderá aferir mais adiante.

Para ilustrar adaptações, foi necessário selecionar um conjunto de paráfrases. Julgou-se que incluir as propostas relativas às 101 analisadas no capítulo 5 deixaria a explanação extensa e cansativa. Portanto, incluímos aqui quatorze das paráfrases analisadas anteriormente, distribuídas em oito verbetes.

A metodologia adotada para aplicar a proposta foi a adaptação das paráfrases originais ao português por meio da tradução e a subsequente adequação do padrão sintático às fórmulas de Farias (2013). Não foram feitas alterações em relação ao conteúdo das paráfrases porque esse tema, conforme já comentado, requer uma investigação exclusiva que fugia do alcance da tese. Além disso, embora não seja o ideal, a tradução é uma opção mais realista e mais facilmente exequível para que se tenham instrumentos *user-friendly* a curto e médio prazos. É pertinente recordar que as significativas dificuldades que os brasileiros enfrentam no que tange aos avanços na sua proficiência em inglês (cf. Introdução) incorrem na demanda nada desprezível por materiais didáticos voltados especificamente para esse público. A tradução permite que obras adaptadas sejam disponibilizadas mais rapidamente comparada com a compilação de um dicionário completamente novo – embora, claro está, as considerações feitas na tese sirvam também para tal fim.

Além das adaptações, amparamo-nos em Longo (2000), Zavaglia e Zavaglia (2000a, b), Zavaglia, C. (2000) e Zavaglia, A. (2016) para incluir uma segunda paráfrase, de natureza sinonímica, para acompanhar a paráfrase principal. Para que o sinônimo não se confundisse com a paráfrase principal, e para deixar claro que se tratava de um equivalente em relação ao lema, empregou-se a fórmula “em português: [*equivalente*]” após a primeira paráfrase.

Outra observação necessária diz respeito ao uso do português nas marcas de uso, como no verbe *pretty*¹ do LWord2 (2008), e para o sistema de remissões (medioestrutura), como nos verbetes *pleasant* e *pretty* (LWord2, 2008). Como as marcas de uso e o sistema de remissões utilizam sintagmas em inglês, a sua apresentação na LM do consulente são justificadas em termos do didatismo do verbe, pois facilitam a compreensão leitora e o acesso rápido à informação.

Quadro 37 – Propostas de aprimoramento das paráfrases explanatórias (inglês): substantivos

Verbetes original	Versão adaptada
<p>landscape /'lændskeɪp/ <i>noun</i></p> <p>1 everything you can see when you look across an area of land: <i>We travelled through the beautiful landscape of northern Scotland.</i> (CoED2, 2014, s.v. <i>landscape</i>)</p>	<p>landscape /'lændskeɪp/ <i>noun</i></p> <p>1 tudo o que se pode ver quando se olha para uma área de terra; em português, <i>paisagem</i>: <i>We travelled through the beautiful landscape of northern Scotland.</i></p>
<p>label¹ /'leɪbəl/ <i>noun</i></p> <p>1 B1 a piece of paper or material that is attached to something and gives you information about it: <i>There are washing instructions on the label.</i></p> <p>2 B1 a company that makes things for sale, mainly clothes: <i>Her favourite designer label (= company that makes expensive clothes) is Armani.</i> (CaED2, 2011, s.v. <i>label</i>¹)</p>	<p>label¹ /'leɪbəl/ <i>noun</i></p> <p>1 B1 pedaço de papel ou tecido que está anexado a alguma coisa e que fornece informações sobre ela; em português, <i>etiqueta</i>: <i>There are washing instructions on the label.</i></p> <p>2 B1 empresa que produz coisas para vender, principalmente roupas; em português, <i>marca</i>: <i>Her favourite designer label (=empresa que produz roupas caras) is Armani.</i></p>

Fonte: CaED2 (2011) e CoED2 (2014). A tradução e adaptação das paráfrases foram feitas pela autora.

Na primeira acepção de *landscape*, percebe-se que os redatores do CoED2 (2014) optaram pela redação de paráfrases em estilo direto que remetem ao consulente/interlocutor. Essa medida parece ser uma estratégia para evitar o uso impessoal (“everything that can be seen when looking across an area of land”), que, no caso em questão, requer o uso de recursos gramaticais mais sofisticados, como a voz passiva e o infinitivo em *-ing*, por exemplo. No caso da versão adaptada ao português, a manutenção de um estilo pessoal de redação geraria a paráfrase “tudo o que você pode ver quando olha para uma área de terra”. Diferente do que ocorre no inglês, a explicitação da segunda pessoa do singular *você* não é comum nos dicionários monolíngues de língua portuguesa; os lexicógrafos brasileiros têm preferência pelo uso de um estilo impessoal. Essa preferência, que se pode chamar de “cultural”, foi mantida na versão adaptada.

Quadro 38 – Propostas de aprimoramento das paráfrases explanatórias (inglês): verbos

Verbete original	Versão adaptada
<p>remind¹ → /rɪ'maɪnd/ verb (reminds, reminding, reminded) 1 to help somebody remember something that they must do: <i>Please remind me to buy some bread on the way home.</i> 2 to make somebody remember somebody or something: <i>She reminds me of her mother.</i></p> <p>(OxED2, 2012, s.v. remind)</p>	<p>remind¹ → /rɪ'maɪnd/ verb (reminds, reminding, reminded) 1 ~ to ajudar [alguém] a lembrar de [algo] que ele/a deve fazer; em português, <i>lembrar [alguém de [algo]]</i>: <i>Please remind me to buy some bread on the way home.</i> 2 ~ of fazer com que [alguém] se lembre de [alguém / algo]; em português, <i>lembrar [alguém de [alguém/algo]]</i>: <i>She reminds me of her mother.</i></p>
<p>rent² → /rent/ verb (rents, renting, rented) 1 to pay to live in a place or to use something that belongs to another person: <i>I rent a house on the edge of town.</i> 2 to let somebody live in a place or use something that belongs to you, if they pay you: <i>He rents rooms in his house to students.</i> ☞ Look at hire</p> <p>(OxED2, 2012, s.v. rent²)</p>	<p>rent² → /rent/ verb (rents, renting, rented) 1 pagar para morar em um [lugar] ou para usar [algo] que pertence a outra pessoa; em português, <i>alugar</i>: <i>I rent a house on the edge of town.</i> 2 permitir que alguém more num [lugar] ou use [algo] que pertence a você, se ele/a pagar; em português, <i>alugar</i>: <i>He rents rooms in his house to students.</i></p>

Fonte: OxED2 (2012). A tradução e adaptação das paráfrases foram feitas pela autora.

Nas versões adaptadas, percebe-se que o uso da própria paráfrase para demarcar a natureza dos complementos dos verbos ajuda a não sobrecarregar o verbete, o que é benéfico para o consulente. Além disso, vale ressaltar que, segundo comentado ao final do capítulo 4, a descrição do comportamento sintático é necessária especialmente em materiais didáticos para bilíngues com baixa proficiência na L2/LE dada a sua dificuldade em identificar e processar esse tipo de informação. Para garantir o aproveitamento desse tipo de recurso, é fundamental que o consulente aprenda a interpretar o papel dos colchetes na paráfrase, seja por intermédio do *front matter* do dicionário, seja pela mediação do professor.

Nos verbetes acima, novamente foi necessário fazer adaptações durante o processo de tradução. É comum que não somente o OxED2 (2012), mas também outros dicionários de aprendizes de inglês, usem o pronome *they* em detrimento dos pronomes *he* e *she*, com gênero gramatical marcado. Como proposta para o português, considerando que, até o momento, a língua não possui um pronome dessa natureza que esteja suficientemente consagrado pelo uso massivo entre seus falantes, empregou-se a solução *ele/a*.

Finalmente, no verbete *rent*², há uma remissão a *hire*. O OxED2 (2012), no entanto, não indica o impulso da remissão. Conforme a base teórica adotada a respeito da medioestrutura, esse tipo de informação deve constar na remissão para que ela seja didática e o consulente possa avaliar a sua pertinência segundo as suas necessidades de consulta. Ao consultar o verbete *hire*, percebe-se que a sua primeira acepção é semanticamente similar à primeira acepção de *rent*: “to pay to use something for a short time” (OxED2, 2012, s.v. *hire*,

1). O significado de *hire* é mais restrito que *rent*, pois se refere somente ao ato de pagar para usar algo (e não para morar em um lugar). Além disso, *hire* se aplica a casos em que o usufruto dura pouco tempo, enquanto em *rent* não há essa restrição. Como a relação entre *rent*² e *hire* envolve somente uma acepção dos dois verbetes, e considerando que ambos contêm duas acepções, a primeira falha do dicionário foi não atrelar a remissão à acepção que é pertinente ao caso. Quanto ao motivo da remissão, é possível apenas especular: pode ter sido a indicação de um sinônimo ou ainda um alerta para a diferença semântica entre as duas ULs. Acreditamos mais na segunda opção, pois o dicionário indica sinônimos por meio de uma etiqueta própria para tal (*same meaning*) e que aparece atrelada à acepção específica envolvida na relação de sinonímia. Em se tratando de uma diferenciação semântica, não é possível entender por que não se especificou a questão por meio de uma nota de uso do tipo *which word?*, conforme previsto pelo próprio OxD2 (2012). No verbete *rent*¹ do CaED2 (2012), por exemplo, há uma nota de uso que esclarece a diferença de significado entre ambas ULs e que chama a atenção para o papel do eixo diatópico (usos britânicos vs. usos americanos). Está claro que, frente à complexidade do tema, uma simples remissão não é suficiente; nesse caso, uma nota de uso seria a melhor opção. Por isso, retiramos a remissão na versão adaptada.

Quadro 39 – Propostas de aprimoramento das paráfrases explanatórias (inglês): adjetivos

Verbete original	Versão adaptada
<p>pleas·ant <i>adjective</i> UK ◀ US ◀</p> <p>1 enjoyable or nice ANTONYM unpleasant</p> <p>◀ <i>The village is a pleasant place to live.</i></p> <p>◀ <i>The weather in June is very pleasant.</i></p> <p>see thesaurus box at NICE</p> <p>2 polite and friendly ANTONYM unpleasant</p> <p>◀ <i>Our boss is always very pleasant to us.</i></p> <p>(LWord2, 2008, s.v. <i>pleasant</i>, 1-2)</p>	<p>pleas·ant <i>adjective</i> UK ◀ US ◀ /'plezənt/</p> <p>1 [algo] que é agradável ou bom; em português, <i>agradável</i>. ANTÔNIMO unpleasant</p> <p>◀ <i>The village is a pleasant place to live.</i></p> <p>◀ <i>The weather in June is very pleasant.</i></p> <p>veja sinônimos em NICE</p> <p>2 [alguém] que é educado e amigável; em português, <i>agradável, simpático</i>. ANTÔNIMO unpleasant</p> <p>◀ <i>Our boss is always very pleasant to us.</i></p>
<p>gorgeous /'gɔ:dʒəs/ adj</p> <p>B1 very beautiful or pleasant: <i>You look gorgeous in that dress!</i></p> <p>(CaED2, 2011, s.v. <i>gorgeous</i>)</p>	<p>gorgeous /'gɔ:dʒəs/ adj</p> <p>B1 [alguém / algo] que é muito bonito ou agradável; em português, <i>lindo, maravilhoso</i>: <i>You look gorgeous in that dress!</i></p>

Fonte: LWord2 (2008) e CaED2 (2011). A tradução e adaptação das paráfrases foram feitas pela autora.

Assim como a valência, as restrições de atribuição dizem respeito às propriedades combinatórias das ULs e são fundamentais para uma escolha adequada de palavras durante tarefas de produção. Note-se que os verbetes originais não descrevem essas restrições de maneira explícita, sendo necessário interpretar tal fato linguístico a partir dos exemplos.

Curiosamente, nem os dicionários Longman e Cambridge de aprendizes avançados de inglês registram nos verbetes *pleasant* e *gorgeous*, respectivamente, as restrições de atribuição em questão (cf. LDOCE6 (2014, s.v. *pleasant*) e CALD4 (2013, s.v. *gorgeous*)). A única linha de dicionários de aprendizes de inglês que fornece esse tipo de informação ao registrar *pleasant* e *gorgeous* é Collins (cf. CCID4, 2018, s.v.s. *gorgeous*; *pleasant*). Ademais, nota-se também que falta uma marca de uso informal no verboete *gorgeous* do CaED2 (2011).

Quadro 40 – Propostas de aprimoramento das paráfrases explanatórias (inglês): preposições e advérbios

Verbete original	Versão adaptada
<p>among /ə'mʌŋ/ <i>preposition</i> 1 in the middle of a group of people or things: <i>There were teenagers sitting among adults.</i> 2 included in a particular group: <i>We discussed it among ourselves.</i> □ <i>Don't worry; you're among friends.</i> 3 used for saying that most people in a group have a particular opinion or feeling: <i>There is a concern among parents about teaching standards.</i></p> <p>(CoED2, 2014, s.v. <i>among</i>, 1-3)</p>	<p>among /ə'mʌŋ/ <i>preposition</i> 1 posição: em meio a um grupo de [pessoas] ou [coisas]; em português, <i>entre</i>: <i>There were teenagers sitting among adults.</i> 2 inclusão: incluído em um determinado grupo; em português, <i>entre</i>: <i>We discussed it among ourselves.</i> □ <i>Don't worry; you're among friends.</i> 3 maioria: indica que a maior parte dos membros de um grupo tem uma determinada opinião ou sentimento; em português, <i>entre</i>: <i>There is a concern among parents about teaching standards.</i></p>
<p>pret·ty¹ <i>adverb</i> UK ◀ US ◀ (spoken) fairly or quite ◀ <i>Their house is pretty big.</i> ◀ <i>It was a pretty quick journey, really.</i> see thesaurus box at FAIRLY</p> <p>(LWord2, 2008, s.v. <i>pretty</i>¹)</p>	<p>pret·ty¹ <i>adverb</i> UK ◀ US ◀ /'prɪtɪ/ (<i>fala</i>) expressa intensidade suficiente, nem pouca nem excessiva; em português, <i>bastante</i>, <i>bem</i>. ◀ <i>Their house is pretty big.</i> ◀ <i>It was a pretty quick journey, really.</i> veja sinônimos em FAIRLY</p>

Fonte: LWord2 (2008) e CaED2 (2011). A tradução e adaptação das paráfrases foram feitas pela autora.

Em *among* (CoED2, 2014), a questão do descritor categorial foi resolvida a partir dos dados sobre as noções gerais do documento Vantage 1990 (2001). Como as paráfrases estavam redigidas em português, as noções relacionadas a cada uma também foram traduzidas ao português. À aceção 1 de *among*, foi associada a noção *relative position* (VANTAGE 1990, p. 61), a qual foi reduzida a *posição*; à aceção 2, foi associada a noção *inclusion* (VANTAGE 1990, p. 69), traduzida como *inclusão*. À aceção 3, a noção mais próxima era *generality* (VANTAGE 1990, p. 68). Seu equivalente em português *generalidade*, embora registrado no Hou (2009), por exemplo, é mais usado no português europeu segundo dados do *corpus* CP (2019). Em vista disso, usou-se o sinônimo *maioria*.

No tocante a *pretty*, optou-se pelo uso concomitante de dois tipos de paráfrases, uma em metalinguagem de signo, tal como proposto por Farias (2013), seguida de outra em metalinguagem de conteúdo, tal como a paráfrase sinonímica do original em inglês. De acordo com Farias (2013, p. 130-131; 197), a utilização de teorias semânticas nem sempre

culminará em paráfrases elucidativas – ao menos aos olhos do lexicógrafo. Foi essa a impressão que tivemos ao empregar a fórmula correspondente aos advérbios. Em prol da facilitação da compreensão do advérbio *pretty*, julgou-se necessário manter a solução sinonímica empregada pelo LWord2 (2008). Em se tratando de línguas diferentes, a sinonímia é compreendida pelo viés dos equivalentes. Assim, a manutenção da relação sinonímica em inglês consistiu no uso de equivalentes na versão em português.

Nas páginas anteriores, foi apresentada uma amostra que continha alguns dos principais ajustes que a utilização do português e a aplicação das fórmulas de Farias (2013) demandaram. Em síntese, a referência ao consulente foi substituída por uma redação impessoal; o pronome *they* foi adaptado para *ele/a*; recorreu-se ao Vantage 1990 (2001) para a indicação da significação categorial das preposições; e, em um caso de paráfrase que seguia sendo pouco elucidativa após a sua revisão, recorreu-se à formulação sinonímica original. De maneira geral, ficou demonstrado que há como adaptar as paráfrases dos quatro dicionários às características e necessidades do usuário brasileiro de proficiência A1 e A2 que aprende inglês no Brasil em prol do didatismo das obras. Os entraves que surgiram – e que sempre surgem quando se trata de paráfrases, não importa a classe de obras de que se trate – podem ser solucionados à medida que as pesquisas em Semântica aplicada à Lexicografia segurem sendo desenvolvidas, em um movimento similar ao de Farias (2013), por exemplo.

O principal objetivo desta seção foi chamar a atenção para a necessidade de se incluir a perspectiva do usuário durante a compilação ou revisão de um dicionário de aprendizes no que diz respeito ao **aspecto formal das paráfrases explanatórias**, mais precisamente à seleção da **língua a ser utilizada**. Conforme exposto no capítulo 1, a Lexicografia de aprendizes contou com grandes avanços ao longo do tempo no que diz respeito ao aprimoramento da descrição da língua. É evidente que esse aprimoramento, em última instância, visava beneficiar o público-alvo das obras; no entanto, faltava uma abordagem direta em relação ao usuário. No caso dos dicionários de aprendizes de inglês A1-B1, delimitar como usuário o aprendiz brasileiro de inglês de níveis A1 e A2 de proficiência que aprende inglês no Brasil requereu uma mudança na língua que veicula as paráfrases. Naturalmente, essa proposta causará um estranhamento em uma tradição que tem o costume de usar somente a língua inglesa em função metalinguística. Porém, insistimos que não há como seguir avançando na área somente pela via da qualidade descrição linguística, pois o viés didático-pedagógico, que tem o usuário como eixo central, é tão importante quanto. Nesse sentido, a busca por um embasamento teórico-metodológico em outras áreas, como a Linguística Aplicada e a Psicolinguística, é fundamental.

Dentre as perspectivas futuras da proposta para o inglês, está, primeiramente, uma pesquisa empírica junto a estudantes brasileiros de inglês para investigar os efeitos de facilitação da redação das paráfrases na LM em comparação com os originais em inglês. Esse tipo de pesquisa ajudaria a complementar as nossas considerações teóricas sobre o usuário, ao mesmo tempo que forneceria mais detalhes sobre ele.

Na parte teórico-metodológica, é de especial interesse uma investigação mais detalhada sobre o potencial de aplicabilidade dos padrões sintáticos de Farias (2013) para a descrição, em português, de um maior número de ULs definidas originalmente em língua inglesa. Além disso, é primordial uma investigação sobre padrões sintáticos para a redação de paráfrases dos substantivos abstratos.

Um último aspecto que merece atenção é o conteúdo das paráfrases explanatórias. A complexidade da seleção do conteúdo das paráfrases explanatórias requer uma pesquisa que trate exclusivamente desse tema a longo prazo.

6.6 Aprimorando as paráfrases explanatórias do dicionário de aprendizes de espanhol

As análises do capítulo 5 demonstraram que é plenamente possível explorar contextos de influências interlinguísticas facilitadoras do português no espanhol para a compilação de materiais didáticos e, em particular, de dicionários de aprendizes. Um dicionário de aprendizes brasileiros de espanhol de níveis A2 e B1 de aprendizagem pode, nesse sentido, manter a redação de suas paráfrases explanatórias em espanhol desde que sejam observados os tipos de definidores usados na redação desses segmentos informativos. Para o caso de se usar o dicionário Señas (2001) como ponto de partida para a compilação de um dicionário de aprendizes nos moldes como se propõe nesta tese, é necessário investir em determinadas estratégias de didatização da descrição metalexiconográfica a fim de aproximar o dicionário do usuário brasileiro. Nesse sentido, propomos o uso de definidores e padrões gramaticais que sejam ou de nível inferior ao da UL que definem, ou que sejam cognatos, e que se evite o uso de falsos amigos entre o espanhol e o português. Outros dicionários de espanhol, tais como os dicionários de sinônimos, podem ser usados como ferramentas complementares durante a proposição de definidores substitutos (cf. Borba (2019b) sobre algumas das obras disponíveis).

Quanto ao equivalente que o Señas (2001) já coloca após o exemplo de cada paráfrase, propomos que esta informação seja incluída da mesma maneira como se propôs para os dicionários de aprendizes de inglês, isto é, como uma segunda paráfrase, de natureza

sinonímica, acompanhando a paráfrase principal. A fórmula que propomos para introdução do equivalente é “en português: [equivalente]”.

Conforme previsto anteriormente, tomou-se por base a proposta de Farias (2013) a respeito do padrão sintático de redação das paráfrases. Dado que fora originalmente elaborada em português, fez-se necessário adaptá-la ao espanhol por meio da tradução dos padrões sintáticos e do ajuste da terminologia à organização da gramática hispânica.

Quadro 41 – Padrões sintáticos de Farias (2013) adaptados ao espanhol

Clase de palabra	Patrón sintáctico
Sustantivos concretos	Hiperónimo (nombre concreto) + Especificador(es)
Sustantivos concretos que designan entidades únicas o varios referentes	“Nombre/Designación/Expresión” + “para referirse a” + alguien/algo
Sustantivos concretos que designan una categoría	“Nombre dado a” / “Nombre que designa” / “Designación para” + Hiperónimo (nombre) + Especificador(es) + Exponentes de la categoría
Sustantivos abstractos de acción	Hiperónimo (“acción/efecto/proceso de”) + Especificador(es)
Verbos intransitivos sin complemento	Verbo transitivo / intransitivo / soporte + Complementos
Verbos transitivos e intransitivos con complemento de régimen	Verbo transitivo / intransitivo / soporte + Complementos + [Valencia del lema]
Adjetivos calificativos	[Restricción de atribución] + oración subordinada adjetiva
Adjetivos relacionales	[Restricción de atribución] + “relativo a” + sintagma nominal
Adverbios	Locución / Oración adverbial
Adverbios modales	“De modo” + Adjetivo
Adverbios cuyo contenido es difícil de explicitar	“Expresa” / “Se usa para expresar” + Noción
Conjunciones	“Expresa” / “Se usa para expresar” + Noción
Preposiciones que son marcadores sintácticos	“Marcador de” + Función sintáctica
Preposiciones con significado categorial	Descriptor + “:” + Significado categorial
Expresiones prepositivas	Verbo / Adverbio + Preposición
Artículos y pronombres personales	Descripción de la clasificación morfológica
Expresiones deícticas	“Se usa para designar” + Entidad
Interjecciones	“Expresa” + Noción

Fonte: Farias (2013). Tradução e adaptação nossas

A alta convergência entre o português e o espanhol, especialmente no que se refere à sintaxe, permitiu que a adaptação ocorresse sem grandes problemas ou ajustes estruturais severos.

Quanto à terminologia, foi necessário revisar as designações *verbos transitivos*, *verbos intransitivos* e *pronomes*.

No caso dos verbos, as mesmas designações são usadas em espanhol em referência a conceitos bem diferentes dos que se usam em português. Farias (2013) ampara-se na gramática de língua portuguesa ao distinguir *verbos transitivos* (exigem complementação, mediada ou não por preposição) e *verbos intransitivos* (não exigem complementação). A gramática hispânica, por outro lado, distingue entre *verbos transitivos*, que exigem objeto direto (complemento obrigatório sem preposição ou complemento obrigatório de pessoa com a preposição *a*), e *verbos intransitivos*, que são definidos em oposição aos anteriores por não exigirem objeto direto. Os *verbos intransitivos* em espanhol se referem a verbos que não requerem nenhum tipo de complementação obrigatória, como *llover* [chover], por exemplo – tal como os *verbos intransitivos* do português. No entanto, são também considerados intransitivos no espanhol os verbos são acompanhados de preposição e um complemento, como *llegar a* [chegar a] sob o significado “5. intr. Alcanzar o producir una determinada acción. *Llegó a reunir una gran biblioteca*” (DRAEe, 2014-2021, s.v. *llegar*, 5. Grifos no original). No português, verbos com esse comportamento são classificados como *transitivos indiretos* (cf. Sac (2010, s.v. *chegar*, 5) enquanto no espanhol continuam sendo rotulados como *intransitivos*. Portanto, além da semelhança designativa com o português, a terminologia hispânica é de difícil compreensão devido à sua aparente contradição ao considerar intransitivos os verbos que exigem complemento preposicionado. Embora a edição integral da *Nueva Gramática de la Lengua Española* (GramLE, 2009, §34.1d) reconheça a “transitividade” dos “verbos intransitivos preposicionados” do espanhol, prefere manter o termo *intransitivo*. Para efeitos da adaptação de Farias (2013), optou-se por usar a designação *verbos intransitivos sin complemento* em referência aos casos de complementação não obrigatória (*verbos intransitivos* no original), e *verbos transitivos e intransitivos con complemento de régimen* em alusão aos casos de complementação obrigatória mediada ou não por preposição (*verbos transitivos* no original).

A categoria “Pronomes (exceto pessoais)” de Farias (2013) também teve de sofrer modificações. No português, os pronomes se subdividem em categorias, como demonstrativos, possessivos, relativos, interrogativos etc. No espanhol, por outro lado, as ULs que são rotuladas como pronomes sob a ótica da gramática do português podem ser designadas não somente como pronomes, mas também como adjetivos e advérbios. Em razão disso, as edições mais atuais da gramática da RAE preferem o termo genérico *expresiones deícticas* [expressões dêíticas], que contempla um variado número de subcategorias

funcionais: *demonstrativos* (MGramLE, 2010, §17), *posesivos* (MGramLE, 2010, §18), *cuantificadores* (MGramLE, 2010, §19, §20), *numerales* (MGramLE, 2010, §21) e *relativos, interrogativos y exclamativos* (MGramLE, 2010, §22). Em outras palavras, houve um processo de independência dos rótulos de pronomes e outros em prol de *expresiones deícticas*. Frente a esse cenário, e para estar de acordo com a perspectiva gramatical hispânica, a categoria “Pronomes (exceto pessoais)” foi adaptada ao espanhol sob o rótulo genérico *expresiones deícticas*.

A metodologia aplicada na revisão das paráfrases contemplou as seguintes etapas:

(1) Substituição dos definidores apontados no Quadro 29 por sinônimos ou por novas formulações de caráter sinonímico caso a adaptação requeresse ajustes mais severos do que uma simples substituição. Os substitutos deveriam cumprir com os critérios de facilitação da compreensão determinados no capítulo 5, isto é, nível CEFR (2001) inferior ao da UL definida e/ou a sua natureza cognata em relação ao português.

(2) Aplicação dos padrões traduzidos de Farias (2013) às paráfrases.

A aplicação da metodologia gerou dois tipos de resultados. Por um lado, estão os verbetes que puderam ser adaptados sem grandes dificuldades. Por outro lado, estão os verbetes que apresentaram alguma dificuldade na sua adaptação ou que precisam de uma revisão no seu conteúdo ou nos definidores. Ambos os grupos serão comentados separadamente.

Quadro 42 – Aprimoramento das paráfrases explanatórias de ULs A2 e B1 do Señas (2001)

Nível	Verbetes original	Versão adaptada
A2	co·pa [...] 2 f. Bebida alcohólica Problema: -	co·pa [...] 2 f. Bebida alcohólica Definidores sugeridos: -
A2	gim·na·sio m. Instalación o establecimiento que dispone de aparatos para hacer ejercicio físico y practicar ciertos deportes Problema: Definidor ausente do PCIC: <i>aparato</i>	gim·na·sio m. Instalación o establecimiento que dispone de máquinas y objetos para hacer ejercicio físico y practicar ciertos deportes Definidores sugeridos: <i>máquina</i> : 100% similar a <i>máquina</i> ; <i>objeto</i> : 100% similar a <i>objeto</i>
A2	pro·pi·na f. Cantidad de dinero que se da voluntariamente para mostrar *satisfacción por el buen trato o el servicio recibido	pro·pi·na f. Cuánta de dinero que se da voluntariamente para mostrar satisfacción por el trato o el servicio recibido

	<p>Problema:</p> <p>Definidores B1: <i>cantidad, buen</i></p>	<p>Definidores sugeridos:</p> <p><i>cuantía</i>: 65% similar a <i>quantia</i>; <i>buen</i>: definidor suprimido</p>
A2	<p>ba·jar 1 intr.-prnl. [de un sitio; a un sitio] Ir de un lugar alto a otro que está más bajo</p>	<p>ba·jar 1 intr.-prnl. ~ de / a Ir de un [lugar] alto a otro que está más bajo</p>
	<p>Problema:</p> <p>Necessidade de adaptação do padrão sintático</p>	<p>Definidores sugeridos: -</p>
A2	<p>re·pa·sar 2. [algo] Examinar o volver a mirar</p>	<p>re·pa·sar 2. Examinar o volver a leer [algo]</p>
	<p>Problema:</p> <p>Definidor A2: <i>mirar</i></p>	<p>Definidores sugeridos:</p> <p><i>leer</i>: A1; <i>algo</i>: 100% similar a <i>algo</i>.</p>
A2	<p>de·ma·sia·do,da 1 adv. c. Superando lo justo o lo necesario; excesivamente</p>	<p>de·ma·sia·do, da 1 adv. c. De modo excesivo, superando la medida justa o necesaria</p>
	<p>Problema:</p> <p>Definidor B2+: <i>lo</i></p>	<p>Definidores sugeridos:</p> <p><i>modo</i>: 100% similar a <i>modo</i>; <i>excesivo</i>: 93% similar a <i>excessivo</i>; <i>superar</i>: 100% similar a <i>superar</i>; <i>medida</i>: 100% similar a <i>medida</i>.</p>
B1	<p>con·tra·se·ña f. Expresión o señal que se dan unas personas a otras para entenderse entre sí, para ser reconocidas o para que les sea permitida una cosa determinada</p>	<p>con·tra·se·ña f. Expresión o gesto que se dan unas personas a otras para entenderse entre sí, para ser reconocidas o para que les sea permitida una cosa determinada</p>
	<p>Problema:</p> <p>Definidor ausente do PCIC: <i>señal</i></p>	<p>Definidores sugeridos:</p> <p><i>gesto</i>: 100% similar a <i>gesto</i>.</p>
B1	<p>huel·ga f. Paro del trabajo que hacen los *obrerros de común acuerdo, generalmente como protesta o para *imponer ciertas condiciones a los dueños de la empresa o al gobierno</p>	<p>huel·ga f. Paro del trabajo que hacen los trabajadores de común acuerdo, generalmente como protesta o para *imponer ciertas condiciones a los dueños de la empresa o al gobierno</p>
	<p>Problema:</p> <p>Definidor B2+: <i>obrero</i></p>	<p>Definidores sugeridos:</p> <p><i>trabajador</i>: A1</p>
B1	<p>re·ba·ja f. Disminución o reducción, especialmente en una cantidad o un precio</p>	<p>re·ba·ja f. Hecho de disminuir o reducir algo, especialmente un número, porción o precio</p>
	<p>Problema:</p> <p>Definidor B1: <i>cantidad</i></p>	<p>Definidores sugeridos:</p> <p><i>hecho</i>: A2; <i>disminuir</i>: 88% similar a <i>diminuir</i>; <i>reducir</i>: 85% similar a <i>reduzir</i>; <i>algo</i>: 100% similar a <i>algo</i>; <i>número</i>: A1; <i>porción</i>: 61% similar a <i>porção</i>.</p>

B1	re·gla [...] 5 Sangre que todos los meses expulsan naturalmente las mujeres y las hembras de ciertos animales	re·gla [...] 5 Sangre que todos los meses expulsan naturalmente las mujeres y ciertos animales de sexo femenino
	Problema: Definidor B1: <i>hembra</i> ;	Definidores sugeridos: <i>sexo</i> : 100% similar a <i>sexo</i> ; <i>femenino</i> (B1): 91% similar a <i>feminino</i> .
B1	en·fa·dar 1 tr.-prnl. [a alguien] Causar disgusto o enfado	en·fa·dar 1 tr.-prnl. Causar disgusto o irritación [a alguien]
	Problema: Definidor B2+: <i>enfado</i>	Definidores sugeridos: <i>irritación</i> : 69% similar a <i>irritação</i> ; <i>alguien</i> : 61% similar a <i>alguém</i> .
B1	li·gar 6 intr.-prnl. fam. [a/con alguien] Establecer relaciones amorosas o sexuales	li·gar 6 intr.-prnl. fam. ~ a / con Establecer relaciones amorosas o sexuales con [alguien]
	Problema: Necessidade de adaptação do padrão sintático	Definidores sugeridos: <i>alguien</i> : 61% similar a <i>alguém</i>
B1	sa·ber 4 intr. Tener sabor	sa·ber 4 intr. ~ a Tener sabor
	Problema: Necessidade de adaptação do padrão sintático	Definidores sugeridos: -
B1	fa·tal [...] 1 adj. Que es muy malo; que no se puede soportar	fa·tal [...] 1 adj. [algo] que es muy malo o que no se puede soportar
	Problema: Necessidade de adaptação do padrão sintático	Definidores sugeridos: -
B1	li·ge·ro, ra [...] 3 (alimento) que se puede *digerir fácilmente	li·ge·ro, ra [...] 3 [alimento] que se puede digerir fácilmente
	Problema: Necessidade de adaptação do padrão sintático	Definidores sugeridos: -
B1	a·no·che adv. t. En la noche de ayer	a·no·che adv. t. En la noche de ayer
	Problema: -	Definidores sugeridos: -
B1	va·go, ga 1 adj.-s. Que no gusta del trabajo	va·go, ga 1 adj.-s. [alguien] a quien no le gusta el trabajo
	Problema: Necessidade de adaptação do padrão sintático e do definidor <i>gustar</i> .	Definidores sugeridos: <i>alguien</i> : 61% similar a <i>alguém</i> . <i>gusta</i> (v.): 78% similar a <i>gosta</i> (v.).

Fonte: Señas (2001) para os verbetes originais; PCIC (2006) para definidores destacados e níveis dos definidores. As propostas de redação foram elaboradas pela autora a partir dos dados do Señas (2001) e do PCIC (2006).

Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do PCIC (2006)

Na amostra acima, houve três tipos de situações: 1) três verbetes sem necessidade de adaptação (*copa, ligero, anoche*); 2) dez verbetes com adaptação dos definidores (*gimnasio, propina, repasar, demasiado, contraseña, huelga, rebaja, regla, enfadar, vago*); e nove verbetes com adaptação do padrão sintático (*bajar, repasar, demasiado, rebaja, enfadar, ligar, saber, fatal, vago*).

No tocante à adaptação dos definidores, a maior parte das soluções envolveu o uso de cognatos. No processo de adaptação dos verbetes B1, o maior número de recursos léxicos correspondentes ao conhecimento prévio de níveis A1 e A2 diminuiu consideravelmente a necessidade de se substituir definidores em comparação com as paráfrases de verbetes A2. Uma única ressalva diz respeito ao verbo *regla*. Embora tenha sido possível proceder com a adaptação, percebeu-se que o definidor *expulsan* (verbo *expulsar*), embora seja cognato e, por isso, não tenha sido alvo das modificações propostas, poderia ser substituído por um verbo mais simples e relacionado aos níveis A1 e A2, como *salir*. Este caso será comentado novamente junto aos casos difíceis de adaptar.

Quanto à aplicação das fórmulas de Farias (2013), em uma boa parcela dos casos, foram necessários apenas ajustes mínimos. Algumas modificações mais sensíveis foram o destaque dos tipos de complemento dos verbos e a descrição das restrições de atribuição dos adjetivos. Ambas as marcações são especialmente úteis para tarefas de produção. Além disso, alguns problemas detectados no capítulo 5, tais como a ausência da preposição *a* no verbo *saber* e o uso do padrão morfossintático formal e pouco usado no espanhol *gustar de* no verbo *vago*.

Outras adaptações mínimas contemplam a retirada do asterisco que acompanhava os definidores *satisfacción* (s.v. *propina*), *obreros* (s.v. *huelga*) e *digerir* (s.v. *ligero*). Esse recurso tipográfico é usado no Señas (2001) para marcar os elementos que não constam na sua lista de definidores. Trata-se de um expediente útil apenas ao lexicógrafo que queira analisar a relação entre os definidores previstos no dicionário e os efetivamente empregados, mas que, sob a perspectiva do consulente, acaba poluindo o verbo de maneira desnecessária.

O seguinte grupo de paráfrases contempla casos em que a adaptação dos definidores foi difícil e/ou precisaria de uma revisão no nível do conteúdo.

Quadro 43 – Paráfrases explanatórias do Señas (2001) difíceis de adaptar

Nível	Verbete original	Versão adaptada
	bi·lle·te 1 m. Papel que imprime y emite un banco y que representa una cantidad de la moneda de un país	bi·lle·te 1 m. Papel que representa una cuantía de la moneda de un país
A2	Problema: Definidor B1: <i>cantidad</i> . Excesso de especificadores.	Definidores sugeridos: <i>cuantía</i> : 65% similar a <i>quantia</i> . Outras adaptações: Simplificação da paráfrase por supressão do especificador desnecessário.

A2	<p>ci·ta 1 f. Fijación del día, la hora y el lugar para encontrarse dos o más personas</p> <p>Problema: Definidores ausentes do PCIC: <i>fijación, persona</i>. O encontro, que é a informação principal, não está em evidência.</p>	<p>ci·ta 1 f. Encuentro entre dos o más individuos en un día, hora y lugar determinados previamente.</p> <p>Definidores sugeridos: <i>encuentro</i>: 84% similar a <i>encontro</i>; <i>individuo</i>: 100% similar a <i>indivíduo</i>; <i>determinado</i> (particípio): 100% similar a <i>determinado</i> (particípio) <i>previamente</i>: 100% similar a <i>previamente</i></p> <p>Outras adaptações: Mudança de foco semântico por meio da substituição do hiperônimo por <i>encuentro</i>.</p>
A2	<p>di·rec·ción [...] 3 Nombre de la población y la calle y número de la casa donde vive una persona</p> <p>Problema: Definidor B1: <i>población</i>; Definidor ausente do PCIC: <i>persona</i>; O especificador “donde vive una persona” exclui outras propriedades do significado de dirección.</p>	<p>di·rec·ción [...] 3 Información que identifica un lugar y permite encontrarlo, como el nombre de la ciudad y de la calle, el número, etc.</p> <p>Definidores sugeridos: <i>información, ciudad, lugar</i>: A1; <i>identifica</i> (v.): 100% similar a <i>identifica</i> (v.)</p> <p>Outras adaptações: Generalização por meio da troca do hiperônimo e supressão de especificadores.</p>
A2	<p>con·tes·tar 1 tr. [algo; a alguien, a algo] Dar una respuesta a lo que se pregunta, dice o escribe</p> <p>Problema: Definidores A2: <i>lo, dice</i>; Especificador redundante “a lo que se pregunta, dice o escribe”</p>	<p>con·tes·tar 1 tr. Responder [algo (a alguien)] o responder [a algo]</p> <p>Definidores sugeridos: <i>responder</i>: 100% similar a <i>responder</i>; <i>algo</i>: 100% similar a <i>algo</i>; <i>alguien</i>: 61% similar a <i>alguém</i>; <i>o</i>: A1.</p> <p>Outras adaptações: Simplificação da paráfrase por supressão do especificador redundante.</p>
A2	<p>in·vi·tar 1 tr. [a alguien] Pedir a una persona que esté en una celebración, comida u otra cosa</p> <p>Problema: Definidor B2+: <i>esté</i> (subjuntivo) Definidor A2: <i>u</i> (variante de <i>o</i>)</p>	<p>in·vi·tar 1 tr. Convidar [a alguien] a una celebración, comida o una otra cosa</p> <p>Definidores sugeridos: <i>o, una</i>: A1; <i>convidar</i>: 100% similar a <i>convidar</i>.</p>
A2	<p>a·bu·rri·do, da 1 adj. Que no tiene interés; que aburre</p>	<p>a·bu·rri·do, da 1 adj. [algo] que causa tedio o insatisfacción</p>

	<p>Problema:</p> <p>Definidor B2+: <i>aburrir</i>.</p> <p>Especificador redundante e pouco claro “que no tiene interés”.</p>	<p>Definidores sugeridos:</p> <p><i>o</i>: A1; <i>causar</i>: 100% similar a <i>causar</i>; <i>tedio</i>: 100% similar a <i>tédio</i>; <i>insatisfacción</i>: 69% similar a <i>insatisfação</i>; <i>algo</i>: 100% similar a <i>algo</i>.</p> <p>Outras adaptações:</p> <p>Simplificação da paráfrase por supressão do especificador redundante.</p>
A2	<p>an·cho 1 adj. Que tiene una extensión de lado a lado excesiva o mayor de lo normal</p>	<p>an·cho 1 adj. [cosa] que tiene una extensión de lado a lado mayor de la medida normal</p>
	<p>Problema:</p> <p>Definidor B2+: <i>lo</i>; Especificador redundante “excesiva”.</p>	<p>Definidores sugeridos:</p> <p><i>medida</i>: 100% similar a <i>medida</i>;</p> <p>Outras adaptações:</p> <p>Simplificação da paráfrase por supressão do especificador redundante.</p>
A2	<p>ma·yor 1 adj. (ser vivo) Que tiene más edad</p>	<p>ma·yor 1 adj. [ser vivo] que tiene más edad que otro.</p>
	<p>Problema:</p> <p>Falta de elemento comparativo.</p>	<p>Outras adaptações:</p> <p>Acréscimo de elemento comparativo.</p>
A2	<p>to·da·ví·a 1 adv. t. Hasta un momento determinado; hasta ahora</p>	<p>to·da·ví·a 1 adv. t. Expresa la continuidad de algo en el tiempo actual o en otro momento</p>
	<p>Problema:</p> <p>Definidor A2: <i>hasta</i>; Ausência de adverbio substituto para <i>hasta</i> e que fosse de nível A1 ou suficientemente similar a uma palavra do português.</p>	<p>Definidores sugeridos:</p> <p><i>expresa</i>: 92% similar a <i>expressa</i>; <i>continuidad</i>: 81% similar a <i>continuidade</i>; <i>algo</i>: 100% similar a <i>algo</i>; <i>actual</i>: 86% similar a <i>atual</i>; <i>otro</i>: 80% similar a <i>outro</i>; <i>la/el, de, en, o, tiempo, momento</i>: A1.</p> <p>Outras adaptações:</p> <p>Aplicação do padrão sintático dos advérbios difíceis de definir.</p>
B1	<p>chis·te 1 m. Historia corta o dibujo que hace reír</p>	<p>chis·te 1 m. Historia corta o dibujo que hace reír</p>
	<p>Problema:</p> <p>Definidor ausente do PCIC: <i>reír</i></p>	<p>Definidores sugeridos: -</p>
B1	<p>ta·pa [...] 2 Alimento ligero que se sirve como acompañamiento de una bebida</p>	<p>ta·pa [...] 2 Aperitivo que se sirve como acompañamiento de una bebida</p>
	<p>Problema:</p> <p>Definidor B1: <i>ligero</i>; O alimento não necessariamente é leve [<i>ligero</i>], mas é servido em porções pequenas.</p>	<p>Definidores sugeridos:</p> <p><i>aperitivo</i>: A2</p> <p>Outras adaptações:</p> <p>Mudança de hiperônimo para um que fosse mais condizente com o significado de <i>tapa</i>.</p>
B1	<p>ro·to, ta [...] 2 Que no sirve, no funciona o no se puede usar</p>	<p>ro·to, ta [...] 2 [utensilio, instrumento o máquina] que no sirve, no funciona o no se puede usar</p>

	<p>Problema: Necessidade de adaptação do padrão sintático</p>	<p>Outras adaptações: <i>utensilio</i>: 100% similar a <i>utensílio</i>; <i>instrumento</i>: 100% similar a <i>instrumento</i>; <i>máquina</i>: 100% similar a <i>máquina</i>.</p>
B1	<p>ma·re·a·do, da adj. Que experimenta una sensación desagradable en la cabeza y en el estómago, generalmente a causa del movimiento de un vehículo o de una nave</p>	<p>ma·re·a·do, da adj. [alguien] que tiene náuseas, generalmente a causa del movimiento de un vehículo o de una embarcación</p>
	<p>Problema: Definidor ausente do PCIC: <i>nave</i>; Definidores que poderiam ser mais simples: <i>experimentar, sensación desagradable en la cabeza y el estómago</i></p>	<p>Definidores sugeridos: <i>embarcación</i>: 71% similar a <i>embarcação</i> Outras adaptações: <i>tener</i>: A2; <i>náusea</i>: 100% similar a <i>náusea</i>.</p>
	<p>re·gla [...] 5 Sangre que todos los meses expulsan naturalmente las mujeres y las hembras de ciertos animales</p>	<p>re·gla [...] 5 Sangre que todos los meses expulsan sale naturalmente las mujeres y ciertos animales de sexo femenino</p>
	<p>Problema: Definidores que poderiam ser mais simples: <i>explulsar</i>.</p>	<p>Outras adaptações: <i>salir</i>: A1</p>

Fonte: Señas (2001) para os verbetes originais; PCIC (2006) para definidores destacados e níveis dos definidores; propostas de redação elaboradas pela autora a partir dos dados do Señas (2001) e do PCIC (2006).
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do PCIC (2006)

Neste segundo grupo de paráfrases, foi necessário adequar definidores avaliados previamente como deficitários em dez casos (*billete, cita, población, contestar, invitar, aburrido, ancho, todavía, tapa, nave*); e ainda adequar o padrão sintático de cinco verbetes (*contestar, invitar, mayor, todavía, roto*).

Quanto à adaptação dos definidores, em apenas uma paráfrase da amostra não foi possível encontrar uma alternativa viável: *reír* [rir], definidor presente em *chiste* [piada]. Trata-se de uma UL ausente do PCIC (2006) e que não é suficientemente similar à representação ortográfica do seu equivalente português.

Quanto a adaptação dos padrões sintáticos, duas situações se destacaram pela dificuldade em aplicar as propostas. Em *roto* [quebrado, estragado], a fórmula dos adjetivos qualificativos requeria que se incluísse entre colchetes um definidor que indicasse a restrição de atribuição da palavra-entrada. A alternativa mais sintética e que melhor aludiria aos tipos de UL aos quais *roto* se aplica é *aparato*; porém, essa UL não está presente no PCIC (2006). Como alternativa, incluímos os definidores “utensilio, instrumento o máquina”, pois, juntos, eram os que mais se aproximavam do significado de *aparato*. Trata-se de uma solução não tão boa quanto *aparato* porque deixou o componente em questão mais extenso; contudo, foi o melhor que conseguimos fazer com os recursos léxicos disponíveis.

Especialmente difícil foi a adaptação da paráfrase de *todavía* [ainda], tanto no que concerne aos definidores como ao padrão sintático. Em se tratando de um advérbio, era necessário, primeiramente, aplicar a fórmula de Faria (2013) relativa a essa classe de palavras. Porém, não havia um substituto de nível A1 para a preposição *hasta*, que está atrelada ao nível A2. Consequentemente, a paráfrase estaria adequada em relação ao padrão sintático, mas com definidores inapropriados ao perfil de usuário estabelecido. Frente a esse impasse, optamos por aplicar o padrão sintático dos advérbios difíceis de definir, pois ele permitia a utilização de definidores que se adequavam ao conhecimento prévio esperado sem comprometer o didatismo da paráfrase em termos de definidores e de redação. É provável que, numa adaptação de paráfrases em larga escala, surjam problemas similares quando se trata de verbetes de advérbios e de preposições.

Um traço marcante das paráfrases explanatórias presentes no quadro anterior são os problemas na dimensão do conteúdo. Ainda que esta dimensão não tenha sido o alvo das nossas análises e propostas, é pertinente comentar certos aspectos para futuros estudos sobre a seleção do conteúdo e dos definidores das paráfrases explanatórias.

O primeiro caso especial é o de especificadores redundantes ou desnecessários para a explicitação do significado. Os verbetes *billete* [nota (dinheiro)], *contestar* [responder], *aburrido* [chato] e *ancho* [largo] apresentam esse problema. Em *billete*, não é necessário indicar a participação dos bancos na emissão do referente (“que imprime y emite un banco”). Em *contestar*, a natureza dos complementos (“lo que se pregunta, dice o escribe”) está implícita no definidor *responder*. Em *aburrido*, a qualificação “que no tiene interés”, além de redundante em relação a “que aburre”, no original, ou a “que causa tedio o insatisfacción”, na versão adaptada, é também vaga. Em *ancho*, a informação original “excesiva o mayor de lo normal” também foi considerada redundante, motivando a supressão de *excesiva* na versão adaptada. Como resultado, obtiveram-se paráfrases explanatórias mais simples e que buscam refletir de maneira mais acurada o significado de suas ULs.

O segundo tipo de caso especial é o das paráfrases que não descrevem as propriedades do significado de maneira satisfatória. Esse problema está presente nos verbetes *cita* [encontro] *dirección* [endereço] e *tapa* [tira-gosto]. Em *cita*, além da necessária substituição do hiperônimo *fijación*, que é um dos pilares sob o qual a paráfrase está construída, foi preciso empregar um definidor cujo significado refletisse melhor o conteúdo semântico da palavra-entrada. O mesmo ocorreu com *dirección*, cuja paráfrase original restringia o conjunto de referentes aos quais a UL em questão se aplica (“la casa donde vive una persona”) ao invés de contemplar as diversas construções e espaços que podem ser

identificados por uma *dirección* (qualquer lugar que tenha uma delimitação e que, por isso, seja identificável). Em *tapa*, o adjetivo *ligero* [leve] não é um dos atributos da palavra-entrada. Conforme o DRAEe (2014-2021, s.v. *tapa*¹, 8), o que caracteriza uma *tapa* é o fato de ser servida em porções pequenas.

O terceiro tipo de caso especial é o das paráfrases de *regla* e *mareado*, que contêm definidores complexos que são passíveis de substituição por alternativas mais simples. Em *regla*, tal como já se havia comentado anteriormente, ao invés de manter o definidor original *expulsar*, poder-se-ia usar a alternativa *salir*, que tem a vantagem de ser de nível A1. Em *mareado*, *tener* é uma alternativa mais simples que o original *experimentar* e apresenta a vantagem de aludir a um padrão combinatório de nível A2, que é “*tener* + [sensação física]” (PCIC, 2006). Já o sintagma “*una sensación desagradable en la cabeza y en el estómago*” é uma sorte de paráfrase da UL *náusea*, que, além de ser direta, é um cognato entre o espanhol e o português. Ao realizar as duas substituições assinaladas, obtém-se uma paráfrase mais simples e que reflete o significado de *mareado* de maneira mais precisa.

Com as propostas apresentadas nesta seção, buscou-se demonstrar que, apesar das dificuldades, é possível empregar os recursos que se tem (PCIC e similaridade ortográfica) para tornar as paráfrases explanatórias do Señas (2001) mais didáticas para os aprendizes brasileiros de espanhol em níveis A2 e B1 de aprendizagem. Em dezessete itens da amostra foi possível realizar as adaptações sem grandes problemas, enquanto outros treze se mostraram mais desafiadores tanto em função dos definidores e padrões sintáticos como de aspectos relativos ao conteúdo. Embora a dimensão do conteúdo não tenha sido o alvo das propostas da tese, as considerações feitas são um simulacro das questões que podem surgir quando se decide investigar a dimensão do conteúdo das paráfrases explanatórias. Conforme já comentado anteriormente, o tema requer uma atenção exclusiva e, por essa razão, não poderíamos tratá-lo nesta oportunidade, não obstante a sua evidente relevância para o aprimoramento da Lexicografia de Aprendizes. Espera-se que, com o avanço de pesquisas em Semântica aplicada à Lexicografia, no futuro se possa determinar e testar parâmetros atinentes à dimensão do conteúdo.

Por fim, e tal como assinalado na seção sobre a adaptação de paráfrases dos dicionários de inglês, o seguinte passo da pesquisa sobre adaptação das paráfrases do Señas (2001) ao público brasileiro é a testagem das propostas junto a esse perfil. Para tanto, é necessário que se faça, antes, uma aplicação das propostas sobre um conjunto mais extenso de ULs e suas paráfrases explanatórias.

O próximo capítulo trata da adaptação das notas de uso ao perfil de usuário da tese.

7 O PERFIL DE USUÁRIO APLICADO AOS DICIONÁRIOS DE APRENDIZES (II)

Este capítulo trata das consequências de se adotar um perfil de usuário específico no que diz respeito às notas de uso de dicionários de aprendizes de inglês e espanhol A1-B1. Para entrar no mérito da questão, é preciso antes discorrer sobre as características e a função desse segmento informativo para então associá-lo ao perfil de usuário.

Assim como ocorre com as paráfrases explanatórias, as notas de uso são constituídas de uma parte formal e de uma contraparte relativa à seleção de seu conteúdo. Porém, diferente das anteriores, não há uma teoria lexicográfica suficientemente desenvolvida sobre as notas de uso. As informações sobre esse segmento informativo são esparsas e relativamente recentes na literatura especializada. Segundo Farias (2013, p. 130), as notas de uso são um tipo de “mecanismo complementar de elucidação do significado”. Conforme a autora, as notas não são primordiais para uma obra lexicográfica cuja ênfase recai sobre o conteúdo das ULs, embora saliente que possam ser extremamente úteis para acrescentar informações que, por um motivo ou outro, não puderam ser incluídas em outros segmentos informativos.

Hartmann e James (2001, s.v. *usage note*) se limitam a assinalar que as notas de uso fornecem informação adicional sobre uma determinada palavra ou sintagma. Os autores citam como exemplo de temas tratados nesse segmento informativo os sinônimos, palavras e sintagmas relacionados, expressões idiomáticas, estilística e restrições de uso. Atkins e Rundell (2008, p. 233-235), por sua vez, desenvolvem mais o tema, identificando dois tipos de notas de uso: as notas que apresentam campos léxicos e as notas relativas a uma determinada palavra-entrada. As notas acerca dos campos léxicos são organizadas por temas e apresentam uma série de ULs e sintagmas relacionados à palavra-entrada, como a nota em *cook*: “There are many ways to cook food. You can **bake** bread and cakes and you can **roast** meat in an **oven**. You can **boil** vegetables in a **saucepan**. You can **fry** fish, eggs, etc. in a **frying pan**” (OxED2, 2012, s.v. *cook*¹. Grifos no original). Embora apareça nos dicionários A1-B1 analisados, esse tipo de nota é mais frequente em dicionários de aprendizes a partir de níveis intermediários, tais como o CCID4 (2018). Junto ao verbete *finance* do CCID4 (2018), por exemplo, há uma nota com um texto sobre finanças pessoais que serve para apresentar ULs e sintagmas relacionados a este tema, tais como *income*, *outgoings*, *cut down on*, *work out* etc. Cada UL e sintagma destacado na nota está lematizado no dicionário e, no verbete que lhe corresponde, há uma remissão à nota em *finance*.

As notas classificadas por Atkins e Rundell (2008, p. 233-235) como sendo relativas a uma palavra-entrada, por sua vez, contêm distinções semânticas entre duas ULs, padrões

combinatórios e erros e dúvidas de aprendizes. Um exemplo desse segundo grupo é a nota em *apologize* no CaED2 (2011): “Remember to use **to** when you mention the person who receives the apology. *I apologized to them immediately. ~~I apologized them immediately.~~*” (CaED2, 2011, s.v. *apology*. Grifos no original). Mais do que chamar a atenção para um padrão combinatório, o objetivo da nota é indicar que a omissão da preposição é um erro comum entre os aprendizes de inglês.

Apesar de elucidativos, os comentários de Atkins e Rundell (2008) acerca das notas de uso consistem somente na sua proposta de classificação e na menção aos diversos tipos de conteúdo veiculados nesse segmento informativo. Não há maiores considerações de cunho metalexiconográfico, isto é, sobre critérios de inclusão e redação das notas de uso; os autores apenas sugerem que os lexicógrafos observem como outros dicionários selecionam conteúdo para suas notas de uso para, então, tomarem suas próprias decisões (ATKINS; RUNDELL, 2008, p. 233). Em outras palavras, na falta de uma teoria, o melhor que se poderia fazer, segundo os autores, é se basear na prática.

Aparentemente, Atkins e Rundell (2008) não se deram conta de que a observação da prática lexicográfica, quando aliada ao que já se sabe na Metalexiconografia, pode auxiliar na teorização sobre as notas de uso. Dada a relativa semelhança com as paráfrases explanatórias em termos de constituição, é perfeitamente possível analisar as notas de uso sob o mesmo prisma que Bugueño Miranda (2009b) propõe e que Farias (2013) desenvolveu na sua teoria geral dos mecanismos explanatórios: a sua função, o seu aspecto formal e a seleção de conteúdo. Este capítulo se fundamenta nesses três eixos para analisar o que há disponível na metalexiconografia, o que a prática lexicográfica oferece e como correlacionar a teoria à prática a fim de propor orientações e metodologias de pesquisa para a elaboração de notas de uso, com especial atenção à adaptação dos dicionários de aprendizes tratados neste trabalho.

7.1 As notas de uso sob a ótica da teoria e da práxis lexicográfica

Conforme já comentado, as considerações disponíveis sobre as notas de uso em dicionários, especialmente nos dicionários de aprendizes, são poucas e insuficientes para que pudéssemos iniciar o capítulo tratando da adaptação dos dicionários que interessam para a tese. Dos três eixos que usamos para a nossa argumentação a respeito das notas de uso, a função é a que encontra respaldo na teoria lexicográfica por meio das contribuições de Farias (2013) e será tratada em 7.1.1.


O aspecto formal das notas, por outro lado, é pouco comentado na teoria. Há alguns poucos trabalhos recentes sobre o particular, como Bugueño Miranda (2019e) e Bugueño Miranda e Borba (2020), que exploram os tipos de orientação idiomática presentes em notas de uso de dicionários. Porém, para tratar do aspecto formal das notas de uso de maneira mais ampla, foi necessário aliar a observação do que já é feito na Lexicografia às considerações teóricas e análises prévias sobre a função e a orientação idiomática. O desenvolvimento do tema em 7.1.2 culmina em propostas a respeito do eixo em questão, com especial atenção à adaptação de dicionários de aprendizes de inglês e espanhol ao público brasileiro A1-B1.

Finalmente, o eixo do conteúdo das notas precisou de tratamento separado devido à sua extensão e ao fato de que os recursos que servem de fonte para a determinação do conteúdo variam conforme a língua de que se trate. A seção 7.2 trata da seleção de conteúdo de notas de uso para dicionários de aprendizes de inglês, enquanto a seção 7.3 trata do mesmo tema relativo aos dicionários de aprendizes de espanhol.

7.1.1 A função das notas de uso

Das três variáveis relativas às notas de uso (função, aspecto formal e seleção de conteúdo), a função está bem desenvolvida na teoria geral dos mecanismos explanatórios de Farias (2013). É possível sintetizar as contribuições da autora em três princípios: a funcionalidade propriamente tal, a articulação com os demais segmentos informativos da microestrutura do verbete e a vinculação a uma teoria geral dos mecanismos explanatórios (FARIAS, 2013, p. 359-360). Para ilustrar cada princípio, serão usados os excertos a seguir.

Quadro 44 – Função das notas de uso

<p>WHICH WORD? Be careful! You cannot buy books from a library. The place where you buy books is called a bookshop or a bookstore.</p> <p>(OxED2, 2012, s.v. <i>library</i>)</p>	<p>equipment  /ɪ'kwɪpmənt/ <i>noun (no plural)</i> special things that you need for doing something: <i>sports equipment</i></p> <p>GRAMMAR Equipment does not have a plural. If you are talking about one item, you say 'a piece of equipment'.</p> <p>(OxED2, 2012, s.v. <i>equipment</i>)</p>
<p>be-a-ˈto, ˈta beáto, ta 1 adj.-s. (persona) Que es muy religioso y pasa mucho tiempo en la iglesia: <i>a la misa de diario sólo van cinco o seis viejas beatas</i>. ☐ Se usa generalmente con intención despectiva. ☐ beato</p> <p>(Señas, 2001, s.v. <i>beato</i>)</p>	<p>SPELLING This word is often spelled wrongly. The correct spelling is: always.</p> <p>(LWord2, 2008, s.v. <i>always</i>)</p>

Fonte: OxED2 (2012, s.v.s. *equipment, library*), Señas (2001, s.v. *beato*), LWord2 (2008, s.v. *amazing*)

Em primeiro lugar, a funcionalidade diz respeito à natureza das informações contidas na nota; nesse sentido, uma nota funcional é aquela que fornece informações discretas e discriminantes (*ibid*). Na nota do verbete *library* do OxD2 (2012), por exemplo, além de indicar as designações para os lugares onde se compra livros, seria interessante salientar que *bookshop* e *bookstore* possuem uma diferença a nível diatópico (a primeira é geralmente usada no Reino Unido enquanto a segunda é mais frequente nos EUA). Trata-se de uma informação discreta que tornaria a nota como um todo mais discriminante para o consulente.

Em segundo lugar, Farias (2013, p. 360) salienta que a nota de uso precisa estar articulada com os demais segmentos informativos da microestrutura. Isso significa que uma nota deve complementar o verbete com peculiaridades linguísticas que não puderam ser incluídas em outros segmentos informativos, seja porque sua natureza difere do tipo de informações previstas para estes componentes, seja porque os deixariam excessivamente longos, prejudicando a localização e compreensão das informações básicas do verbete (*ibid*). Em suma, segundo esse princípio, as notas de uso são um recurso a ser reduzido ao mínimo, evitando-se a repetição desnecessária de informações.

No entanto, não é difícil encontrar notas que divergem do princípio da articulação nos dicionários de aprendizes de inglês e de espanhol analisados na tese. Por vezes, as obras incluem notas com informações que poderiam estar em um segmento informativo já previsto na microestrutura abstrata. Tome-se como exemplo a nota de uso do Señas (2001, s.v. *beato*). A informação diafásica *despectivo* poderia estar presente no verbete como um segmento informativo de marca de uso, prescindindo do uso da nota e ocupando menos espaço no verbete. Outro exemplo de informação repetida desnecessariamente é a indicação da flexão na nota de *equipment* (OxD2, 2012), pois o verbete já traz esse dado no comentário de forma. A nota poderia conter somente a parte sobre o uso do recurso léxico quantificador “a piece of”.

Por outro lado, há também um número considerável de casos em que as notas contêm informações já presentes no verbete. Essa situação é relativamente comum nos dicionários de aprendizes de inglês analisados, como se pode observar na nota em *always*: “This word is often spelled wrongly. The correct spelling is: **always**.” (LWord2, 2008, s.v. *always*. Grifos no original). A representação ortográfica está contida previamente no lema do verbete, de modo que a sua repetição seria, em princípio dispensável. No entanto, mais do que simplesmente repetir essa informação, a nota adverte o consulente a respeito do que considera como um erro comum entre os aprendizes de inglês. Nesse sentido, a repetição da representação é válida porque o foco da nota é o erro dos aprendizes.

A vinculação das notas de uso a uma teoria geral dos mecanismos explanatórios, por sua vez, é justificada por Farias (2013, p. 360) a partir da necessidade de se determinar “em que situações é preciso empregar mecanismos de elucidação do significado alternativos ou complementares às paráfrases definidoras, e qual recurso é mais apropriado em cada caso”. Ao propor a sua teoria, a autora cita cinco tipos de elementos que se beneficiariam de notas de uso nos verbetes: artigos, pronomes, verbos de ligação, verbos-suporte e informações extraordinárias, isto é, eventuais peculiaridades de uma UL (FARIAS, 2013, p. 291; 360).

Nos dicionários de aprendizes de inglês e de espanhol analisados, é comum encontrar notas junto a verbetes das classes de palavras mencionadas por Farias (2013). No entanto, as obras atribuem especial atenção às “informações extraordinárias”, como nos exemplos a seguir:

Quadro 45 – Notas de uso com informações extraordinárias

<p>cu-tis kútis <i>m.</i> Piel de la cara de las personas: <i>esta crema suaviza mucho el ~; las mujeres tienen el ~ terso.</i> □ El plural es <i>cutis</i>. □ tez</p> <p>dam-ni-fi-car da^mnifikár <i>tr.</i> [a alguien] Causar daño: <i>sin querer hemos damnificado a nuestros compañeros.</i> □ Se conjuga como <i>l</i>. □ prejudicar (Señas, 2001, s.vs. <i>cutis</i>, <i>damnificar</i>)</p>	<p>because /br'kɔ:z/ conjunction</p> <p>Common mistake: because Learners often spell because wrong. Remember that it has au in the middle, like cause.</p> <p>A1 used to give a reason for something: <i>I'm calling because I need to ask you something.</i></p> <p>(CaED2, 2011, s.v. <i>because</i>)</p>
<p>alcoholic 0→ /,ælkə'hɒlɪk/ <i>adjective</i> containing alcohol: <i>an alcoholic drink</i></p> <p>WORD BUILDING Drinks without alcohol are called soft drinks.</p> <p>stare 0→ /steə(r)/ <i>verb</i> (stares, staring, stared /steəd/) to look at somebody or something for a long time: <i>Everybody stared at her hat.</i> ♦ <i>He was staring out of the window.</i></p> <p>WHICH WORD? Stare or gaze? You usually stare at somebody or something if you are surprised or shocked. It is rude to stare at people. You often gaze at somebody or something that you love or admire.</p> <p>bow¹ /baʊ/ <i>verb</i> (bows, bowing, bowed /baʊd/)</p> <p>PRONUNCIATION With this meaning, the word bow sounds like now.</p> <p>to bend your head or body forward to show respect: <i>The actors bowed at the end of the play.</i></p> <p>(OxED2, 2012, s.vs. <i>alcoholic</i>, <i>asleep</i>, <i>bow</i>¹)</p>	<p>angry /'æŋɡri/ <i>adjective</i> (angrier, angriest) feeling a strong emotion when someone has done something bad, or someone has treated you unfairly: <i>We are very angry about the decision to close the school.</i> □ <i>An angry crowd gathered.</i></p> <p>LANGUAGE HELP If someone is very angry, you can say they are furious. If they are less angry, you can say they are annoyed or irritated.</p> <p>(CoED2, 2014, s.v. <i>angry</i>)</p> <p>be-lieve /bə'li:v/ <i>verb</i></p> <p>KEY PATTERNS believe something/someone believe that believe someone/something to be something</p> <p>1 to feel sure that something is true: <i>No one believed his story about what had happened.</i> • <i>I can't believe they're brothers.</i> • <i>The police believed that they had found the killer.</i> • <i>His family believed him to be dead.</i></p> <p>(LWord2, 2008, s.v. <i>believe</i>)</p>

Fonte: Señas (2001, s.vs. *cutis*, *damnificar*), CaED2 (2011, *because*), OxED2 (2012, s.vs. *alcoholic*, *stare*, *bow*), CoED2 (2014, s.v. *angry*) e LWord2 (2008, s.v. *believe*)

No Señas (2001), são muito comuns as notas de uso de cunho morfológico, principalmente para indicar a flexão de número e a flexão verbal, como se vê em *cutis* e *damnificar*. No CaED2 (2011), por outro lado, as notas de uso são introduzidas pela etiqueta “common mistake”, ou seja, tratam de erros e dúvidas dos aprendizes de inglês, como se pressupõe do alerta em *because* a respeito da sua representação ortográfica. No OxE2 (2012), por sua vez, grande parte das notas de uso se dedica a sugerir ULs do mesmo campo léxico, como em *alcohol*, e a esclarecer diferenças semânticas entre ULs sinônimas, como em *stare*. Além destas, um diferencial do OxE2 (2012) é recorrer a macetes para facilitar a pronúncia de algumas ULs, como em *bow*¹. No CoED2 (2014), também há notas a respeito de informações de diferentes níveis de organização da linguagem, com predominância de notas que sugerem e diferenciam sinônimos, como se vê em *angry*. Por fim, no LWord2 (2008), além de haver notas sobre diversos níveis de organização da linguagem, ganham destaque as notas de uso etiquetadas “key patterns”, como se vê em *believe*. Trata-se de notas situadas entre o comentário de forma e o comentário semântico que mostram uma prévia de padrões combinatórios destacados nos exemplos de cada acepção.

Há, portanto, muito mais a dizer sobre o que há de “extraordinário” no léxico de uma língua. Em alguns casos, as notas parecem ter finalidade puramente descritiva, pois apresentam certas propriedades que se poderiam considerar intrínsecas às palavras-entradas, como a flexão (*cutis*, *damnificar*) e os padrões combinatórios (*believe*). Em outros casos, a motivação parece estar além da descrição, tal como comentado anteriormente; trata-se de um alerta para erros e dúvidas considerados comuns, como a ortografia correta (*believe*) e a pronúncia (*bow*). Para um terceiro grupo de notas, referentes a campos léxicos e diferenciação semântica (*alcoholic*, *stare*, *angry*), as informações apresentadas não permitem saber qual a motivação das notas, isto é, se o lexicógrafo julgou pertinente descrever essas informações com base em erros, dúvidas ou outra razão.

7.1.2 O aspecto formal das notas de uso

Entendemos por aspecto formal das notas de uso a sua redação e *layout*. No âmbito teórico, nem Farias (2013) nem Atkins e Rundell (2008) desenvolveram o tema. Em vista disso, para discriminar as propriedades referentes aos dois aspectos, observamos e comparamos amostras de notas de uso dos cinco dicionários tratados na tese. Mediante essa observação, foi possível destacar seis propriedades regularmente presentes nas notas de uso: 1) *layout*: recursos (ícono)gráficos para sinalizar a nota de uso; 2) redação e *layout*: etiquetas, seja para introduzir

a nota, seja para indicar a que se refere o conteúdo da nota de uso; 3) redação: língua em que as informações são apresentadas na nota; 4) orientação: modalidade de orientação; 5) construção: padrão sintático; e 6) redação e *layout*: recursos tipográficos. Esses elementos não necessariamente estão presentes em todos os dicionários analisados, mas aparecem em, ao menos, dois deles.

7.1.2.1 Recursos (ícono)gráficos

Os recursos (ícono)gráficos das notas de uso compreendem os símbolos que as introduzem, como se vê nas notas de OxED2 (2012) e do Señas (2001), bem como o fundo em cor diferenciada que os quatro dicionários de aprendizes de inglês empregam nas suas notas. O fundo em cor diferenciada é, inclusive, sugerido por Farias (2013, p. 362) para dar mais destaque a nota e diferenciá-la dos demais segmentos informativos. De fato, ao se observar o verbete *beato* do Señas (2001), a nota passa quase despercebida por não contar com recursos (ícono)gráficos suficientes para chamar a atenção do consulente. O ícone \triangle produz um baixo efeito de saliência pela semelhança com o ícone \square , usado para introduzir o equivalente em língua portuguesa (cf. Quadro 45, s.vs. *cutis* e *damnificar*).

Dos dicionários de inglês analisados, o único que conta com um ícone é o OxED2 (2012), que emprega uma lupa (\binocular). Frente a esse cenário, propomos que, no caso das obras que não contam com um ícone, ou quando este possui baixo efeito de saliência, que se use o símbolo \blacklozenge (disponível na fonte Webdings do Microsoft Word). No tocante ao fundo, seguimos a sugestão de Farias (2013), bem como a práxis lexicográfica dos quatro dicionários de aprendizes de inglês analisados, e reforçamos que se utilize um fundo em cor diferenciada nas notas de uso.

7.1.2.2 Etiquetas

As etiquetas para indicar o conteúdo das notas de uso são usadas somente pelos quatro dicionários de aprendizes de inglês analisados. Há dois padrões de uso de etiquetas. Em um deles, a etiqueta introduz a nota e é igual para todas. É o caso do CaED2 (2011), que usa a etiqueta “common mistake” com a especificação da(s) UL(s) envolvidas, e do CoED2 (2014), que usa “language help”. O uso de uma mesma etiqueta para todas as notas serve apenas de sinalizador da nota de uso, tal como os recursos iconográficos. Em casos assim, se o dicionário já emprega um fundo colorido, como o CaED2 (2011) e o CoED2 (2014), a etiqueta torna-se

supérflua porque não agrega informação nem desempenha uma função especial. O OxED2 (2012) e o LWord2 (2008), por outro lado, adotam diferentes etiquetas conforme a natureza do conteúdo.

Quadro 46 – Etiquetas das notas de uso do OxED2 (2012) e do LWord2 (2008)

Tipo de conteúdo	Etiqueta do OxED2 (2012)	Etiqueta do LWord2 (2008)
ortografia	<i>spelling</i>	<i>spelling</i>
pronúncia	<i>pronunciation</i>	<i>pronunciation</i>
diferença semântica entre duas ULs	<i>which word?</i>	<i>word choice</i>
palavras relacionadas: sinônimos e campos léxicos	<i>word building</i>	<i>thesaurus</i>
alternativa mais frequente e com registro informal	<i>speaking</i>	-
registro diatópico, diaprágmatico, entre outros	<i>culture</i>	-
flexão	-	<i>grammar</i>
erros comuns de aprendizes	-	<i>usage</i>

Fonte: Elaboração da autora com base nos dados do OxED2 (2012) e do LWord2 (2008)

A vantagem de usar etiquetas que discriminam o conteúdo da nota é que o consulente já pode identificá-lo e julgar se o assunto é do seu interesse ou não sem que tenha de ler o texto. Embora não haja orientações teóricas sobre o particular, a decisão de incorporar tais etiquetas parece acertada. A única desvantagem é que, como cada obra emprega um sistema próprio de etiquetagem, é preciso se familiarizar com o sistema de cada um e com o tipo de informação que veicula, especialmente em casos que abrem margem para várias interpretações, como *word building*, *speaking*, *culture*, *grammar* e *usage*. Os erros comuns dos aprendizes que o LWord2 (2008) discrimina em notas de uso específicas, por exemplo, poderiam ser tratados nas outras categorias de notas, prescindindo de uma rubrica especial para esse fim.

Para o consulente que precisa se familiarizar com o sistema de classificação das notas, é mais interessante que haja um sistema com uma variedade menor de etiquetas que abarquem um escopo mais amplo de tipos de conteúdo. Por essa razão, e como o objetivo é contribuir com propostas suficientemente genéricas para o aprimoramento tanto dos dicionários de aprendizes que não empregam etiquetas como daqueles que empregam um ou mais tipos de etiquetas, elaboramos um sistema com quatro distinções básicas: *ortografia*, *pronúncia*, *vocabulário* e *uso*. Esse sistema foi baseado no OxED2 (2012) e no LWord2 (2008), bem como nos níveis de organização da linguagem, com o objetivo de permitir ao consulente julgar se o conteúdo da nota é do seu interesse ou não. Como a etiqueta está em função metalinguística, a escolha do idioma em que aparecerá depende de cada caso, seguindo as mesmas decisões que se tomarem

em relação ao idioma das paráfrases explanatórias (cf. 6.3 e 6.4). Portanto, para a adaptação de um dicionário de inglês, as etiquetas permaneceriam em português; já para a adaptação de um dicionário de espanhol, as etiquetas apareceriam em espanhol (*ortografía, pronunciación, vocabulario e uso*).

As rubricas *ortografía* e *pronúncia* são autoexplicativas em relação ao tipo de informação referente aos níveis de organização da linguagem correspondentes. Sob a rubrica *vocabulário*, enquadramos questões semânticas, como os casos de diferenças entre duas ULs, sinônimas ou não (cf. Quadro 45, s.v.s. *stare* (OxED2, 2012) e *angry* (CoED2 (2014))), e os campos léxicos (cf. Quadro 45, s.v. *alcoholic* (OxED2, 2012)). Alguns dos dicionários de aprendizes de inglês analisados especificam já na etiqueta as ULs que estão envolvidas na diferenciação semântica, como em “WHICH WORD? **Stare** or **gaze**?” (OxED2, 2012, s.v. *stare*. Grifos no original). Trata-se de uma estratégia que agiliza a leitura do verbete na medida que permite que o consulente decida se a nota é relevante para si ou não. Na nossa proposta, a adaptação da etiqueta usada como exemplo “VOCABULÁRIO: **stare** ou **gaze**?”.

A rubrica *uso*, por sua vez, abarca informações de registros de uso, flexão e padrões combinatórios que não possam ser tratados em outros segmentos informativos no verbete. Também se enquadram nesse grupo as informações de cunho pragmático que sejam necessárias para a compreensão das expectativas comunicativas e comportamentais, para a identificação do referente ao qual se aplica uma UL ou ainda para a adequação conforme o interlocutor em uma determinada situação comunicativa. Alguns exemplos desse tipo de informação são a observação em *stare*: “It is rude to **stare** at people” (OxED2, 2012, s.v. *stare*. Grifo no original); e em *English*: “Be careful! The people of **Scotland** (the **Scots**) and the people of **Wales** (the **Welsh**) are **British**, not **English**” (OxED2, 2012, s.v. *English*. Grifo no original).

7.1.2.3 Língua utilizada na nota

As conclusões parciais da tese no que diz respeito à língua empregada nas paráfrases explanatórias se aplica também às notas de uso sob a justificativa de torna-las mais didáticas. Portanto, para fins de adaptação ao público brasileiro de proficiência A1 e A2, seguimos propondo o uso do português para as notas de um dicionário de aprendizes de inglês A1-B1 e a manutenção do uso do espanhol no caso do Señas (2001).

7.1.2.4 Modalidade de orientação e padrão sintático

A quarta e quinta propriedades que aparecem nas notas de uso são a modalidade de orientação e o respectivo padrão sintático que a manifesta. A relação entre esses dois elementos é parecida com a dos tipos de paráfrases explanatórias no que diz respeito aos seus padrões sintáticos. Porém, diferente do caso anterior, há poucas investigações sobre o tema na literatura. Farias (2013, p. 361-362), por exemplo, sugere apenas que as notas de artigos, pronomes e verbos suporte e de ligação precisam tomar a forma de compêndios gramaticais. Bogueño Miranda (2019f) e Bogueño Miranda e Borba (2020), por outro lado, desenvolvem melhor o tema mediante a observação da práxis lexicográfica. Após analisar as modalidades de orientação linguística presentes em notas de uso em dicionários de diferentes tradições lexicográficas, os autores propõem a seguinte taxonomia:

Figura 6 – Modalidades de orientação linguística



Fonte: Bogueño Miranda e Borba (2020, p. 162)

No âmbito das orientações linguísticas, a de natureza indicativa corresponde à descrição ou reforço de algum aspecto que geralmente é alvo de erros ou dúvidas, como as notas em *equipment* e *beato*: “**Equipment** does not have a plural. If you are talking about one item, you say ‘a piece of equipment’” (OxED2, 2012, s.v. *equipment*); “Se usa generalmente con intención despectiva” (Señas, 2001, s.v. *beato*). Na contraparte exortativa, chama-se a atenção do consulente para um fato de língua que é ou que pode ser objeto de erros ou dúvidas. Sob o enfoque prescriptivo, há um lembrete ou ordem direta para que se use a língua de uma determinada maneira, como se lê em *apologize*: “Remember to use **to** when you mention the person who receives the apology” (CaED2, 2011, s.v. *apologize*). O enfoque proibitivo, por outro lado, indica um uso errôneo e orienta o aprendiz a não proceder daquela forma. Nessa rubrica, enquadram-se as notas em *amazing* e *library*: “Don’t say ‘It’s very amazing.’ Just say *It’s amazing* or *It’s absolutely amazing*” (LWord2, 2008, s.v. *amazing*. Grifos no original) e

“Be careful! You cannot buy books from a **library**. The place where you buy books is called a **bookshop** or a **bookstore**” (OxED2, 2012, s.v. *library*. Grifos no original). A orientação extralinguística, por sua vez, refere-se à relação entre uma designação e o(s) referente(s) extralinguísticos aos quais se aplica.

Segundo constatam Bugueño Miranda e Borba (2020), os dicionários de aprendizes de inglês fazem uso dos diferentes tipos de orientação descritos. A distinção entre eles é necessária não para que se advogue por um ou outro, mas, sim, para que o lexicógrafo, ciente da variação de opções disponíveis, possa fazer uma escolha consciente do tom que deseja dar às orientações das notas de uso. Nesta tese, para embasar a escolha da modalidade orientativa, classificamos a natureza da informação das notas de uso entre “nova” e “reiterada”. Uma informação nova é a que, embora seja julgada como relevante tanto para o perfil de usuário como para a classe de dicionário de que se trate, não está presente em outros segmentos informativos do verbete porque estes não conseguem contemplá-la. Para esse tipo de informação, julgamos pertinente empregar a orientação indicativa, já que é um fato novo, tal qual o são as informações dos demais segmentos informativos. Por outro lado, uma informação reiterada é aquela que, estando já presente em um segmento informativo do verbete, precisa ser reforçada por alguma razão, de preferência porque é alvo de dúvidas ou erros por parte do público-alvo do dicionário. Nesse tipo de situação, a simples repetição da descrição do fato linguístico via orientação indicativa poderá parecer uma obviedade ou até passar despercebida. Por isso, convém empregar uma das modalidades de orientação exortativa, pois justificará a repetição do fato linguístico.

Portanto, quando a nota trata de um aspecto não descrito previamente no verbete, propomos a orientação indicativa; por outro lado, quando a informação contida na nota reforça algum aspecto que já está descrito em outra parte do verbete, propomos o uso da orientação exortativa, seja a prescritiva (“faça assim”), seja a proibitiva (“não faça assim”).

A escolha de uma ou outra modalidade de orientação acarreta a adoção de um determinado padrão sintático correspondente. É difícil saber se as pesquisas sobre a redação de notas de uso alcançarão padrões de redação tal como Farias (2013) propôs para as paráfrases explanatórias. O que se pode fazer no momento é continuar observando a prática, isto é, averiguando que aspectos gramaticais normalmente se manifestam em cada modalidade de orientação e, a partir disso, estabelecer um formalismo que auxilie na redação das notas de uso.

A orientação indicativa geralmente descreve fatos linguísticos usando o presente do indicativo em orações afirmativas. Há várias notas com esse tipo de orientação reproduzidas nos Quadros 44 e 45, como, por exemplo, “Drinks without alcohol are called **soft drinks**” (OxED2, 2012, s.v. *alcohol*. Grifo no original) e “El plural es *cutis*” (Señas, 2001, s.v. *cutis*).

Grifo no original). Além da voz ativa, também se pode verificar o uso da voz passiva associado a uma determinada tradição lexicográfica, como a de língua inglesa: “This word is often spelled wrongly. The correct spelling is: **always**” (LWord2, 2008, s.v. *always*. Grifos no original). Nessa nota, assim como em *bow*¹ do OxE2 (2012), por exemplo, usa-se a orientação indicativa para repetir uma informação prévia no verbete. Outras notas de orientação indicativa reproduzidas anteriormente são *stare* (OxE2, 2012), *angry* (CoED2, 2014) e *damnificar* e *beato* (Señas, 2001). Não é comum que notas indicativas tenham exemplos para ilustrar o fato linguístico descrito. Trata-se de um componente aparentemente dispensável; as notas indicativas dos Quadros 44 e 45, não contêm exemplos e estes tampouco fazem falta.

A orientação exortativa prescritiva também pode descrever fatos linguísticos nos moldes da orientação indicativa, mas se diferencia pelo uso do imperativo afirmativo. Em alguns casos, esse modo verbal é usado para chamar a atenção do consulente – a expressão “Be careful!” por vezes aparece introduzindo notas de uso exortativas prescritivas em dicionários de aprendizes e inglês e pode ser seguida de uma orientação explicativa, como em “Be careful to chose the orct peposition after **depend**. *I might go – it depends on the weather [...]*” (CaED2, 2011, s.v. *depend*. Grifos no original). Em outros casos, usa-se o imperativo para introduzir uma recomendação propriamente tal, como em “Use the short form **n’t** when you are speaking English. For example, ‘didn’t’ is short for ‘did not’” (CoED2, 2014, s.v. *not*. Grifo no original). O uso de exemplos, por sua vez, é eventual. Em *depend* (CaED2, 2011), o exemplo é necessário porque contém o objeto da orientação, isto é, a preposição que deve ser usada com o verbo; porém, em *not* (CoED2, 2014), não há uma frase ilustrativa além da menção ao uso com o auxiliar *did*.

A orientação exortativa proibitiva, por sua vez, é um alerta sobre como não usar a língua. Algumas notas manifestam esse alerta de maneira direta ao empregar o imperativo negativo, como em “Don’t say ‘It’s very amazing.’ Just say *It’s amazing* or *It’s absolutely amazing*” (LWord2, 2008, s.v. *amazing*. Grifos no original) e “No se debe decir *diabetis*” (Señas, 2001, s.v. *diabetes*. Grifo no original). Em outros casos, a negativa de forma indireta sobre um determinado uso, como em “Be careful! You cannot buy books from a **library**. The place where you buy books is called a **bookshop** or a **bookstore**” (OxE2, 2012, s.v. *library*. Grifos no original) e “Remember that you **cannot** say that you ‘avoid to do something’” (CoED2, 2014, s.v. *avoid*. Grifo no original). Note-se que as duas notas anteriores estão redigidas no imperativo afirmativo, tal como na orientação prescritiva; porém, ao invés de recomendarem um determinado uso, o proibem. Nas orientações proibitivas em geral, é comum que haja uma segunda orientação a respeito do uso correto, seja prescritiva (*amazing*) ou indicativa (*library*).

No tocante aos exemplos, novamente se observa que sua presença é eventual, podendo até mesmo formar parte da instrução da nota de uso, como acontece s.v. *amazing* (LWord2, 2008). No caso de *library* (OxED2, 2012), o exemplo é prescindível, sendo facultativo em *avoid* (CoED2, 2014).

Da mesma forma que na orientação prescritiva, só nos foi possível fazer julgamentos subjetivos a respeito dos exemplos. São necessários mais estudos sobre o tema para desenvolvê-lo. A decisão mais sensata é que, em caso de dúvida sobre a pertinência de um exemplo, o melhor a fazer é incluí-lo.

7.1.2.5 Recursos tipográficos

O último elemento relativo ao aspecto formal das notas de uso são os recursos tipográficos empregados para destacar palavras usadas em função metalinguística e os eventuais exemplos presentes nas notas. O Señas (2001) usa apenas o tipo itálico para destacar palavras em função metalinguística e nos exemplos. Os dicionários de aprendizes de inglês analisados, por outro lado, empregam uma variedade maior de tipos e a variação entre eles é pequena. Em geral, ULs usadas em função metalinguística nas notas aparecem em negrito e os exemplos são apresentados em itálico. No caso de exemplos errôneos, como os do CaED2 (2011), usa-se o tipo tachado. Algumas vezes, as obras usam aspas para indicar os exemplos errôneos em função metalinguística, como em “Don’t say ‘It’s very amazing.’ [...]” (LWord2, 2008, s.v. *amazing*), bem como os sintagmas corretos, como em “If you are talking about one item, you say ‘a piece of equipment’” (OxED2, 2012, s.v. *equipment*).

A decisão mais adequada a se tomar em relação à tipografia é a de que se uniformize e simplifique o uso dos recursos tipográficos. Sendo assim, propomos que o negrito seja usado primeiramente nas ULs e sintagmas corretos que sejam objeto da nota e que apareçam em função metalinguística nas instruções. Em segundo lugar, propomos que o tipo negrito também seja usado para destacar a UL ou sintagma que é alvo da nota de uso quando este aparece nos exemplos (se a nota os tiver). No tocante ao tipo itálico, e seguindo o padrão da maior parte dos dicionários analisados, propomos que esse recurso siga sendo usado nos exemplos para diferenciá-los da instrução da nota. Eventuais palavras e sintagmas que constem nas instruções da nota em função metalinguística, mas que não sejam o elemento principal desta, podem estar em itálico também. Por fim, e no que concerne ao tipo tachado, propomos que siga sendo usado para marcar ULs e sintagmas errôneos, seja nas instruções da nota, seja nos exemplos.

A seguir, há um resumo das propostas que fazemos sobre o aspecto formal das notas:

Quadro 47 – Propostas para o aspecto formal das notas de uso

Elemento	Descrição
Recursos iconográficos	Ícone: Para o OxD2 (2012), pode-se manter a opção empregada pela obra. Para os demais, sugere-se o ícone ✎. Fundo: em cor diferenciada em relação ao fundo do verbete.
Etiquetas	ORTOGRAFIA (dic. de inglês) / ORTOGRAFÍA (dic. de espanhol) PRONÚNCIA (dic. de inglês) / PRONUNCIACIÓN (dic. de espanhol) VOCABULÁRIO (dic. de inglês) / VOCABULARIO (dic. de espanhol): diferença semântica entre duas ULs; sinônimos; campos léxicos. USO (dic. de inglês) / USO (dic. de espanhol): registros de uso, flexão e padrões sintáticos que não possam ser tratados em outros segmentos informativos; outras informações para a compreensão do significado ou identificação do referente.
Língua	Português em notas de dicionários de inglês Espanhol em notas de dicionários de espanhol
Modalidade de orientação e padrão sintático	Orientação indicativa: para informações novas. Verbos no presente do indicativo em orações afirmativas; uso de voz ativa ou passiva. Exemplo facultativo. Orientação exortativa prescritiva e proibitiva: repetição da descrição de um fato linguístico. Uso do imperativo seguido de orientação indicativa. Exemplo sempre que possível.
Recursos tipográficos	Negrito: elementos que são objeto da nota, seja em função metalinguística nas instruções, seja para destacá-los nos exemplos. <i>Itálico:</i> exemplos; outros elementos em função metalinguística que não são objeto principal da nota. Tachado: elementos errôneos em qualquer parte da nota.

Fonte: Elaboração da autora

As propostas resumidas acima têm como objetivo facilitar a localização das notas de uso e a compreensão de seu conteúdo, de modo a tornar esse segmento informativo mais didático. No caso dos dicionários de espanhol, o espanhol pode ser mantido nas etiquetas que introduzem a nota em função da convergência ortográfica e semântica entre o espanhol e o português¹²⁰. A exemplificação das propostas acima será feita mais adiante em seções específicas sobre os dicionários de aprendizes de inglês e espanhol.

Na presente seção 7.1, tratou-se da função e do aspecto formal das notas de uso em dicionários de aprendizes. O último aspecto a ser observado nas notas de uso é o seu conteúdo. Conforme já comentado e exemplificado, uma nota pode conter informações novas ou repetir algo já descrito no verbete. Nas notas com informações novas, o foco recai sobre a língua-objeto. Por outro lado, nas notas que repetem outras informações já presentes no verbete, o foco

¹²⁰ Os pares espanhol-português *ortografía-ortografia*, *vocabulario-vocabulário* e *uso-uso* são 100% similares entre o espanhol e o português, o que permite que se use as etiquetas em espanhol. O mesmo ocorre com o par *pronunciación-pronúncia*, que é 73% similar segundo o NIM (2013).

não recai sobre a descrição da língua – embora esse aspecto esteja intrinsecamente presente –, mas sim sobre o usuário. Trata-se de uma resposta às necessidades desse usuário, que não são outras que as suas dúvidas e erros cometidos ao usar a língua-objeto. Esse segundo tipo de conteúdo só pode ser identificado e se tornar uma informação discriminante se estiver amparado em um perfil de usuário específico e se houver fontes de dados sobre esse perfil. Considerando que a tese versa sobre o perfil de usuário, atemo-nos à seleção de conteúdo para responder às necessidades desse perfil.

No caso dos aprendizes brasileiros, ao menos três fontes podem ser usadas para coletar dados: 1) os *corpora* de aprendizes, nos quais é possível analisar as dificuldades a partir do registro dos erros dos aprendizes; 2) os fóruns de aprendizagem de línguas disponíveis na internet, nos quais é possível analisar as dúvidas de falantes de determinada LM em relação a uma certa LE; e, na ausência dos recursos anteriores, ou frente a algum tipo de limitação, 3) uma comissão de professores com experiência no ensino das LEs em questão ao público-alvo do dicionário, tal como já mencionado em capítulos anteriores. Como a disponibilidade de recursos varia conforme a língua, não é possível estabelecer uma metodologia geral de coleta e análise de informações para constar nas notas de uso.

As seções a seguir tratam das notas de uso adaptadas ao usuário brasileiro adulto A1-B1 de forma separada no que diz respeito aos dicionários de aprendizes de inglês (7.2) e espanhol (7.3). Primeiramente, apresentamos algumas fontes de dados úteis para a coleta de informações acompanhadas de sugestões de metodologia para extração de dados. Em seguida, coletam-se amostras de dados. Ao final, são aplicadas as propostas feitas anteriormente a respeito da redação e *layout* das notas.

7.2 Notas de uso para dicionários de aprendizes brasileiros de inglês

As propostas relativas à adaptação de um dicionário de aprendizes de inglês A1-B1 ao público brasileiro dizem respeito tanto ao didatismo como à inclusão de informações discriminantes para esse público. O didatismo está atrelado à redação em português, enquanto o fato discriminante corresponde ao menos às dúvidas e erros dos aprendizes brasileiros.

Como o aspecto formal das notas de uso dos dicionários de aprendizes de inglês envolve a língua a ser usada na descrição, e tendo em vista que foi proposta a utilização do português, julgou-se pertinente tratar, primeiramente, da adaptação de uma pequena amostra de notas de uso já presentes nas obras consultadas. Em seguida, entra-se no mérito da seleção do conteúdo.

7.2.1 Adaptação das notas de uso já existentes com vistas ao público brasileiro

Para ilustrar as adaptações das notas já existentes, foram tomados como exemplo os seguintes verbetes:

Quadro 48 – Notas de uso em dicionários de aprendizes de inglês

<p>already /ə:'redi/ <i>adverb</i> 1 used for showing that one thing happened before another thing: <i>The meeting had already finished when we arrived.</i> 2 used for showing that a situation exists now or that it started earlier than expected: <i>Most of the guests have already left.</i> LANGUAGE HELP already or yet? In British English, you usually use already and yet with the perfect tenses. <i>I've already seen that film. Have they arrived yet?</i> In American English, people often use a past tense. <i>She already told us about the party. I didn't have any breakfast yet.</i></p> <p style="text-align: right;">(CoED2, 2014, s.v. <i>already</i>)</p>	<p>arrive /ə'raɪv/ <i>verb (arriving, arrived)</i> A2 to get to a place: <i>We arrived in Paris at noon.</i> ° <i>It was dark when we arrived at the station.</i></p> <p style="text-align: center;">Common mistake: arrive somewhere</p> <p>Be careful to choose the correct preposition after arrive. You arrive at a place such as a building. <i>We arrived at the hotel just after ten o'clock.</i> You arrive in a town, city, or country. <i>They arrived in Tokyo on Wednesday.</i> <i>When did David arrive in Australia?</i> You arrive home, here, or there. You do not use a preposition when arrive is used before these words. <i>We arrived home yesterday.</i> <i>I had a lot of problems when I first arrived here.</i></p> <p style="text-align: right;">(CaED2, 2011, s.v. <i>arrive</i>)</p>
<p>cook¹ /kɒk/ <i>verb (cooks, cooking, cooked)</i> to prepare and heat food: <i>I have to go and cook dinner.</i> LANGUAGE HELP Cooking verbs: You roast meat in an oven. You bake bread and cakes. You can boil vegetables in hot water. You can fry meat and vegetables in oil. You can also grill meat directly under or over a flame.</p> <p style="text-align: right;">(CoED2, 2014, s.v. <i>cook</i>¹)</p>	<p>glad /glæd/ <i>adjective</i> happy about something ⇔ SAME MEANING pleased: <i>He was glad to see us.</i> VOCABULARY BUILDING You are usually glad or pleased about a particular event or situation: <i>I'm glad that he's feeling better.</i> ◇ <i>I'm pleased to say that you've passed your exam.</i> You use happy to describe a state of mind: <i>I always feel happy when the sun is shining</i> or before a noun: <i>a happy child</i></p> <p style="text-align: right;">(OxED2, 2012, s.v. <i>glad</i>)</p>

Fonte: CaED2 (2011), OxED2 (2012), CoED2 (2014)

Em todas as notas acima, há um acréscimo de informação nova ao verbete no que diz respeito à descrição linguística. Outra característica em comum é a utilidade dessas informações para tarefas de produção. Para a sua adaptação ao português, em alguns casos, é possível empregar a tradução sem grandes dificuldades, como em *already*. Nos demais casos, porém, as orientações usam determinadas ULs e combinatórias em função metalinguística encapsuladas no próprio texto da orientação. Nota-se que esse estilo de redação está presente de maneira especial no caso das combinatórias léxicas; uma simples tradução da nota do verbete *cook*, por exemplo, apagaria a descrição do fato linguístico objeto da nota. Frente a essas situações, foi

necessário elaborar uma solução alternativa, na qual o padrão sintático “you + (verbo auxiliar) + [UL em função metalinguística]”, encontrado em *arrive*, *cook* e *glad*, foi adaptado ao português por meio de “usa(-se) + [UL em função metalinguística] + com [substantivo] / ao [verbo e complementos] / para se referir a”. Caso algum padrão combinatório não fosse contemplado pela tradução, era explicitado em inglês, como em *cook* (fry **in** oil) e *arrive* (arrive + [lugar]). A seguir, ilustramos as notas já modificadas juntamente com o restante do verbete a que pertencem, a fim de fornecer também uma visão de conjunto das adaptações propostas:

Quadro 49 – Propostas de aprimoramento das notas de uso (inglês)

<p>already /ə:'redi/ <i>adverb</i></p> <p>1 usa-se para expressar que uma coisa aconteceu antes de outra; em português, <i>já</i>: <i>The meeting had already finished when we arrived.</i></p> <p>2 usa-se para expressar que uma dada situação existe agora ou que iniciou antes do esperado; em português, <i>já</i>: <i>Most of the guests have already left.</i></p> <p>✎ VOCABULÁRIO: ALREADY OU YET?</p> <p>Em inglês britânico, normalmente se usa already e yet com tempos perfeitos. <i>I've already seen that film. Have they arrived yet?</i> Em inglês americano, as pessoas frequentemente usam o passado. <i>She already told us about the party. I didn't have any breakfast yet.</i></p>	<p>arrive /ə'raɪv/ <i>verb (arriving, arrived)</i></p> <p>A2 chegar em um lugar; em português, <i>chegar a/em</i>: <i>We arrived in Paris at noon.</i> ◦ <i>It was dark when we arrived at the station.</i></p> <p>✎ USO: arrive + [lugar]</p> <p>Cuide para escolher a preposição correta após arrive. Use arrive at com locais, tais como um prédio.</p> <p><i>We arrived at the hotel just after ten o'clock.</i></p> <p>Use arrive in com cidades ou países.</p> <p><i>They arrived in Tokyo on Wednesday.</i></p> <p><i>When did David arrive in Australia?</i></p> <p>Use arrive com as palavras <i>house</i>, <i>here</i> ou <i>there</i>. Não se usa preposição quando arrive está junto dessas palavras.</p> <p><i>We arrived home yesterday.</i></p> <p><i>I had a lot of problems when I first arrived here.</i></p>
<p>cook¹ /kʊk/ <i>verb (cooks, cooking, cooked)</i></p> <p>preparar e esquentar comida; em português, <i>cozinhar</i>: <i>I have to go and cook dinner.</i></p> <p>✎ VOCABULÁRIO</p> <p>Verbos de culinária: usa-se to roast ao cozinhar carne em um forno. Usa-se to bake ao cozinhar pães e bolos. Usa-se to boil ao cozinhar vegetais em água quente. Usa-se to fry ao cozinhar carne e vegetais em óleo (<i>fry in oil</i>). Usa-se também to grill ao cozinhar carne sob ou sobre o fogo.</p>	<p>glad /glæd/ <i>adjective</i></p> <p>[pessoa] que está feliz por causa de algo; em português, <i>contente</i> ⇨ MESMO SIGNIFICADO pleased: <i>He was glad to see us.</i></p> <p>♀ VOCABULÁRIO: GLAD, PLEASED OU HAPPY?</p> <p>Normalmente se usa glad ou pleased para se referir a um determinado acontecimento ou situação: <i>I'm glad that he's feeling better.</i> ◊ <i>I'm pleased to say that you've passed your exam.</i></p> <p>Usa-se happy para descrever um estado de espírito: <i>I always feel happy when the sun is shining</i> ou antes de um substantivo: <i>a happy child</i></p>

Fonte: Elaboração da autora a partir dos dados do CaED2 (2011), OxEED2 (2012), CoED2 (2014) e LWord2 (2008)

Todas as modificações nas notas obedecem ao padrão de redação proposto em 7.1 em relação aos recursos iconográficos, etiquetas, idioma, modalidades de orientação, padrão sintático e recursos tipográficos. No tocante à tradução, a fórmula “usa(-se) + [UL em função metalinguística]” demonstrou ser apropriada para os casos em que a redação da nota usa a

segunda pessoa do singular *you* para incluir ULs e combinatórias em função metalinguística. Ainda assim, ainda é preciso testar a adequação da fórmula de redação proposta junto a uma amostra maior de notas e, eventualmente, elaborar soluções de redação alternativas.

As seções seguintes tratam da seleção de conteúdo para notas de uso a partir dos dados de brasileiros no *Cambridge Learner Corpus* (CLC, 2019). Dado que se trata de um *corpus* pouco conhecido em função do seu acesso restrito, julgamos pertinente dedicar uma seção exclusiva à apresentação dessa base de dados.

7.2.2 O Cambridge Learner Corpus (2019)

No que diz respeito ao conteúdo das notas de uso, a adaptação de qualquer um dos quatro dicionários de inglês analisados aos usuários brasileiros em níveis A2 e B1 de aprendizagem consiste na inclusão de notas em resposta às necessidades desse público, isto é, suas dúvidas e erros. Esse procedimento perpassa pela consulta a um *corpus* com dados de aprendizes brasileiros de inglês em níveis iniciais de proficiência. É importante salientar que os *corpora* de aprendizes não são comuns como os *corpora* gerais de língua, conforme se pode constatar a partir dos registros da Learner Corpus Association (<https://www.learnercorpusassociation.org>), entidade que se encarrega de divulgar *corpora* de aprendizes disponíveis para as mais diversas línguas. Mais raros ainda são os *corpora* de aprendizes brasileiros de inglês e de espanhol e que não estejam compostos somente por um gênero textual (<https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>). Por esse motivo, e ante a escassez de instrumentos dessa natureza, também consideramos válido recorrer à experiência de professores que dão aulas de inglês a alunos brasileiros com o perfil mencionado quando necessário. Para tanto, teria de ser conduzida uma pesquisa empírica junto aos docentes (sob a forma de questionários, por exemplo) para que, com base na sua percepção e experiência, respondessem quais erros e dúvidas são mais frequentes entre os alunos, se possível discriminando os dados por etapas de aprendizagem.

O único *corpus* de que temos ciência que contém produções de aprendizes brasileiros de inglês de níveis iniciais do CEFR (2001) é o Cambridge Learner Corpus (CLC, 2019), baseado nas provas de proficiência de Cambridge Assessment e processado pelo software Sketch Engine. O acesso ao CLC (2019) é restrito a indivíduos previamente autorizados. Contudo, é possível preencher um formulário com uma petição de acesso ao *corpus*, que será avaliada pelos responsáveis pela ferramenta. Afortunadamente, ao enviarmos uma petição, foi-nos concedido acesso ao CLC (2019), o que nos permitiu verificar quais eram os erros comuns

de aprendizes brasileiros que alcançaram os níveis A1, A2 e B1 de proficiência em inglês. Os dados gerais do *corpus*, bem como algumas das informações específicas pertinentes para esta tese estão dispostos a seguir. Os *documentos* equivalem a cada texto escrito nas provas de proficiência de Cambridge Assessment. As *palavras* são uma letra ou sequência de letras entre dois espaços, incluindo nomes próprios, e a exceção das abreviaturas.

Tabela 3 – Dados gerais e do Brasil no CLC (2019)

Filtro	Especificação	Documentos	Palavras (tokens)	Sinais gráficos (tokens)
País	Todos*	259.321	31.260.641	35.821.138
	Todos*, A1-C2	220.701	25.879.399	29.654.848
	Brasil	23.263 (9% do CLC; <i>ranking</i> geral: 3º)	2.220.291 (7% do CLC; <i>ranking</i> geral: 4º)	2.544.202 (7,1% do CLC; <i>ranking</i> geral: 3º)
	Brasil, A1-C2	21.431	1.960.748	2.246.795
País + Idade	Todos os países*, crianças** A1-C2	125.988	10.345.303	11.854.541
	Todos os países*, adultos*** A1-C2	93.931	15.420.194	17.669.788
	Brasil, crianças** A1-C2	15.649	1.122.198	1.285.913
	Brasil, adultos*** A1-C2	5.710	831.878	953.241

Fonte: CLC (2019)

Nota: *De um total de 199 nacionalidades de candidatos que prestaram provas de proficiência Cambridge. **Crianças: 9-17 anos exceto intervalo de idade “16-20”. ***Adultos: 18+ anos exceto intervalo “16-20”.

Legenda: Documentos = textos escritos nas provas de proficiência de Cambridge Assessment; Palavras = letra ou sequência de letras entre dois espaços, exceto abreviaturas; Sinais gráficos = qualquer sinal gráfico entre dois espaços, abrangendo de palavras a números e signos de pontuação. *Tokens* = número total de elementos em um *corpus* ou *subcorpus*.

Foram considerados adultos os indivíduos com idade a partir de dezoito anos, enquanto os indivíduos considerados “crianças” foram aqueles com idades entre nove (menor registro disponível) a dezessete anos. No CLC (2019), o filtro “idade” apresenta tanto opções de números isolados (17, 18, 19, assim por diante) como intervalos de idade (16-20, 20-25 e 21-30, por exemplo). Dado que o intervalo “16-20” extrapola os limites das categorias “adultos” e “crianças” estabelecidas para a análise de dados, as produções atreladas a essa característica foram desconsideradas. Trata-se de uma fração mínima de dados que não foram considerados.

No que diz respeito aos grupos da rubrica “País + Idade”, foram incluídos os dados apenas de documentos certificados, isto é, que foram classificados em algum nível de proficiência. Há documentos do CLC (2019) que não foram enquadrados em nenhum nível de proficiência por não obedecerem aos requisitos de certificação. Como o nível de proficiência

(ou performance, para usar os termos do CLC (2019)) é uma variável relevante para a tese, preferimos analisar somente os dados provenientes de documentos nivelados.

Como se pode observar do *ranking* gerado pelo próprio CLC (2019), parte significativa dos documentos do *corpus* corresponde a produções de brasileiros, o que é benéfico para as pesquisas acerca dos aprendizes brasileiros. As duas primeiras posições são ocupadas, respectivamente, por Itália (1º) e Espanha (2º). Entre os documentos de brasileiros certificados (A1-C2), observa-se que a proporção de documentos dos adultos é três vezes menor que a de crianças. Tal disparidade, no entanto, não afeta a qualidade da nossa pesquisa da produção de adultos brasileiros no CLC (2019), pois a média de palavras¹²¹ de um grupo para outro apresenta uma diferença significativamente menor: 145,6 palavras/documento no caso dos adultos frente a 71,7 palavras/documento no caso das crianças. Para compreender esses valores, há que se considerar que o volume e a extensão dos textos solicitados nas provas de proficiência de Cambridge Assessment naturalmente aumentam conforme a progressão de níveis da escala do CEFR (2001). Textos mais extensos são próprios de níveis superiores de proficiência, como B2 em diante, o que indica que os adultos brasileiros realizam provas e atingem tais níveis de proficiência com mais frequência do que as crianças brasileiras.

A proficiência atrelada a cada documento, aliás, é um dos critérios de busca que pode ser correlacionado aos filtros de país e idade. Os dados de produções de adultos brasileiros conforme o nível de performance atingido estão dispostos a seguir.

Tabela 4 – Dados de brasileiros adultos A1-C2 no CLC (2019)

Nível de proficiência	Documentos	Palavras	Sinais gráficos
A1	1.072	50.702	58.099
A2	1.884	155.875	178.616
B1	789	87.228	99.954
B2	821	163.917	187.831
C1	620	184.200	211.073
C2	524	189.956	217.668
Total absoluto	5.710	831.878	953.241
Subtotal A1-B1	3.745 (65,5%)	293.805 (35,3%)	336.669 (35,3%)
Subtotal B2-C2	1.965 (34,4%)	538.073 (64,6%)	616.572 (64,6%)

Fonte: CLC (2019)

Novamente, não foram incluídos na contagem os documentos (e respectivas palavras e sinais gráficos) aos quais não foi aferido um nível performance. O volume relativamente

¹²¹ O CLC (2019) considera *palavra* como uma letra ou sequência de letras entre dois espaços, exceto abreviaturas.

crescente de palavras e sinais gráficos conforme avança a progressão de níveis de proficiência, ainda que o volume de documentos diminua, confirma a relação diretamente proporcional entre o nível de proficiência atingido por cada documento e a extensão desses documentos. Consequentemente, e segundo indicam os dados acima, a maior quantidade de dados disponíveis de adultos brasileiros está atrelada ao intervalo de níveis B2-C2. O fato de que o intervalo que interessa para a tese (A1-B1) conta com uma parcela menor de palavras e sinais gráficos não é, portanto, necessariamente um demérito, mas sim, reflete a realidade do que se é capaz de produzir em cada nível e que foi possível coletar mediante as provas de proficiência de Cambridge Assessment. A disparidade em relação aos níveis superiores acaba sendo compensada pela diversidade de informantes – contabilizados pelo número de documentos – responsáveis por gerarem os dados que se tem.

Além de conter um número considerável de dados de produções aprendizes de inglês, o que torna o CLC (2019) um *corpus* de especial importância para a *Learner Corpus Research* (antiga Linguística Contrastiva) é o seu sistema de anotação de erros. Tal sistema, quando associado às funcionalidades do Sketch Engine, permite contabilizar os erros gerais do *corpus* e os que são atrelados a públicos específicos. A tabela a seguir reúne dados de referência e dados filtrados conforme o país, a performance atingida e a idade.

Para compreender o valor dos números absolutos dos erros em cada caso, foram adotados dois tipos de valores de referência: comparação dos erros de grupos adjacentes, por um lado, e comparação dos erros de um grupo com o total de sinais gráficos desse mesmo grupo, por outro lado, conforme dados da Tabela 4.

Tabela 5 – Estatísticas de erros em sinais gráficos no CLC (2019)

Perfil de aprendizes [país, performance, idade]	Nº absoluto de erros	Comparação com perfil sobrejacente	Comparação com nº de sinais gráficos do perfil
Todo o CLC	4.050.955	-	11,3% de todos os sinais gráficos
Todo o CLC, A1-C2	2.778.279	68,5% dos erros do CLC	9,3% dos sinais gráficos do perfil
Todos os países, A1-C2, 18+	1.579.548	56,8% dos erros do CLC A1-C2	8,9% dos sinais gráficos do perfil
Brasil	250.811	6,2% dos erros do CLC	9,8% dos sinais gráficos do perfil
Brasil, A1-C2	218.739	7,8% dos erros do CLC A1-C2	9,7% dos sinais gráficos do perfil
Brasil, A1-C2, 18+	84.300	38,5% dos erros do Brasil A1-C2	8,8% dos sinais gráficos do perfil
Brasil, A1, 18+	7.880	9,3%	13,5%

		dos erros do Brasil A1-C2, 18+	dos sinais gráficos do perfil
Brasil, A2, 18+	21.084	25% dos erros do Brasil A1-C2, 18+	11,8% dos sinais gráficos do perfil
Brasil, B1, 18+	9.845	11,6% dos erros do Brasil A1-C2, 18+	9,8% dos sinais gráficos do perfil
Brasil, B2, 18+	16.958	20% dos erros do Brasil A1-C2, 18+	9% dos sinais gráficos do perfil
Brasil, C1, 18+	16.352	19,3% dos erros do Brasil A1-C2, 18+	7,7% dos sinais gráficos do perfil
Brasil, C2, 18+	12.180	14,4% dos erros do Brasil A1-C2, 18+	5,5% dos sinais gráficos do perfil
Brasil, A1-B1, 18+	38.809	46% dos erros do Brasil A1-C2, 18+	11,5% dos sinais gráficos do perfil
Brasil, B2-C2, 18+	45.490	54% dos erros do Brasil A1-C2, 18+	7,3% dos sinais gráficos do perfil

Fonte: CLC (2019)

Nas comparações entre os dados de todo o CLC (2019) e do Brasil, percebe-se que a proporção de erros no total de sinais gráficos de cada perfil é muito semelhante. No caso dos adultos que atingiram níveis A1-C2, a diferença entre os dados de todo o *corpus* e os do Brasil é de apenas 0,1%. Isso indica que os brasileiros, em especial os adultos, seguem a tendência mundial de erros cometidos por adultos no CLC (2019).

Ainda mais interessante é a distribuição dos erros dos brasileiros adultos entre os seis níveis de proficiência. Os dados do CLC (2019) confirmam que, nesse caso em específico, há uma diminuição da incidência de erros conforme avançam os níveis de proficiência. Na comparação entre os grupos de brasileiros adultos A1-B1 e B2-C2, o primeiro responde por uma quantidade ligeiramente menor dos erros que o segundo (46% vs. 54%). Essa diferença se deve muito provavelmente à maior massa de sinais gráficos relativos ao grupo B2-C2, ao qual correspondem 64,6% de todos os sinais gráficos produzidos pelos brasileiros adultos A1-C2 (cf. Tabela 4).

Por outro lado, mais reveladores que os valores absolutos dos erros dos dois grupos são os valores relativos na comparação com o total de sinais gráficos produzidos por cada um. A proporção de erros no total de sinais gráficos do grupo de adultos brasileiros A1-B1 (11,5%) é maior que a proporção de erros tanto do grupo de adultos brasileiros B2-C2 (7,3%) como a do grupo de adultos brasileiros em geral A1-C2 (8,8%). Portanto, o grupo de aprendizes brasileiros

adultos A1-B1 é o que mais tem necessidade de materiais didáticos de referência que os auxiliem na resolução de dúvidas e correção de erros. Tais dados revelam o rumo que precisam tomar as pesquisas e iniciativas editoriais no que diz respeito à elaboração de materiais didáticos de ensino de inglês para adultos no Brasil e corroboram a hipótese de trabalho da tese de que os níveis A2 e B1 de aprendizagem estão entre os mais necessitados de materiais de apoio.

As estatísticas comentadas anteriormente só foram possíveis porque o CLC (2019) é um *corpus* com anotação de erros e porque a ferramenta que o manipula, o Sketch Engine, permite que se colem dados dessa natureza. Em termos de análise de erros, a anotação facilita a localização de erros e permite que se identifique e analise uma quantidade maior de dados de erros em menos tempo se comparado aos corpora sem anotação de erros, como o CAES (2020) do espanhol. É necessário observar que não é óbvio que um *corpus* de aprendizes seja anotado porque este não é um procedimento automático, mas sim manual, que envolve muito tempo e esforço, além dos altos custos financeiros (GRANGER, 2015b, 2017; DEL RÍO; MENDES, 2018). Até o momento, a identificação e anotação manual de erros é a única opção metodológica disponível para compilação de *corpora* de aprendizes com anotação de erros.

A anotação, por sua vez, pressupõe uma taxonomia de erros. Cada *corpus* anotado de aprendizes costuma elaborar uma taxonomia própria, embora, segundo Del Ríó e Mendes (2018, p. 4116), não haja diferenças significativas entre as propostas disponíveis. Os responsáveis pelo CLC (2019) elaboraram uma classificação com um total de 77 tipos de erros, que correspondem à combinação dos seguintes subconjuntos: a) tipo geral de erro (forma, ausência de algo, necessidade de substituição, redundância, flexão e derivação); b) classe de palavras envolvida (pronomes, conjunções, determinantes, adjetivos, substantivos, quantificadores, preposições, verbos e advérbios); c) erros de pontuação; d) erros de contabilidade de substantivos; erros de concordância; f) outros erros (estrutura argumental, colocação, fraseologia, registro diafásico-diastrático, ordenação de palavras, frases negativas, tempos verbais, ortografia, erros difíceis de classificar).

As buscas por erros no CLC (2019) podem ser conduzidas sob dois critérios: por tipos de erros (indiretas) e por erros em itens específicos (diretas). As duas seções a seguir tratam desses tipos de busca e a sua contribuição para a Lexicografia e para esta tese.

7.2.3 Seleção de conteúdo para notas de uso: buscas por tipos de erros no CLC (2019)

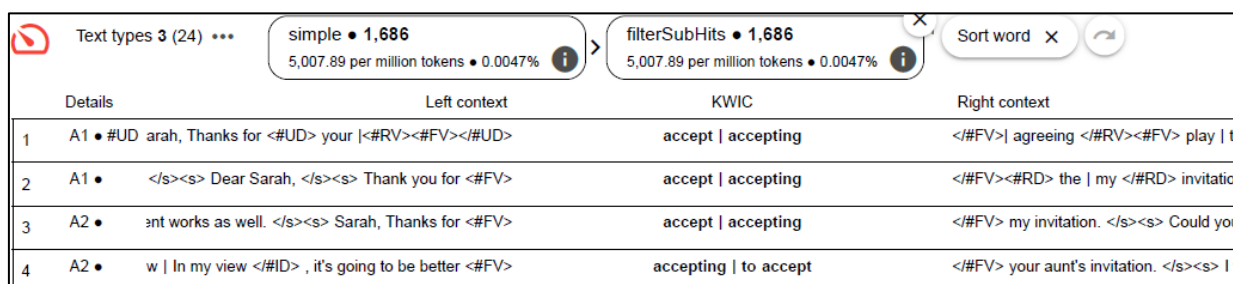
Na *Learner Corpus Research*, trabalhos sobre tipos de erros em *corpora* de aprendizes não são tão comuns muito provavelmente devido à baixa profusão de *corpora* com anotação de

erros – que, por sua vez, deve-se aos elevados custos em termos de recursos humanos e financeiros. Conforme já comentado no capítulo 4, a *Contrastive Interlanguage Analysis* (CIA) é a metodologia mais usada em LCR, pois se aplica à análise de *corpora* de aprendizes sem anotação de erros. A sua variante *Computer-aided Error Analysis* (CEA), por outro lado, é menos popular e se refere à análise de *corpora* de aprendizes com anotação de erros como o CLC (2019).

Dentre os trabalhos sobre análise de erros em *corpora* anotados de aprendizes, como Can (2017), MacDonald (2016) e Granger (2003), por exemplo, é comum que o objetivo da pesquisa seja apenas a geração de *rankings* de tipos de erros mais comuns. Nesses trabalhos, a interpretação dos dados não ultrapassa a constatação da frequência de cada tipo de erro e a comparação com dados de outras fontes, de modo que as constatações feitas são demasiado vagas e superficiais. Do ponto de vista prático do desenho e elaboração de materiais didáticos de ensino de línguas, de pouco adianta a informação de que erros de ortografia, pontuação e seleção lexical, por exemplo, são os mais frequentes em um determinado perfil de aprendizes. Acreditamos que a metodologia de análise dos trabalhos sobre tipos de erros precisa progredir do simples ranqueamento para a análise dos recursos e contextos linguísticos envolvidos nesses tipos de erros.

Ao fazer buscas por tipos de erros no CLC (2019) por meio do Sketch Engine, assinala-se o código correspondente a esse erro e selecionam-se os filtros pertinentes (país, nível de proficiência atingido, idade). Como resultado, são obtidos os alinhamentos nos quais todos os sinais gráficos anotados como manifestações do erro consultado estão ao centro (coluna *KWIC*), acompanhados da forma correta imediatamente à sua direita e do restante do alinhamento à esquerda (coluna *left context*) e à direita (coluna *right context*). A figura a seguir reproduz parte da tela de resultados de uma busca pelo tipo de erro relativo à forma nominal de verbos:

Figura 7 – Resultados do CLC (2019) na busca por tipos de erros



Details	Left context	KWIC	Right context
1	A1 • #UD arah, Thanks for <#UD> your <#RV><#FV><#UD>	accept accepting	<#FV> agreeing <#RV><#FV> play t
2	A1 • </s><s> Dear Sarah, </s><s> Thank you for <#FV>	accept accepting	<#FV><#RD> the my </RD> invitatio
3	A2 • ant works as well. </s><s> Sarah, Thanks for <#FV>	accept accepting	<#FV> my invitation. </s><s> Could yo
4	A2 • w In my view <#ID> , it's going to be better <#FV>	accepting to accept	<#FV> your aunt's invitation. </s><s> I

Fonte: CLC (2019)

A partir dos dados do CLC (2019), elaborou-se um *ranking* parcial dos 10 tipos erros mais frequentes cometidos por brasileiros adultos que atingiram níveis A1-B1 de performance/proficiência. Também foram incluídos outros dados sobre os erros computados no CLC (2019) para fornecer um panorama mais detalhado dos resultados.

Tabela 6 – Dez erros mais frequentes de brasileiros adultos A1-B1 no CLC (2019)

Tipo de erro	Nº absoluto	Comparação com erros
Total	38.809	46% dos adultos/BR A1-C2
1. RP: Substituir pontuação	4.198	10,8% dos adultos/BR A1-B1
2. S: Ortografia	2.566	6,6% dos adultos/BR A1-B1
3. RT: Substituir preposição	2.559	6,6% dos adultos/BR A1-B1
4. TV: Tempo verbal incorreto	2.417	6,2% dos adultos/BR A1-B1
5. RV: Substituir verbo	2.341	6,0% dos adultos/BR A1-B1
6. MP: Falta de pontuação	1.890	4,8% dos adultos/BR A1-B1
7. FV: Forma nominal do verbo incorreta	1.686	4,3% dos adultos/BR A1-B1
8. UP: Pontuação desnecessária	1.341	3,4% dos adultos/BR A1-B1
9. MA: Falta de elemento anafórico	1.241	3,2% dos adultos/BR A1-B1
10. MD: Falta de determinante	1.222	3,1% dos adultos/BR A1-B1

Fonte: CLC (2019)

Nota: As siglas indicativas dos erros são aquelas usadas pelo CLC (2019) para catalogá-los.

Chama a atenção o fato de que tópicos de pontuação ocupam três das dez primeiras posições do *ranking* de erros do perfil de aprendizes analisado: (1) necessidade de substituir a pontuação, (6) falta de pontuação e (8) pontuação desnecessária. A pontuação abarca os signos de pontuação (32%), o uso de maiúsculas e minúsculas (67%) e os símbolos monetários e grafia dos números (1%) (CLC, 2019). Um *ranking* completo dos erros se encontra no Anexo 14.

Erros de pontuação, bem como alguns dos erros relativos às anáforas (MA) e à seleção do tempo verbal adequado (TV), só poderiam ser incluídos no dicionário como compêndios léxico-gramaticais em algum dos textos externos à macroestrutura, ou seja, o *front matter* ou o *back matter*. As razões para isso são o espaço limitado nos verbetes e a dificuldade em atrelar esses temas a um ou outro verbete específico. Embora o tratamento dos aspectos supracitados não seja obrigatório nem constitua a razão de ser de um dicionário, nos dicionários de aprendizes de inglês, essa é uma prática comum e, na nossa opinião, benéfica para o aprendiz – cf. Borba (2020a), por exemplo, sobre a alta relevância das informações do *back matter* do CCID4 (2018) sob o ponto de vista do aprendiz brasileiro de inglês. É altamente recomendável, portanto, que o desenho de um dicionário de aprendizes brasileiros de inglês A1-B1 inclua essas informações em compêndios gramaticais.

Os demais tipos de erros elencados no *ranking* correspondem a informações que ou já costumam constar nos verbetes de dicionários de aprendizes de inglês ou que, em alguma medida, podem ser incluídas nos verbetes caso não estejam ainda presentes.

O fato de que os erros de ortografia ocupam a segunda posição no *ranking* é um importante indicador da necessidade de materiais didáticos de referência, em especial os dicionários, para servir de apoio em relação a esse nível de organização da linguagem. Esses erros serão comentados mais detalhadamente nas próximas páginas.

Na terceira e quinta posições, estão os erros na seleção de preposições e verbos. Normalmente, os erros envolvendo a seleção léxica equivocada possuem duas dimensões complementares: sinonímia e combinatórias léxicas. Por um lado, a confusão com ULs sinônimas revela que o aprendiz não percebe as nuances de significado de cada UL envolvida, como é o caso da preposição *for* ao invés de *about* (24 registros) e do verbo *to use* empregado no lugar de *to wear* (35 registros) (CLC, 2019). Seja por influência da LM, seja por falta de armazenamento de determinados traços de significado no sistema de representações semânticas da memória, ou ambas as razões, o aprendiz termina usando uma UL no lugar de outra. De maneira muitas vezes complementar ao anterior, estão as combinatórias léxicas dessas ULs, que, muitas vezes, demarcam as nuances de significado ou de função de cada uma. Os erros envolvendo *for* e *about* ocorreram junto a adjetivos que expressam emoções, precisamente *happy* (11 registros), *sorry* (7 registros), *excited* e *pleased* (3 registros cada), quando manifestados com o padrão (*to be*) [adjetivo emocional] $\text{f}\text{æ}\text{r}$ | *about* [algo] (CLC, 2019). Os erros envolvendo *to use/to wear* dizem respeito à seleção do verbo equivocado para acompanhar roupas e acessórios. Outro exemplo no qual o significado está atrelado a determinadas combinatórias léxicas é o uso errôneo do verbo **to say* ao invés de *to tell* (15 registros). No inglês, *to say* se refere ao ato de se manifestar com palavras, ocorrendo comumente sob o padrão *to say* [algo]; já *to tell* se refere ao conteúdo dessa manifestação, ocorrendo sob o padrão *to tell* (alguém) [algo] (cf. LWord2 (2008, s.v. *tell*)).

Na quarta posição, encontram-se os erros relativos ao uso de tempos verbais. Esse tipo de erro pode revelar dois problemas: escolha do tempo verbal equivocado, como **come* ao invés de *came* (14 ocorrências) e **I'll go* ao invés de *I'm going* (17 ocorrências) (CLC, 2019); e erros relativos ao uso de verbos auxiliares e modais, como **Do you like* ao invés de *Would you like* (7 ocorrências) e **can* no lugar de *could* (48 ocorrências) (CLC, 2019). Esse tipo de erro parece ser produto do conhecimento vacilante em relação aos tempos verbais e aos verbos auxiliares e modais. No que diz respeito à forma, os dicionários de aprendizes de inglês já indicam a flexão no comentário de forma (3ª pessoa do singular, gerúndio, pretérito e

particípio), além de incluir os pretéritos e particípios irregulares na macroestrutura. No que concerne aos tempos verbais, como não estão relacionados a ULs específicas, o único lugar possível para esse tipo de informação em um dicionário de aprendizes é no *back matter*, sob a forma de compêndios gramaticais. No caso dos auxiliares e modais, é possível inclui-los como verbetes e discriminar suas funções por metalinguagem de signo nas acepções, além de incluir outras informações sob a forma de notas de uso, conforme sugere Farias (2013) para essas subclasses de palavras.

Na sétima posição, estão os erros envolvendo as formas nominais dos verbos. Novamente, a falta de conhecimento sobre combinatórias léxicas tem um papel na incidência de erros envolvendo o uso de infinitivos, particípios e gerúndios. Vinte e um casos de diferentes verbos no infinitivo, por exemplo, foram catalogados como errados porque a expressão “Thank you for” (e suas variações morfológicas), que os antecedia, requer que o verbo que a acompanha esteja flexionado no gerúndio (*Thank you for writing*, e não *Thank you for *write*) (CLC, 2019). Portanto, só é possível analisar esse tipo de erro a partir das ULs que o acompanham.

Na nona posição, os elementos anafóricos aos quais o erro se refere são o uso de pronomes pessoais, possessivos, relativos, interrogativos, demonstrativos e indefinidos (quantifiers).

Na décima posição, encontram-se os erros relativos à omissão de determinantes, isto é, artigos (*a, the*), que correspondem a 94% dos erros desse tipo (CLC, 2019); e pronomes possessivos (*her, his, your, our*) e demonstrativos (*this, that, these, those*), que correspondem aos 6% restantes (CLC, 2019). Os erros de omissão de pronomes terminam, em parte, distribuídos entre as categorias de falta de elemento anafórico e falta de determinante.

Para a presente oportunidade, não havia tempo hábil para analisar todas as categorias de erros do CLC (2019) referentes aos brasileiros adultos com performance A1-B1, pois isso já constituiria um trabalho acadêmico *per se*. Estamos cientes de que a principal tarefa da tese ao utilizar o CLC (2019) não era analisar exaustivamente a massa de dados disponível, mas sim testar uma opção metodológica no marco da LCR que fosse pertinente para fins lexicográficos. Para tanto, foi escolhido o tipo de erro S (ortografia), que é o segundo mais frequente e passível de tratamento sob a forma de notas de uso. Como metodologia de coleta e análise de dados, adotou-se a perspectiva *corpus-driven*, cuja premissa, segundo Tognini-Bonelli (2001, p. 85), é a de que os dados do *corpus* são analisados para elaboração *a posteriori* de descrições e teorias linguísticas. A seguir, estão dispostos os casos mais frequentes de representações ortográficas grafadas erroneamente pelos brasileiros adultos A1-B1. Para auxiliar na análise dos dados, foi adicionado o nível atribuído pelo EVP (2015). Houve casos em que uma mesma representação

ortográfica apareceu como diferentes ULs, que, por sua vez, estavam atreladas a diferentes níveis de referência no EVP (2015). A representação ortográfica *comfortable*, por exemplo, é classificada como A2 quando usada para descrever roupas e mobília (EVP, 2015); quando se refere à personalidade de alguém, é B1 (EVP, 2015); quando alude a um estado de espírito em que não há preocupações, é B2 (EVP, 2015); e quando se refere a uma situação financeira que permite uma boa qualidade de vida, é C2 (EVP, 2015). Como o objeto de análise eram as representações ortográficas independentemente do significado que veiculavam, tomou-se como referência o nível mais baixo (no caso de *comfortable*, A2) porque consideramos que se trata de uma propriedade a ser usada desde a aprendizagem do primeiro significado atrelado a ela.

Tabela 7 – Dez erros ortográficos mais frequentes dos brasileiros adultos A1-B1 no CLC (2019)

Representação ortográfica e nível EVP	Nº total de erros	Nº de erros por nível	Grafias errôneas mais frequentes
<i>comfortable</i> A2	59	A1: 20	18x * <i>comfortable</i>
		A2: 23	23x * <i>comfortable</i>
		B1: 16	16x * <i>comfortable</i>
<i>which</i> A1	48	A1: 4	4x * <i>wich</i>
		A2: 23	12x * <i>wich</i>
		B1: 21	19x * <i>wich</i>
<i>beautiful</i> A1	41	A1: 4	2x * <i>beautifull</i>
		A2: 23	10x * <i>beautifull</i> ; 13x * <i>beatiful</i>
		B1: 14	4x * <i>beautifull</i> ; 10x * <i>beatiful</i>
<i>because</i> A1	30	A1: 9	5x * <i>becouse</i>
		A2: 15	11x * <i>becouse</i> ; 3x * <i>beacause</i>
		B1: 6	2x * <i>becouse</i> ; 3x * <i>becaus</i>
<i>believe</i> A2	27	A1: 2	1x * <i>bilieve</i> ; 1x * <i>bealive</i>
		A2: 18	15x * <i>belive</i>
		B1: 7	5x * <i>belive</i>
<i>different</i> A1	25	A1: 1	1x * <i>diferent</i>
		A2: 17	15x * <i>diferent</i>
		B1: 7	6x * <i>diferent</i>
<i>hi</i> A1	21	A1: 12	7x * <i>hai</i> ; 4x * <i>hy</i>
		A2: 6	2x * <i>hai</i> ; 4x * <i>hy</i>
		B1: 3	2x * <i>hy</i>
<i>address</i> A1	20	A1: 10	10x * <i>adress</i>
		A2: 6	6x * <i>adress</i>
		B1: 4	4x * <i>adress</i>
<i>together</i> A1	17	A1: 2	2x * <i>togheter</i>
		A2: 11	6x * <i>togheter</i> ; 3x * <i>toghether</i>
		B1: 4	4x * <i>togheter</i>
<i>disappoint</i> B1	17	A1: 0	-
		A2: 17	7x * <i>dissapoint</i> ; 4x * <i>disapoint</i>
		B1: 0	-

Fonte: Elaboração própria da tabela com base nos dados do CLC (2019) e do EVP (2015)

Nota: Os números de erros se referem somente aos erros de ortografia

É possível supor duas motivações para os erros acima e que podem, inclusive, estar interrelacionadas em alguns casos: influências interlinguísticas do português (*comfortable, beautiful, different, disappoint, hi, address*); a baixa correspondência ortográfico-fonológica do inglês (*which, beautiful, believe, hi*); e um meio-termo entre as motivações anteriores (*because, together*).

O mais evidente candidato a influências interlinguísticas do português é **comfortable* [comfortable], pela influência do padrão de uso da letra *m* antes das letras *p* e *b*. Além disso, embora a representação ortográfica *comfortable* seja usada pelos aprendizes de inglês em geral a partir do nível A2, a similaridade ortográfica de 62% (NIM, 2013) em relação a *confortável* do português é um forte indicativo para explicar o uso a partir do nível A1 por parte dos brasileiros. Em **beautiful* [beautiful], a ausência da sequência de letras *eau* nas palavras do português é uma hipótese para justificar a omissão da letra *u*. Em **hai* [hi], por sua vez, emprega-se a associação entre fonemas e letras do português. Em *different, address* e *disappoint*, os problemas envolvem a duplicação de consoantes cuja duplicação, à exceção do dígrafo *ss*, é alheia aos padrões do português. Observe-se que a grafia **adress* parece revelar uma maior familiaridade com a duplicação do *s* em função do conhecimento prévio do português. Na grafia **dissappoint*, por outro lado, além da falta de duplicação do *p*, observa-se que foi estabelecida uma relação entre o fonema /s/ do inglês e a sua representação ortográfica em português pelo dígrafo *ss*.

Entre os casos de baixa convergência ortográfico-fonológica, por um lado, são omitidas letras que não têm respaldo sonoro, como o *h* no dígrafo *wh* em **wich* [which] e o *e* mudo em **belive* [believe]. Por outro lado, em **hy* e **beautiful*, há uma vacilação em relação à diversidade de representações de um mesmo fonema ou sequência de fonemas. Tendo em vista que as sequências de fonemas /ai/ e /fol/ de *hi* e *beautiful* são representadas ortograficamente com outras letras em palavras como *by* e *full*, tal variação produz o contexto perfeito para confusões e influências entre umas e outras representações ortográficas dos mesmos fonemas.

Finalmente, e a meio caminho entre as influências interlinguísticas do português e a baixa correspondência ortográfico-fonológica, estão as grafias **becouse* [because] e **togheter* [together]. A vacilação na ortografia pode se dever ao fato de que os fonemas /ɔ/ e /ɒ/ podem ser representados no inglês pela letra *o* (e.g. *course, lot*) e pelo fato de que o fonema /g/ pode ser representado tanto pelo dígrafo *gh* (e.g. *ghost*) como pela letra *g* (e.g. *get*). Também é possível que os brasileiros tenham feito analogias com o português. No caso de dos fonemas /ɔ/ e /ɒ/, estes podem ter sido associados à correspondência letra-fonema da letra *o* no português. Já a sequência fonêmica /ge/ pode ter sido associada à sua representação em português por meio

do dígrafo *gue* (e.g. *guerra*); nesse caso, a inserção do *h* no inglês “supriria” a falta da intermediária *u* associada à produção de /ge/.

As considerações e hipóteses *ad supra* sobre os erros ortográficos dos brasileiros são úteis sob a perspectiva didático-pedagógica de sala de aula, isto é, para que os professores de inglês tenham ciência de algumas das dificuldades mais corriqueiras do público adulto brasileiro A1-B1 e possam estar preparados quando essas dificuldades surgirem. No que diz respeito à lexicografia, as buscas por ULs mais comumente sujeitas a um determinado tipo de erro são úteis para a compilação de obras especializadas em erros e dúvidas, como o *Common Errors in English* (WALTER; WOODFORD, 2015), que tem uma seção especial para erros de ortografia, bem como seções sobre pontuação, tempos verbais, entre outros os temas alvo de erros de aprendizes de inglês.

No entanto, um dicionário de aprendizes não trata de tipos de erros como o faz um dicionário de dúvidas e dificuldades como o de Walter e Woodford (2015); um dicionário de aprendizes descreve ULs e sintagmas e, eventualmente, os erros associados a eles. Nesse sentido, os dados absolutos das ULs (como *comfortable*, *hi* etc.) que comumente são alvo de um certo tipo de erro (como a ortografia) fornecem apenas um panorama parcial sobre certas ULs, ao invés de um diagnóstico integral das necessidades dos aprendizes em relação a cada UL. Por exemplo: embora *hi* integre o grupo das ULs mais sujeitas a erros ortográficos, com 21 registros errôneos, esse dado perde relevância ao se ter em vista que o total de ocorrências de *hi* no grupo dos brasileiros adultos A1-B1 é de 892. Em outras palavras, se, do ponto de vista ortográfico, os erros em *hi* são altamente relevantes, em uma perspectiva de frequência de ocorrências, esse mesmo índice de erros perde relevância frente aos demais 871 registros. Frente a esse cenário, é evidente que não é interessante acrescentar uma nota de uso em *hi* a respeito da vacilação ortográfica dos aprendizes brasileiros. Acrescenta-se ainda o fato de que não se sabe se *hi* terá outros tipos de erros ou manifestações errôneas específicas que sejam mais frequentes – e, por isso, mais relevantes para uma nota de uso – que os erros ortográficos.

Portanto, embora útil para outras classes de dicionários, a busca e análise de tipos de erros, quando desacompanhada de outros dados de frequência, não responde às necessidades de descrição de erros em um dicionário de aprendizes sob a forma de notas de uso. Em vista disso, as dez representações ortográficas mais sujeitas a erros por parte dos brasileiros foram analisadas novamente sob o prisma da frequência de uso por parte dos brasileiros adultos A1-B1 (cf. tabela na próxima página).

Tabela 8 – Análise relativa dos dez erros ortográficos mais frequentes

Item	Erros / frequência absolutos	Erros / frequência relativos
<i>comfortable</i> A2	59/109 (54%)	A1: 20/31 (64%)
		A2: 23/47 (49%)
		B1: 16/31 (51,6%)
<i>which</i> A1	48/353 (13,5%)	A1: 4/15 (26,6%)
		A2: 23/156 (14,7%)
		B1: 21/182 (11,5%)
<i>beautiful</i> A1	41/348 (11,7%)	A1: 4/47 (8,5%)
		A2: 23/216 (10,6%)
		B1: 14/85 (16,4%)
<i>because</i> A1	30/1.702 (1,7%)	A1: 9/356 (2,5%)
		A2: 15/940 (1,6%)
		B1: 6/406 (1,4%)
<i>believe</i> A2	27/170 (15,8%)	A1: 2/16 (12,5%)
		A2: 18/97 (18,5%)
		B1: 7/57 (12,2%)
<i>different</i> A1	25/287 (8,7%)	A1: 1/21 (4,7%)
		A2: 17/171 (10%)
		B1: 7/95 (7,3%)
<i>hi</i> A1	21/892 (2,3%)	A1: 12/322 (3,7%)
		A2: 6/483 (1,2%)
		B1: 3/87 (3,4%)
<i>address</i> A1	20/42 (47,6%)	A1: 10/14 (71,4%)
		A2: 6/17 (35,2%)
		B1: 4/11 (36,3%)
<i>together</i> A1	17/166 (10,2%)	A1: 2/18 (11%)
		A2: 11/102 (10,7%)
		B1: 4/46 (8,6%)
<i>disappoint</i> B1	17/117 (14,5%)	A1: 0
		A2: 17/113 (14,5%)
		B1: 0/4

Fonte: Elaboração própria da tabela com base nos dados do CLC (2019) e do EVP (2015)

Nota: Os números de erros se referem somente aos erros de ortografia

Os dados absolutos da razão entre erros e frequência de uso dos itens analisados permitem reavaliar a dimensão dos erros cometidos pelo grupo de brasileiros adultos A1-B1. Por um lado, há erros de ortografia que são claramente relevantes pela sua porcentagem próxima ou superior a 50% da frequência total de uso, como são os erros em *comfortable* e *address*. Por outro lado, há também erros claramente irrelevantes quando comparados ao número de vezes em que as representações ortográficas em questão foram utilizadas. Trata-se dos erros ortográficos com *because* e *hi*, que apresentam as porcentagens mais baixas.

Os demais itens analisados, por sua vez, encontram-se em uma “zona cinza”, pois, sem um critério de referência, não se pode necessariamente julgá-los como relevantes ou irrelevantes. Uma opção seria a adoção da porcentagem dos erros de ortografia (n. 3.002) no

subcorpus de adultos brasileiros A1-B1 (n. 336.669), isto é, 0,9%, como ponto de referência sob a premissa de que esse valor indicaria o que se pode esperar em termos de erros de ortografia, isto é, o que é “normal” ou “aceitável”. Sob essa perspectiva, proporções de erros ortográficos superiores a 0,9% seriam anômalas e dignas de atenção lexicográfica.

No entanto, e no que diz respeito aos dados dos brasileiros adultos A1-B1, o índice de 0,9% abarca uma enorme quantidade de erros ortográficos, incluindo casos como *because* e *hi*. Se, por um lado, essa seria uma alternativa a ser melhor investigada e avaliada no marco da LCR, para a Metalexigrafia, a medida culminaria em um sem-fim de notas de uso. Falta, pois, um critério para avaliação do que é mais ou menos relevante para fins lexicográficos. Considerando que nem os trabalhos em LCR nem os compiladores de dicionários de aprendizes de inglês esclarecem o que é um relevante, a decisão mais segura seria incluir os erros relacionados a *comfortable* e *address* sob a forma de notas de uso sob a justificativa de que os registros ortográficos errôneos se aproximam ou superam a marca de 50% da frequência de uso.

Dos quatro dicionários de aprendizes de inglês considerados na tese, três incluem notas de uso que chamam a atenção para a representação ortográfica de *comfortable* ou *address* – nenhum dicionário apresentava notas para os dois casos.

Quadro 50 – Notas de uso sobre a ortografia de *comfortable* e *address* nos dicionários

<p>Common mistake: comfortable</p> <p>Learners often spell comfortable wrong. Remember that it begins with com.</p> <p>(CaED2, 2011, s.v. <i>comfortable</i>)</p>	<p>SPELLING</p> <p>Remember! You spell address with DD and SS.</p> <p>(OxED2, 2012, s.v. <i>address</i>)</p>	<p>SPELLING</p> <p>This word is often spelled wrongly. The correct spelling is: address.</p> <p>(LWord2, 2008, s.v. <i>address</i>)</p>
--	--	---

Fonte: CaED2 (2011), OXED2 (2012) e LWord2 (2008)

As notas do CaED2 (2011) e do OxED2 (2012) contêm orientações exortativas prescritivas, enquanto a nota do LWord2 (2008) apresenta uma orientação indicativa. Dado que em ambos os casos há uma forma ortográfica errônea que se destaca nas produções dos brasileiros adultos A1-B1 segundo o CLC (2019), seria interessante mencioná-la como exemplo de erro comum do público considerado na tese. A seguir, revisamos as três notas conforme as propostas realizadas no início do capítulo.

Quadro 51 – Adaptação das notas de uso de *comfortable* e *address*

<p>ORTOGRAFIA</p> <p>Os brasileiros muitas vezes escrevem <i>comfortable</i>. Lembre-se de que a palavra inicia com com.</p>	<p>ORTOGRAFIA</p> <p>Lembre-se! A palavra address se escreve com DD e SS, e não <i>adress</i>.</p>	<p>ORTOGRAFIA</p> <p>A escrita correta é address. Não escreva <i>adress</i>.</p>
--	--	--

Fonte: Elaboração da autora a partir dos dados do CaED2 (2011), OxED2 (2012) e LWord2 (2008)

Há muitas formas de adaptar as notas anteriores. Como a representação ortográfica é uma informação já presente no verbete, sugerimos que todas as notas contenham orientação prescritiva ou proibitiva. A menção à grafia equivocada ajuda o consultante a entender melhor por que a nota se fez necessária e para que tipo de manifestação errônea ela deseja alertar. Naturalmente, a adaptação dos dicionários requer não somente a modificação das notas existentes, mas também a inclusão de notas sobre a ortografia de *comfortable* ou *address* caso não apareçam no dicionário que se decida adaptar.

Finalmente, os valores relativos de erros e frequência de uso calculados por nível são úteis para, opcionalmente, discriminar na nota de uso em que nível ou níveis o erro em questão costuma ser cometido. Do ponto de vista da LCR, é interessante notar que a distribuição proporcional dos erros de cada item entre os três níveis assinalados varia consideravelmente. Em alguns casos, os erros se concentram majoritariamente em um determinado nível, como se nota em *different* (A2), *address* (A1) e *disappoint* (A2). Em outros casos, há uma distribuição mais homogênea, com menos diferenças entre os três níveis, como ocorre com os erros em *believe* e *together*. Um terceiro cenário são os erros crescentes (*beautiful*) e decrescentes (*which*) entre os três níveis de proficiência. A investigação desses casos no marco da CEA seria de grande valia, principalmente sob a perspectiva pedagógica.

7.2.4 Seleção de conteúdo para notas de uso: buscas por itens específicos no CLC (2019)

Na LCR, a busca por erros em determinadas ULs ou representações ortográficas se propõe a identificar os usos e erros que são próprios de falantes de uma certa LM ou comuns a falantes de diversas LMs por meio da comparação de dados entre grupos (GILQUIN; GRANGER, 2015; DE COCK; GRANGER, 2005). Outra perspectiva de análise é a investigação de casos *overuse* como de *underuse* de ULs, combinatórias e estruturas, nos quais se compara a frequência no *corpus* de aprendizes com a frequência em um *corpus* geral de língua (GRANGER, 2017; DE COCK; GRANGER, 2005). Para este trabalho, a perspectiva de

análise é a dos erros cometidos pelo grupo de brasileiros adultos A1-B1, sem estabelecer comparativos com outros grupos de aprendizes.

Há duas modalidades de busca e análise de dados do CLC (2019) via Sketch Engine. A primeira é de viés *corpus-driven*. Trata-se da geração de listas de palavras mais usadas por um determinado perfil de aprendizes, seguida de buscas individuais de cada uma. A segunda modalidade é a busca por itens selecionados por critérios externos ao *corpus*, como os intervalos lexicais analisados no capítulo 5. Trata-se de uma perspectiva de cunho *corpus-based*, na qual o *corpus* não é a origem do trabalho de descrição linguística, mas um meio para tal. Segundo Tognini-Bonelli (2001, p. 65), no viés *corpus-based*, os dados de um *corpus* são um meio para “[...] expor, testar ou exemplificar teorias e descrições que foram formuladas antes de que grandes corpora estivessem disponíveis [...]”¹²².

No meio acadêmico, existe uma forte oposição entre as perspectivas *corpus-driven* e *corpus-based*. Os defensores mais ferrenhos da Linguística de *Corpus*, como Tognini-Bonelli (2001, p. 65-66), por exemplo, condenam a abordagem *corpus-based* por considerarem que ela “rebaixa” a área para uma mera metodologia e ajusta os dados a uma teoria, ao invés de ajustar a teoria aos dados. Acerca dos adeptos da perspectiva *corpus-driven*, a autora comenta que “[...] a atenção dispensada aos *corpora* é bastante recente e específica; eles estão interessados principalmente no quão bem as suas teorias explicam os dados”¹²³ (TOGNINI-BONELLI, 2001, p. 67).

Para fins lexicográficos, no entanto, ambas as perspectivas ajudam a fornecer uma descrição mais acurada da língua em uso. Se, por um lado, a perspectiva *corpus-driven* vem auxiliando na compilação de dicionários, a perspectiva *corpus-based* ajuda na sua revisão e atualização, de modo que ambas são complementares.

Para ilustrar as duas modalidades de busca e análise, analisou-se um conjunto estrito de itens léxicos, já que este trabalho não almeja analisar exaustivamente os dados do CLC, mas sim indicar opções de consulta visando a redação de notas de uso em dicionários. Uma análise mais extensa de itens, além de demandar um tempo maior do que o disponível para a tese, desviaria sobremaneira o foco do presente capítulo.

No caso da primeira modalidade, que parte de listas de palavras mais frequentes, tomaram-se os cinco substantivos mais frequentemente usados pelos brasileiros adultos A1-B1,

¹²² [to expound, test or exemplify theories and descriptions that were formulated before large corpora became available].

¹²³ [Their attention to corpora is fairly recent and rather specific; they are primarily interested in how well their theories account for the data]

indicados a seguir juntamente com a frequência de ocorrência no grupo em questão: *house* (n. 1.056), *friend* (n. 922), *time* (n. 816), *shop* (n. 630) e *place* (n. 618) (CLC, 2019). Por outro lado, para exemplificar a segunda modalidade, tomou-se como ponto de partida a lista de ULs analisadas no capítulo 5 e selecionaram-se os cinco itens mais usados pelos brasileiros adultos A1-B1: *already* (n. 96), *glad* (n. 92), *rent* (verbo, n. 35), *towards* (n. 34) e *appointment* (n. 32) (CLC, 2019). Antes de reproduzir os resultados, é pertinente fazer algumas observações comuns às duas modalidades de consulta no que diz respeito ao processo de análise de dados.

O CLC (2019) permite que os resultados sejam organizados por critérios para facilitar a análise. Foram escolhidos o nível de *performance* e os erros, de forma que o conjunto de alinhamentos se organizava primeiramente entre os níveis, com o total de dados A1 primeiro, seguidos pelos dados A2 e finalizando com os dados B1. Em cada uma dessas sequências de dados, os alinhamentos se dividiam em dois grupos: primeiro apareciam os alinhamentos sem código de erro atrelado ao item objeto da busca (usos corretos do item); as posições finais da sequência de alinhamentos de cada nível, por sua vez, eram ocupadas pelas manifestações errôneas do item analisado.

Antes da consulta, pressupôs-se que todas as ocorrências errôneas de um determinado item estariam agrupadas e discriminadas claramente. Porém, dois fatos constatados durante a análise provaram que a identificação e a contabilidade dos erros constituem um processo mais complexo. O primeiro fato é que nem todas as ocorrências acompanhadas de um código de erro estavam ao final da progressão; algumas estavam diluídas entre as demais. Isso ocorria nos casos em que o item alvo da busca fora usado incorretamente no lugar de outro. Por exemplo: na busca pelo substantivo *time*, os alinhamentos em que *time* era usado erroneamente no lugar de *turn* (“~~time~~ | turn”) (CLC, 2019) estavam etiquetados, mas distribuídos entre os alinhamentos ao invés de constar na sequência de alinhamentos errôneos identificados pelo Sketch Engine ao final da progressão de cada nível. O *software* organiza em sequência somente os alinhamentos errôneos nos quais o item pesquisado é a alternativa indicada como correta no CLC (2019). Por exemplo: nas buscas por *time*, os usos errôneos de *hour* no lugar de *time* (“~~hour~~ | time”) (CLC, 2019) estão agrupados junto a outros erros em que *time* é apontada como a opção correta, independentemente do tipo de erro de que se trate.

O segundo aspecto que chamou a atenção durante a análise foi que os erros envolvendo combinatórias muitas vezes só podem ser percebidos e contabilizados mediante uma análise manual dos alinhamentos – embora não tenhamos dúvidas de que as marcações dos erros satélites a um item são de sumamente úteis para identificar erros de combinatória. Um exemplo recorrente é a seleção da preposição equivocada para acompanhar uma UL ou sintagma, como

“to be generous”, que repetidas vezes foi sucedida erroneamente por *with*. Parte considerável dos erros de seleção de preposição não são casos isolados, mas erros nos quais se selecionou uma preposição equivocada para acompanhar uma UL ou sintagma. Sua identificação é, sem dúvidas, sutil, pois exige interpretação dos dados a partir de uma perspectiva sintagmática, e não somente paradigmática.

Portanto, nem sempre se poderá chegar ao número exato de erros relativos a um item. Além disso, o processo de identificação e contabilização de erros depende, em grande medida, da relação entre o volume de dados e o tempo disponível para análise.

Após as considerações iniciais, reproduzem-se a seguir os resultados das buscas sob as perspectivas *corpus-driven* e *corpus-based*. Dado que a análise dos alinhamentos em ambas consiste no registro dos tipos de erros e sua frequência, os resultados foram reunidos em uma única tabela. Dentre os erros de cada item, foram detalhados os mais recorrentes.

Tabela 9 – Incidência de erros sobre itens A2-B1

Item	Erros (aprox.)	Erros mais frequentes e sua correção no CLC (2019)
<i>house</i> (f = 1.056)	≈ 198	109x: Combinatória: in at + [possessive determiner] + house 26x: Ordenação de palavras incorreta. Não são próprios de <i>house</i> .
<i>friend</i> (f = 922)	≈ 106	15x: Ordenação de palavras incorreta. Não são próprios de <i>friend</i> . 12x: Substantivo desnecessário: dear + friend + [proper name] 12x: Concordância incorreta. Não é próprio de <i>friend</i> .
<i>time</i> (f = 816)	≈ 178	31x: Substituir substantivo: casos variados. 6x hour time 23x. Substituir pontuação: casos variados. 7x lunch time lunchtime; 7x some times sometimes
<i>shop</i> (f = 630)	≈ 107	54x: Substituir substantivo. 38x shop shopping centre 21x: Ordenação de palavras incorreta. Não são próprios de <i>shop</i> .
<i>place</i> (f = 618)	≈ 78	24x: Substituir substantivo: casos variados. 8x: Ordenação de palavras incorreta. Não são próprios de <i>place</i> .
<i>already</i> (f = 96)	≈ 34	10x: Tempo verbal incorreto: 4x already + simple past; 4x already + simple present 7x: Substituir advérbio: casos variados. 5x: Erro de posição sintática: <i>already</i> após o núcleo do predicado
<i>glad</i> (f = 92)	≈ 11	7x: Substituir adjetivo. 2x glad grateful; 2x glad pleased
<i>rent</i> (f = 35)	≈ 9	5x: Forma nominal do verbo incorreta: rent renting
<i>towards</i> (f = 34)	≈ 32	26x: Substituir preposição: “[to be] generous with towards [someone]”

<i>appointment</i> (f = 32)	≈ 14	10x: Ortografia: casos variados. 6x falta de duplicação da letra -p-; 4x acréscimo de vogal após o 1º -t-
--------------------------------	------	--

Fonte: Elaboração própria a partir dos dados do CLC (2019)

Para a seleção dos dados que compõem a Tabela 9, optou-se por elencar até três tipos de erros e manifestações errôneas pontuais mais frequentes dentre os resultados obtidos. Conforme comentado anteriormente, na ausência de estudos estatísticos e de critérios de referência no campo da LCR, não há como avaliar que frequência ou porcentagem de frequência qualificaria um erro como (muito/pouco) relevante para uma nota de uso de dicionário. Portanto, abstemo-nos de avaliar a relação entre o total de erros aproximado de cada item (sinalizado pelo símbolo ≈), bem como a relação entre cada tipo de erro frente ao total de erros de cada item.

Os resultados obtidos permitem dividir os dez itens acima em dois grupos. Por um lado, estão os itens com erros pontuais nos quais predomina(m) uma ou mais manifestação(ões) errônea(s) em relação ao total de erros do item léxico em questão, como ocorre com *house*, *friend*, *shop*, *already*, *towards* e *appointment*. Tais casos podem ser tratados em uma nota de uso de dicionário. Por outro lado, estão os itens que não apresentam um tipo de erro ou manifestação errônea sobressalente, ou ainda os erros referentes a regras gramaticais, isto é, que não são próprios ou exclusivos de um determinado item, como é o caso dos erros de concordância em *friend*. Trata-se de erros cuja distribuição ou natureza não justifica uma nota de uso. Alocadas neste segundo grupo, encontram-se *time*, *place*, *glad* e *rent*.

A seguir, serão comentados os resultados dos itens do primeiro grupo, isto é, os erros cuja natureza e predominância os tornam candidatos para notas de uso em um dicionário. Paralelamente, averigua-se se o fenômeno linguístico e/ou o erro estão contemplados nos quatro dicionários de aprendizes de inglês considerados na tese. Após as considerações necessárias, propõe-se um exemplo de nota de uso a ser acrescentada a um deles para fins de ilustração.

Dentre os erros relativos ao item léxico *house*, o que mais se destaca é a seleção da preposição equivocada para se referir a *house* como ponto de localização, como em “I left my book ~~in~~ | at your house” (CLC, 2019. Grifos nossos). Esse erro demonstra que, ao menos no uso de *house*, os brasileiros confundem as funções das preposições *in*, que alude ao interior de algo (no caso, uma casa), e *at*, que indica localização. Trata-se, portanto, de uma dificuldade relativa às preposições *in* e *at*. Por essa lógica, pode-se argumentar que os verbetes dessas preposições são os que precisam de chamadas de atenção por meio de notas de uso. Porém, ante o fato de que grande parte dos erros que os brasileiros A1-B1 cometem ao usar *house* se refere

à seleção da preposição equivocada, é mais efetivo e proveitoso para esse usuário que se faça menção ao tipo de erro em questão por meio de uma nota de uso no substantivo. Dos dicionários de aprendizes analisados, apenas o OxED2 (2012) e o LWord2 (2008) registravam esse padrão, embora ambos o façam sob a forma de exemplos e sem destaque tipográfico: “We’re having dinner at Jane’s house tonight” (OxED2, 2012, s.v. *house*, 1) e “We spent the evening at Harriet’s house” (LWord2, 2008, s.v. *house*, 1).

Para ilustrar como seria a redação de uma nota de uso sobre o erro em questão, tomou-se como base o verbete do OxED2 (2012):

Quadro 52 – Proposta de nota de uso em *house* para o OxED2 (2012)

Verbetes original	Proposta de adaptação
<p>house ↔ /haʊs/ <i>noun</i> (plural houses /ˈhaʊzɪz/) 1 a building where a person or a family lives: <i>How many rooms are there in your house?</i> ◇ <i>We’re having dinner at Jane’s house tonight.</i></p> <p>WORD BUILDING</p> <p>A small house is called a cottage. A house with only one floor is called a bungalow: <i>My grandparents live in a bungalow near the sea.</i> A tall building where lots of people live is called a block of flats or an apartment block: <i>They live on the third floor of an apartment block.</i> 2 a building that has a special use: <i>an opera house.</i></p> <p>(OxED2, 2012, s.v. <i>house</i>)</p>	<p>house ↔ /haʊs/ <i>noun</i> (plural houses /ˈhaʊzɪz/) 1 um imóvel onde vive uma pessoa ou uma família; em português, <i>casa</i>: <i>How many rooms are there in your house?</i> ◇ <i>We’re having dinner at Jane’s house tonight.</i></p> <p>USO</p> <p>Use at para indicar um endereço: <i>We’re having dinner at Jane’s house tonight.</i> Use in para falar sobre o que há dentro de uma casa: <i>How many rooms are there in your house?</i></p> <p>VOCABULÁRIO</p> <p>Uma casa pequena se chama cottage. Uma casa de um piso apenas se chama bungalow: <i>My grandparents live in a bungalow near the sea.</i> Uma construção alta onde vivem muitas pessoas se chama block of flats ou apartment block: <i>They live on the third floor of an apartment block.</i> 2 um imóvel usado para um fim especial; em português, <i>casa</i>: <i>an opera house.</i></p>

Fonte: À esquerda, verbete do OxED2 (2012, s.v. *house*); à direita, proposta da autora com base nos dados do OxED2 (2012, s.v. *house*)

O verbete original já continha uma nota de uso com alguns elementos do campo léxico de *house*. Julgamos mais pertinente incluir a nota de uso referente aos padrões sintáticos “in [somebody’s] house” e “at [somebody’s] house” antes da observação sobre o campo léxico (vocabulário) por ser a primeira uma informação menos prescindível que a segunda.

A respeito do tipo de orientação, havia mais de uma opção conforme se considerasse que as funções das preposições *at* e *in* estivessem ou não incluídas. Por um lado, é possível argumentar que os exemplos em si não auxiliam a diferenciar ou reforçar os usos de *at* e *in*, nem representam uma descrição das funções de ambas as preposições quando estão junto de *house*. Portanto, frente a uma informação nova no verbete, o tipo de orientação adequado é o

indicativo. A nota iniciaria com uma menção ao erro cometido pelos brasileiros, para então apresentar a descrição do uso no inglês e aproveitaria os exemplos já presentes na acepção 1: “Os brasileiros costumam confundir **in** e **at** quando usam a palavra **house**. Para indicar um endereço, usa-se **at**: *We’re having dinner at Jane’s house tonight*. Para falar sobre o que há dentro de uma casa, usa-se **in**: *How many rooms are there in your house?*”.

Por outro lado, pode-se argumentar também que o exemplo não seria uma casualidade, mas estaria presente justamente para descrever o uso da preposição. Nesse caso, a falta de destaque tipográfico ou de qualquer outro recurso para chamar a atenção do consulente é um ônus, pois exige um nível de interpretação de informações que o consulente de níveis iniciais pode não ter. Sendo assim, a nota conteria uma repetição de uma informação anterior, já presente nos exemplos. Também se pode considerar que a diferença de usos de *in* e *at* é uma informação repetida porque forma parte do conhecimento prévio do nível A1 segundo o EVP (2015). Essa foi a perspectiva adotada na adaptação, que culminou com uma nota exortativa.

No que se refere a *friend*, dentre os três tipos de erros mais frequentes, o único passível de ser contemplado em uma nota de uso é o caso em que o substantivo é usado de maneira desnecessária ao iniciar uma correspondência escrita, como em “Dear ~~friend~~ Robbie” (CLC, 2019. Grifo nosso). Na verdade, o erro “*dear* + *friend* + [proper name]” diz muito mais respeito ao fato de que o adjetivo *dear*, na sua função de saudação, não se combina concomitantemente com *friend* e um nome próprio. Portanto, caso se incluía uma nota de uso sobre o particular, faz mais sentido que ela esteja presente no verbete *dear*.

Ao consultar os verbetes *dear* e *friend* nos dicionários de aprendizes, verificou-se que a impossibilidade de se combinar *dear* com *friend* não está incluída em nenhum. Três deles fazem, no máximo, uma menção na paráfrase explanatória sobre o tipo de palavra que acompanha *dear*, com exemplos: 1) “used at the beginning of a letter, before the name of the person you are writing to: *Dear Amy* ◦ *Dear Sir/Madam*” (CaED2, 2011, s.v. *dear*¹, 1. Grifos no original); 2) “a word that you use before person’s name at the beginning of a letter: *Dear Mr Carter*, ... ◊ *Dear Sir or Madam*, ...” (OxED2, 2012, s.v. *dear*¹, 1. Grifos no original); e 3) “**Dear Sir, Dear Mr Smith, Dear Mrs Jones, etc** (written) used before the name of the person you are writing to: *Dear Miss Patterson, I am writing to you about your son*” (LWord2, 2008, s.v. *dear*³, 1. Grifos no original).

Tendo em vista o erro cometido pelos brasileiros, é possível adotar uma das seguintes perspectivas a respeito do conteúdo das paráfrases explanatórias anteriores. Por um lado, é possível argumentar que os próprios dicionários já restringiriam as possibilidades de combinação de *dear*, reduzindo-as aos nomes próprios acompanhados ou não por substantivos

como *sir, madam, Mr, Mrs, Ms, etc.* – e nada mais. Nesse sentido, qualquer outra combinação que não se enquadra nesse padrão, como **dear friend*, estaria automaticamente descartada, não sendo necessária uma nota de uso sobre o particular. Por outro lado, também é possível argumentar que a impossibilidade de combinação com *friend* não estaria suficientemente clara, ou que o usuário não teria condições de deduzir que não há outras possibilidades de combinação além dos nomes próprios. Nesse sentido, seria necessário um reforço sob a forma de nota de uso para que o dicionário contemple as necessidades dos usuários adultos brasileiros A1-B1.

Somos adeptos da segunda perspectiva, favorável à nota de uso, pois esta retira do consulente o ônus da inferência de informações. No entanto, vale ressaltar que não há uma perspectiva certa ou errada. Na ausência de critérios para avaliar a pertinência da nota, cabe ao lexicógrafo decidir com base na sua interpretação e, no caso do suporte impresso, na viabilidade em termos de espaço disponível e efeito de saliência na página em que se encontraria a nota, pois quanto mais notas reunidas em verbetes próximos na mesma página, menor o efeito de saliência.

A seguir, ilustra-se a proposta de inclusão de nota de uso:

Quadro 53 – Proposta de nota de uso em *dear* para o CaED2 (2011)

Verbete original	Proposta de adaptação
<p>dear¹ /diə^r/ adj</p> <p>A1 used at the beginning of a letter, before the name of the person you are writing to: <i>Dear Amy</i> ◦ <i>Dear Sir/Madam</i> [...]</p> <p>(CaED2, 2011, s.v. <i>dear</i>¹)</p>	<p>dear¹ /diə^r/ adj</p> <p>A1 usado no começo de uma carta, antes do nome da pessoa a quem se escreve; em português, <i>querido</i> (informal), <i>prezado</i> (formal): <i>Dear Amy</i> ◦ <i>Dear Sir/Madam</i></p> <div style="background-color: #f9e79f; padding: 5px; border: 1px solid #ccc;"> <p>USO</p> <p>Ao iniciar uma carta para um amigo, não se usa friend: <i>Dear friend Amy</i></p> </div>

Fonte: À esquerda, verbete do CaED2 (2011, s.v. *dear*¹); à direita, proposta da autora com base nos dados do CaED2 (2011, s.v. *dear*¹)

O tipo de orientação adotado foi o indicativo, baseado na perspectiva de que a impossibilidade de combinatória com *friend* não estava explícita no verbete. Como *friend* era o objeto da nota de uso, optou-se pela grafia em negrito, enquanto o tachado invalida o seu uso. O exemplo utilizado está baseado em um dos exemplos do próprio verbete.

O seguinte caso a ser analisado é o de *shop*. O erro mais frequente e, em princípio, passível de ser registrado em nota é o uso errôneo de *shop* em lugar de *shopping centre*, como em “Yesterday, I visited a new shop | shopping centre that opened in my city” (CLC, 2019). Não é possível entender, contudo, por que o sintagma *shopping centre* foi sugerido como

correção, ao invés de outras opções como o acréscimo de um especificador, como *coffee/clothes/gift shop*. Aparentemente, trata-se de algo relacionado às exigências da tarefa que tinham de cumprir, e não de algo estritamente linguístico, pois os registros errôneos em questão estavam em textos do gênero carta nos quais o remetente se dirigia a um amigo a respeito de uma nova loja ou centro comercial que havia descoberto, com o objetivo de convidá-lo a que fossem a esse lugar. Seja como for, parece tratar-se de uma correção por fatores extralinguísticos, o que não justificaria uma nota de uso.

Os principais erros relativos a *already*, por sua vez, dizem respeito ao tempo verbal usado com esse advérbio e à sua posição sintática.

Quanto ao tempo verbal, foram classificadas como errôneas as construções de tipo “already + simple present”, como “She ~~already ask~~ | has already asked me to visit her home” (CLC, 2019. Grifos nossos). Também foram classificadas como errôneas as construções de tipo “already + simple past”, como “I ~~already chose~~ | have already chosen the colour” (CLC, 2019. Grifos nossos).

No tocante ao uso de *already* com verbos no presente, Leech e Svartvik (2002, §156) esclarecem que esse padrão ocorre quando o verbo que o acompanha é de estado ou apresenta o aspecto progressivo (*present progressive*)¹²⁴, como em “He already has an income of 800,000 a year” (BNC, 2007, *already*). As funções de *already* no padrão combinatório em questão são a de enfatizar que algo existe no presente (cf. OALD8e (2010, s.v. *already*, 3)) e são relativas ao nível B1 de proficiência segundo o EVP (2015). Dentre os resultados errôneos encontrados no CLC (2019), todos consistiam no uso de *already* com verbos de estado conjugados no presente, mas para narrar um evento situado no passado: “[...] she ~~already ask~~ | has already asked me to visit her home” (CLC, 2019. Grifos nossos); “[...] he ~~already work~~ | has already worked in a factory” (CLC, 2019. Grifos nossos); “The parking problems we ~~already sort~~ | have already sorted [...]” (CLC, 2019. Grifos nossos); “[...] really surprised that you ~~already get~~ | have already got a job [...]” (CLC, 2019. Grifos nossos). Dos dicionários de aprendizes analisados, somente o CoED2 (2014, s.v. *already*, 2) e o LWord2 (2008, s.v. *already*, 2) incluem uma acepção a respeito da função de enfatizar a existência de algo no momento presente, mas não descrevem o padrão sintático que a manifesta.

¹²⁴ Segundo Leech e Svartvik (2002, §114), são considerados de estado aqueles verbos que denotam “um estado de coisas que continua no decorrer de um período e não precisa ter um começo e um fim bem definidos, como *ser/estar, permanecer, conter, saber/conhecer* [...]” [a state of affairs which continues over a period, and does not need to have a well-defined beginning and end, e.g. *be, remain, contain, know* [...]].

No que concerne ao uso de *already* acompanhado de verbos em referência a eventos do passado, Leech e Svartvik (2002, §156) assinalam que tais verbos se apresentam com o aspecto perfeito (*present perfect, past perfect*) no inglês britânico. Por outro lado, e segundo os autores, no inglês norte-americano, é mais usual a combinação de *already* com o *simple past*, sem o aspecto perfeito. Portanto, os registros no padrão “*already + simple past*”, considerados errôneos no CLC (2019), não o são sob a perspectiva norte-americana. Sem informações sobre a tarefa à qual os textos com trechos errôneos respondiam, não é possível julgar a pertinência da avaliação. Ainda assim, a variação diatópica no padrão de uso em questão é uma propriedade a ser descrita nos dicionários de aprendizes que contemplam o nível A2, pois, em qualquer tarefa de produção linguística, a adequação à situação comunicativa requer necessariamente que se esteja consciente sobre os tempos verbais com os quais *already* se combina e em que regiões geográficas. Dos dicionários de aprendizes de inglês analisados, o único que faz menção à variação do tempo verbal é o CoED2 (2014), por meio de uma nota de uso: “In British English, you usually use **already** and **yet** with the perfect tenses. [...] In American English, people often use a past tense. [...]” (CoED2, 2014, s.v. *already, language help*. Grifos no original). O OxED2 (2012, s.v. *already, which word?*) e o LWord2 (2008, s.v. *already, word choice*) mencionam apenas que *already* é usado em frases afirmativas, em contraste com seu sinônimo *yet*, que se usa nas interrogativas e negativas.

Uma terceira parcela dos erros mais comuns no uso do advérbio *already* diz respeito à sua posição sintática, que varia conforme esteja acompanhado de um verbo principal (*simple present, simple past*) ou de uma locução com verbo auxiliar (*present perfect, past perfect*), tal como nos casos relatados acima, ou ainda quando se trata do verbo de ligação *be*. Entre os dicionários e gramáticas a que tivemos acesso, a obra de referência que descreve de forma mais clara e completa a variação da posição sintática de *already* é o LDOCE6 (2014):

Figura 8 – Posições sintáticas de *already*

GRAMMAR: Word order

- **Already** often comes at the end of a sentence:
I knew that already.
I was tired already.
I have booked the tickets already.

As well as at the end of a sentence, **already** can be used in the following ways:

- **Already** can come before a main verb:
I already knew that.
- **Already** comes after 'be' when it is the main verb:
I was already tired.
✗ Don't say: I already was tired.
- **Already** comes after the first auxiliary verb:
I have already booked the tickets.
✗ Don't say: I already have booked the tickets.

Fonte: LDOCE6 (2014, s.v. *already*, *grammar*)

Em síntese, além da possibilidade de ocupar a última posição sintática em uma frase, o advérbio pode ainda anteceder tanto o verbo principal como o adjetivo ou locução adjetiva que complementa o verbo *be*. Em vista desse padrão de uso, no CLC (2019), estão sinalizadas como incorretas construções como “I ~~have already~~ | already have a meeting with a new supplier [...]” (CLC, 2019. Grifos nossos), “I ~~already was~~ | was already nervous” (CLC, 2019, grifos nossos) e “We had some problems with our staff last week, but they have ~~been already~~ | already been solved” (CLC, 2019), por exemplo. Chama a atenção que tais padrões não estejam descritos em nenhum dos quatro dicionários analisados.

De fato, estranha que esta e as demais propriedades básicas de *already*, sumamente necessárias para fins de produção, não estejam descritas nos dicionários que se dirigem ao público A2 e B1 – níveis que, segundo o EVP (2015), são justamente aqueles nos quais os diferentes usos de *already* se consolidam entre os aprendizes de inglês como LE.

Quadro 54 – Proposta de nota de uso em *already* para o LWord2e (2008)

Verbetes original	Proposta de adaptação
<p>al·read·y <i>adverb</i> UK ◀ US ◀</p> <p>1 before ◀ <i>I've seen that film already.</i> ◀ <i>He's already won three titles this year.</i> ◀ <i>When they went in the shop, they had already decided what to buy.</i></p> <p>2 now, rather than later ◀ <i>I can't wait for Christmas, I'm getting excited already.</i> ◀ <i>It was only 11.30, but Pat was already feeling hungry.</i></p> <p>WORD CHOICE already or yet? You use already in positive statements: <i>She has already visited the US three times.</i> You usually use yet in negative sentences and questions: <i>I haven't finished my homework yet. Are you in bed yet?</i> Yet usually comes at the end of a sentence.</p> <p>(LWord2, 2008, s.v. <i>already</i>)</p>	<p>al·read·y <i>adverb</i> UK ◀ US ◀</p> <p>1 anteriormente; em português, <i>já</i>. ◀ <i>I've seen that film already.</i> ◀ <i>He's already won three titles this year.</i> ◀ <i>When they went in the shop, they had already decided what to buy.</i></p> <p>✎ USO No Reino Unido, usa-se already com <i>present perfect</i> e <i>past perfect</i>: <i>He's already won three titles this year. When they went in the shop, they had already decided what to buy.</i> Nos EUA, geralmente se usa already com <i>simple past</i>: <i>Everyone already knew how to play soccer except me.</i> Quando already se refere ao passado, não se usa verbos no presente: <i>She already ask has already asked me to visit her home.</i></p> <p>2 agora, e não depois; em português, <i>já</i>. ◀ <i>I can't wait for Christmas, I'm getting excited already.</i> ◀ <i>It was only 11.30, but Pat was already feeling hungry.</i></p> <p>✎ USO Em todos os casos, already é usado sempre à esquerda do verbo principal ou ao final da frase: <i>He's already won three titles this year. I can't wait for Christmas, I'm getting excited already.</i> Quando o verbo principal é be, sem auxiliares, usa-se already à esquerda do adjetivo que o acompanha ou ao final da frase: <i>I'm already depressed about it. Yes, most are vaccinated already.</i></p> <p>✎ VOCABULÁRIO already or yet? Usa-se already em frases afirmativas: <i>She has already visited the US three times.</i> Normalmente se usa yet em negações e perguntas: <i>I haven't finished my homework yet. Are you in bed yet?</i> Yet normalmente está ao final da frase.</p>

Fonte: À esquerda, verbete do LWord2e (2008, s.v. *already*); à direita, proposta da autora com base nos dados do LWord2 (2008, s.v. *already*)

Todas as orientações propostas são de cunho indicativo devido à sua ausência no verbete original. Para separar as notas sobre uma acepção específica daquelas que se aplicavam a todas as acepções, as orientações relativas somente à acepção 1 foram incluídas em nota após dita acepção. Para a observação a respeito dos usos no Reino Unido, foram reaproveitados os exemplos já presentes na acepção 1, enquanto o exemplo “Everyone already knew how to play soccer except me”, que acompanha os usos nos EUA, foi retirado do *Corpus of Contemporary American English* (COCA, 2020). Após a acepção 2, foram situadas as notas que se referem a *already* de maneira mais ampla, com observações válidas para as duas acepções. Na nota “uso”,

foram novamente reaproveitados os exemplos do verbete. Para ilustrar os usos com adjetivos, os exemplos “I’m already depressed about it” e “Yes, most are vaccinated already” foram extraídos do BNC (2007). Ao final, consta a versão adaptada da nota do verbete original do dicionário.

Towards é a penúltima UL que pertence ao grupo das que apresentam registros errôneos passíveis de tratamento em uma nota de uso. A maior parte dos registros (32/34) continha o mesmo padrão errôneo: “[to be] generous with | towards [someone]”, como em “Remember that she is always generous with | towards you” (CLC, 2019. Grifo nosso). Há uma grande possibilidade de que essa manifestação errônea esteja presente em textos redigidos em resposta a uma mesma tarefa, pois todos os registros do padrão em questão foram produzidos por indivíduos de nível A2 de proficiência.

Há duas considerações importantes a respeito dessa estrutura. Primeiramente, o foco do erro não é a preposição *towards*, mas sim o adjetivo *generous*, na acepção “desejoso de dar” [willing to give] (EVP, 2015, *generous*). Trata-se de um erro de combinatória léxica no qual os aprendizes brasileiros, ao empregarem a UL *generous*, não selecionaram a preposição correta ao complementá-la com um adjunto de pessoa (ou beneficiário da generosidade). Isso não significa, contudo, que a combinatória *generous with* não seja usada no inglês. Segundo o OALD8 (2010, s.v. *generous*, 1), o LDOCE6 (2014, s.v. *generous*, 1) e o MEDAL2 (2007, s.v. *generous*, 1), a estrutura “generous with [something]” é usada no inglês em alusão ao objeto com o qual se praticou a ação considerada generosa: “And he was generous with his money” (BNC, 2007). É pertinente considerar que a construção “generous *with [someone]” é um forte candidato a ter sido produto de uma influência interlinguística da combinatória do português “generoso com [alguém]” (cf. padrão descrito no Sac (2010, s.v. *generoso*, 1)).

Em segundo lugar, o cenário se torna ainda mais complexo quando se analisa a incidência da combinatória “[to be] generous towards [someone]” em um *corpus*. Uma consulta por *generous* no BNC (2007) revelou que, dos 2.234 registros resultantes, apenas seis eram referentes ao uso de *towards* no padrão em questão. Mais recorrente era o uso da preposição *to* nesse padrão, que conta com 61 registros: “By contrast, he was very generous to people who had less than he did [...]” (BNC, 2007). Além disso, apenas *generous to* é descrita como combinatória no OALD8 (2010, s.v. *generous*), no LDOCE6 (2014, s.v. *generous*) e no MEDAL (2007). Os dados do BNC (2007) e dos dicionários de aprendizes avançados levam a crer que a correção *towards* presente no CLC (2019) não é própria da combinatória de *generous* em alusão a “desejoso de dar”. Portanto, em um dicionário de aprendizes, o mais adequado é incluir a combinatória com o candidato mais frequente e reconhecido como par de *generous*,

isto é, *to*. Também se poderia questionar se esse tipo de construção não seria pouco frequente no inglês a ponto de não ser pertinente para um dicionário de aprendizes iniciantes da língua. No momento, não é possível responder a essa pergunta. O EVP (2015), por exemplo, não especifica se as combinações de “*generous (willing to give)*” com preposições são igualmente atreladas ao nível B1.

Em um caso como este, o critério para se decidir pela inclusão de uma nota com os diferentes usos de *generous* junto das preposições *to* e *with* é justamente a sua relevância para o usuário em questão, isto é, para os brasileiros adultos entre os níveis A1 e B1 de proficiência. Se guardadas as ressalvas acerca da possibilidade de que os registros errôneos do CLC (2019) estejam enviesados por responderem a uma mesma tarefa, a recorrência do erro no uso de *generous* permite supor que há uma tendência a que os brasileiros associem *generous with* a um adjunto de pessoa e cometam esse erro em outras situações comunicativas. Essa tendência, portanto, justifica a pertinência de uma nota de uso sobre o particular.

No que diz respeito aos quatro dicionários de aprendizes de inglês A1-B1 analisados, somente o LWord2 (2008, s.v. *generous*) inclui a combinatória em questão, utilizando-se tanto do quadro *Key patterns*, que está localizado após o comentário de forma, como de um destaque tipográfico para o verbo *to be* e para a preposição *to* nos exemplos (cf. quadro a seguir). A inclusão dos padrões combinatórios de *generous* em outro dicionário, como o CoED2 (2014), por exemplo, poderia ser feita da seguinte maneira:

Quadro 55 – Proposta de nota de uso em *generous* para o CoED2 (2014)

Verbetes do LWord2	Verbetes original do CoED2	Proposta de adaptação
<p>gen·e·rous <i>adjective</i> UK ◀ US ◀</p> <p>KEY PATTERNS</p> <p>it is generous of someone to do something</p> <p>be generous to someone</p> <p>someone who is generous gives a lot of money, presents, or help to other people ANTONYM stingy, mean <i>BrE</i></p> <p>◀ <i>It was really generous of Jack to take us all on holiday.</i></p> <p>◀ <i>My parents have always been very generous to me.</i></p> <p>– generously <i>adverb</i> [...]</p> <p>(LWord2e, 2008, s.v. <i>generous</i>)</p>	<p>generous /'dʒenərəs/<i>adjective</i></p> <p>giving you more than you expect of something: <i>He is generous with his money.</i></p> <p>▶ generosity [...]</p> <p>(CoED2, 2014, s.v. <i>generous</i>)</p>	<p>generous /'dʒenərəs/<i>adjective</i></p> <p>[alguém] que dá [algo] em quantidade maior do que o esperado; em português, <i>generoso</i>: <i>He is generous with his money.</i></p> <p>↗ USO</p> <p>Para dizer que alguém é generoso com outra pessoa, usa-se generous to: <i>He was always friendly and generous to all classes of people.</i></p> <p>Para dizer que alguém é generoso em algo, usa-se generous with: <i>People have been very generous with their donations.</i></p>

Fonte: À esquerda, verbete do LWord2 (2008, s.v. *generous*); ao centro, verbete do CoED2 (2014, s.v. *generous*); à direita, proposta da autora com base nos dados do CoED2 (2014, s.v. *generous*)

Como a nota inclui informações que não estão descritas em outras partes do verbete, a orientação empregada na sua redação foi a indicativa. Os exemplos foram extraídos do BNC (2007) para fins de ilustração, mas provavelmente a editora usaria seu *corpus* privado para tal.

Finalmente, o último caso a ser analisado é o de *appointment*, cuja representação ortográfica é um desafio para os brasileiros adultos de níveis A1 a B1. Dentre as dez manifestações com ortografia equivocada, dois padrões se destacaram como os mais frequentes: a falta de duplicação da letra -p- (5x **apointment*; 1x **apontiment*); e o acréscimo de vogal após a primeira letra -t- (2x **appointement*, 1x **apointiment*, 1x **appoitament*). Como a duplicação da letra -p- e o uso da letra -t- no final de sílaba não são próprios do português – à exceção de empréstimos, tais como *cappuccino* e *impeachment* –, é plausível pensar que os dois padrões errôneos tenham origem em influências interlinguísticas do português.

Dentre os dicionários de aprendizes analisados, todos contêm a representação ortográfica de *appointment* no próprio lema. Conforme já comentado, o que justifica a inclusão de uma nota de uso que repete uma informação anterior é a constatação de que os aprendizes cometem erros numa frequência que seja considerada significativa pelo lexicógrafo. No que diz respeito a *appointmet*, embora os erros ortográficos representem a terça parte (n. 10) da frequência de uso (n. 32) dessa UL pelos brasileiros adultos A1-B1, há que se considerar também que dita frequência poderia ser considerada insuficiente para que se justificasse o acréscimo de uma nota. Novamente, na ausência de critérios na Metalexicografia e na Linguística de *Corpus* para julgar que erros são mais relevantes para assinalar em um dicionário, fica a cargo do editor da obra e do lexicógrafo responsável estabelecer critérios próprios de julgamento conforme as suas possibilidades (recursos humanos disponíveis, tempo etc.). Caso se decida pela inclusão da nota, esta poderia ser redigida da seguinte maneira:

Quadro 56 – Proposta de nota de uso em *appointment* para o CaED2 (2011)

Verbetes original	Proposta de adaptação
<p>appointment /ə'pɔɪntmənt/ noun A2 a time you have arranged to see someone: <i>I've made an appointment with the dentist.</i></p> <p>(CaED2, 2011, s.v. <i>appointment</i>)</p>	<p>appointment /ə'pɔɪntmənt/ noun A2 um horário reservado para ver alguém; em português, <i>hora marcada</i> (com profissionais em geral), <i>consulta</i> (com profissionais da saúde): <i>I've made an appointment with the dentist.</i></p> <p>ORTOGRAFIA</p> <p>Os brasileiros costumam escrever appointment errado. Lembre-se de que há pp e que não há vogal após o t: appointment</p>

Fonte: À esquerda, verbete do CaED2 (2011, s.v. *appointment*); à direita, proposta da autora com base nos dados do CaED2 (2011, s.v. *appointment*)

Em se tratando da repetição de uma informação já presente no verbete, a modalidade de orientação correspondente segundo a nossa proposta é a exortativa. Optou-se pela orientação exortativa prescritiva, mais especificamente, porque este estilo de redação é relativamente comum no dicionário, como ocorre, por exemplo, em “Learners often spell **beautiful** wrong. Remember that it starts with **beau** and has one **l** at the end” (CaED2, 2011, s.v. *beautiful*. Grifos no original).

Como corolário da presente seção, concluímos que os dados do CLC (2019) e as propostas acerca das notas de uso podem tornar mais didáticos os dicionários de aprendizes de inglês A1-B1 tendo em vista as necessidades do público brasileiro adulto. O CLC (2019) tem muito a oferecer para o aprimoramento e até mesmo confecção de materiais didáticos, especialmente no que diz respeito ao estabelecimento das necessidades do público brasileiro, que é um dos grupos que mais conta com registros. No entanto, o acesso restrito a essa ferramenta é um obstáculo para uma possível aplicação das propostas de adaptação que fazemos na tese. A competição mercadológica entre as editoras de dicionários de aprendizes de inglês é o que motiva a restrição de acesso aos seus *corpora*. É evidente que a Cambridge University Press não cederia acesso à HarperCollins, por exemplo, para que esta aprimorasse o CoED2 (2014). A explanação sobre o CLC (2019) e a testagem de metodologias de análise, simulando uma situação de adaptação de qualquer um dos quatro dicionários de inglês, tiveram como objetivo lançar luz sobre o aproveitamento das contribuições da LCR para fins lexicográficos. O fato de que o CLC (2019) é um *corpus* de acesso restrito e os *corpora* de aprendizes com anotação de erros são pouco comuns faz com que os trabalhos acerca da utilidade desse tipo de recurso para fins de análise e descrição linguística e lexicográfica sejam ainda incipientes na literatura especializada. Com as simulações descritas anteriormente, foi possível relacionar mais claramente as perspectivas de investigação da Linguística de *Corpus*, em geral, e da LCR, em particular, às funcionalidades do CLC (2019) no Sketch Engine. Ao confrontar o que havia disponível na literatura com tais funcionalidades, percebeu-se que há uma importante questão teórico-metodológica em aberto: faltam critérios para o julgamento dos dados do CLC (2019) – e mesmo de outros *corpora* de aprendizes – no que diz respeito à determinação de graus de relevância dos erros e inadequações produzidos pelos aprendizes. Trata-se de um aspecto inerente ao efetivo aproveitamento da LCR na Lexicografia, pois não há espaço suficiente em um dicionário para a descrição de todos os erros cometidos por um perfil de aprendizes; na realidade, não é tarefa do dicionário de aprendizes descrever a totalidade dos usos (errôneos ou não) realizados por dito perfil. Uma matriz de critérios mais bem fundamentada facilitaria o julgamento dos dados e a identificação das manifestações pertinentes para fins lexicográficos.

No que diz respeito aos critérios de inclusão e redação de notas de uso, a principal contribuição deste trabalho diz respeito à escolha da modalidade de orientação, a partir da contribuição de Bugueño Miranda e Borba (2020). Os próximos passos das investigações dizem respeito aos padrões de redação em língua portuguesa das notas já existentes e à necessária testagem do modelo de nota proposto junto aos brasileiros adultos A1-B1.

A maneira como a Lexicografia se correlaciona com a LCR dependerá sempre dos *corpora* de aprendizes disponíveis, das possibilidades de acesso a eles e de quais são as suas funcionalidades. A próxima seção trata dos recursos em LCR para a língua espanhola que contemplam os aprendizes brasileiros A1-B1.

7.3 Notas de uso para dicionários de aprendizes brasileiros de espanhol

O processo de adaptação do Señas (2001) ao público brasileiro A1-B1 requer procedimentos metodológicos deferentes dos que foram propostos para o inglês. No que se refere ao aspecto formal das notas, o diferencial está no uso do espanhol em função metalinguística. No que se refere ao conteúdo, a adaptação requer, primeiramente, uma revisão das notas preexistentes em relação à língua em uso, já que a obra tem mais de vinte anos. Além disso, como o Señas (2001) está baseado em um dicionário de aprendizes de espanhol que não tinha um público-alvo bem delimitado, o conteúdo de algumas notas de uso é dispensável para os brasileiros. Um exemplo disso é a nota de uso do verbete *nombre* relativa à combinatória *nombre propio*, que salienta que “La primera letra de los nombres propios se escribe con mayúscula” (Señas, 2001, s.v. *nombre*). Trata-se de uma informação discreta, isto é, um fato de norma, mas não discriminante para o brasileiro, pois, na sua LM, que compõe o seu conhecimento prévio, os nomes próprios também são escritos com maiúscula. Na verdade, a nota seria mais relevante para um falante de línguas germânicas, por exemplo, pois, nesse caso, os substantivos comuns, assim como os nomes próprios, são escritos com inicial maiúscula.

Porém, o principal aspecto da adaptação do Señas (2001) no que se refere ao conteúdo das notas de uso diz respeito às fontes de onde se pode extrair dados acerca das necessidades dos aprendizes brasileiros adultos A1-B1. Algumas dessas fontes são o *Corpus de Aprendices de Español* (CAES, 2020), o *Nuevo Diccionario de Falsos Amigos Español/Portugués* (NDiFAEP, 2014) e fóruns de discussão de aprendizes lusófonos de espanhol, como o que oferece o site WordReference (<https://forum.wordreference.com/forums/português-español.55/>). Os dois primeiros foram compilados a partir de dados de *corpus*, o primeiro de um *corpus* de produções escritas de aprendizes de espanhol e o segundo de um *corpus*

específico para os fins de compilação do dicionário. Já o terceiro é uma interessante fonte de registro de dúvidas dos aprendizes de espanhol que falam português. Ao cruzar os dados dessas três fontes com o inventário do PCIC (2006), é possível determinar quando incluir notas de uso e que informações incluir de acordo com os recursos linguísticos atinentes aos níveis de aprendizagem A2 e B1. As seções páginas estão dedicadas à descrição dessas fontes e à exposição da metodologia de consulta e de uma amostra de resultados obtida a partir de buscas-piloto. O objetivo é ilustrar metodologias de consulta e extração de dados para fins lexicográficos. Após a apresentação das ferramentas e dos resultados, serão elaborados esboços de notas de uso com base nos erros e dúvidas coletados.

7.3.1 Seleção de conteúdo para notas de uso: CAES (2020)

A primeira fonte a ser comentada é o CAES (2020). A consulta aos dados ocorre por meio do preenchimento de um dos critérios considerados obrigatórios: a) elemento (palavra flexionada, como *jugó*); b) lema (palavra não flexionada, como *jugar*); ou c) por etiqueta (classe de palavra, como *verbo*). Há também critérios adicionais de busca, como o nível de proficiência, a nacionalidade e a língua materna. Por exemplo: substantivos usados por aprendizes brasileiros com nível A1 de proficiência.

A parcela de *tokens* e *types* produzidos por falantes de português como língua materna é de 130.492 e 7.381, respectivamente (dados obtidos excluindo-se os registros de números e símbolos). Os textos que serviram de base para a compilação do *corpus*, por sua vez, foram tomados de provas respondidas por alunos do Instituto Cervantes no período de dois anos. Trata-se de um *corpus* relativamente pequeno, se comparado com a parcela do CLC (2019) correspondente a dados de brasileiros, que soma 2.544.202 *tokens*. Por essa razão, as possibilidades de investigação sobre a produção em língua espanhola por brasileiros são sensivelmente limitadas.

Outro fator ao qual os estudiosos de linguística devem atentar ao cogitar o uso do CAES (2020) para fins de pesquisa acadêmica é o fato de que não há anotação de tipos de erros. Trata-se de uma limitação para a qual Llorián (2017, p. 19-21) já havia chamado a atenção. A falta de anotação causa um sensível atraso nas investigações, pois é necessário analisar cada alinhamento do *corpus* sob a perspectiva de cada um dos níveis de organização da linguagem. Além disso, sem uma anotação de tipos de erros, é impossível estabelecer um panorama dos aspectos que mais causam dificuldades nos brasileiros e, nesse âmbito, as ULs mais frequentemente envolvidas. Nesse cenário, é curioso notar que a ausência de anotação de erros

não costuma ser mencionada em artigos científicos que apresentam ou comentam sobre o CAES (2015-2020), como Parodi (2015) e Ferreira Cabreira e Elejadle Gómez (2017, p. 514), entre outros. O mesmo acontece com artigos que usaram alguma versão desse *corpus* para fins de análise de erros, como Soler Montes (2017) e Malaver (2019), por exemplo.

Tendo em vista o escopo de possibilidades de consulta ao CAES (2020), e buscando ilustrar como o *corpus* pode ser usado como fonte do conteúdo das notas de uso de um dicionário de aprendizes brasileiros de espanhol, aplicou-se a seguinte metodologia de consulta:

- (1) Critério obrigatório de busca selecionado: lema. O campo foi preenchido com as ULs A2 e B1 que foram alvo das análises do Señas (2001) em 5.2.
- (2) Filtros de busca: país (Brasil) e idade (a partir de 18 anos).

Optou-se por não especificar os níveis de proficiência devido ao tamanho reduzido do CAES (2020) como um todo e da parcela de produções de brasileiros em particular. Ao não selecionar um determinado nível de proficiência, e considerando que os erros no uso das ULs objeto de busca ocorriam também em outros níveis além de A1 e A2, essa metodologia permitiu obter uma imagem mais clara dos erros dos aprendizes brasileiros e do quanto eles podiam perdurar no decorrer dos demais níveis de proficiência.

Nem todas as ULs consultadas tinham registro no CAES (2020); algumas, como *propina*, *ancho* e *reparar* não estavam registradas. A seguir, estão alguns dos erros encontrados.

Quadro 57 – Erros de aprendizes brasileiros de espanhol no CAES (2020)

<p>dirección Erro ortográfico: *<i>dirección</i> (5/28) (CAES, 2020, <i>dirección</i>, linhas 26, 27, 28, 29, 30)</p>	<p>gimnasio Erro ortográfico: *<i>gimnasio</i> (3/16) (CAES, 2020, <i>gimnasio</i>, linhas 14, 15, 16)</p>
<p>contestar Erro semântico: uso com o significado de <i>perguntar</i> (1/11) (CAES, 2020, <i>contestar</i>, linha 6)</p>	<p>invitar Erro sintático: ausência da preposição <i>a</i> junto ao objeto direto de pessoa (3/15) (CAES, 2020, linhas 10, 12, 13)</p>
<p>mayor Erro ortográfico: *<i>major</i> (8/58) (CAES, 2020, <i>mayor</i>, linhas 2-9) Erro semântico: <i>más viejo</i> ao invés de <i>mayor</i> nos registros de <i>viejo</i> (9/20) (CAES, 2020, <i>viejo</i>, linhas 1, 3, 5, 7, 13, 15, 16, 17, 19)</p>	<p>demasiado Erro semântico: uso com palavras de conotação positiva, como <i>admirar</i>. Nesses casos, é mais adequado usar <i>muy/mucho</i>. (2/17) (CAES, 2020, <i>demasiado</i>, linhas 1, 10, 11)</p>

todavía Erro ortográfico: * <i>todavía</i> (19/59) (CAES, 2020, <i>todavía</i> , linhas 1-19)	vago Erro semântico: uso no lugar de <i>libre</i> (1/2) (CAES, 2020, <i>vago</i> , linha 2)
Erro semântico: uso no lugar de nexos adversativos (3/59) (CAES, 2020, <i>todavía</i> , linhas 7, 23, 38)	ligar Erro semântico: uso de <i>ligar</i> no lugar de <i>llamar [por teléfono]</i> (1/2) (CAES, 2020, <i>ligar</i> , linha 2)

Fonte: Elaboração da autora com base nos dados do CAES (2020)

Erros de ordem semântica e ortográfica foram relativamente comuns na amostra analisada. Porém, não são os únicos. Durante a leitura dos alinhamentos do CAES (2020), percebeu-se que alguns tipos de erros de natureza mais genérica ocorriam com certa frequência na produção de aprendizes brasileiros, tais como acentuação desnecessária (**mí trabajo*), ausência de acentuação (**proximo*), ausência da preposição *a* nos objetos diretos de pessoa (**invitando el chico*), concordância, posição pronominal do pronome indicativo de objeto direto (acusativo) e beneficiário (dativo) (**puedes me contestar*), entre outros casos. Em geral, trata-se de erros que se aplicam a um grande número de ULs e que podem ser facilmente revisados por meio da consulta a um dicionário. Os problemas de acentuação em unidades léxicas totalmente similares ortograficamente com o português, como é o caso de *próximo*, chamam a atenção. Uma hipótese que se pode levantar é que o aprendiz percebe diferenças nos padrões de acentuação do espanhol, mas não consegue discernir em quais casos a acentuação se aplica pelas mesmas regras do português. Outra hipótese é a de que o aprendiz já apresente problemas de acentuação em sua LM; devido à similaridade ortográfica com o português, os padrões errôneos de acentuação do português acabam sendo aplicados também ao processamento do espanhol. Seja como for, são dados importantes para a compilação de qualquer material didático de ensino de espanhol para brasileiros.

Um problema mais severo, no entanto, não diz respeito aos erros dos aprendizes, mas sim a deficiências de ordem técnica do CAES (2020) no que concerne às contrações e amálgamas. Os dois casos de contrações no espanhol envolvendo a preposição *a* e o artigo *el* e a preposição *de* e o artigo *el* aparecem separados nos alinhamentos (**jugar a el fútbol*; **de el mes*). Ao expandir o contexto de diversos alinhamentos, percebeu-se que essa separação errônea é fruto do CAES (2020), pois muitas vezes não corresponde àquilo que o aprendiz escreveu (*jugar al fútbol*; *del mes*). Também as amálgamas entre verbos e pronomes, como (**Pido te*) aparecem separadas, mesmo que o aprendiz as tenha empregado no original (**Pidote*, que ainda carece de acento). Trata-se de uma falha na apresentação das informações dos alinhamentos

que atrasa consideravelmente a análise de dados, pois, para que se saiba se a separação foi realmente produzida pelo aprendiz, é necessário expandir um a um os alinhamentos do *corpus* que contenham contrações e/ou amálgamas.

7.3.2 Seleção de conteúdo para notas de uso: Fórum Português-Español (WR, 2020)

A segunda fonte de dados que pode auxiliar na elaboração de notas de uso de um dicionário de aprendizes brasileiros de espanhol são os fóruns de aprendizagem de línguas disponíveis na internet, como o “Português-Español” do site WordReference (WR, 2020)¹²⁵. A ferramenta dá acesso às mais de 19.000 dúvidas que não somente os aprendizes lusófonos de espanhol têm, mas que também têm os aprendizes hispânicos de português. Durante a consulta, não existe a opção de filtrar somente as dúvidas dos lusófonos que aprendem espanhol. Dos critérios de organização dos resultados, o que pareceu mais relevante foi a organização dos tópicos em ordem descendente por número de visualizações. Dessa forma, tem-se acesso às dúvidas mais recorrentes do público relevante para a pesquisa. A seguir, estão elencadas algumas dessas dúvidas.

Quadro 58 – Dúvidas de aprendizes brasileiros de espanhol no WordReference

<p>hacia x hasta (A2-B2) Dúvida quanto aos traços que diferenciam essas preposições ao indicar direção. (WR, 2020, <i>hacia, hacia, hasta?</i>)</p>	<p>a lo mejor + [modo verbal] (B1) Dúvidas sobre qual modo verbal acompanhava essa expressão – indicativo ou subjuntivo. (WR, 2020, “<i>a lo mejor</i>”: <i>subjuntivo o indicativo</i>)</p>
<p>pretérito perfecto simple x compuesto (A2) Dúvidas sobre o que cada tempo verbal denota e quais são os seus contextos de uso. (WR, 2020, <i>Pretérito perfecto simples e compuesto</i>)</p>	<p>solero x acostumbrar (B1) Dúvidas quanto à diferença semântica entre as duas ULs no espanhol. (WR, 2020, <i>Yo “suelo”??; solero vs. acostumbrar</i>)</p>
<p>mucho más + substantivo (A1) Dúvida sobre concordância de <i>mucho más</i> em função adjetiva ante substantivos. (WR, 2020, <i>mucho más / mucha más</i>)</p>	<p>Pronombre reflexivo (A1-B2) Dúvidas quanto à posição do pronome reflexivo em relação ao verbo a que se refere. (WR, 2020, <i>Pronome reflexivo em espanhol</i>)</p>

Fonte: Fórum Português-Español do site Wordreference. Consultado em 20 nov. 2020.

São comuns as dúvidas de ordem sintática e semântica. Também é comum a busca por esclarecimentos de questões gramaticais, tais como a distinção entre tempos verbais. É difícil atrelar temas dessa natureza a um ou outro verbete específico; talvez seja por isso que, nos

¹²⁵ Disponível em: <https://forum.wordreference.com/forums/portugués-español.55/>. Consultado durante os meses de novembro e dezembro de 2020.

dicionários de aprendizes, é comum encontrar compêndios de tópicos gramaticais no *front matter* ou no *back matter*.

7.3.3 Seleção de conteúdo para notas de uso: NDiFAEP (2014)

Quando o erro ou dúvida detectado no CAES (2020) ou no WR (2020) envolve falsos amigos entre o espanhol e o português, uma fonte oportuna de esclarecimentos é o NDiFAEP (2014). O dicionário se destaca por descrever a complexidade do fenômeno dos falsos amigos. O grande diferencial da obra é que, ao invés de salientar somente as diferenças de ordem semântica entre o português e o espanhol, o NDiFAEP (2014) apresenta os falsos amigos como palavras que, paralelamente, também podem apresentar convergência de significações. A seguir, reproduz-se um verbete do dicionário:

Quadro 59 – Excerto do NDiFAEP (2014)

academia	academia
I. f instituição científica / literária / artística / militar.	I. f institución científica / literária / artística / militar.
	II. f 1. for establecimiento de enseñanza superior. 2. establecimiento privado donde se practica gimnasia.
III. f estabelecimento de ensino especializado.	

Fonte: NDiFAEP (2014, s.v. *academia* / *academia*)

A metade esquerda do verbete contém o lema em espanhol e as significações usadas em espanhol e redigidas em língua portuguesa para facilitar a compreensão do consulente lusófono; a metade direita, por sua vez, contém o lema em português redigido em língua espanhola para facilitar a compreensão do consulente hispânico. Há três grupos de acepções, sinalizados por números romanos. No grupo I, estão os significados comuns às duas línguas; nos grupos II e III, estão os significados próprios do português e do espanhol, respectivamente. Sob a ótica saussureana que vem sendo adotada ao longo da tese, em espanhol, há duas ULs: por um lado, *academia* como instituição científica / literária / artística / militar; por outro lado, *academia* como estabelecimento de ensino especializado. Destas duas ULs, uma é cognata em relação ao português, enquanto a outra, não. Ao invés tratar somente das semelhanças ou das diferenças

semânticas entre o português e o espanhol, o NDiFAEP (2014) oferece um panorama mais aprofundado da complexidade das relações entre as duas línguas, o que o converte em uma ferramenta de grande relevância para o ensino-aprendizagem de espanhol a brasileiros e de português a hispânicos.

7.3.4 Esboço de notas de uso para adaptação do *Señas* (2001)

A partir dos erros coletados no CAES (2020) e das dúvidas do fórum WordReference, esboçamos notas de uso que contemplem os problemas e dúvidas dos brasileiros com o uso do espanhol.

Quadro 60 – Propostas de aprimoramento das notas de uso (espanhol)

<p>di·rec·ción dire^kθi^on [...]</p> <p>✎ ORTOGRAFÍA Se escribe siempre con doble c: dirección. ¡No escribas dirección!</p>	<p>gim·na·sio xi^mnásio [...]</p> <p>✎ ORTOGRAFÍA Se escribe siempre con i: gimnasio. ¡No escribas gymnasio!</p>
<p>ma·yor mayór </p> <p>1 adj. [ser vivo] que tiene más edad que otro; en portugués, <i>mais velho</i>: <i>su hermano mayor vendrá con nosotros</i>.</p> <p>✎ VOCABULARIO Cuando se quiere decir que una persona tiene más edad que otra, se usa mayor, no más viejo.</p> <p>2 [...]</p> <p>✎ ORTOGRAFÍA Se escribe siempre con y: mayor. ¡No escribas major!</p>	<p>de·ma·sia·do, da demasiádo, ða </p> <p>1 adj.-pron. indef. [algo] que supera lo justo o lo necesario; que es excesivo; en portugués, <i>demais</i>: <i>tiene demasiadas necesidades</i>.</p> <p>✎ USO Expresa algo negativo: <i>hay gente que usa demasiado perfume</i>. No se usa para intensificar algo positivo. En estos casos se usa mucho: <i>tengo demasiadas muchas aficiones</i>.</p>
<p>vie·jo biéxo, xa </p> <p>1 adj.-s. [ser vivo] que tiene una edad avanzada; en portugués, <i>velho</i>: <i>mi perro es muy viejo</i>. ⇨ <i>anciano</i>.</p> <p>✎ VOCABULARIO Cuando se quiere decir que una persona tiene más edad que otra, se usa mayor, no más viejo.</p>	<p>li·gar liyár </p> <p>[...]</p> <p>✎ VOCABULARIO 1-2 Para contactar a alguien por teléfono, se usa el verbo llamar (por teléfono): <i>cuando esté a camino de casa te ligaré llamaré</i>.</p>
<p>so·ler solér </p> <p>1 intr. Tener costumbre; en portugués, <i>costumar</i>: <i>tu padre suele venir los sábados</i>. ⇨ <i>acostumbrar</i>.</p> <p>✎ VOCABULARIO: ¿SOLER O ACOSTUMBRAR? Cuando se trata de describir un hábito o costumbre ya adquiridos, soler y acostumbrar son sinónimos. Para referirse a la necesidad u obligatoriedad de crear un hábito o costumbre, se usa solo acostumbrar: <i>Es mejor que te acostumbres a levantarte temprano</i>.</p>	<p>va·go báyo, ya </p> <p>1 adj.-s. [persona] a quien no le gusta el trabajo; en portugués, <i>vagabundo</i>: <i>Sonia es una vaga</i>: <i>se pasa el día tumbada en el sillón</i>.</p> <p>2 [cosa] que está poco clara; que no está determinada: <i>tan sólo tengo una idea vaga de lo que quieres decir</i>. ⇨ <i>impreciso</i>.</p>

2 [...]	<p>✎ VOCABULARIO Para decir que un espacio no está ocupado por algo o alguien, se usa la palabra libre: <i>había un sitio vago libre para sentarse.</i></p>
<p>to·da·ví·a toðaβía </p> <p>1 adv. t. Hasta un momento determinado; hasta ahora; en portugués, <i>ainda: todavía está durmiendo.</i> ⇨ <i>aún.</i></p> <p>- 2 adv. Indica mayor intensidad en las comparaciones; en portugués, <i>ainda: Juan es todavía más listo que su hermano.</i></p> <p>3 Con todo eso; a pesar de ello; en portugués, <i>ainda assim: aunque sé que es malo, todavía lo quiero.</i></p> <p>✎ VOCABULARIO 1-3 No se usa para indicar oposición. La oposición se expresa en español por nexos como sin embargo: <i>No soy fumador. Todavía Sin embargo, fumo algunas veces en la calle o en bares.</i></p>	<p>has·ta asta </p> <p>1 prep. Límite: indica el punto del cual no se pasa en cuanto al tiempo, el espacio o la cantidad; en portugués, <i>até: iremos hasta Madrid en avión y luego cogeremos el tren.</i></p> <p>✎ USO Puede sustituir a la preposición <i>a</i> para indicar dirección: <i>nos acercaremos a/~ tu casa.</i></p> <p>- 2 conj. Indica la posibilidad de llegar a cierto límite en cuanto al tiempo, al espacio o a la cantidad; en portugués, <i>até: podría pagar por ello 3000 pts y hasta más.</i></p> <p>✎ VOCABULARIO: ¿HACIA O HASTA? 1-2 Hacia se usa para referirse a un límite aproximado, poco preciso. Hasta, por otro lado, se usa para indicar un límite exacto, preciso.</p>
<p>mu·cho mútʃo, tʃa </p> <p>1 adj.-pron. indef. [seres vivos o cosas] en gran número o cantidad; en portugués, <i>muito: hay mucha gente en la sala.</i> ⇨ <i>bastante.</i></p> <p>✎ USO La expresión comparativa mucho más concuerda en género y número con el sustantivo a que se refiere: <i>Mi hermano tiene mucha más paciencia que yo.</i></p>	<p>con·tes·tar kontestár </p> <p>1 tr. Responder a [algo] que se pregunta, dice o escribe; en portugués, <i>responder: tengo que contestar su carta.</i></p> <p>✎ VOCABULARIO No tiene el significado de «hacer preguntas». El hecho de «hacer preguntas» se expresa por verbos como preguntar, cuestionar, etc.</p> <p>2 [...]</p>
<p>me·jor mexór </p> <p>[...]</p> <p>■ a lo mejor, <i>fam.</i>, Expresa posibilidad e incerteza; en portugués, <i>talvez: a lo mejor puedo venir a tu fiesta, aún no lo sé.</i></p> <p>✎ USO Los verbos que acompañan esta expresión deben estar en indicativo. No se usa el subjuntivo.</p>	<p>in·vi·tar imbitár 1 tr. Pedirle [a alguien] para estar en una celebración, comida o una otra cosa; en portugués, <i>convidar: la invitó a comer en su casa. Me gusta mucho invitar a mis amigos para mirar al fútbol.</i></p>

Fonte: As notas de uso e os destaques tipográficos nas transcrições fonológicas foram elaborados pela autora. As demais informações foram retiradas do Señas (2001).

As notas foram redigidas conforme as diretrizes propostas no início do capítulo (7.1). Para ilustrar a interação das notas com o restante do verbete, ou quando a nota era de tipo exortativo, algumas acepções foram incluídas e adaptadas a partir do original do Señas (2001) e de acordo com as propostas do capítulo 6 a respeito do comentário de forma e das paráfrases explanatórias. Além disso, optou-se por reduzir as informações ao mínimo necessário para não sobrecarregar o verbete, de forma que os exemplos das acepções foram reduzidos a um e as informações extra, como os antônimos, foram retiradas. Como produto, obtêm-se verbetes mais limpos, que facilitam a localização das notas de uso e a compreensão de informações presentes

tanto nas notas como nas acepções. Em última instância, cumpre-se com os pressupostos do didatismo e da inclusão de informações discretas e discriminantes.

Por fim, o único caso em que não se incluiu uma nota de uso foi em *invitar*, a respeito do objeto direto de pessoa. Como não se trata de uma característica exclusiva desse verbo, mas de qualquer verbo que aceita objeto direto de pessoa, optamos por incluir um novo exemplo, que consiste na adaptação de um registro do CAES (2020) produzido por um estudante brasileiro.

Como corolário da seção sobre inclusão de notas de uso no Señas (2001), concluímos que há ferramentas suficientes para determinar as dúvidas e dificuldades do público brasileiro A1-B1 em relação à aprendizagem do espanhol. O fato de que sejam suficientes não significa que tais ferramentas possa oferecer uma imagem tão próxima do perfil de usuário em questão como poderia um *corpus* de aprendizes mais extenso e com anotação de erros. Não se pode, no entanto, desconsiderar o custo de compilação de um *corpus* de aprendizes em termos de recursos financeiros, humanos e de tempo, nem o fato de que os dados (*i.e.* as produções de aprendizes) são limitados em termos de quantidade e, muitas vezes, ainda precisam ser gerados para esse fim – ao contrário dos *corpora* gerais de língua, onde os dados já estão disponíveis previamente. Ante a diversidade de desafios inerentes à compilação de um *corpus* de aprendizes, é necessário fazer o melhor com o que já se tem e promover iniciativas para a construção e manutenção de bases de dados dessa natureza. A divulgação da sua utilidade para pesquisas e produção de materiais didáticos é, talvez, o primeiro passo para o seu aprimoramento.

No que diz respeito à metodologia de coleta e análise dos dados do CAES (2020) e do WR (2020), permanece a questão do que se poderia considerar como mais (ou menos) relevante em termos de frequência de erros e dúvidas. Novamente, ressaltamos que mais investigações sobre esse tema, bem como a interseção com a Lexicografia, são necessárias para uma avaliação mais consciente e um pouco menos subjetiva a respeito dos graus de relevância dos dados.

Finalmente, e tal como já foi comentado na seção sobre o inglês, as propostas relativas à redação das notas de uso só poderão se consolidar como orientações metalexigráficas mediante a sua testagem junto ao público-alvo. Trata-se de um procedimento posterior à pesquisa que relatamos na tese.

Este capítulo encerra a explanação acerca da determinação do perfil de usuário em dicionários de aprendizes. Por se tratar de um tema transversal na tese, continuar-se-á recorrendo às considerações sobre o perfil de usuário nos demais capítulos. O seguinte pilar que sustenta as propostas de aprimoramento dos dicionários de aprendizes são as combinatórias

léxicas – tema que já emergiu durante os comentários aos excertos de verbetes dos dicionários analisados.

8 AS COMBINATÓRIAS LÉXICAS NO DICIONÁRIO DE APRENDIZES

As palavras não são unidades autossuficientes em uma língua; ao contrário, elas se combinam. A constatação desse fenômeno é, talvez, um dos pontos de interseção que une diferentes ramos da Linguística, em especial os que se preocupam da Sintaxe e da Semântica, da Lexicologia e da Linguística de *Corpus*. As estimativas acerca da presença de combinatórias léxicas em uma língua são várias. Com base na sua experiência com *corpus*, Sinclair (2004, p. 4) chega a afirmar que, em “grandes porções de texto” [a large proportion of text], há uma predominância de 50% de combinações de palavras em comparação com palavras isoladas – considerando *palavra* genericamente como sequência de letras dotada de significado e ladeada por espaços em branco. Erman e Warren (2000, p. 37), por sua vez, ao conduzirem um estudo sobre a presença de combinatórias léxicas em textos do inglês oral e escrito, encontraram uma predominância de 52,3% de combinatórias nas amostras de textos escritos e de 58,6% nas amostras de textos orais, totalizando uma média de 55,3% nos dois subgrupos. Os resultados dos autores são comumente mencionados como referência para a presença de combinatórias léxicas em uma língua (independentemente de qual seja), como o fazem Sánchez Rufat (2018, p. 261) e Mel’čuk (2015, p. 359). Embora mencione os dados de Erman e Warren (2000), Mel’čuk (2015, p. 298), por outro lado, prefere defender uma estimativa própria e subjetiva de que as combinatórias estariam em uma proporção de 10:1 em comparação com unidades léxicas consideradas “livres” (cf. conceito mais adiante).

Se as combinatórias importam às áreas “duras” da análise e descrição linguística, são também, por conseguinte, objeto de interesse das áreas aplicadas, como a Linguística Aplicada ao Ensino de Línguas e a (Meta)Lexicografia. Já há algum tempo se sabe que ensinar e aprender uma língua implica ter consciência da maneira como as suas palavras se combinam. Martinez (2014, p. 184), ao revisar os resultados e conclusões de estudos empíricos a respeito do papel das combinatórias léxicas no ensino de inglês como LE, observa que há um consenso sobre o fato de que esse tipo de fenômeno desempenha “[...] um papel essencial na aquisição, compreensão e produção de língua”¹²⁶, tanto da LM como da LE. Cowie (2008, p. 163), por sua vez, é categórico ao afirmar que a “[...] proficiência nativa em uma língua depende necessariamente do conhecimento de um estoque de unidades pré-fabricadas [...]”¹²⁷. À (Meta)Lexicografia, por sua vez, cabe desenhar e compilar dicionários que descrevam as combinatórias léxicas de maneira didática, isto é, associando os avanços nos estudos teóricos

¹²⁶ [an essential role in language acquisition, comprehension, and production]

¹²⁷ [native-like proficiency in a language depends crucially on knowledge of a stock of prefabricated units]

sobre o tema às características do perfil de usuário. Na condição de área de estudos *aplicada*, a (Meta)Lexicografia não se vincula unicamente a uma ou outra corrente teórica, mas é responsável por aproveitar as propostas que melhor se ajustam tanto à descrição lexicográfica da língua como ao perfil de usuário. À (Meta)Lexicografia de aprendizes e, em particular, a esta tese, estando definido o usuário de um dicionário de aprendizes a partir do conhecimento prévio de etapas anteriores de aprendizagem e da sua língua materna, bem como a partir do conhecimento da etapa de ensino-aprendizagem na qual se encontra, cabe explorar as principais contribuições de diferentes ramos da Linguística para propor uma descrição lexicográfica e didática das combinatórias. Conforme se constatou nos capítulos 1, 6 e 7, embora os cinco dicionários de aprendizes de inglês e de espanhol de que se preocupa este trabalho descrevam combinatórias léxicas, a descrição oferecida é, muitas vezes, pouco clara e pouco uniforme.

O presente capítulo executa o referido movimento de análise de perspectivas e síntese propositiva a respeito da descrição de combinatórias léxicas em dicionários de inglês e de espanhol para usuários brasileiros A1-B1. Primeiramente, são comentadas as principais correntes de estudos das combinatórias léxicas. Em seguida, são elencados alguns dos tipos de combinatórias identificados na teoria e estabelece-se um paralelo com a descrição de cada fenômeno nas cinco obras lexicográficas de que nos ocupamos. Como síntese, propõem-se estratégias de descrição lexicográfica dos fenômenos analisados tendo em vista o desenho de dicionários de aprendizes de inglês e de espanhol, e, em especial, a adaptação das obras que interessam a este trabalho.

8.1 Breve panorama dos estudos sobre combinatórias léxicas

Embora seja arriscado precisar exatamente quando se começou a estudar as combinações que as palavras estabelecem entre si, é comum encontrar referências na literatura que consideram Firth (1962) como um marco ao qual se seguiu um crescente interesse pelo fenômeno em questão (cf. TARP, 2008, p. 250; WRAY, 2013, p. 320; STEYER, 2015, p. 280). Firth (1962) se insere em uma perspectiva de análise linguística denominada *contextualismo britânico*, ou *contextualismo inglês*. Em linhas gerais, essa corrente se caracterizava por se restringir às manifestações físicas (observáveis) de uma língua e propunha que esta fosse analisada sob uma perspectiva holística e social (FIRTH, 1962, p. 2-7).

Por um lado, para Firth (1962) e os demais contextualistas, um estudo somente poderia ser circunscrito ao âmbito da linguística se estivesse baseado nas manifestações concretas da língua, isto é, em textos. Nesse sentido, nenhuma afirmação sobre o funcionamento de uma

língua poderia ser feita sem a comprovação empírica. A importância desse posicionamento reside no fato de que o empirismo por meio de *corpora* é uma marca registrada dos linguistas britânicos presente ainda nos dias de hoje. Não é coincidência que os primeiros grandes *corpora* linguísticos e a própria Linguística de *Corpus* tenham surgido em solo britânico, a começar pela iniciativa COBUILD (cf. detalhes em 8.1.1).

Por outro lado, ao tratar a língua como um sistema holístico, Firth (1962, p. 4-5; 7) rechaçava as dicotomias estruturalistas. Para o autor, cada língua deveria ser concebida como um sistema de relações internas (eixos sintagmático e paradigmático) e externas (eixo diafásico-diastrático) presentes em diferentes níveis de organização da linguagem (FIRTH, 1962, p. 5; 9-10). Dentre as relações internas de uma língua, as que mais interessam a Firth (1962) são aquelas que se estabelecem no eixo sintagmático. Tanto é assim que o autor reserva duas seções para tratar de combinatórias léxicas, proferindo, em um dado momento, a famigerada frase “You shall know a word by the company it keeps!” [Você conhecerá uma palavra pelas companhias que ela mantém] (FIRTH, 1962, p. 11).

O interesse pelas combinatórias léxicas cresceu tanto após o trabalho de Firth (1962) que três perspectivas a respeito das combinatórias léxicas começaram a tomar forma e a se estruturar geograficamente: a estatística, na Inglaterra, e a lexicológica e a gramatical, na Europa. As seções a seguir discorrem acerca de cada uma.

8.1.1 Viés estatístico sobre as combinatórias léxicas: Sinclair (1991, 2004) e Hoey (2005)

O viés estatístico se desenvolveu primordialmente na Inglaterra e tem Sinclair (1991, 2004) como o seu principal expoente. O advento dos primeiros *corpora* informatizados proporcionou o acesso a um volume de dados exponencial, que precisava ser processado de maneira cada vez mais automática. Nesse contexto, o viés estatístico surge com o objetivo de estudar metodologias para a identificação (semi-)automática das combinatórias léxicas. Segundo Evert (2009, p. 1243), por exemplo, há mais de cinquenta¹²⁸ opções de fórmulas para extrair um dos tipos de combinação de palavras (as colocações).

Um importante marco da abordagem estatística foi o projeto COBUILD, uma parceria entre a editora Collins e a Universidade de Birmingham nos anos 1980, liderado por Sinclair. O projeto culminou na compilação do maior *corpus* eletrônico de língua inglesa do mundo à época, o Collins Corpus, bem como do primeiro dicionário com tais proporções de dados (THE

¹²⁸ Algumas das fórmulas mais conhecidas são as que aparecem citadas no glossário da ferramenta Sketch Engine: logDice, log-likelihood, Mutual Information e T-score (<https://www.sketchengine.eu/guide/glossary/>).

HISTORY of Cobuild, 2021). Adotando a perspectiva do contextualismo inglês, Sinclair pôs em prática os pressupostos de Firth (1962) tanto no que diz respeito à descrição linguística a partir de conjuntos de textos autênticos como ao estudo da maneira como as palavras se combinam.

Longe de se reduzir a fórmulas matemáticas para identificação e extração automática de dados de *corpora*, a abordagem estatística tem, curiosamente, uma base que se poderia chamar de semântica. Sinclair (2004), ao comentar a origem do seu interesse pelas combinatórias, inicia pela dimensão do significado das palavras e como esse significado se manifesta linguisticamente. O autor comenta que discordava da perspectiva de relação 1:1 entre palavra e significado, argumentada pela semântica lexical, bem como da maneira como essa perspectiva concebe a polissemia e as ditas “palavras gramaticais” na veiculação do significado (SINCLAIR, 2004, p. 1-2). Para Sinclair (2004, p. 2), primeiramente, a simples constatação de que havia palavras com mais de um significado atrelado (polissemia) e que dita pluralidade não acarretava problemas de compreensão na comunicação não era suficiente para entender o fenômeno da polissemia. Além disso, Sinclair (2004, p. 1-2) também questionava se o único caso de desvio da norma 1:1, culminando em uma manifestação sintagmática do significado, seria o das fraseologias [idioms] – unidades sintagmáticas com significado não composicional. Por fim, o autor não compreendia por que as palavras gramaticais eram excluídas dos estudos semânticos, pois, para ele, essa categoria de palavras tinha, sim, participação na veiculação do significado.

A experiência de Sinclair com o estudo da língua inglesa a partir do Collins Corpus permitiu ao autor responder às três questões que o intrigavam na concepção do significado pela via da semântica lexical (SINCLAIR, 2004, p. 2; 4). No tocante à polissemia, Sinclair (2004, p. 4) argumenta que “a pesquisa com corpus oferece uma explicação completa dessa situação ao indicar que o significado é criado não sobre cada palavra isolada, mas sobre muitas palavras juntas”¹²⁹. Mais precisamente, uma palavra polissêmica não causava problemas de comunicação porque “as outras palavras ao redor tendem a focar a mente em apenas um dos significados disponíveis”¹³⁰ (SINCLAIR, 2004, p. 2). Por conseguinte, e conforme Sinclair (2004, p. 4), a construção sintagmática do significado não se reduz às fraseologias [idioms], mas contempla um amplo espectro de casos. O autor segue ainda assinalando que tais casos incluem não somente palavras de classe “aberta”, como afirmavam os semanticistas, mas

¹²⁹ [Corpus research offers a full explanation of this situation by pointing out that meaning is created, not over each single word, but over several words together]

¹³⁰ [the other words around tend to focus the mind on only one of the available meanings]

também palavras de classes ditas “fechada”, ou gramaticais, como os *phrasal verbs* (verbo + preposição/advérbio), por exemplo (SINCLAIR, 2004, p. 4). No português e no espanhol, as regências são um exemplo de combinatória que, de forma similar aos *phrasal verbs*, inclui uma palavra gramatical (no caso, uma preposição) que tem relação direta com a constituição do significado da regência.

As constatações relatadas em Sinclair (2004) repercutiram na maneira ímpar com que hoje os verbetes polissêmicos são desenhados nos dicionários de aprendizes de inglês. As acepções de diferentes dicionários costumam ser acompanhadas de um padrão morfossintático que sinaliza a manifestação formal da distinção de cada significado e/ou dos complementos que costumam aparecer na manifestação de cada significado, especialmente nos dicionários de aprendizes de tipo *intermediate* e *advanced*.

Quadro 61 – Distinção de acepções em verbetes polissêmicos nos dicionários de inglês

<p>fight¹ /faɪt/ verb (fights, fighting, fought)</p> <p>1 to try to hurt someone with words or by using physical force: 'Stop fighting!' Mum shouted. □ Susan fought a lot with her younger sister.</p> <p>2 to take part in a war: He fought in the war and was taken prisoner.</p> <p>3 to try very hard to stop something</p> <p>unpleasant: It is very hard to fight forest fires.</p> <p>4 to try very hard to get something: Lee had to fight hard for his place on the team.</p> <p>5 to argue (informal): Robert's parents fight all the time.</p> <p>(CoED2, 2014)</p>	<p>fight /faɪt/ (fights, fighting, fought /fɔ:t/)</p> <p>1 V-T & N-COUNT If you fight something, or fight against something, you try in a determined way to prevent it or stop it happening. A fight against something is an attempt to stop it. □ The police are working hard to fight crime. □ I've spent a lifetime fighting against racism. □ ...the fight against cancer. 2 V-T/V-I & N-COUNT If you fight for something, or fight to do something, you try in a determined way to get it or achieve it. A fight for something is an attempt to get it. □ The vets fought to save the most popular horse after he fell during a race. □ ...the fight for justice. 3 V-I If a person or army fights, they oppose each other with weapons. □ I will never fight for this country. □ The two countries have been fighting for more than two years. • fighting N-UNCOUNT □ More than nine hundred people have died in the fighting. 4 V-T/V-I & N-COUNT If one person fights with another, if</p> <p>(CCID4, 2018)</p>
--	---

Fonte: CoED2 (2014, s.v. *fight*¹) e CCID4 (2018, s.v. *fight*)

Nos verbetes acima, nota-se que apenas o CCID4 (2018), de nível intermediário, indica as preposições dos complementos de *fight* em cada acepção, enquanto o seu correspondente de níveis iniciais, não. Além de constituírem um traço distintivo dos dicionários de aprendizes intermediários e avançados de inglês, as informações combinatórias junto a cada acepção auxiliam tanto na recepção, ao permitirem que o usuário identifique a acepção correspondente à sua busca a partir do padrão morfossintático, como também, e de maneira especial, na produção de língua, ao descreverem como cada significado se manifesta formalmente. Porém, essa descrição é reduzida ao mínimo, quando não omitida, quando se trata de dicionários de aprendizes A1-B1. Aparentemente, na tentativa de simplificar os verbetes a fim de deixá-los

mais palatáveis e fáceis de compreender para um consultante de níveis iniciais de proficiência, os editores do CoED2 (2014) abriram mão da descrição dos padrões combinatórios, reduzindo em grande medida o potencial de aplicação das obras à função de recepção.

Ao relacionar os significados de palavras polissêmicas ao seu comportamento sintagmático, e ao perceber que esse comportamento não era suficientemente contemplado pelos estudos linguísticos (SINCLAIR, 2004, p. 2), Sinclair (1991) propõe uma designação para se referir aos fenômenos de natureza sintagmática: o *idiom principle* [princípio idiomático]. Segundo esse princípio,

[...] o usuário da língua tem à sua disposição um grande número de sintagmas pré- e semiconstruídos que constituem escolhas únicas, embora possam parecer analisáveis em segmentos. [...] Na sua forma mais simples, o princípio idiomático pode ser visto na escolha aparentemente simultânea de duas palavras, por exemplo, *é claro*. Este sintagma opera efetivamente como uma única palavra [...] (SINCLAIR, 1991, p. 110)¹³¹.

Sinclair (1991, p. 111) faz uma série de constatações acerca da considerável variação formal dos sintagmas que constituem o princípio idiomático. A seguir, reproduzimos algumas delas com exemplos em português para facilitar a sua compreensão.

a) “Muitos sintagmas têm uma extensão indeterminada”¹³² (SINCLAIR, 1991, p. 111). Um exemplo disso é o caso de *imposto de renda* (f=19828) (CP, 2019). Esse sintagma pode ser considerado bem delimitado até o momento em que se pode estendê-lo para *declaração do imposto de renda* (f=2485), que, por sua vez, pode se estendido para *fazer a declaração do imposto de renda* (f=173) (CP, 2019).

b) “Muitos sintagmas permitem variação léxica interna”¹³³ (SINCLAIR, 1991, p. 111), ou seja, pode haver um paradigma de opções para um de seus componentes. Por exemplo: *ficar/estar/continuar/assistir/esperar de braços cruzados* (*ficar*: f=510; *estar*: f=140; *continuar*: f=40; *assistir*: f=38; *esperar*: f=18) (CP, 2019).

c) “Muitos sintagmas permitem variação na ordenação das palavras”¹³⁴ (SINCLAIR, 1991, p. 111), isto é, pode haver alteração na posição sintática dos elementos de um sintagma.

¹³¹ [...] a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. [...] At its simplest, the principle of idiom can be seen in the apparently simultaneous choice of two words, for example, *of course*. This phrase operates effectively as a single word [...]

¹³² [Many phrases have an indeterminate extent]

¹³³ [Many phrases allow internal lexical variation]

¹³⁴ [Many phrases allow some variation in word order]

Por exemplo: os componentes de *cair a ficha* podem aparecer nesta mesma ordem, como em “Ainda não caiu a ficha” (CP, 2019), ou invertidos, como em “Demorou para a ficha cair” (CP, 2019).

d) “Muitas palavras e sintagmas têm a tendência de coocorrer com certas opções gramaticais”¹³⁵ (SINCLAIR, 1991, p. 112), isto é, podem estar associados a um tempo, modo ou aspecto verbal específico, ou a uma determinada forma nominal, por exemplo. É o caso de *perder o amigo, mas não perder a piada* (f=20) sintagma em que o verbo *perder* é usado somente no presente segundo o CP (2019).

e) “Muitas palavras e sintagmas têm a tendência de ocorrer em um certo contexto semântico”¹³⁶ (SINCLAIR, 1991, p. 112), isto é, se associam a palavras que têm um traço semântico em comum, como *comemorar* + [acontecimento feliz], por exemplo.

Apesar das suas diversas constatações a respeito da variedade de sintagmas que ilustram o princípio idiomático, Sinclair (1991) não aprofunda o tema nem almeja fornecer um panorama classificatório de tipos de combinatórias léxicas e suas propriedades. Seu interesse recai sobre um único tipo de combinatória que ele usa para ilustrar manifestações do princípio idiomático: as colocações (SINCLAIR, 1991, p. 115). A primeira menção do fenômeno no âmbito dos estudos contextualistas ingleses foi a de Firth (1962, p. 11-13; 14), que as define como “palavras reais em companhia habitual” [actual words in habitual company] (FIRTH, 1962, p. 14). Sinclair (1991, p. 170) desenvolve o conceito definindo *colocação* como

[...] a ocorrência de duas ou mais palavras a uma curta distância uma da outra. A medida comum de proximidade é um máximo de quatro palavras interpostas. As colocações podem ser dramáticas e interessantes porque inesperadas, ou podem ser importantes na estrutura léxica da língua porque repetidas com frequência [...]¹³⁷ (SINCLAIR, 1991, p. 170).

A definição de Sinclair (1991, p. 170) a respeito das colocações diz respeito aos critérios para identificação de estruturas com certas características de frequência e localização nos alinhamentos de um *corpus*. Observe-se que, assim como Firth (1962), o autor não distingue entre palavras lexicais ou de classe aberta (substantivos, verbos, adjetivos e advérbios) e palavras gramaticais ou de classe fechada (preposições, pronomes, verbos modais e auxiliares

¹³⁵ [Many uses of words and phrases show a tendency to co-occur with certain grammatical choices]

¹³⁶ [Many uses of words and phrases show a tendency to occur in a certain semantic environment]

¹³⁷ [the occurrence of two or more words within a short space of one another. The usual measure of proximity is a maximum of four words intervening. Collocations can be dramatic and interesting because unexpected, or they can be important in the lexical structure of the language because of being frequently repeated]

etc.). Em outros termos, e como sintetiza Legallois (2012, p. 39), para Firth (1962) e Sinclair (1991), uma colocação é “[...] um fenômeno de associação entre palavras lexicais [...], entre palavras lexicais e gramaticais [...], entre palavras gramaticais [...]”¹³⁸. Sob essa ótica, as combinações de *fight* com preposições, como *fight against*, *fight for* e *fight with*, descritas no CCID4 (2018, s.v. *fight*) e reproduzidas anteriormente, são todas colocações, assim como os *phrasal verbs*.

Em outro ponto da sua exposição, Sinclair (1991, p. 115) ainda esclarece que, na busca pelas colocações de uma palavra em um *corpus*, a palavra de referência é chamada de *nódulo* [node], enquanto as palavras que aparecem frequentemente combinadas com a base são chamadas de *colocados* [collocates]. Dessa forma, se o ponto de partida da busca de uma combinação é *fight*, esse será o *nódulo* da colocação, enquanto *against*, *for* e *with* serão os colocados. Ao contrário, se o objeto da busca for a preposição *for*, esta será o *nódulo*, e *fight* e outras palavras serão consideradas como colocados.

O número estândar de componentes de uma colocação é dois, mas Sinclair (1991, p. 170) não exclui possibilidades de colocação com três ou mais constituintes. A frequência com que duas ou mais palavras aparecem a ponto de enquadrá-las no fenômeno das colocações, por sua vez, é chamada de *força de colocação* [collocational strength] (LEHECKA, 2015), sendo medida por fórmulas estatísticas, conforme já mencionado. A distância entre os componentes de uma colocação diz respeito ao limite de palavras interpostas para que a coocorrência de duas palavras possa ser considerada como colocação. Trata-se da *janela colocacional* [collocational window], que, conforme Lehecka (2015, p. 3), varia segundo a perspectiva de cada pesquisador. Tanto é assim que o CP (2019), por exemplo, o mecanismo de busca permite que se determine de zero a nove palavras interpostas à direita e/ou à esquerda da palavra-chave da busca. Finalmente, o aspecto “inesperado” das colocações não é desenvolvido por Sinclair (1991) em sua definição, apesar de ser o mais instigante. O viés lexicológico é o que se ocupa em investigar tal efeito por meio da via semântica.

As pesquisas sobre colocações iniciadas por Firth (1962) e desenvolvidas por Sinclair (1991) e outros contextualistas tinham uma finalidade bem delimitada: a descrição lexicográfica da língua inglesa. Nota-se que tanto Sinclair (2004, p. 4-5; 10) como Firth (1962, p. 26-27) manifestaram claramente o seu interesse pela aplicação lexicográfica suas análises linguísticas. De fato, na linguística inglesa, *corpus* e dicionário estão intrinsecamente relacionados, de forma que o primeiro alimenta o segundo. Tanto é assim que as mesmas instituições e organizações

¹³⁸ [um phénomène d’association entre mots lexicaux [...], entre mots lexicaux et grammaticaux [...], entre mots grammaticaux]

detentoras de *corpora* cobijados por qualquer estudioso de Linguística de/com *Corpus* são as que publicam dicionários que descrevem as línguas com dados desses *corpora*. Na linguística inglesa, o dicionário constitui um elo entre os dados “crus” do *corpus* e o usuário comum, com a função de apresentar tais dados sobre a língua em uso de uma maneira didática. No caso dos dicionários de aprendizes de inglês, particularmente daqueles que se intitulam intermediários e avançados, é comum encontrar quadros com uma relação de colocados de determinadas palavras-entrada. Observem-se os exemplos a seguir.

Quadro 62 – Descrição de colocações nos dicionários de aprendizes de inglês

<p>COLLOCATIONS</p> <p>verbs</p> <p>to make a decision <i>I've made my decision: I'm leaving.</i></p> <p>to take a decision <i>BrE</i> (=to make an important or formal decision) <i>The company took the decision to close the factory.</i></p> <p>to come to/reach a decision (=to make a decision after a lot of thought) <i>The jury took three days to reach a decision.</i></p> <p>adjectives</p> <p>an important/major decision <i>Don't make any important decisions without talking to me.</i></p> <p>a big decision (=an important decision) <i>It was a big decision and I needed some time to think about it.</i></p> <p>a difficult/hard/tough decision <i>She was facing a difficult decision.</i></p> <p>a good/bad decision <i>I admit I've made some bad decisions in the past.</i></p> <p>the right/wrong decision <i>I'm sure you've made the right decision.</i></p> <p>(LAS5, 2010, s.v. <i>decision</i>)</p>	<p>Word partners for decision</p> <p><i>make/take a decision • come to/reach a decision • overturn/reverse a decision • a difficult/hard/tough decision • a big/crucial/important decision • a conscious decision • a hasty/snap decision • a decision about/on sth</i></p> <p>(CALD4, 2013, s.v. <i>decision</i>)</p> <p>de-ci-sion <i>dr'si:ʒn/ noun</i></p> <p>1 [C] ~ (on/about sth) ~ (to do sth) a choice or judgment that you make after thinking and talking about what is the best thing to do: <i>to make a decision</i> (= to decide) ◊ (BrE) <i>to take a decision</i> (= to decide) ◊ <i>We need a decision on this by next week.</i> ◊ <i>Who took the decision to go ahead with the project?</i> ◊ <i>He is really bad at making decisions.</i> ◊ <i>We finally reached a decision</i> (= decided after some difficulty). ◊ <i>We must come to a decision about what to do next by tomorrow.</i> ◊ <i>a big</i> (= an important) <i>decision</i> ◊ <i>The final decision is yours.</i> ◊ <i>It's a difficult decision for any doctor.</i> ◊ <i>The editor's decision is final.</i> ◊ <i>Mary is the decision-maker in the house.</i> 2 (also de-cisive-ness) [U] the ability to decide sth clearly and quickly: <i>This is not a job for someone who lacks decision.</i> OPP indecision 3 [U] the process of deciding sth: <i>The moment of decision had arrived.</i></p> <p>(OALD8, 2010, s.v. <i>decision</i>)</p>
---	---

Fonte: LAS5 (2010, s.v. *decision*), CALD4 (2013, s.v. *decision*), OALD8 (2010, s.v. *decision*)

No LAS5 (2010) e no CALD4 (2013), que são dicionários de tipo intermediário e avançado, respectivamente, os colocados de *decision* são indicados em quadros após as acepções e organizados de acordo com a sua classe gramatical. No OALD8 (2010), por outro lado, optou-se por incluir exemplos com algumas das colocações de *decision* e destacá-las em negrito. Nota-se que não raras vezes há colocações com mais de dois componentes, como *the right decision* e *to come to a decision*. Há também um certo grau de opções paradigmáticas, como em *a difficult/hard/tough decision*, e *a good/bad decision*. A concepção contextualista inglesa a respeito das colocações é suficientemente ampla para abarcar essa diversidade de combinações de palavras.

O uso de recursos para descrever e destacar colocações nos verbetes, tais como os quadros do LAS5 (2010) e do CALD4 (2013), assim, como a marcação tipográfica das colocações no OALD8 (2010) parecem ter ficado restritos aos dicionários de aprendizes de inglês. Na tradição hispânica, e, mais precisamente, no Señas (2001), as combinatórias em geral são esparsas e as colocações, mais esparsas ainda. Se se toma como exemplo o lexema *decisión*, que é o equivalente em espanhol a *decision*, percebe-se que a sua descrição no Señas (2001) quase não abarca padrões combinatórios.

Figura 9 – Verbetes *decisión* do Señas (2001)

<p>de·ci·sión deθisjón 1 f. Determinación ante una cosa dudosa: <i>todavía no he tomado ninguna ~ al respecto. Llegaron a la ~ de que no iban a tener hijos.</i> ⇔ <i>indecisión.</i> • decisão.</p> <p>2 Valor o firmeza para decidir; seguridad: <i>Ángel se levantó con ~ y la invitó a bailar.</i> ⇔ <i>determinación.</i> ⇔ <i>indecisión.</i> • decisão.</p>

Fonte: Señas (2001, s.v. *decisión*)

Pelos exemplos na acepção 1, infere-se as combinações “tomar [determinante] *decisión*” e “llegar a [determinante] *decisión*”. Ao consultar o DComP (2011, s.v. *decisión*), um dicionário de combinatórias do espanhol, dentre as diversas palavras que acompanham *decisión*, estão os adjetivos *grande*, *difícil*, *cruda*, *arriesgada*, *oportuna*; as locuções substantivas *poder de* e *capacidad de*; e os verbos *adoptar* e *madurar*, entre muitos outros. Há mais de cem palavras e sintagmas registrados pelo DComP (2011) que se combinam com *decisión*, o que torna a escassez de informações dessa natureza um aspecto débil do Señas (2001) – ainda mais pelo fato de que, na sua condição de dicionário de aprendizes, a obra se propõe a auxiliar na produção linguística.

Embora as informações sobre colocações enriqueçam a descrição linguística nos dicionários de aprendizes de inglês, a sua apresentação se restringe ao aspecto formal, sem tratar da dimensão semântica das combinatórias. Se os múltiplos significados de uma palavra são organizados semanticamente por meio de acepções para viabilizar a sua localização e facilitar a diferenciação entre uns e outros, as combinatórias, enquanto fenômeno que também envolve a dimensão do significado, precisariam receber uma atenção similar no tocante à sua organização no verbete. A origem dessa limitação é conceitual, e não lexicográfica. Apesar de útil para orientar a busca por colocações em um *corpus*, a abordagem de Sinclair (1991) não explora as características das colocações enquanto fenômeno léxico nem justifica por que esses

(e não outros) critérios de busca são os que resultam em colocações. A definição do autor é suficientemente genérica para abrigar uma variedade de configurações de combinações de palavras sem que se entenda com precisão o que são realmente as colocações e o lugar que ocupam no princípio idiomático. Deduz-se apenas que se trata de combinações com significado composicional – diferente das fraseologias [idioms], consideradas como portadoras de um significado não composicional.

As imprecisões da perspectiva contextualista não passaram despercebidas pelos adeptos do viés lexicológico. Sánchez Rufat (2010, p. 293), por exemplo, observa que

[...] o critério da coocorrência frequente [...] é insuficiente para estabelecer a existência de uma colocação, já que não é possível dar um conteúdo homogêneo compreensível conceitualmente em termos teóricos ao variado grupo de fenômenos que compartilham o critério com o qual o termo se identifica¹³⁹ (SÁNCHEZ RUFAT, 2010, p. 293).

Em outras palavras, a autora chama a atenção para a falta de elementos suficientes para elaborar o conceito de colocação. Sánchez Rufat (2010, p. 304) complementa sua observação argumentando que “quando se identifica a colocação pelo critério de coaparição frequente, não se está designando um conceito propriamente linguístico, ao mesmo tempo que não é um traço exclusivo dessa unidade e, por isso, não é discriminatório”¹⁴⁰. Portanto, falta definir o conceito de colocação a partir de critérios linguísticos. Em vista de críticas como essa, autores como Siepmann (2005, p. 411) preferem usar o termo *coocorrência* ao invés de *colocação*.

Como corolário das contribuições de Sinclair (1991, 2004), percebe-se que o autor promoveu um avanço no que diz respeito à constatação de que o significado de uma palavra é, muitas vezes, construído sintagmaticamente. Outro aspecto importante consiste na ênfase por uma descrição linguística e lexicográfica baseada em dados de *corpora*, que conferem prestígio e confiabilidade às pesquisas e produtos originários de dados empíricos. Porém, quando se trata das colocações, o viés semântico é deixado de lado em prol da sua identificação (semi-) automática em um *corpus*, de modo que o fenômeno é definido e analisado de maneira parcial.

No decorrer do tempo, o contextualismo inglês atribuiu especial atenção à participação de elementos gramaticais nas combinações de palavras. Conforme Legallois (2012) e Gyllstad (2014, p. 1), passou-se a usar os termos *colocação lexical* e *colocação gramatical* em referência

¹³⁹ [el criterio de la coocurrencia frecuente [...] resulta insuficiente para establecer la existencia de una colocación, ya que no es posible dar un contenido homogéneo comprensible conceptualmente en términos teóricos al variado grupo de fenómenos que comparten este criterio con el que se identifica el término]

¹⁴⁰ [cuando se identifica la colocación con el criterio de coaparición frecuente, no se está designando un concepto propiamente lingüístico al tiempo que no es un rasgo exclusivo de esta unidad, por lo que no es discriminatorio]

às combinações que envolviam somente palavras lexicais (*deficiente visual*, por exemplo) e as que envolviam ao menos uma palavra gramatical (*desistir de*, por exemplo), respectivamente. Ao se considerarem os exemplos de colocação com *decision*, reproduzidos anteriormente, percebe-se que a presença de palavras gramaticais pode ser algo comum. As combinações *the right/wrong decision* e *come to a decision*, descritas no LAS5 (2010), são exemplos disso.

Paralelamente, os contextualistas desenvolveram outro conceito de combinatória abordado por Firth (1962): as coligações. Segundo Hoey (2005, p. 43), Legallois (2012, p. 39), Gyllstad (2014, p. 1) e Lehecka (2015, p. 5-6), atualmente, uma coligação é considerada como a associação entre uma palavra (lexical ou gramatical) e uma categoria gramatical. Por categoria gramatical, os autores se referem a concepções abstratas, e não a palavras propriamente ditas, como nas colocações. Tais concepções podem ser veiculadas por palavras, como “pronomes possessivos”, ou não, como “função de sujeito”. Hoey (2005, p. 43) é um dos estudiosos que mais desenvolveu o tema, propondo três tipos de coligações:

- 1) Item léxico associado a (ou evitando) um contexto gramatical, que pode ser um item pertencente a uma determinada classe de palavras (pronomes possessivos, por exemplo), oração subordinada, voz, forma nominal, tempo, modo ou aspecto etc.
- 2) Item léxico associado a (ou evitando) uma função sintática (sujeito, objeto, complemento ou adjunto).
- 3) Item léxico associado a (ou evitando) uma posição em um sintagma, frase, texto ou (início/final de frase, anteposição ao verbo etc.).

A fim de exemplificar cada tipo de coligação, a seguir reunimos verbetes de dicionários compilados com base em *corpus* e que, cientes ou não do fenômeno, descrevem suas manifestações. Foram escolhidos exemplos em português para facilitar a compreensão.

Quadro 63 – Exemplos de coligações em dicionários

Item léxico + contexto gramatical	Item léxico + função sintática	Item léxico + posição sintática
EMBORA em·bo·ra Conj [Subordinativa. Concessiva] 1 ainda que; posto que; conquanto: <i>A platina é o metal mais precioso que existe e é praticamente indestrutível, embora seja relativamente leve</i>	ELE e·le Pron [Pessoal] [...] 3 precedido de preposição, funciona como complemento ou adjunto: <i>Ana deu a ele uma rosa vermelha. Ergueu para ele olhos curiosos; Quem saiu com ele? [...]</i> (DUPC, 2004, s.v. <i>ele</i>)	ENTÃO en·tão Adv [...] 7 iniciando frase interrogativa instiga alguém a uma resposta decidida: <i>Deu uma gargalhada e perguntou: –Então, essa história de fugir com Laura é verdade ou não? [...]</i> (DUPC, 2004, s.v. <i>então</i>)

[...] * Em 1 é usada com o verbo no Subj. (DUPC, 2004, s.v. <i>embora</i>)		
--	--	--

Fonte: DUPC (2004, s.v. *embora, ele, então*)

Os verbetes anteriores contêm as coligações: “*embora* + [verbo no subjuntivo]”, que envolve uma palavra gramatical associada a um verbo no modo subjuntivo; “[*preposição*] + *ele* + [função de complemento/adjunto]”, que pode ser considerada como uma coligação “complexa”, associada concomitantemente a um contexto gramatical e a uma função sintática; e “*então* + [posição inicial de frase interrogativa]”, que também envolve uma amálgama de tipos de coligação, isto é, entre o contexto gramatical (frase interrogativa) e a posição sintática (início).

A atenção dispensada às coligações por meio da perspectiva contextualista pode ser observada em diversos dicionários de aprendizes de inglês, em especial os de nível intermediário e avançado. Por outro lado, os dicionários de aprendizes de níveis iniciais apresentam uma descrição ainda tímida desse fenômeno, por vezes omitindo-o. Observe-se os casos a seguir:

Quadro 64 – Coligações em dicionários de aprendizes de inglês

<p>! Common mistake: already</p> <p>Warning: check your word order!</p> <p>Already usually goes directly before the main verb in a sentence:</p> <p><i>We have already enough supermarkets in this town.</i></p> <p><i>We already have enough supermarkets in this town.</i></p> <p>But if the main verb is am/is/are/was/were, already usually goes directly after it:</p> <p><i>There are already enough supermarkets here.</i></p> <p>(CALD4, 2013, s.v. <i>already</i>)</p>	<p>already /ɔ:l'redi/ adv</p> <p>1 A2 before now, or before a time in the past: <i>I've already told him.</i></p> <p>2 B1 used to say that something has happened earlier than you expected: <i>I can't believe you've already finished!</i></p> <p>(CaED2, 2011, s.v. <i>already</i>)</p>
<p>USAGE</p> <p>another</p> <p>Don't use 'another' in front of a plural noun or an uncountable noun. Don't say, for example, 'They arrange things better in another countries'. Say 'They arrange things better in other countries'. □ Other people had the same idea.</p> <p>(CCID4, 2018, s.v. <i>another</i>)</p>	<p>another /ə'nʌðə/ adjective</p> <p>1 one more person or thing of the same type: <i>We're going to have another baby.</i></p> <p>• another pronoun 'These biscuits are delicious.' – 'Would you like another?'</p> <p>2 a different person or thing: <i>I'll deal with this problem another time.</i></p> <p>• another pronoun He said one thing and did another.</p> <p>one another used for showing that each member of a group does something to or for the other members: <i>These children are learning to help one another.</i></p> <p>(CoED2, 2014, s.v. <i>another</i>)</p>

<p>enjoy 0- [T] /ɪn'dʒɔɪ/ verb 1 [T] to get pleasure from sth: ~ sth We thoroughly enjoyed our time in New York. ◊ Thanks for a great evening. I really enjoyed it. ◊ ~ doing sth I enjoy playing tennis and squash. 2 [T] ~ yourself to be happy and get pleasure (OALD8, 2010, s.v. enjoy)</p>	<p>enjoy 0- [T] /ɪn'dʒɔɪ/ verb (enjoys, enjoying, enjoyed /ɪn'dʒɔɪd/) to like something very much: I enjoy playing football. ◊ Did you enjoy your dinner? enjoy yourself to be happy; to have a good time: I really enjoyed myself at the party. (OxED2, 2012, s.v. enjoy)</p>
--	--

Fonte: CALD4 (2013, s.v. *already*), CaED2 (2011, s.v. *already*), CCID4 (2018, s.v. *another*), CoED2 (2014, s.v. *another*), OALD8 (2010, s.v. *enjoy*), OxED2 (2012, s.v. *enjoy*).

Em primeiro lugar, nos verbetes dos dicionários Cambridge, os padrões coligacionais “already + [posição sintática anterior ao verbo principal]” e “already + [posição sintática anterior ao substantivo ou locução substantiva que complementa ‘there be’]”, descritos em uma nota de uso do CALD4 (2013, s.v. *already*), não são mencionados no CaED2 (2011, s.v. *already*). Em segundo lugar, nos verbetes dos dicionários Collins, os padrões coligacionais “another + [~~substantivo plural~~]” e “another + [~~substantivo não contável~~]” são descritos na nota de uso do CCID4 (2018, s.v. *another*), mas omitidos da sua edição para iniciantes CoED2 (2014, s.v. *another*). Em terceiro lugar, os padrões coligacionais “enjoy + [algo]” e “enjoy + [verbo no gerúndio] + [algo]”, presentes no OALD8 (2010, s.v. *enjoy*, 1), tampouco são descritos no OxED2 (2012, s.v. *enjoy*). Como já se comentou anteriormente, é muito provável que as edições para iniciantes dos dicionários de aprendizes de inglês não descrevam certos padrões porque não os julguem pertinentes ou mesmo compreensíveis para o público A1-B1. As coligações são um fenômeno majoritariamente formal que, estando relacionadas à materialização do significado, são muito mais pertinentes para tarefas de produção (=materialização do significado). Portanto, a omissão dessa propriedade combinatória, seja pela razão que for, converte os dicionários A1-B1 em ferramentas pouco eficientes no que tange ao auxílio em tarefas de produção linguística.

No que diz respeito à lexicografia hispânica de aprendizes, mais precisamente ao dicionário Señas (2001), percebe-se que combinações de tipo coligacional estão pouco presentes nos verbetes. Seja porque esse dicionário não foi compilado segundo a perspectiva do contextualismo britânico, seja por outra razão, o fato é que uma propriedade combinatória recorrente como a coligação termina sendo, não raras vezes, negligenciada. Observem-se os excertos a seguir:

Quadro 65 – Coligações no Señas (2001)

o·fre·cer [ofreθér] 1 tr. [algo] Presentar y dar voluntariamente: <i>nos ofreció un café y bombones; cuando supo que venía, me ofreció su casa.</i> [...]	pa·ra [para] 1 prep. Indica finalidad o utilidad: <i>fue al cine ~ distraerse; compró tela ~ un vestido.</i> • para 2 Introduce un objeto indirecto añadiendo la idea de finalidad: <i>compraremos un juguete ~ el niño; traigo una carta ~ tu madre.</i> ⇨ a. • para [...] 7 Indica causa o razón: <i>¿ ~ qué te levantas tan temprano? – Para hacer ejercicio.</i> • para [...]
do·min·go [domín̄go] m. Séptimo y último día de la semana, dedicado generalmente al descanso: <i>el sábado me acosté muy tarde, así que el ~ me levanté casi a mediodía.</i> • domingo	

Fonte: Señas (2001, s.v. *ofrecer, domingo, para*)

Primeiramente, a primeira acepção do verbete *ofrecer* omite a coligação relativa ao beneficiário da ação verbal. O padrão “ofrecer [algo] [a alguien]” aparece 42 vezes entre os primeiros 100 alinhamentos de *ofrecer* no Corpes (2021) (de um total de 103.158). Apesar de haver um número considerável de registros dentre os alinhamentos analisados continha comente o objeto direto, não se poderia ignorar os usos com beneficiário. Trata-se de um componente que até aparece nos exemplos da acepção 1 sob a forma dos pronomes oblíquos *nos* e *me*, mas ainda falta uma indicação explícita a respeito dessa coligação.

Em segundo lugar, s.v. *domingo*, assim como nos verbetes dos demais dias da semana, aparentemente não haveria coligação. No entanto, uma busca no Corpes (2021) revelou que, dentre os 100 primeiros alinhamentos de *domingo* (de um total de 43.043), 50 correspondiam à coligação “[artigo definido] + domingo”. Portanto, essa informação deveria ter sido incluída no verbete de forma explícita, para além do exemplo que acompanha a acepção e contém o padrão em questão.

Em terceiro lugar, dentre os usos de *para* indicando finalidade, falta a indicação da coligação com verbos no modo subjuntivo, isto é, “para + que + [verbo no subjuntivo]”. Uma busca no Corpes (2021) revelou que, dos 100 primeiros alinhamentos do sintagma “para que”, 75 aparecem com o padrão em questão (de um total de 207.524). Os 25 usos restantes eram atinentes ao significado descrito na acepção 7 de *para*, que se apresenta sob o padrão “para + qué + [verbo no indicativo]” – outro padrão que, aliás, seria pertinente descrever.

Em suma, as coligações são um tipo de padrão combinatório que não deve ser ignorado pela Lexicografia. A originalidade dessa noção proposta por Firth (1962) e desenvolvida por Hoey (2005) se torna ainda mais evidente quando se nota que essa perspectiva de análise não está contemplada por outras abordagens acerca das combinatórias léxicas, como a lexicológica e a gramatical (cf. próximas seções). No que diz respeito à Lexicografia de Aprendizes, enquanto os dicionários da tradição inglesa se projetaram como protagonistas na descrição das

coligações, a tradição hispânica, mais precisamente o dicionário Señas (2001), ainda tem muito pouco a apresentar no que concerne à descrição desse fenômeno na língua espanhola.

As coligações, juntamente com as colocações lexicais e gramaticais, constituem o principal aporte conceitual da abordagem estatística (ou, sob uma perspectiva mais ampla, do contextualismo inglês) a respeito das combinatórias léxicas. Essa abordagem é um marco nos estudos das combinatórias, pois alertou de maneira contundente que a dimensão do significado precisava ser analisada sintagmaticamente nos estudos linguísticos, impulsionando as pesquisas sobre essa dimensão da linguagem. Além disso, com a possibilidade de se contar com grandes bases de dados de registros linguísticos, pertence aos contextualistas ingleses o mérito pelo incentivo no tocante ao uso de *corpora* para que se tivesse um maior empirismo na teorização sobre a língua.

No que concerne às noções combinatórias propostas no marco dessa abordagem, embora o conceito de colocação seja demasiado vago, os estudos acerca das coligações lançam luz sobre fenômenos combinatórios recorrentes em uma língua e que não necessariamente se manifestam com palavras. Para além das medições estatísticas, a atenção dispensada à dimensão gramatical, seja pelo viés das coligações, seja pelo reconhecimento de que há colocações gramaticais, é uma importante contribuição do contextualismo inglês para a descrição das palavras que se combinam, dado que, em outras abordagens, essa dimensão costuma ser negligenciada.

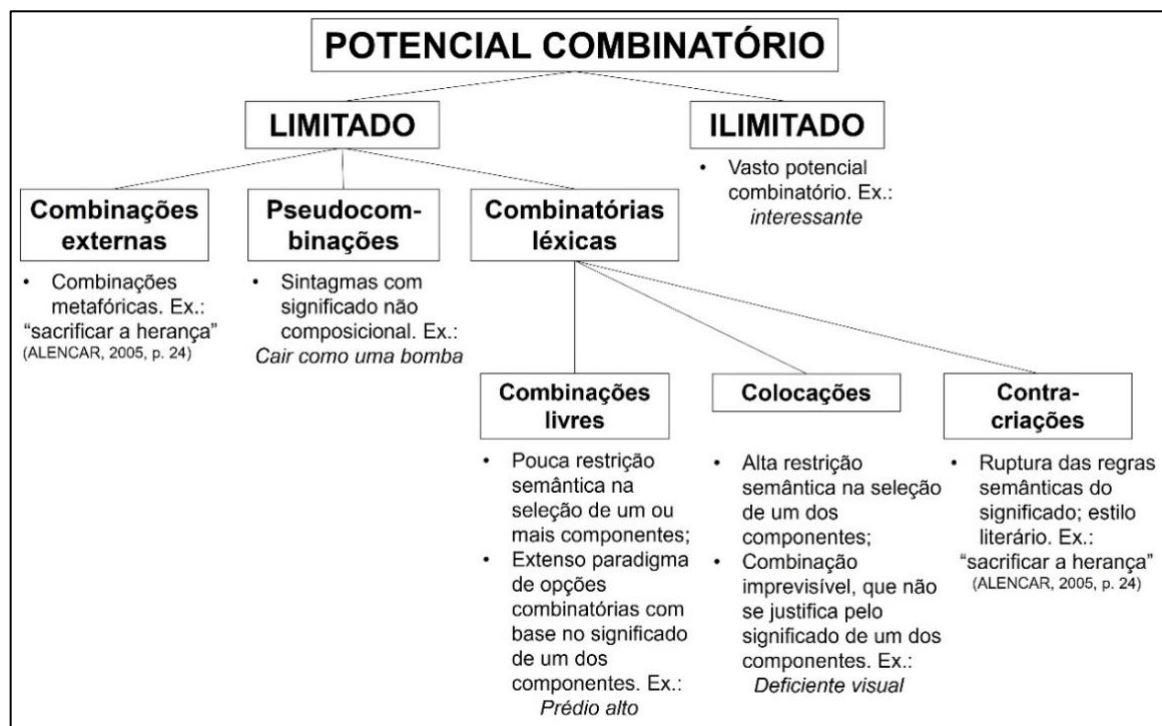
A pedra de tropeço da abordagem estatística, no entanto, é a falta de análise das relações semânticas entre os componentes de uma combinatória léxica. As críticas comentadas anteriormente sobre o particular nada mais são que reclamações quanto à superficialidade com que as combinatórias léxicas são analisadas, em especial as colocações lexicais. Motivadas por essa limitação, e para fazer frente à perspectiva ultramarina, foram propostas abordagens lexicológicas e gramaticais acerca das combinatórias léxicas. O viés lexicológico inclui a dimensão semântica em suas análises e possui como finalidade o tratamento lexicográfico desse fenômeno. Duas contribuições basilares no marco dessa abordagem são Hausmann (2007[1984]) e Mel'čuk (2012; 2015). O viés gramatical, por sua vez, ampara-se em conceitos do gerativismo para descrever as combinatórias léxicas, também com finalidade lexicográfica. Seu principal expoente é Bosque (2001). As próximas seções tratam das contribuições de cada autor e a sua relevância para a descrição de combinatórias em dicionários de aprendizes.

8.1.2 Viés lexicológico sobre as combinatórias léxicas I: Hausmann (2007 [1984])

Hausmann (2007)¹⁴¹ é considerado o precursor do viés lexicológico. Assim como Sinclair (1991), o autor trata das combinatórias léxicas com ênfase no fenômeno das colocações. Porém, segundo Tarp (2008, p. 250), Hausmann (2007) é um contestador do viés contextualista. Hausmann (2007, p. 49-51) justifica a importância das combinatórias léxicas e das colocações, em particular, a partir de alguns exemplos de tradução nos quais os padrões combinatórios de uma língua não foram respeitados. O autor também dedica boa parte de seu trabalho à argumentação sobre a relevância das combinatórias léxicas para o ensino-aprendizagem de línguas, criticando o ensino de vocabulário de viés estritamente paradigmático. Hausmann (2007, p. 61) não nega a importância do ensino de sinônimos, antônimos, hipônimos e hiperônimos, mas sim, ressalta o caráter essencial da aprendizagem do perfil sintagmático das palavras, em particular para a função de produção de língua.

O autor propõe uma tipologia de combinatórias léxicas, a qual esquematizamos e exemplificamos a seguir:

Figura 10 – Tipologia de Hausmann (2007)



Fonte: Elaboração própria do esquema e dos comentários com base em Hausmann (2007, p. 50-55). Exemplos em português incluídos por iniciativa da autora, sendo um deles de autoria de Alencar (2005, p. 24).

¹⁴¹ Publicado inicialmente em 1984. Segundo o autor, a edição de 2007 é uma tradução do original com pequenas modificações.

Hausmann (2007) distingue, em um primeiro nível, entre o potencial combinatório limitado e ilimitado das palavras. Por potencial ilimitado, o autor entende determinadas palavras, em especial alguns adjetivos, têm a capacidade de se combinar com um conjunto numeroso de elementos a ponto de ser difícil discriminá-los (HAUSMANN, 2007, p. 50-51). Como exemplo, sugerimos o adjetivo *importante*, que, segundo o CP (2019), é usado em combinação com unidades léxicas tão variadas quanto *papel* (f=6299), *aviso* (f=6073), *passo* (f=4223), *parte* (f=3110), *ponto* (f=2412), *momento* (f=1865), *fator* (f=1733), *jogo* (f=1174), entre muitas outras.

Por outro lado, Hausmann (2007) identifica casos em que o potencial combinatório é, de alguma forma, restrito. Sob essa rubrica, se encontram três distinções, sendo a primeira delas a das “combinações externas” [combinaisons externes], isto é, metáforas empregadas no estilo literário que fogem ao padrão combinatório de uma unidade léxica. Como exemplo, citamos a combinação “sacrificar a herança”, presente na obra *Senhora* de José de Alencar (2005).

A segunda categoria é a das “pseudocombinações” [pseudocombinaisons], que são sintagmas cujo significado não é composicional. Em língua portuguesa, costuma-se referir a essas unidades como *fraseologias* ou *expressões idiomáticas* (no inglês, *idioms*). O sintagma *cair como uma bomba* é um exemplo de pseudocombinação, ou sintagma não composicional, que conta com 443 registros no CP (2019) – incluídos os casos em que o verbo *cair* está flexionado.

A terceira categoria de combinações limitadas, por sua vez, se subdivide entre o que Hausmann (2007) chama de “combinação livre” [combinaison libre] ou “cocriação” [co-création], “colocação” [collocation] e “contra-criação” [contre-création]. Por combinação livre, o autor entende os casos em que as combinações reúnem elementos com pouca restrição semântica, tais como os adjetivos de cores junto a substantivos (“uma mala vermelha” [une valise rouge]) (HAUSMANN, 2007, p. 53-54). Segundo Hausmann (2007, p. 53), as combinações livres se constituem com base no que se pode inferir a partir do significado de cada elemento. No exemplo do autor, *mala*, enquanto objeto para guardar pertences, especialmente roupas e artigos pessoais, pode ser acompanhado de uma série de opções de adjetivos qualificativos. O mesmo ocorre com *vermelho/a*, que pode ser aplicado a qualquer substantivo com dimensões físicas que sejam perceptíveis aos olhos. No esquema anterior, usou-se *prédio alto* como exemplo de combinação livre, pois o adjetivo *alto* pode se combinar com uma vasta gama de opções de substantivos que contem com dimensões físicas (de estruturas a sons), assim como *prédio* pode se combinar com diversos adjetivos qualificativos.

As colocações, por outro lado, são combinações que Hausmann (2007, p. 53) define como portadoras de um alto nível de restrição semântica e que são recuperadas da memória como um todo – diferentemente das combinações livres, que não possuem essas características. O autor ainda salienta que a união entre os componentes de uma colocação não se justifica apenas pela via semântica, mas que há uma “inclinação” para que se combinem, uma classe de “afinidade” misteriosa que não pode ser explicada e que está consagrada pelo uso (HAUSMANN, 2007, p. 53). Nesse sentido, ao invés de um amplo paradigma de opções combinatórias e de uma liberdade maior de combinação, como ocorre com as combinações livres, há um paradigma restrito, às vezes com uma única opção. No esquema anterior, usou-se como exemplo a combinação *deficiente visual*, com 2307 registros no CP (2019), sendo a combinatória mais recorrente de *deficiente*. Outras opções que constituem o paradigma de combinações mais frequentes da UL em questão são *físico* (f=1916), *auditivo* (f=925) e *mental* (f=544) (CP, 2019).

Hausmann (2007, p. 55-56) apresenta uma proposta inovadora para a época (anos 1980) ao estabelecer que uma colocação se constitui de dois componentes: o primeiro, chamado *base*, seria hierarquicamente superior, semanticamente autônomo e selecionaria o outro componente, chamado *colocado*. Os termos *base*, cunhado pelo autor, e *nódulo*, utilizado no contextualismo inglês, são aparentemente similares, mas guardam uma considerável diferença conceitual entre si. Para os contextualistas, o nódulo é apenas o ponto de partida para as buscas de combinações. Por isso, qualquer palavra pode receber o *status* de nódulo. No exemplo *deficiente visual*, tanto *deficiente* como *visual* podem ser considerados indistintamente como *nódulo* ou *colocado*, a depender de qual delas parte a consulta a um *corpus*. Para Hausmann (2007), por outro lado, a base é hierarquicamente superior ao colocado. O autor também assinala que não todas as classes de palavras podem ser consideradas como base de uma colocação e que, entre as que podem desempenhar essa função, há novamente uma relação hierárquica. Segundo Hausmann (2007), o substantivo é, por excelência, a base das colocações, pois é sobre os substantivos que se diz “toda sorte de coisas” [toutes sortes de choses] (HAUSMANN, 2007, p. 55). O autor observa que a hierarquia entre base e colocado e entre as classes de palavras a ocuparem cada posição reflete a função de produção textual: “[...] quando se formula um texto, a progressão vai da base ao colocado e não o contrário. Ela vai, portanto, do substantivo ao verbo e ao adjetivo, pois são os substantivos que exprimem o que se vai dizer mediante o uso de verbos e adjetivos”¹⁴²

¹⁴² [qu’em formulant un texte, la progression va de la base au collocatif et non dans le sens inverse. Elle va donc du substantif au verbe et à l’adjectif, car ce sont les substantifs qui expriment ce sur quoi il y a des choses à dire à l’aide des verbes et des adjectifs]

(HAUSMANN, 2007, p. 56). Nesse sentido, os adjetivos, verbos e, por extensão, os advérbios, são os que costumam ocupar a posição de colocado: *deficiente visual* (base: substantivo *deficiente*; colocado: adjetivo *visual*); *tirar um cochilo* (base: substantivo *cochilo*; colocado: verbo *tirar*). Quando não há um substantivo, subentende-se que combinações como *amar incondicionalmente* e *ser tolerante* têm como base o verbo *amar* e o adjetivo *tolerante*.

A concepção de Hausmann (2007) a respeito das é um marco nos estudos combinatórios e propostas de viés lexicológico e até mesmo gramatical. A hierarquia semântica entre base e colocado tinha uma finalidade eminentemente prática, que era estabelecer uma orientação teórico-metodológica sobre onde incluir uma colocação em um verbete de dicionário. Para Hausmann (2007, p. 56), as colocações deveriam ser incluídas no verbete relativo à base.

Finalmente, a última categoria de combinatórias léxicas da tipologia de Hausmann (2007) são as contra-criações. O autor fornece poucos detalhes sobre essa categoria, de forma que é difícil compreender a sua dimensão conceitual. Aparentemente, há uma sobreposição entre as contra-criações e as combinações externas, pois o autor usa, em parte, os mesmos exemplos para ilustrar ambas as categorias.

Das distinções propostas por Hausmann (2007) em sua tipologia, as mais relevantes para o estudo do potencial combinatório das palavras com vistas ao ensino-aprendizagem e à descrição lexicográfica são as pseudocombinações (*cair como uma bomba*), as combinações livres (*prédio alto*) e as colocações (*deficiente visual*). Destas, percebe-se que o autor prioriza as colocações, dedicando a elas um espaço consideravelmente maior em sua exposição (sete páginas) se comparado com as outras duas (aproximadamente meia página para cada uma). Segundo Bartsch (2004, p. 39) a razão para isso reside no fato de que, à época da publicação da primeira versão do trabalho (anos 1980), o autor estava envolvido na compilação de um dicionário de colocações. Uma definição não tão esmerada das pseudocombinações e das combinações livres, com comentários apenas pontuais sobre ambas, prejudica a adoção da tipologia de Hausmann (2007) como referência para o estudo das propriedades de cada tipo de combinatória com vistas à aplicação lexicográfica.

Essas e outras limitações tampouco passaram despercebidas entre os adeptos do contextualismo inglês. Tarp (2008, p. 250-251), por exemplo, critica o alto grau de subjetividade na definição das combinações livres e das colocações e chega a afirmar que são impossíveis de se colocar em prática. Bartsch (2004, p. 36), por sua vez, critica o fato de que Hausmann, em seus diversos trabalhos sobre combinatórias léxicas, considera somente as “palavras de conteúdo” ou “de classe aberta” (substantivos, verbos, adjetivos e advérbios) como constituintes de colocações. Segundo a autora, Hausmann ignora a participação de constituintes

funcionais (preposições, conjunções, artigos, pronomes) nesse tipo de combinatória, como *a* e *of* em *a pack of dogs* [matilha], no exemplo da autora (BARTSCH, 2004, p. 36). No português, poder-se-ia citar a preposição *de* em *controle de qualidade*. Ainda no âmbito das combinações com palavras funcionais, outros casos, como as regências, como *candidatar-se a* (português), *postular a* (espanhol) *apply for* (inglês), também terminam não tendo lugar na tipologia de Hausmann (2007).

8.1.3 Viés gramatical sobre as combinatórias léxicas: Bosque (2001) e Sánchez Rufat (2010)

Ignacio Bosque é um gramático espanhol que esteve envolvido na compilação de duas importantes gramáticas da língua espanhola: a *Gramática Descriptiva de la Lengua Española* (GDLE, 1999) e a *Nueva Gramática de la Lengua Española* (GramLE, 2009), ambas atreladas à Real Academia Española. É em função dessa experiência de estudos e trabalho que Bosque (2001) trata das combinatórias léxicas sob uma perspectiva gramatical. A proposta do autor, aliás, serviu de base teórica para a compilação de dois dicionários de combinatórias dirigidos por ele, o *Redes* (2006) e o *DComP* (2011).

Assim como Sinclair (1991, 2004) e Hausmann (2007), Bosque (2001) também tem nas colocações o seu objeto de interesse. O diferencial do autor é o arcabouço teórico no qual se ampara para definir o fenômeno, que é a relação entre um predicado e seus argumentos proveniente do gerativismo. A concepção de predicado com a qual o autor trabalha é a de que “os predicados são categorias que designam estados, ações, propriedades ou processos nos quais intervêm um ou vários participantes. [...] Esses participantes que intervêm na noção predicativa costumam ser denominados ARGUMENTOS”¹⁴³ (MGramLE, 2010, §1.6.1c). Trata-se, portanto, de palavras que são complementadas por outras, chamadas argumentos. A classe de palavras que mais comumente é considerada como predicado são os verbos, enquanto as palavras que ocupam as funções sintáticas de sujeito e objeto, por exemplo, são seus argumentos. Em um verbo como *proponer*, em “Ferreira propôs ao CFM uma campanha que levantasse o número de médicos” (CP, 2019), os argumentos são 3: um que ocupa a função sintática de sujeito (*Ferreira*), outro que ocupa a função sintática de objeto (*uma campanha que levantasse o número de médicos*) e um terceiro que ocupa a função de beneficiário (*CFM*). Não somente os verbos podem ser predicados, mas

¹⁴³ [los predicados son categorías que designan estados, acciones, propiedades o procesos en los que intervienen uno o varios participantes. [...] estos participantes que intervienen en la noción predicativa se suelen denominar ARGUMENTOS.]

Também os substantivos, os adjetivos, os advérbios e as preposições podem ter argumentos, isto é, constituintes exigidos pelo seu significado léxico. Com raras exceções, como os auxiliares, quase todos os verbos têm estrutura argumental. Ao contrário, somente um pequeno conjunto das categorias restantes citadas exige semanticamente um ou mais argumentos que completem o seu significado¹⁴⁴ (MGramLE, 2010, §1.6.1d).

A relação entre predicados e argumentos constitui, conforme Bosque (2001, p. 2), um fenômeno gramatical de seleção léxica, também conhecida na literatura como *selectional restrictions* [restrições de seleção] (ATKINS; RUNDELL, 2008, p. 302). Sob essa perspectiva, os predicados são o pivô responsável por selecionar seus argumentos. Em termos práticos, essa concepção inverte a direção tradicional da hierarquia entre base e colocado. Se, para Hausmann (2007), exemplos como *tirar um cochilo* estão constituídos de uma base substantiva (*cochilo*) que seleciona um colocado verbal (*tirar*), para Bosque (2001), é o colocado *tirar* que seleciona a base *cochilo*, em conjunto com outras bases argumentais. A seleção léxica ocorre, portanto, a partir do colocado (predicado) em direção à base (argumento).

Outra distinção fundamental de Bosque (2001, p. 8) diz respeito ao objeto da seleção léxica dos colocados (predicados). Segundo o autor, os colocados selecionam não bases específicas, mas conjuntos de bases que compartilham um traço semântico em comum. A tais conjuntos, Bosque (2001) se refere por *classes léxicas*. A seguir, reproduzem-se exemplos de Bosque (2001) para ilustrar essa distinção. Unidades léxicas em espanhol que desempenhavam função metalinguística (marcadas em itálico) foram mantidas no original.

Não são as peças léxicas *aceptar* ou *difundir* as que aparecem selecionadas individualmente pelo advérbio *universalmente*, mas sim uma série de classes léxicas formadas pelos verbos que denotam aceitação (*aceptar, admitir, acoger...*), valoração positiva (*aplaudir, celebrar, reconocer...*) ou negativa (*condenar, detestar, repudiar...*), difusão (*publicar, trascender, extenderse...*), conhecimento (*conocer, saber...*) e alguma outra mais. De forma análoga, o advérbio *poderosamente* modifica verbos que denotam atração e influência (*llamar la atención, influir, marcar, atraer, incidir, afectar, influenciar...*), ajuda e contribuição (*contribuir, ayudar, revitalizar, favorecer, respaldar...*), crescimento, proeminência ou desenvolvimento (*destacar, sobresalir, crecer, aumentar, prosperar...*), semelhança e diferença (*contrastar, parecerse, semejar...*) e algumas outras classes léxicas [...]¹⁴⁵ (BOSQUE, 2001, p. 8. Grifos no original).

¹⁴⁴ [También los substantivos, los adjetivos, los adverbios y las preposiciones pueden tener argumentos, es decir, constituyentes exigidos por su significado léxico. Con escasas excepciones, como los auxiliares, casi todos los verbos tienen estructura argumental. En cambio, solo un pequeño conjunto de las restantes categorías citadas exige semánticamente uno o más argumentos que completen su significado]

¹⁴⁵ [No son las piezas léxicas *aceptar* o *difundir* las que aparecen seleccionadas individualmente por el adverbio *universalmente*, sino más bien una serie de clases léxicas formadas por los verbos que denotan aceptación (*aceptar, admitir, acoger...*), valoración positiva (*aplaudir, celebrar, reconocer...*) o negativa (*condenar, detestar, repudiar...*), difusión (*publicar, trascender, extenderse...*), conocimiento (*conocer, saber...*) y alguna más. Análogamente, el adverbio *poderosamente* modifica a verbos que denotan atracción e influencia (*llamar la atención, influir, marcar, atraer, incidir, afectar, influenciar...*), ayuda y contribución (*contribuir, ayudar,*

As combinações entre colocados e classes léxicas constituem uma perspectiva igualmente válida para tratar das colocações, a começar pela proposta de uma nova definição conceitual ao termo. Há um grau de subjetividade na identificação e determinação de classes léxicas, com a desvantagem de que Bosque (2001) não tem um estudo mais aprofundado sobre esse expediente nem referências ou orientações metodológicas de como empregá-lo.

A perspectiva de Bosque (2001) acerca das colocações motivou, anos mais tarde, a elaboração de uma proposta de classificação de tipos de combinatórias léxicas que tem como pilar a definição de colocação proposta pelo autor. Trata-se da tipologia de Sánchez Rufat (2010).

Figura 11 – Tipologia de Sánchez Rufat (2010)



Fonte: Elaboração própria do esquema com base em Sánchez Rufat (2010, p. 297-300). Exemplos no original, sendo um retirado do DComP (2001, s.v. *pele*)

O conceito de *colocação* proposto por Bosque (2001) está contemplado pelas categorias de *restricções sistemáticas*, que consistem na seleção de classes léxicas de argumentos a partir de fatores semânticos. Sánchez Rufat (2010, p. 298-299) aplica uma distinção já presente em Bosque (2001) entre as seleções léxicas orientadas pelo conhecimento de mundo

revitalizar, favorecer, respaldar...), crecimiento, prominencia o desarrollo (*destacar, sobresalir, crecer, aumentar, prosperar...*), semejanza y diferencia (*contrastar, parecerse, semejar...*) y algunas clases léxicas más]

(extralinguísticas) e as seleções léxicas orientadas por um princípio idiomático (intralinguísticas). As primeiras seriam universais, ou comuns a todas as línguas; já as segundas não teriam caráter universal. O emprego dessas noções é arriscado, pois nem Sánchez Rufat (2010) nem Bosque (2001) empreenderam um estudo comparativo entre línguas para comprovarem empiricamente que as restrições classificadas como extralinguísticas são realmente universais. Além disso, para fins lexicográficos, e em especial para a lexicografia de aprendizes, essa subdivisão é irrelevante.

O outro eixo sobre o qual se apoia a proposta de Sánchez Rufat (2010) é o das *restrições não sistemáticas*. A autora se refere a elas como combinatórias que não podem ser justificadas pelo viés semântico, podendo ser fruto de fatores sociais, culturais ou históricos (SÁNCHEZ RUFAT, 2010, p. 297-298). Por um lado, Sánchez Rufat (2010) reúne sob a categoria *unidades fraseológicas* uma série de combinatórias envolvendo enunciados, tanto completos como incompletos. Essa ramificação, no entanto, não recebe uma definição mais detalhada nem exemplos ilustrativos dos fenômenos que a autora nomeia como representantes de unidades fraseológicas (provérbios, citações, fórmulas rotineiras, locuções). Trata-se do ponto mais frágil da sua proposta de classificação.

Por outro lado, as restrições não sistemáticas contemplam o que Sánchez Rufat (2010) chama de preferências léxicas. Muito similares às colocações, as preferências léxicas contemplam, segundo a autora, os casos em que não todos os argumentos/bases de uma classe léxica são selecionados por um colocado, mas apenas alguns, e sem que se saiba ao certo por quê. Sánchez Rufat (2010, p. 298) exemplifica o fenômeno com os predicados *fúnebre* e *mortuorio*, que, embora sejam sinônimos e se combinem com argumentos relativos a lugar, transporte e evento social, alguns desses argumentos se combinam apenas com uma ou com outra opção: *cámara/esquela mortuoria; elogio/coche/honras fúnebre(s)* (SÁNCHEZ RUFAT, 2010, p. 298).

Por fim, assim como os autores da perspectiva fraseológica, Sánchez Rufat (2010) não inclui a possibilidade de combinações entre elementos lexicais e gramaticais em sua classificação. A autora segue a mesma linha de raciocínio de Bosque (2001, p. 18), que tampouco trata das combinatórias com palavras gramaticais.

A tipologia de Sánchez Rufat (2010) apresenta como ponto positivo o desdobramento do conceito amplo de colocação cunhado por Bosque (2001) em subcategorias específicas. Os fenômenos mais interessantes da classificação no que diz respeito à Lexicografia de aprendizes são as restrições sistemáticas e as preferências léxicas. A autora, no entanto, não desenvolve toda a sua classificação com o mesmo esmero; a pouca atenção dispensada ao que denomina

unidades fraseológicas é um exemplo disso. A distinção entre fenômenos intralinguísticos e extralinguísticos é outro ponto frágil da proposta, pois não se sabe qual é a sua utilidade prática (ao menos para fins lexicográficos). Também faz falta uma quantidade maior de exemplos para os fenômenos de restrição não sistemática. Tais problemas não permitem que a proposta possa ser aproveitada na tese.

No que diz respeito à noção de colocação de Bosque (2001) e sua relevância para a tese, é de se questionar o quão próxima essa abordagem está do contexto e das demandas de ensino aprendizagem de língua estrangeira e, por conseguinte, da Lexicografia de Aprendizes. Mais precisamente, questionamos se o que pode vir a ser objeto de dúvida ou desconhecimento dos aprendizes seriam as bases/argumentos. Na mesma esteira de Hausmann (2007), parece-nos que, para se chegar a uma colocação como *influenciar poderosamente*, por exemplo, é mais provável que se parta do verbo, ou o “objeto do dizer”, e não tanto o contrário, para buscar informações em uma obra lexicográfica. Arriscamos dizer que o mesmo vale para outros casos comentados anteriormente, como *deficiente visual* e *tirar um cochilo*. Em uma tarefa de produção, parece incomum que se tenha conhecimento dos colocados *visual* e *tirar*, sem que se saiba quais bases se combinam com ambos a fim de expressar o conteúdo semântico de *deficiente visual* e *tirar um cochilo*. Mais razoável seria uma situação em que *deficiente* e *cochilo* sejam o tema sobre o qual se quer expressar algo, mas não se saiba com que colocados tais palavras comumente se combinam.

A proposta de Bosque (2001), portanto, não demonstra ser útil para descrever colocações em um dicionário de aprendizes, posto que prevê uma lógica não usual para um aprendiz, isto é, que as colocações sejam descritas no verbete do colocado potencialmente “desconhecido”, e não da base potencialmente “conhecida”.

Contudo, a perspectiva do autor revela um fato linguístico relevante para a Lexicografia: os conjuntos de entidades, ou classes léxicas, às quais se aplica uma paráfrase explanatória em sua dimensão extensional, e que denotam um tipo de restrição de seleção semântica. O próprio autor assinala que “[...] muitas das informações que representam o contorno da definição lexicográfica constituem ‘bases’ de relações colocativas”¹⁴⁶ (BOSQUE, 2001, p. 10), e isso realmente acontece. O Señas (2001), por exemplo, no verbete *tener*, acepção 6, contém a paráfrase “experimentar o sentir”, incluindo exemplos de combinatórias envolvendo a dimensão física (*tener hambre*) e a psíquica (*tener prisa*). A seleção de argumentos/bases de *tener* manifestando parte desse conteúdo semântico está prevista pelo PCIC (2006) no apartado

¹⁴⁶ [muchas de las informaciones que representan el contorno de la definición lexicográfica constituyen ‘bases’ de relaciones colocativas]

léxico relativo ao nível A2, no qual, sob a categoria “sensaciones y percepciones físicas”, são incluídos os sintagmas “tener calor/frío/sueño/sed/hambre/dolor” (PCIC, 2006, A2, Nociones específicas §2.3). O diferencial da proposta de Bosque (2001) consiste em chamar a atenção para a dimensão extensional do significado, manifestada pelo fenômeno das seleções léxicas, de modo que se invista em formulações de paráfrases que prevejam e salientem tal dimensão.

Uma última observação acerca da abordagem de Bosque (2001) é a potencial confusão que o termo *colocação*, com o conceito proposto pelo autor, pode gerar no momento de se referir à sua perspectiva. Bosque (2001) trata as colocações como manifestação das seleções léxicas, termo que tampouco deixa claro que, mais que uma simples seleção, trata-se de uma restrição na seleção de classes léxicas. O termo *restrição*, aliás, é usado por Sánchez Rufat (2010) para descrever o eixo das colocações na sua classificação. A fim de evitar confusões designativas e precisar melhor a que tipo de fenômeno estamos nos referindo na tese, adotamos a designação *restrição de seleção* em referência ao conceito de colocações de Bosque (2001). Trata-se de uma alternativa terminológica já presente na literatura e usada por Atkins e Rundell (2008, p. 302).

8.1.4 Viés lexicológico sobre as combinatórias léxicas II: Mel’čuk (2015)

Mel’čuk (2015) reúne elementos do estruturalismo e da lógica para propor uma teoria que reserva às combinatórias léxicas um lugar de prestígio: a Teoria Sentido-Texto (TST).

Primeiramente, Mel’čuk e Milicévic (2014, p. 33-35) propõem que se adicione a sintaxe como um terceiro elemento da concepção saussureana de signo linguístico, ao lado do significante e do significado. Segundo os autores, o componente sintático (que, na verdade, é concebido por eles como um componente morfossintático) diz respeito às propriedades combinatórias dos signos e possui natureza idiossincrática. Mel’čuk (1997, p. 4-6) entende que a (morfo)sintaxe é fundamental na concepção de signo linguístico porque intermedia a correspondência entre o nível semântico (a contraparte do *sentido* da TST) e as realizações de uma língua (a contraparte do *texto* da TST). Conforme o autor, o objetivo da TST é fornecer um modelo de descrição lexicográfica exaustiva de uma UL em todos os seus níveis de representação (fonológico, morfológico, sintático e semântico).

A teoria se chama *sentido-texto* em função de que o principal interesse de Mel’čuk (2015) e Mel’čuk e Milicévic (2014, p. 78-79) recai sobre a direção dos sentidos aos textos, isto é, sobre como o componente semântico se manifesta concretamente na língua, de modo a

descrever exaustivamente as possibilidades de realização dos sentidos. Os autores defendem essa perspectiva, entre outras razões, porque o processo de síntese de sentidos (*i.e.* a função de produção) requer um conhecimento linguístico mais robusto que a análise de sentidos (*i.e.* a função de compreensão). Em outras palavras, e ainda segundo os autores, não é possível produzir algo que não se conhece, mas, em tarefas de compreensão, o contexto pode auxiliar a superar a falta de conhecimento de uma determinada unidade léxica. Mel'čuk e Milicévic (2014, p. 78-79) ainda justificam a primazia da direção sentido-texto argumentando que as colocações são um tipo de combinatória léxica que, pela sua natureza composicional, são fáceis de compreender, mas difíceis de produzir, especialmente no que diz respeito a uma língua estrangeira.

A contraparte formal da TST é o seu modelo de representação, chamado Modelo Sentido-Texto (MST). Esse modelo corresponde a princípios lógico-matemáticos utilizados para delimitar os conceitos propostos na teoria. O emprego desse expediente é, aliás, o diferencial da abordagem de Mel'čuk (1997, 2021, 2015) em comparação com as que foram comentadas anteriormente.

No que diz respeito às combinatórias léxicas, há três conceitos básicos da TST que fundamentam esse fenômeno: sintagma [*phrase*], sintagma livre [*free phrase*] e frasema [*phraseme*]. Mel'čuk (2012, p. 32) define *sintagma* (tradução nossa do original *phrase*) como “[...] uma expressão linguística formada por vários (ao menos dois) lexemas [*i.e.* unidades léxicas] sintaticamente vinculados de forma regular”¹⁴⁷. O autor identifica dois tipos de sintagmas: os sintagmas livres [*free phrases*] e os frasemas [*phrasemes*], ou sintagmas não livres [*non-free phrases*]. Os primeiros são definidos da seguinte forma:

Um sintagma é **livre** se, e somente se (=sse) cada um dos seus componentes léxicos L_i é selecionado pelo falante de **forma** linguisticamente **não restrita** – isto é, cada L_i é selecionado estritamente em função do seu significado e em conformidade com as suas propriedades linguísticas, mas independentemente da identidade léxica de outros componentes (MEL'ČUK, 2012, p. 33. Grifos no original)¹⁴⁸.

Um sintagma livre é, portanto, constituído de componentes que são selecionados livremente e que obedecem apenas a regras gramaticais de flexão, por exemplo. Para ilustrar essa categoria de sintagmas, Mel'čuk (2015, p. 300) fornece o exemplo “This dictionary has

¹⁴⁷ [a linguistic expression formed by several (at least two) lexemes syntactically linked in a regular way]

¹⁴⁸ [A phrase is **free** if and only if (=iff) each of its lexical components L_i is selected by the speaker in a linguistically **non-constrained way** – that is, each L_i is selected strictly for its meaning and in conformity with its linguistic properties but independently of the lexical identity of other components]

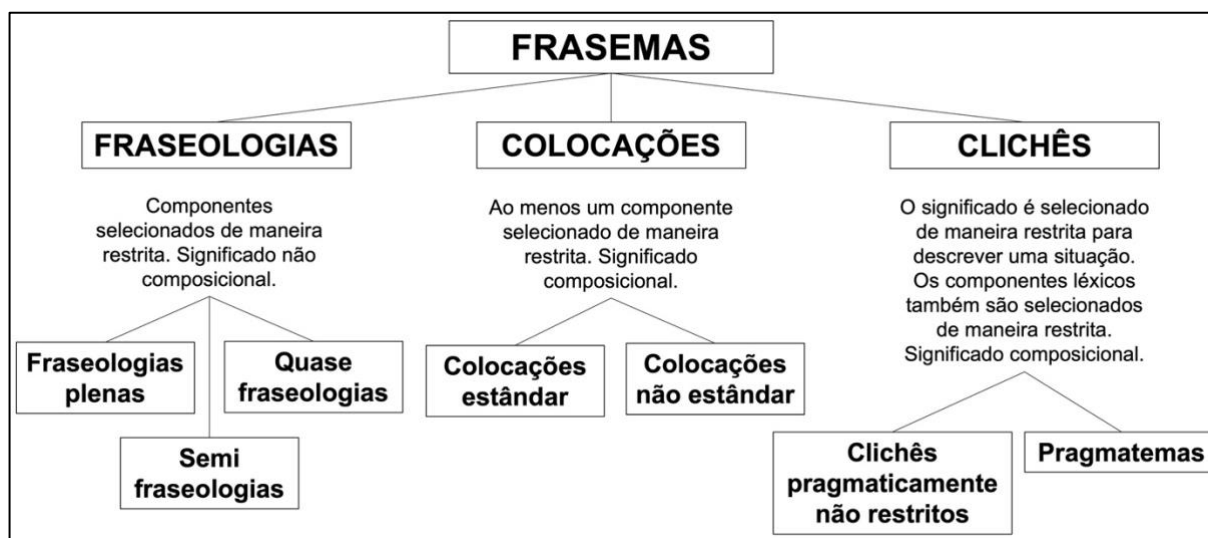
been compiled by many people” [Este dicionário foi compilado por muitas pessoas], no qual o autor argumenta que cada componente léxico fora selecionado livremente, obedecendo apenas a regras gramaticais, mas sem que um componente interfira semanticamente na seleção dos demais. Mel’čuk (2015, p. 295) argumenta que uma maneira de se aferir se um sintagma é livre é pela substituição de um dos seus componentes por um sinônimo. Segundo o autor, um sintagma livre, como o exemplo anterior, permite que seus componentes sejam substituídos (ou parafraseados) por outros sem alterar o significado do sintagma, de forma a se gerar outros sintagmas com o mesmo significado, como “The present dictionary has been developed by a large group of lexicographers” [O presente dicionário foi desenvolvido por um grande grupo e lexicógrafos] (MEL’ČUK, 2015, p. 295).

Os frasemas, por outro lado, são definidos por Mel’čuk (2012, p. 33) da seguinte forma: “Um sintagma é não livre (=fraseologizado) sse [*i.e.* se e somente se] ao menos um dos seus componentes léxicos é selecionado pelo falante de forma linguisticamente restrita – isto é, como uma função da identidade léxica de outro(s) componente(s)”¹⁴⁹ (MEL’ČUK, 2012, p. 33. Grifos no original). Em outras palavras, um frasema é um sintagma composto de unidades que são selecionadas em função de outra(s) para expressar um determinado significado. Nesse caso, segundo o autor, o significado não se mantém mediante a substituição sinonímica de um dos elementos. Mel’čuk (2012, p. 33) emprega o exemplo *kick the bucket* [bater as botas] para ilustrar a restrição presente nos frasemas, argumentando que, ao substituir o verbo *kick* [chutar] pelo sinônimo *boot* [chutar com força], o sintagma já não expressa mais o significado “morrer”.

É, portanto, sobre os frasema que recai o interesse de Mel’čuk (1997, 2012, 2015), para quem esse tipo de sintagma constitui a unidade básica de estudos da Fraseologia. O autor identifica dos tipos de restrição que podem constar nos frasemas, bem como a presença ou ausência de significado composicional. Com base nessas propriedades, propõe uma tipologia com três tipos de frasemas: fraseologias [idioms], colocações [collocations] e clichês [clichés].

¹⁴⁹ [A phrase is **non-free** (=phraseologized) iff at least one of its lexical components L_i is selected by the speaker in a **linguistically constrained way** – that is, as a function of the lexical identity of other component(s)]

Figura 12 – Tipologia de Mel’čuk (2012, 2015)



Fonte: Elaboração própria do esquema a partir das informações em Mel’čuk (2012, 2015)

Cada tipo de frasema será objeto das subseções a seguir.

8.1.4.1 Fraseologias [*idioms*]

Conforme Mel’čuk (2012, p. 37; 2015, p. 306), as fraseologias (tradução que adotamos para *idioms*) têm dois traços fundamentais e distintivos. O primeiro deles é a não composicionalidade do significado. O autor define o fenômeno da seguinte forma: “um frasema semanticamente não composicional $\mathbf{AB} = \langle 'C'; /AB/; \Sigma_{AB} \rangle$, onde $'C' \neq 'A \oplus B'$, é chamado fraseologia” (MEL’ČUK, 2015, p. 305. Grifo no original)¹⁵⁰. Nessa formulação, o autor aplica a sua concepção tríplice de signo linguístico, segundo a qual as letras “A” e “B” em negrito representam o frasema em si; a letra “C” entre aspas simples (‘C’) representa o significado do frasema; as letras “A” e “B” juntas entre barras (/AB/) representam o significante; e o símbolo Σ acompanhado das letras “A” e “B” representa o comportamento sintático dos componentes do frasema, que, no caso, segue as regras gramaticais de combinação de “A” e “B”. Nesse sentido, um frasema como “chutar o balde”, definido como “ter atitude explosiva; descontrolar-se emocionalmente” (DUPC, 2004, s.v. *chutar*) é uma fraseologia segundo a concepção de Mel’čuk (2012, 2015) porque o seu significado não corresponde à associação dos significados de *chutar* e *balde*. Ao se adotar a linguagem lógico-matemática de Mel’čuk (2015), o frasema

¹⁵⁰ [A semantically non-compositional phraseme $\mathbf{AB} = \langle 'C'; /AB/; \Sigma_{AB} \rangle$, where $'C' \neq 'A + B'$, is called an idiom]

em questão é descrito da seguinte forma: **chutar o balde** = ⟨‘atitude explosiva; descontrole emocional’; /chutar o balde/; $\Sigma_{\text{CHUTAR,BALDE}}$ ⟩, segundo o qual se indica que o significado do fraseo é alheio aos significados de seus componentes, mas o significante e o padrão sintático se mantêm.

O segundo traço distintivo de uma fraseologia é a seleção livre do significado. Segundo Mel’čuk (2015, p. 306), “o significado ‘C’ de uma fraseologia **AB** é selecionado livremente pelo Falante [...], mas sua expressão – o significante da fraseologia – não é selecionado livremente: sua seleção é completamente limitada pelo significado”¹⁵¹ (grifo no original). Essa característica auxilia na diferenciação das fraseologias em relação aos clichês, que são fraseos cujo significado é selecionado de maneira restrita por uma determinada situação comunicativa, como se comentará mais adiante.

Além das propriedades fundamentais das fraseologias, Mel’čuk (2012, p. 37) salienta que há um traço não distintivo presente em algumas delas: a variação no grau de transparência e opacidade desse tipo de fraseo. Essa constatação motivou o autor a criar uma subclassificação de fraseologias: fraseologias plenas [full idioms], semifraseologias [semi-idioms] e quase fraseologias [quasi-idioms] (MEL’ČUK, 2012, p. 37; 2015, p. 306-309).

As fraseologias plenas [full idioms] se caracterizam por apresentarem um significado que não inclui o significado de nenhum de seus componentes (MEL’ČUK, 2012, p. 37). No formalismo lógico-matemático do autor, esse tipo de fraseologia é definido da seguinte maneira: “Uma fraseologia $AB = \langle ‘C’; /AB/; \Sigma_{AB} \rangle$ é chamada forte ou plena sse ‘C’ $\not\supseteq$ ‘A’ and ‘C’ $\not\supseteq$ ‘B’” (MEL’ČUK, 2015, p. 306)¹⁵². O fraseo *chutar o balde* é um exemplo de fraseologia plena, pois o seu significado não contém ($\not\supseteq$) o significado de *chutar* nem de *balde*. Outros exemplos de fraseologias plenas são *olho vivo*, sob o significado de “atenção redobrada” (DUPC, 2004, s.v. *olho vivo*); *ir para o olho da rua*, sob o significado de “ser expulso, despedido ou demitido” (Sac, 2010, s.v. *olho*); e *prata da casa*, sob o significado de “recursos ou indivíduos próprios, do próprio grupo ou comunidade, e não alheios” (Sac, 2010, s.v. *prata*).

As semifraseologias e as quase fraseologias, por outro lado, apresentam o significado de um dos seus componentes (semifraseologias) ou de todos (quase fraseologias). Os fraseos *fermento biológico* (“fungo usado para fermentar massas” (definição *ad hoc*)) e *prato feito* (“refeição barata, [sic] cujos componentes são escolhidos pelo dono do estabelecimento, e não pelo cliente, servida já pronta no prato; almoço comercial” (Sac, 2010, s.v. *prato*)) são exemplos

¹⁵¹ [The signified of an idiom **AB** ‘C’ is selected freely by the Speaker [...] but its expression – the idiom’s signifier – is not selected freely: its selection is completely bound by the signified]

¹⁵² [An idiom $AB = \langle ‘C’; /AB/; \Sigma_{AB} \rangle$ is called strong, or full, iff ‘C’ $\not\supseteq$ ‘A’ and ‘C’ $\not\supseteq$ ‘B’]

de semifraseologia e quase fraseologia, respectivamente. Em *fermento biológico*, o significado de *fermento* está contido no frasema (“fermentar”), mas não é o elemento principal. Já em *prato feito*, os significados de *prato* e de *feito* estão contidos no frasema, mas tampouco são o elemento semântico principal.

Um conceito fundamental para compreender a semifraseologias e as quase fraseologias é o de *pivô semântico*. Mel’čuk (2012) define o conceito a partir da noção de argumentos e predicados: “Tome-se o significado ‘ σ ’ que se divide em duas partes, ‘ σ_1 ’ e ‘ σ_2 ’ (‘ σ ’ = ‘ σ_1 ’ \oplus ‘ σ_2 ’). A parte ‘ σ_1 ’ do significado é chamada de **pivô semântico** de ‘ σ ’ sse a outra parte ‘ σ_2 ’ é um predicado do qual ‘ σ_1 ’ é o argumento: ‘ σ ’ = ‘ σ_2 ’(‘ σ_1 ’)” (MEL’ČUK, 2012, p. 36. Grifo no original)¹⁵³. O autor exemplifica o conceito de argumento com o frasema *take a shower* [tomar banho], que significa “wash oneself under a shower” [lavar-se sob um chuveiro] (MEL’ČUK, 2012, p. 36). No exemplo, segundo Mel’čuk (2012, p. 36), *shower* é o pivô semântico, isto é, o argumento que carrega a parcela mais importante do significado do frasema *take a shower*, enquanto o significado *wash* é, ainda segundo o autor, o predicado. Mel’čuk (2015, p. 305-306) comenta que, no caso das fraseologias, sejam elas de tipo *pleno*, *semi-* ou *quase*, o pivô semântico não coincide com o significado dos seus constituintes. No caso da fraseologia plena *chutar o balde*, o pivô semântico é “atitude explosiva”, isto é, um elemento alheio ao significado dos seus constituintes.

A partir da noção de pivô semântico, Mel’čuk (2015) define as semifraseologias da seguinte forma: “uma fraseologia **AB** = $\langle \text{‘C’}; /AB/; \Sigma_{AB} \rangle$ de L é chamada de semifraseologia sse as Condições 1 e 2 são simultaneamente satisfeitas: 1. ‘C’ \supset ‘A’, e ‘C’ $\not\supset$ ‘B’. 2. ‘A’ não é o pivô semântico de ‘C’” (MEL’ČUK, 2015, p. 307)¹⁵⁴. Em outras palavras, para que um frasema seja uma semifraseologia, o significado ‘C’ deve conter (\supset) o significado ‘A’, mas não o de ‘B’ ($\not\supset$); e o significado ‘A’ não deve ser o pivô semântico do significado ‘C’. Em *fermento biológico*, cujo significado é “fungo usado para fermentar massas” o significado do componente *fermento* está presente no significado do frasema (“fermentar”), enquanto o de *biológico*, não; o pivô semântico, por sua vez, é o argumento *fungo*, que não está presente no significado nem de *fermento* nem de *biológico*. Outro exemplo de semifraseologia é *fonte de alimentação*, definida como “Dispositivo que fornece corrente elétrica a um circuito” (Sac, 2010, s.v. *alimentação*). O significado do componente *fonte* está contido nessa definição por meio do

¹⁵³ [Let there be meaning ‘ σ ’ that is divided into two parts, ‘ σ_1 ’ and ‘ σ_2 ’ (‘ σ ’ = ‘ σ_1 ’ + ‘ σ_2 ’). The part ‘ σ_1 ’ of meaning ‘ σ ’ is called the **semantic pivot** of ‘ σ ’ iff the other part ‘ σ_2 ’ is a predicate of which ‘ σ_1 ’ is the argument: ‘ σ ’ = ‘ σ_2 ’(‘ σ_1 ’)]

¹⁵⁴ [An idiom **AB** = $\langle \text{‘C’}; /AB/; \Sigma_{AB} \rangle$ of L is called a semi-idiom iff Conditions 1 and 2 are simultaneously satisfied: 1. ‘C’ \supset ‘A’, and ‘C’ $\not\supset$ ‘B’. 2. ‘A’ is not the semantic pivot of ‘C’.]

verbo *fornecer*, enquanto o que *alimentação*, não. O pivô semântico, por sua vez, é o argumento *dispositivo*.

No que concerne às quase fraseologias, Mel'čuk (2015) define esse frasema da seguinte forma: “Uma fraseologia $AB = \langle 'C'; /AB/; \Sigma_{AB} \rangle$ é chamada de fraseologia fraca, ou quase fraseologia, sse as Condições 1 e 2 são simultaneamente satisfeitas: 1. ‘C’ = ‘A’ \oplus ‘B’ \oplus ‘D’ | ‘D’ $\not\subset$ ‘A’ \oplus ‘B’. 2. O pivô semântico de ‘C’ é um elemento de ‘D’: ‘d’ \in ‘D’” (MEL'ČUK, 2015, p. 308. Grifo no original)¹⁵⁵. Nessa notação, ‘D’ é um significado externo ao dos componentes A e B do frasema, e ‘d’ é o pivô semântico de AB. Em outras palavras, o significado dos quase frasemas contém os significados dos seus constituintes A e B e um significado extra ‘D’, que constitui o pivô semântico do frasema AB. O exemplo *prato feito*, usado anteriormente, ilustra essa questão. O significado desse frasema é (“refeição barata, [*sic*] cujos componentes são escolhidos pelo dono do estabelecimento, e não pelo cliente, servida já pronta no prato; almoço comercial” (Sac, 2010, s.v. *prato*)). Os significados de *prato* e *feito* estão contidos na definição (“cujos componentes são escolhidos pelo dono do estabelecimento, e não pelo cliente, servida já pronta no prato”), mas o seu pivô semântico é o elemento externo “refeição barata”. Outro exemplo de quase fraseologia é *chefe de cozinha*, definido como “profissional de hotel ou de restaurante responsável pela confecção dos alimentos destinados ao consumo dos hóspedes ou fregueses” (Sac, 2010, s.v. *chefe*). Nesse frasema, os significados de chefe e de cozinha estão presentes na definição por meio de “responsável pela confecção dos alimentos destinados ao consumo dos hóspedes ou fregueses”, mas o pivô semântico é “profissional de hotel ou de restaurante”.

A subclassificação das fraseologias proposta por Mel'čuk (2012, 2015) estabelece um panorama da diversidade de manifestações desse fenômeno, o que é útil no momento de se avaliar manifestações tão variadas. Embora essa subdivisão não caiba na descrição linguística em um dicionário de aprendizes, o conhecimento de cada subtipo é válido no momento de se julgar o que é fraseologia e o que não é, com vistas à aplicação de um tratamento lexicográfico adequado às características desse tipo de frasema. Há quatro propriedades que resumem o conceito de fraseologia e que são pertinentes para a Lexicologia e a Metalexigrafia: 1) a seleção de um significado de maneira não restrita a um tipo de situação comunicativa; 2) a seleção restrita de componentes para representar esse significado em um frasema; 3) a presença de um significado não composicional com variação no grau de transparência e opacidade; e 4)

¹⁵⁵ [An idiom $AB = \langle 'C'; /AB/; \Sigma_{AB} \rangle$ is called a weak idiom, or quasi-idiom, iff Conditions 1 and 2 are simultaneously satisfied: 1. ‘C’ = ‘A’ + ‘B’ + ‘D’ | ‘D’ $\not\subset$ ‘A’ + ‘B’. 2. The semantic pivot of ‘C’ is an element of ‘D’: ‘d’ \in ‘D’]

a presença de um pivô semântico (argumento) externo ao significado dos componentes do frasema e que constitui o principal elemento semântico da fraseologia. A partir dessas propriedades, é possível, em primeiro lugar, identificar mais claramente que frasemas são fraseologias com base em conceitos bem definidos. Em segundo lugar, e no que diz respeito ao conceito de pivô semântico, esse expediente permite averiguar que tipos de dificuldades as fraseologias oferecem ao estudante de língua estrangeira. Em se tratando de um elemento semântico hierarquicamente superior e ausente do significado dos componentes do frasema, a conclusão a que se chega é que o acesso às fraseologias é um problema tanto de compreensão como de produção linguística. Em terceiro lugar, e em decorrência do anterior, as fraseologias requerem um tratamento lexicográfico específico que tenha em vista não somente o auxílio na compreensão do seu significado – pois isso já ocorre em maior ou menor medida nos dicionários –, mas que também considere o acesso do consulente a esse expediente linguístico para desempenhar tarefas de produção. Em outras palavras, faz-se necessário descrever as fraseologias de uma forma tal que elas não sejam somente compreendidas, mas também produzidas.

8.1.4.2 Colocações

Na TST, as colocações são frasemas de dois elementos que possuem duas propriedades fundamentais: significado composicional e seleção restrita de um de seus componentes (MEL'ČUK, 2012, p. 38-39). Mel'čuk (2015, p. 320) define esse tipo de frasema da seguinte forma:

Um frasema composicional $\mathbf{AB} = \langle 'A \oplus B'; /A \oplus B/ \Sigma_{A \oplus B} \rangle$ é chamado de colocação da UL [unidade léxica] \mathbf{A} sse satisfaz simultaneamente as condições 1-3: 1. A UL \mathbf{A} é selecionada pelo Falante pelo seu significado 'A' de maneira não restrita – i.e., independentemente de \mathbf{B} . 2. A UL \mathbf{B} não é selecionada de maneira não restrita: é selecionada pelo significado 'B', mas como uma função de \mathbf{A} . 3. O significado de \mathbf{AB} inclui o significado de \mathbf{A} como seu pivô semântico: ' \mathbf{AB} ' = ' $A \oplus B$ ' (MEL'ČUK, 2015, p. 320. Grifos no original)¹⁵⁶.

A natureza composicional do significado das colocações é expressa na fórmula acima entre aspas simples no componente do significado do signo linguístico (' $A \oplus B$ '). A primeira

¹⁵⁶ [A compositional phraseme $\mathbf{AB} = \langle 'A \oplus B'; /A \oplus B/ ; \Sigma_{A \oplus B} \rangle$ is called a collocation of the LU \mathbf{A} iff it satisfies simultaneously Conditions 1–3: 1. The LU \mathbf{A} is selected by the Speaker for its signified 'A' non-restrictedly—i.e., independently of \mathbf{B} . 2. The LU \mathbf{B} is not selected non-restrictedly: it is selected for the signified 'B', but as a function of \mathbf{A} . 3. \mathbf{AB} 's signified includes the signified of \mathbf{A} as its semantic pivot: ' \mathbf{AB} ' = ' $A \oplus B$ '.]

e a segunda condições dizem respeito ao grau de restrição na seleção dos seus componentes, também chamado de semirrestrição por Mel’čuk (2015, p. 320). Segundo o autor, o componente A, que corresponde à base da colocação, é selecionado de maneira livre, independentemente da sua contraparte B e a serviço somente do desejo do falante de expressar o significado ‘A’. O componente B, por outro lado, que corresponde ao colocado, é selecionado em função de A, isto é, a serviço da UL A para que se possa expressar um significado que A não é capaz de manifestar sozinha. Finalmente, Mel’čuk (2015, p. 320) salienta que, diferente das fraseologias, o pivô semântico das colocações corresponde ao significado de um de seus componentes – a base, ou A. Um exemplo de colocação é o frasema *intoxicação alimentar* (“intoxicação por alimentos”), que é composicional e no qual *intoxicação* é a base e também o pivô semântico; *alimentar*, por sua vez, é o colocado, que se manifesta sob o significado de “que ocorre por meio de alimento” (DUPC, 2004, s.v. *alimentar*, 10).

De acordo com Mel’čuk (2015, p. 323), “as colocações constituem a maior parte dos frasemas em uma língua e são um desafio para qualquer teoria fraseológica”¹⁵⁷. É em virtude dessa constatação que o autor desenvolveu um modelo teórico-metodológico para descrever colocações: as funções léxicas de tipo sintagmático. Uma função léxica sintagmática é um elemento semântico que, quando associado a uma UL A, produz um sintagma composto por A e B (AB). O autor emprega a notação das funções matemáticas $f(x) = y$ para representar as funções léxicas. Por essa lógica, a colocação AB (*intoxicação_A alimentar_B*, por exemplo) é produto da aplicação de uma função léxica f à base A (*intoxicação*), representada como $f(A)$. Essa operação resulta em um frasema AB que está composto da base e de um segundo elemento léxico, o colocado B (*alimentar*). Nesse sistema de relações, o colocado está vinculado à função léxica f , pois sua presença é “requisitada” junto a A em virtude de f – e não por liberdade expressiva do falante, o que lhe confere o caráter restrito. Uma colocação como *intoxicação alimentar* é representada no Modelo Sentido-Texto da seguinte maneira: $f(\textit{alimentar}) = \textit{intoxicação}$.

A originalidade da proposta de Mel’čuk (2015) a respeito das colocações está na investigação e sistematização de tipos de funções léxicas. Conforme ele mesmo afirma, as “FLs [funções léxicas] são projetadas para representar, de maneira formal e sistemática, relações léxicas convencionalizadas” (MEL’ČUK, 2015, p. 166). O objetivo do autor ao propor um modelo de funções é viabilizar a descrição dos padrões sintagmáticos (colocações) e paradigmáticos (derivação e outros fenômenos) em qualquer língua (MEL’ČUK, 2015, p. 262).

¹⁵⁷ [Collocations constitute the majority of phrasemes in a language and represent a challenge for any theory of phraseology]

Tendo em vista a revisão bibliográfica feita neste capítulo com as principais referências nos estudos sobre combinações de palavras, o MST é uma das propostas teóricas mais consistentes e desenvolvidas a respeito das combinatórias léxicas em geral e das colocações em particular. Em termos práticos, o modelo fornece uma base sólida que facilita uma descrição lexicográfica de colocações que seja orientada semanticamente. Frente a essas vantagens, o MST será tomado como referência para parte das nossas propostas a respeito da descrição de combinatórias léxicas, das quais trataremos mais adiante neste capítulo.

Mel'čuk (2015) apresenta um conjunto de mais de trinta funções léxicas (FLs) sintagmáticas, as quais denomina *estândar* e serviram de critério para a ramificação das colocações estândar em sua tipologia de frasemas (cf. Figura 12). Conforme o autor, as FLs estândar se aplicam a um amplo conjunto de bases (não há uma especificação de quantas), e, em virtude disso, manifestam-se por meio de distintos colocados (MEL'ČUK, 2015, p. 173). Outra característica das FLs estândar é a sua manifestação em diferentes línguas (MEL'ČUK, 2015, p. 174; 262-263).

As FLs não estândar, ao contrário, têm caráter idiomático e se aplicam a poucas ULs, manifestando-se em um número pequeno de colocados (MEL'ČUK, 2015, p. 176; 179). Dado o seu caráter relativo, as FLs não estândar são um recurso menos desenvolvido que as anteriores, de modo que a sua identificação e inventariação é incerta.

Para fins desta tese, convém tratar das FLs consagradas na literatura da TST, isto é, as FLs estândar, pois elas servirão de base para as propostas de descrição que serão feitas mais adiante. Com base nas suas características semânticas, organizamos grande parte das mais de trinta FLs em cinco grupos: 1. FLs qualificativas; 2. FLs de localização; 3. FLs de verbos suporte, copulativos e de realização; 4. FLs de processo; e 5. FLs variadas. Cinco FLs estândar (*Pred*, *Liqu*, *Perm*, *Excess* e *Sympt*) não serão tratadas nesta oportunidade pela sua complexidade.

8.1.4.2.1 FLs sintagmáticas qualificativas

O primeiro grupo de FLs está constituído majoritariamente de adjetivos e uma pequena parcela de advérbios, ambos aplicados a substantivos. As FLs qualificativas são: *Epit*, *Magn*, *Plus*, *Minus*, *Ver*, *Bon* e *Pos*. Neste e nos demais quadros ilustrativos das FLs, usa-se como referência a notação “ $f(A) = B$ ”, na qual “ f ” é a função, “ A ” é a palavra base e “ B ” é o colocado. O exemplo *intoxicação alimentar*, citado anteriormente para exemplificar a notação matemática das FLs, é descrito como $Epit^{restr}(intoxicação) = \sim alimentar$. Por fim, ao indicar cada FL no

quadro, usou-se a mesma fonte (Courier New) que Mel'čuk (2015) usa neste e em outros trabalhos ao citar as FLs.

Quadro 66 – FLs qualificativas

Função léxica	Valor semântico	Manifesta-se por (B)	Aplica-se a (A)	Exemplo
Epit ^{restr}	especificação	adjetivo	substantivo altamente polissêmico	Epit ^{restr} (corrente) = ~ elétrica Epit ^{restr} (conta) = ~ bancária
Epit ^{qual}	reforço; redundância	adjetivo	substantivo	Epit ^{qual} (tempestade) = forte ~ Epit ^{qual} (mês) = ~ do ano
Magn	intensificação	adjetivo, advérbio	UL passível de gradação	Magn(partida) = ~ acirrada Magn(preço) = ~ exorbitante Magn(chuva) = ~ torrencial
AntiMagn	atenuação	adjetivo, advérbio	UL passível de gradação	AntiMagn(partida) = ~ amistosa AntiMagn(custar) = ~ uma pechincha
Plus	comparação de alta gradação	adjetivo	UL passível de gradação	Plus(pressão) = ~ alta Plus(autoestima) = alta ~; ~ elevada
Minus	comparação de baixa gradação	adjetivo	UL passível de gradação	Minus(pressão) = ~ baixa Minus(autoestima) = baixa ~
Ver	condição (culturalmente) esperada	adjetivo; advérbio	UL de qualquer classe gramatical	Ver(argumento) = ~ convincente Ver(família) = ~ unida
AntiVer	violação de condição (culturalmente) esperada	adjetivo; advérbio	UL de qualquer classe gramatical	AntiVer(argumento) = ~ fraco AntiVer(faca) = ~ cega
Bon	juízo de valor positivo	adjetivo	substantivo	Bon(sorriso) = ~ sincero Bon(reputação) = boa ~
AntiBon	juízo de valor negativo	adjetivo	substantivo	AntiBon(sorriso) = ~ amarelo, ~ falso AntiBon(reputação) = má ~
Pos	avaliação positiva	adjetivo	substantivos que estejam sujeitos a avaliação	Pos(parecer) = ~ favorável Pos(crítica) = ~ construtiva
AntiPos	avaliação negativa	adjetivo	substantivos que estejam sujeitos a avaliação	AntiPos(parecer) = ~ desfavorável AntiPos(crítica) = ~ negativa

Fonte: Elaboração própria a partir das informações em Mel'čuk (2015, p. 206-213)

Em geral, percebe-se que o valor semântico das FLs qualificativas nada mais é do que a descrição de propriedades conhecidas de adjetivos e advérbios. Mel'čuk (2015, p. 211) salienta que as FLs podem se combinar umas com as outras, formando o que ele chama de FLs complexas. Segundo o autor, no caso das FLs que classificamos como qualificativas, a combinação mais recorrente é com a FL *Anti*, que é paradigmática e denota negação.

Mel'čuk (2015, p. 212) reconhece ainda que “as FLs Magn, Ver, Bon e Pos são semanticamente próximas umas das outras, e com algumas palavras-chave [*i.e.* bases] podem ser indistinguíveis”¹⁵⁸. A subjetividade termina sendo um traço intrínseco às propostas de descrição linguística de caráter semântico, o que, no entanto, não diminui a sua utilidade nem inviabiliza a descrição de colocações.

8.1.4.2.2 FLs sintagmáticas de localização

As FLs que classificamos como sendo de localização dizem respeito ao espaço físico, temporal e, subjetivamente, à parte mais importante ou central de algo. Trata-se de alguns dos poucos casos de FLs que envolvem preposições no MST.

Quadro 67 – FLs de localização

Função léxica	Valor semântico	Manifesta-se por (B)	Aplica-se a (A)	Exemplo
LOC _{in}	localização física	preposição	substantivo	LOC _{in} (aeroporto) = no ~ LOC _{in} (airport) = at the ~, into the ~
LOC _{ad}	movimento para uma direção	preposição	substantivo	LOC _{ad} (aeroporto) = ao ~, para o ~, no ~ LOC _{ad} (airport) = to the ~, towards the ~, into the ~
LOC _{ab}	movimento a partir de um local	preposição	substantivo	LOC _{ab} (aeroporto) = do ~ LOC _{ab} (airport) = from the ~, off the ~
LOC ^{temp}	localização temporal	preposição	substantivo	LOC ^{temp} (manhã) = de ~, pela ~ LOC ^{temp} (mañana) = por la ~, a/en la ~ del [día de la semana]
Centr	localização central ou parte mais importante	substantivo	substantivo concreto ou abstrato	Centr(floresta) = o coração da ~ Centr(carreira) = o auge da ~ Centr(onda) = crista da ~

Fonte: Elaboração própria a partir das informações em Mel'čuk (2015, p. 205-206; 213-214)

¹⁵⁸ [The LFs Magn, Ver, Bon and Pos are semantically close to each other, and with some keywords they can be indistinguishable]

As funções de tipo *Loc* se expressam no português e no espanhol por meio de poucas preposições. Por isso, parecem ser pouco relevantes na descrição dessa propriedade em ambas as línguas. Por outro lado, há línguas, como o inglês, em que a função de localização se manifesta por meio de um conjunto de preposições que possuem regras de uso mais específicas, como ocorre com *in*, *at* e *on*, e que são objeto recorrente de dúvidas da parte dos aprendizes de inglês (WALTER; WODFORD, 2015, p. 84-88). Nesses casos, a FL *Loc* é bastante relevante para a descrição lexicográfica.

No que concerne às combinações de funções, Mel'čuk (2015, p. 206) observa que a FL *Centr* geralmente se combina com uma das variações de *Loc*.

8.1.4.2.3 FLs sintagmáticas de verbos suporte, copulativos e de realização

Ao introduzir as principais FLs verbais, Mel'čuk (2015) distingue entre FLs de verbos suporte e copulativos (*Copul*, *Oper*, *Func* e *Labor*), que podem ser semanticamente vazios, e FLs de verbos de realização (*Real*, *Fact* e *Labreal*), que mantêm seus traços semânticos. Nem sempre é possível distinguir um caso do outro; há verbos classificados pelo autor como suporte que podem ser facilmente confundidos com verbos de realização por manterem suas características. Alguns exemplos que ilustram os limites difusos entre FLs de suporte e de realização são “*Oper*₁(*support*_(N)) = *give, lend, offer* [~ to N_Y]” e “*Oper*₂(*support*_(N)) = *receive* [~ from N_X]” (MEL'ČUK, 2015, p. 220), cujos colocados são classificados por Mel'čuk (2015) como verbos suporte; e “*Real*₁(*duty*) = *discharge, do, fulfill* [ART ~]”, cujo colocado é classificado como verbos de realização.

Quadro 68 – FLs de verbos suporte, copulativos e de realização

Função léxica	Valor semântico	Manifesta-se por (B)	Aplica-se a (A)	Exemplo
Copul	estado	verbo de ligação	adjetivo; substantivo (geralmente abstrato)	Copul(<i>exemplo</i>) = <i>ser um ~, servir de ~</i> Copul(<i>gerente</i>) = <i>ser um ~, trabalhar como ~, tornar-se ~, atuar como ~</i>
Oper	suporte	verbo suporte	substantivo [objeto de B]	Oper(<i>controle</i>) = <i>ter ~ [de/sobre], estar sob ~</i> Oper(<i>suporte</i>) = <i>dar ~ [para alguém], receber ~ [de alguém]</i>
Real	realização	verbo	substantivo [objeto de B]	Real(<i>demanda</i>) = <i>suprir a/uma ~, atender ~</i>

				Real(<i>contrato</i>) = <i>preencher o/um ~, assinar o/um ~, firmar o/um ~, estabelecer um ~, cumprir o/um ~, negociar o/um ~</i>
AntiReal	realização	verbo	substantivo [objeto de B]	AntiReal(<i>contrato</i>) = <i>descumprir o/um ~</i>
Func	suporte	verbo suporte	substantivo [sujeito de B]	Func(<i>chuva</i>) = <i>~ cair</i> Func(<i>boato</i>) = <i>~ correr, circular</i>
Fact	realização	verbo	substantivo [sujeito de B]	Fact(<i>sonho</i>) = <i>~ virar realidade</i> Fact(<i>promessa</i>) = <i>~ cumprir-se</i>
Labor	suporte	verbo suporte	substantivo [adjunto adv. ou objeto indireto de B]	Labor(<i>interrogatório</i>) = <i>convocar [alguém] para um ~</i> Labor(<i>herança</i>) = <i>receber [algo] de ~; deixar [algo] de ~</i>
Labreal	realização	verbo	substantivo [adjunto adv. ou objeto indireto de B]	Labreal(<i>varal</i>) = <i>pendurar [algo] no ~</i> Labreal(<i>asfalto</i>) = <i>cobrir [algo] com ~</i>
Prepar	preparação de A via B para ação C	verbo	substantivo	Prepar(<i>carro</i>) = <i>abastecer o ~</i> Prepar(<i>programa</i>) = <i>instalar o/um ~, carregar o/um ~</i>
Obstr	dificuldade de funcionamento ou realização	verbo	substantivo	Obstr(<i>economia</i>) = <i>~ estagnar</i>
Son	som (característico) de A	verbo	substantivo	Son(<i>cachorro</i>) = <i>~ latir, ~ uivar, ~ rosnar</i>

Fonte: Elaboração própria a partir das informações em Mel'čuk (2015, p. 215-223; 227-231)

A principal distinção entre os pares de FLs análogas Oper e Real, Func e Fact e Labor e Labreal é a função sintática da base (A). Em Oper/Real, A é objeto de B; em Func/Fact, A é sujeito de B; e em Labor/Labreal, A pode ser ou adjunto adverbial de B ou o seu objeto indireto. Na proposta original de Mel'čuk (2015), as FLs de verbos suporte e de verbos de realização são acompanhadas por índices que fazem referência aos actantes envolvidos (Oper₁, Real₂, etc.). Trata-se de um nível de especificidade na descrição linguística que não será tratado nesta oportunidade porque requer um nível de aprofundamento na descrição linguística que foge aos interesses da tese.

Copul, por sua vez, é uma FL que serve especialmente para distinguir os casos em que a cópula ocorre por meio de verbos não classificados tradicionalmente como copulativos/de ligação, tais como *atuar e trabalhar como* em “atuar/trabalhar como gerente”. Essa FL também auxilia a distinguir quais verbos copulativos costumam acompanhar uma UL.

Além das FLs de verbos suporte, copulativos e de realização, acrescentamos *Prepar*, *Obstr* e *Son*, que, a nosso ver, têm valor semântico próximo ao dos verbos de realização.

8.1.4.2.4 FLs sintagmáticas de processo

O quarto grupo de FLs sintagmáticas corresponde àquelas que denotam uma etapa em um processo. As FLs inicialmente classificadas por Mel'čuk (2015) como sendo de processo são *Incep*, *Cont* e *Fin*; acrescentamos *Prox*, *Stop* e *Degrad* a este grupo por semelhança semântica.

Quadro 69 – FLs de processo

Função léxica	Valor semântico	Manifesta-se por (B)	Aplica-se a (A)	Exemplo
Prox	acontecimento iminente de A	verbo	substantivo	ProxOper(<i>colapso</i>) = <i>estar à beira do / de um ~</i>
Incep	início	verbo	substantivo	IncepReal(<i>contrato</i>) = <i>aderir ao / a um ~, acertar um ~</i> IncepFact(<i>contrato</i>) = <i>~ entrar em vigor, ~ entrar em vigência</i>
Cont	continuação	verbo	substantivo	ContOper(<i>contrato</i>) = <i>retomar o/um ~, renovar o/um ~</i> ContReal(<i>prazo</i>) = <i>prorrogar o ~</i>
Stop	interrupção	verbo	substantivo	Stop(<i>contrato</i>) = <i>suspender o ~</i> Stop(<i>porta</i>) = <i>~ emperrar</i>
Degrad	degradação	verbo	substantivo	Degrad(<i>leite</i>) = <i>~ azedar, ~ coalhar</i> Degrad(<i>saúde</i>) = <i>~ piorar, ~ deteriorar</i>
Fin	finalização	verbo	substantivo	FinReal(<i>contrato</i>) = <i>rescindir o/um ~, anular o/um ~, cancelar o/um ~, extinguir o/um ~</i> FinFact(<i>contrato</i>) = <i>~ terminar, ~ encerrar-se</i>

Fonte: Elaboração própria a partir das informações em Mel'čuk (2015, p. 223-224; 231-233)

Segundo Mel'čuk (2015, p. 224; 231), *Prox*, *Incep*, *Cont* e *Fin* são normalmente usadas com uma FL de verbo suporte, enquanto *Stop* e *Degrad* não apresentam esse prerequisite.

8.1.4.2.5 FLs sintagmáticas variadas

Por fim, as FLs variadas são um conjunto de funções ao qual Mel'čuk (2015) não atribuiu um denominador semântico comum.

Quadro 70 – FLs variadas

Função léxica	Valor semântico	Manifesta-se por (B)	Aplica-se a (A)	Exemplo
Caus	causa	verbo	substantivo	CausOper(<i>crise</i>) = <i>provocar uma ~</i>
Propt	causa	preposição	substantivo	Propt(<i>medo</i>) = <i>de ~; por ~ de</i>
Instr	meio	preposição	substantivo	Instr(<i>e-mail</i>) = <i>por ~</i>
Involv	abrangência	verbo	substantivo	Involv(<i>fogo</i>) = <i>~ alastrar-se</i>
Manif	manifestação de A	verbo	substantivo	CausManif(<i>arrependimento</i>) = <i>demonstrar ~</i>

Fonte: Elaboração própria a partir das informações em Mel'čuk (2015, p. 214-215; 224-227; 230-231)

As FLs variadas encerram o conjunto de FLs sintagmáticas estandar que selecionamos para esta oportunidade. Segundo Mel'čuk (2015, p. 267), esse conjunto é suficientemente exaustivo para descrever “séries regulares de colocações” [regular series of collocations] de uma língua. Nesse sentido, as colocações que eventualmente não possam ser descritas por uma das FLs anteriores (incluídas as cinco que omitimos da exposição) constituiriam casos à parte limitados numericamente por conterem FLs não estandar. Alguns exemplos disso são as colocações *café preto* e *ano bissexto*, que, segundo Mel'čuk (2015, p. 174-175; 353), não podem ser descritas de maneira satisfatória por uma FL. A única FL cujo valor semântico se aproxima dos colocados em questão é *Epit^{restr.}*; porém, *café* e *ano* não cumprem com as condições de manifestação dessa FL porque não são bases altamente polissêmicas. Conforme o autor, trata-se de casos à parte porque os colocados *preto* (“sem adição de leite”) e *bissexto* (“com 366 dias”) se combinam exclusivamente com *café* e *ano*, não havendo uma série regular de colocações em que esses colocados, com os significados em questão, estejam presentes. Para casos assim, Mel'čuk (2015, p. 267) afirma que a única saída é a descrição via uma FL não estandar, tais como “sem adição de leite” e “com 366 dias”.

Durante as buscas no CP (2019) por exemplos de colocações para ilustrar cada FL, deparamo-nos com outros casos de colocações que, aparentemente, também são de tipo não estandar porque não se enquadram totalmente nos critérios de manifestação das FLs sintagmáticas. Trata-se das colocações adjetivais *contrato de trabalho* e *contrato de aluguel*; e

das colocações verbais *mudar o/um contrato*, *alterar o/um contrato* e *rever o/um contrato*. As FLs *Epit* (qualificativa) e *Real* (verbo de realização) parecem não contemplar as relações entre base e colocado dos exemplos acima. Por outro lado, as colocações verbais mencionadas têm colocados que se combinam com outras bases, não parecendo tão restritas quando *preto* e *bissexto*, por exemplo. Somente mediante um estudo mais aprofundado e uma aplicação sobre um extenso número de colocações poderia fornecer respostas mais claras.

Apesar das dúvidas que certamente surgem ao se tentar empregar as FLs na descrição de colocações, os benefícios do modelo se sobrepõem aos percalços. Conforme já mencionado, as vantagens de se descrever colocações segundo a proposta de Mel'čuk (2015) estão na definição conceitual mais bem elaborada e nas FLs como recurso para descrever as colocações e, principalmente, organizá-las por um critério semântico ao invés de apenas dispô-las conforme a classe gramatical da base e do colocado – como o fazem os dicionários de aprendizes de inglês de níveis intermediário e avançado, por exemplo.

Note-se que uma das consequências de se tomar a definição de Mel'čuk (2015) a respeito das colocações e de outros fenômenos combinatórios é a mudança na classificação tradicional de algumas combinatórias. Segundo já comentado, na proposta de Mel'čuk (2015), a pedra de toque no momento de se identificar um tipo de frasema é o tipo de manifestação do pivô semântico no frasema. Se o pivô semântico não está manifestado por um dos constituintes do frasema, encontrando-se implícito, trata-se de uma fraseologia; ao contrário, se o pivô semântico está manifestado por um dos constituintes do frasema, trata-se de uma colocação. Essa diferença conceitual implica uma nova classificação de alguns frasemas tradicionalmente concebidos como fraseologias ou *idioms*, tais como *chover/estar chovendo canivete*, *llover a cántaros* e *[to be] raining cats and dogs*. Sob a ótica da TST, os três frasemas são colocações, pois o pivô semântico é um dos seus constituintes— *chover*, *llover* e *to rain*. A constatação da FL Magn reforça a constatação, pois o colocado é um intensificador da base. Portanto, $Magn(chover) = \sim canivete$; $Magn(llover) = \sim a cántaros$; e $Magn(rain_{(v)}) = \sim cats and dogs$. A mudança de perspectiva da TST tem uma implicância direta na maneira como esses frasemas precisam ser descritos em um dicionário. Essa questão será retomada mais adiante, na seção 8.2.

Outra consequência de se tomar a perspectiva de Mel'čuk (2012, 2015) a respeito das combinatórias léxicas consiste na obrigatoriedade de se adaptar as FLs a uma linguagem compreensível a um consulente não especialista em linguística e/ou não conhecedor da teoria de TST e do MST. Trata-se de um desafio para o qual ainda há poucas propostas de solução. Alonso Ramos (2006, p. 85), por exemplo, propõe que as FLs sejam adaptadas ao espanhol por

meio de *glosas*, isto é, uma adaptação, ou “tradução”, como prefere a autora, na qual as FLs são parafraseadas por ULs e expressões da língua comum. Sua proposta se insere em um contexto lexicográfico, pois a autora é a responsável pela compilação do *Diccionario de Colocaciones del Español* (DiCE, 2004), um dicionário de colocações compilado à luz da TST e que descreve, majoritariamente, substantivos do campo dos sentimentos.

Na mesma esteira de Alonso Ramos (2006) no que diz respeito às glosas, em Borba (2018), propusemos uma primeira tentativa de adaptação do nome de algumas FLs para a linguagem comum com o auxílio de uma lista de definidores de dicionário. O trabalho analisou as FLs que se manifestavam junto à UL *hambre* [fome] do espanhol e se baseou em uma lista de definidores que constam no dicionário Señas (2001) para adaptar os nomes das funções. Nessa oportunidade, propôs-se que os nomes adaptados das FLs fossem usados como etiquetas para introduzir os conjuntos de colocações relativas a cada FL. Também se sugeriu a inclusão de frases de exemplo junto de cada colocação.

A principal conclusão a que se chega a partir das propostas de adaptação das FLs é que esse procedimento é possível, mas requer uma certa flexibilidade por parte do lexicógrafo. Segundo Alonso Ramos (2006, p. 74; 77), uma mesma FL pode revelar nuances semânticas diferentes conforme a base à qual se aplica. Portanto, uma adaptação das FLs de Mel’čuk (2015) culmina não em uma glosa para cada função, mas em um conjunto de glosas a depender, em alguns casos, da relação entre a FL e a base da colocação. Esse tema será retomado mais adiante na tese.

Para finalizar as considerações a respeito da tipologia de frasemas de Mel’čuk (2012, 2015), o terceiro tipo de frasemas, após as fraseologias e as colocações, são os clichês.

8.1.4.3 Clichês

Assim como as colocações, os clichês são combinações de palavras com significado composicional (MEL’ČUK, 2015, p. 324). Segundo Mel’čuk (2015, p. 325), a diferença em relação às colocações está no nível de restrição: todos os componentes léxicos de um clichê são selecionados de maneira restrita, assim como o seu significado, que também está engessado. Formalmente, na TST os clichês são definidos da seguinte maneira:

Um frasema semanticamente composicional $\mathbf{AB} = \langle 'A \oplus B'; /A \oplus B/ \Sigma_{A \oplus B} \rangle$ é chamado de clichê se todos os seus semantemas e lexemas são selecionados de modo

restrito no que diz respeito a uma dada representação conceitual – isto é, o Falante seleciona **AB** como um todo (MEL'ČUK, 2015, p. 325. Grifos no original)¹⁵⁹.

O autor cita como exemplos de clichês os provérbios (ou ditados populares), como “Há males que vêm para o bem”, “Antes tarde do que nunca” e “Deus ajuda quem cedo madruga”; e fórmulas ritualísticas, como “Bom dia”, “Proibido estacionar” e “Olha por onde anda!”. Mel'čuk (2015, p. 324) salienta que nem todos os provérbios e fórmulas ritualísticas se enquadram na categoria de clichês, mas somente aqueles que têm significado composicional. Nesse sentido, provérbios como “Água mole em pedra dura tanto bate até que fura” e “A mentira tem perna curta” não são considerados clichês na TST; em função do seu significado não composicional, essas expressões se enquadram na categoria das fraseologias [idioms]. Embora os conceitos da TST possam causar estranhamento ao separarem elementos comumente abrigados sob a mesma categoria, como os provérbios, a separação entre o que é composicional, parcialmente composicional e não composicional promove uma descrição linguística adequada ao nível de dificuldade de compreensão e de produção que cada tipo de combinatória léxica pode oferecer (considerando que as combinatórias não composicionais oferecem mais dificuldade, e as composicionais, mais facilidade).

Mel'čuk (2012, p. 42; 2015, p. 332) divide os clichês em duas categorias: clichês pragmaticamente não restritos e pragmatemas. O fator que motiva essa divisão é o tipo de exigência da situação comunicativa. Os clichês pragmaticamente não restritos obedecem, em certa medida, à vontade do falante em se expressar de uma determinada maneira, não tendo vinculação direta com a situação comunicativa em que são proferidos. Em outras palavras, não oferecem restrições a nível pragmático (Mel'čuk, 2012, p. 40-41). Um exemplo desse tipo de clichê são os provérbios citados anteriormente, que não estão vinculados a uma determinada situação comunicativa, mas podem ser usados a critério. Os pragmatemas, por outro lado, são clichês cujo uso responde a uma determinada situação comunicativa. O exemplo “Proibido estacionar” é encontrado em placas de trânsito, ou seja, em um contexto bem delimitado. A fórmula “Olha por onde anda!” é uma resposta ao descuido de alguém que, ao se locomover, esbarra no falante. Finalmente, a fórmula “Bom dia” é para demonstrar educação e respeito ao encontrar uma pessoa conhecida (um vizinho, por exemplo), ao passar por funcionários de ambientes privados (um recepcionista de prédio comercial, por exemplo) ou ao dividir certos espaços com estranhos (uma recepção de consultório médico ou um elevador, por exemplo).

¹⁵⁹ [A semantically compositional phraseme $\mathbf{AB} = \langle 'A \oplus B'; /A \oplus B/ \Sigma_{A \oplus B} \rangle$ is called a cliché iff al of its semantemes and lexemes are selected restrictedly with respect to a given conceptual representation – that is, the Speaker selects **AB** as a whole]

O estudo dos clichês é, segundo Mel'čuk (2012, p. 40), ainda incipiente no âmbito da Lexicologia. O autor observa que, os clichês são “os mais difíceis de localizar e coletar e, atualmente, os menos estudados”¹⁶⁰ (MEL'ČUK, 2012, p. 40). Sob a perspectiva (meta)lexicográfica, é difícil determinar um denominador comum ao qual reduzir os clichês. De maneira análoga ao pivô semântico das fraseologias e colocações, Mel'čuk (2012, p. 41) propõe o conceito de “âncora léxica” [lexical anchor], que consiste na redução do significado do clichê a uma UL, registrando-o no verbete correspondente. Trata-se de uma redução do significado que visa a função de produção. Nesse sentido, o clichê “Olha por onde anda!” poderia ser incluído sob o verbete *atenção* e “Proibido estacionar”, sob o verbete *estacionar*; no entanto, no caso de “Há males que vêm para o bem”, é difícil determinar o que poderia ser a âncora léxica. A identificação dessa propriedade parece até mais subjetiva que a identificação do pivô semântico de uma fraseologia.

Nos dicionários de aprendizes de inglês e de espanhol, é raro encontrar clichês pragmaticamente não restritos, muito provavelmente porque a sua estrutura semanticamente complexa dificulta o procedimento de redução. Quanto às fórmulas ritualísticas, os dicionários de aprendizes de inglês intermediários e avançados costumam incluir exemplares desse tipo de combinatória no *middle matter* (páginas que interrompem a progressão macroestrutural na metade do dicionário) ou no *back matter* (páginas finais, após a macroestrutura). As situações comunicativas que normalmente aparecem são as que envolvem o ambiente acadêmico, laboral e de exames de proficiência. No OALD8 (2010) e no CALD4 (2013), que são dicionários avançados, essas situações são tratadas sob a forma de gêneros textuais, tais como redação, apresentações orais, relatórios, currículo, e-mails formais e informais, resenhas de livros e filmes, entre outros. Outro dicionário que descreve de forma exaustiva uma série de situações e funções comunicativas e as respectivas fórmulas ritualísticas (dentre outros recursos léxicos) que costumam aparecer nesses contextos é o CCID4 (2018), um dicionário intermediário (cf. Borba (2020a) para uma análise mais exaustiva da obra).

Em suma, os clichês constituem um fenômeno lexicológico altamente relevante que requer mais investigações para que possa ser mais bem dimensionado. Além dos *corpora*, os dicionários de aprendizes são uma fonte de dados que pode auxiliar nas pesquisas a respeito dos clichês.

Após o breve panorama das perspectivas estatística, lexicológica e gramatical acerca das combinatórias léxicas, a próxima seção reúne as principais contribuições de cada

¹⁶⁰ [the most difficult to pinpoint and collect and, at the current time, the least studied]

perspectiva e como cada uma pode ser aproveitada para fins de descrição de combinatórias léxicas em dicionários de aprendizes de inglês e de espanhol.

8.2 Proposta para a descrição de combinatórias léxicas em dicionários de aprendizes

Uma das grandes vantagens da (Meta)Lexicografia é a possibilidade de aproveitamento de contribuições de distintas perspectivas teóricas conforme os objetivos que se tem, sem a necessidade de filiação a uma única teoria. No caso das combinatórias léxicas, é pertinente ter em vista que uma única perspectiva, por mais desenvolvida que seja, não cobre a diversidade de fenômenos combinatórios; haverá sempre algo relevante que, não sendo contemplado por uma abordagem, será objeto de interesse da abordagem vizinha. Ante o cenário de pontos fortes e fracos dos vieses estatístico, gramatical e lexicológico exposto na seção anterior, tomamos o que havia de mais sólido em cada um para determinar alguns dos tipos de fenômenos combinatórios existentes, identificar as suas propriedades e propor estratégias com vistas à sua descrição lexicográfica em dicionários de aprendizes de inglês e de espanhol de níveis A2 e B1 de aprendizagem. Não se trata de ensinar aos aprendizes os diferentes tipos de combinatórias léxicas, seus conceitos e terminologia, pois, além de irrelevantes para o usuário almejado pela tese, os limites fluidos entre as combinatórias e os casos de amálgama de fenômenos combinatórios dificultam a sua categorização em certos casos (cf. seções a seguir e Borba (2020b)). Ao contrário, nossa proposta é que se use a base teórica disponível para descrever fenômenos combinatórios da maneira mais clara possível a partir das suas propriedades.

Há, primeiramente, um mínimo de cinco fenômenos combinatórios que foram identificados na literatura: coligações (HOEY, 2005), restrições de seleção (BOSQUE, 2001), fraseologias (MEL'ČUK, 2015), colocações (SINCLAIR, 1991; MEL'ČUK, 2015) e clichês (MEL'ČUK, 2015). As coligações e as colocações gramaticais dos contextualistas têm o potencial de contribuir com a discriminação de acepções de verbetes polissêmicos; as restrições de seleção de Bosque (2001) podem auxiliar na redação de paráfrases explanatórias, em particular de verbos, no que diz respeito à especificação dos tipos de complemento que os acompanham; as fraseologias e as colocações da perspectiva de Mel'čuk (2015) lançam luz sobre a variedade e singularidade das combinações de uma UL com outras; e os clichês de Mel'čuk (2015) tratam da dimensão social das combinatórias.

Dentre os cinco fenômenos listados, foram elaboradas propostas de descrição lexicográfica para os quatro primeiros, tratadas nas seções a seguir. Não obstante a sua relevância para os estudos combinatórios, os clichês foram excluídos desse apanhado porque requerem mais dados de pesquisas sobre a sua identificação.

Considerando que o perfil de usuário é uma variável transversal na tese, as considerações dos capítulos anteriores também repercutem sobre as propostas de descrição de padrões combinatórios. Sendo assim, o material linguístico que serviu de base para ilustrar as propostas relacionadas a cada padrão foram os documentos sobre o ensino-aprendizagem de inglês e espanhol como LE, EP (2011) e PCIC (2006), bem como os dados analisados nos capítulos 6 e 7.

8.2.1 *Propostas para a descrição de coligações e colocações gramaticais (contextualismo inglês)*

Sob a perspectiva do contextualismo inglês de Sinclair (1991, 2004) e de Hoey (2005), as coligações e colocações gramaticais são combinatórias que ajudam a distinguir ULs (*i.e.* as associações entre a palavra-entrada e cada acepção) em um verbete de dicionário. Por essa razão, o local mais apropriado para a inclusão desses dois tipos de combinatória são as acepções que lhes correspondem – embora não se possa descartar a necessidade de se recorrer a notas de uso para descrever certos padrões coligacionais mais complexos, como é o caso das posições sintáticas de *already* e do tipo de sentença em que esse advérbio é usado (afirmativas) (cf. Quadro 64).

No que diz respeito ao ensino-aprendizagem, a informação a respeito do padrão sintático sob o qual cada UL (ou acepção) se apresenta é relevante tanto em termos de compreensão linguística (localização de um significado a partir de seu padrão sintático) como em termos de produção linguística (manifestação de um significado). Em termos de descrição lexicográfica, cada função implica um posicionamento distinto da coligação ou colocação gramatical em relação à paráfrase explanatória. Ao se adotar a perspectiva da compreensão, a solução lógica é incluir colocações gramaticais e coligações em uma posição anterior à paráfrase explanatória: [índice da acepção + categoria morfológica + padrão sintático + paráfrase + exemplo]. Essa perspectiva é adotada por Farias (2013), por exemplo, para indicar regências em sua proposta. Por outro lado, a posição condizente com a perspectiva da produção é a posterior à paráfrase: [índice da acepção + categoria morfológica + paráfrase + padrão sintático + exemplo]. Essa

perspectiva está presente no verbete *enjoy* do OALD8 (2010), no qual os exemplos da acepção 1 são introduzidos pelos padrões coligacionais *~sth* e *~doing sth* (cf. Quadro 64).

Uma terceira alternativa é a de se adotar ambas as perspectivas, de forma a repetir informações no verbete. Essa foi a solução adotada nas adaptações do capítulo 6, como a de *remind* (Quadro 38), na qual a acepção 1 registra uma colocação gramatical antes da acepção nos exemplos, com destaque tipográfico: “1 **~to** ajudar [alguém] a lembrar de [algo] que ele/a deve fazer; em português, *lembrar [alguém] de [algo]: Please **remind me to buy some bread on the way home***”. Embora deixe o verbete mais denso, a repetição desse tipo de informação pode ser benéfica em se tratando de aprendizes com níveis A1 e A2 de proficiência, pois constitui o primeiro contato desses aprendizes com o fenômeno da combinação de palavras em outra língua. Além disso, o reforço os ajuda a perceber desde cedo que o léxico não é somente uma lista vertical de palavras isoladas, mas que também há uma dimensão horizontal que contempla as palavras em suas combinações. À medida que o perfil de usuário de um dicionário de aprendizes pressupõe níveis de proficiência mais altos, diminui a necessidade de se repetir informações combinatórias.

Para ilustrar a descrição de colocações gramaticais e coligações, foram tomadas as seguintes combinatórias: *to think of/about [doing something]* (EVP, 2015, A2, *think* [consider doing]) e *to think about [someone/something]* (EVP, 2015, B1, *think* [use brain]) do inglês; *haber + [artículo indefinido]* (PCIC, 2006, A2, gramática) e *parecerse a [alguem]* (PCIC, 2006, B1, nociones específicas) do espanhol.

Quadro 71 – Proposta de descrição de colocações gramaticais e coligações

Verbetes original	Versão adaptada
<p>think /θɪŋk/ verb (past tense and past participle thought /θɔ:t/) KEY PATTERNS think about something think carefully think (that) what do you think of someone/something?</p> <p>1 to use your mind, for example to make a decision or to imagine what something else is like: <i>Alison thought about what he had said.</i> • <i>Think carefully before you make any decisions.</i> • <i>I need some time to think.</i> [...] PHRASAL VERBS think about/of to consider doing something: <i>We're thinking of going to Florida this year.</i> • <i>Have you ever thought about marriage?</i> • <i>I've been thinking about joining the army</i></p> <p>LWord2 (2008, s.v. <i>think</i>)</p>	<p>think /θɪŋk/ verb (past tense and past participle thought /θɔ:t/) USO think about something think carefully think (that) what do you think of someone/something?</p> <p>1 (about) usar a mente para tomar uma decisão ou imaginar como algo se parece; em português, <i>pensar sobre/em: Alison thought about what he had said.</i> • <i>Think carefully before you make any decisions.</i> • <i>I need some time to think.</i> [...] PHRASAL VERBS think about/of [doing something] considerar fazer algo: <i>We're thinking of going to Florida this year.</i> • <i>Have you ever thought about marriage?</i> • <i>I've been thinking about joining the army</i></p>

<p>ha·ber aβér [...] 2 unipers. Existir o estar disponible: <i>hay amores que matan; había fruta en la nevera; habrá cinco camareros.</i> [...]</p>	<p>ha·ber aβér [...] 2 unipers. (un/os, una/s) Existir o estar disponible; en portugués, <i>haver: hay amores que matan; había fruta en la nevera; hay una continua influencia entre los dos.</i> [...]</p>
<p>pa·re·cer pareθér [...] - 5 parecerse prnl. Ser igual; mostrar características iguales o comunes: <i>el padre y su hijo se parecen; dicen que tú y yo nos parecemos; esa foto se aprece a una que tiene Pedro; los dos caballos se parecían, pero uno era más fuerte.</i> [...]</p>	<p>pa·re·cer pareθér [...] - 5 parecerse (a) prnl. Ser igual; mostrar características iguales o comunes; en portugués, <i>parecer-se: en el fondo te pareces a mí; el padre y su hijo se parecen.</i> [...]</p>

Fonte: LWord2 (2008), Señas (2001) e Corpes (2020)

No verbete *think*, algumas mudanças mínimas foram feitas para reforçar a descrição dos casos de colocação gramatical e coligação retirados do EVP (2015). Na aceção 1, acrescentou-se a preposição *about* entre parênteses para indicar a possibilidade da sua ocorrência com *think* para expressar o conteúdo semântico daquela aceção. Mais adiante, na lista dos *phrasal verbs*, incluiu-se o padrão coligacional *doing something* à colocação gramatical *think about/of*. Optou-se por incluir *doing something* entre colchetes para indicar que essa coligação não é de uso facultativo, mas necessário.

No verbete *haber*, como o verbo se manifesta tanto com a coligação “haber + artículo indefinido” como sem ela, os artigos indefinidos foram colocados entre parênteses. Além disso, como não havia exemplos que explicitassem o padrão coligacional em questão, buscou-se um alinhamento do Corpes (2021) para ilustrá-lo (“hay una continua influencia entre los dos”). Também se optou por reduzir a quantidade de exemplos ao mínimo necessário.

No verbete *parecer*, por sua vez, incluiu-se a preposição *a* na quinta aceção. Como o significado em questão poderia se manifestar tanto com a preposição como sem ela, utilizamos o recurso dos parênteses para indicar que o seu uso, dependendo da situação, não era obrigatório. Além disso, como não havia um exemplo ilustrativo da colocação gramatical *parecerse a*, novamente foi incluído um exemplo extra a partir de um alinhamento do Corpes (2021) (“en el fondo te pareces a mí”). Finalmente, assim como na adaptação anterior, optou-se por reduzir a quantidade de exemplos ao mínimo necessário.

8.2.2 Propostas para a descrição de restrições de seleção (BOSQUE, 2001)

De acordo com o exposto em 8.1.3, sob a perspectiva de Bosque (2001), as restrições de seleção auxiliam na especificação da extensão de um significado na paráfrase explanatória. Trata-se de uma combinatória cuja descrição é especialmente necessária para tarefas de produção. Nossa proposta para o fenômeno em questão consiste na especificação das classes léxicas de complemento em substituição aos pronomes *something* ou *algo* que aparecem nas paráfrases de dicionários de aprendizes de inglês e de espanhol e que não permitem saber a quais entidades uma dada UL se aplica.

Para ilustrar a proposta, foram selecionadas as seguintes restrições de seleção: *to fold* [*paper, cloth, body parts*] (EVP, 2015, B1, *fold*) e *to go* + [*doing sport*] (EVP, 2015, *go running/swimming, etc.*) do inglês; e *tener* + [*percepción física*] (PCIC, 2006, A2, nociones específicas) e *ponerse* + [*estado de ánimo*] (PCIC, 2006, B1, nociones específicas) do espanhol.

Quadro 72 – Proposta de descrição de restrições de seleção

Verbetes original	Versão adaptada
<p>fold¹ ↔ /fəʊld/ verb (folds, folding, folded) (also fold up) 1 to bend something so that one part is on top of another part: <i>I folded the letter and put it in the envelope.</i> ◇ <i>Fold up your clothes.</i> ⇨ OPPOSITE unfold [...]</p>	<p>fold¹ ↔ /fəʊld/ verb (folds, folding, folded) (também fold up) 1 dobrar [papel, tecido, parte do corpo] de modo que uma parte fique em cima da outra: <i>I folded the letter and put it in the envelope.</i> ◇ <i>Fold up your clothes.</i> ◇ <i>She felt cold and folded her arms across her chest.</i> ⇨ OPPOSITE unfold [...]</p>
<p>go¹ ↔ /gəʊ/ verb (goes, going, went /went/, has gone /gɒn/) [...] 2 to travel to a place to do something: <i>Paul has gone shopping.</i> ◇ <i>Are you going to Dave's party?</i> ◇ <i>I'll go and make some coffee.</i> [...]</p>	<p>go¹ ↔ /gəʊ/ verb (goes, going, went /went/, has gone /gɒn/) [...] 2 (~doing) ir a um lugar para fazer uma [atividade] ou um [esporte]: <i>Paul has gone shopping.</i> ◇ <i>Are you going to Dave's party?</i> ◇ <i>I'll go and make some coffee.</i> ◇ <i>After leaving the chapel, they went walking in the gardens.</i> [...]</p>
<p>te•ner tenér [...] 6 Experimentar o sentir: <i>tengo calor; tengo hambre; tengo dolor de cabeza; tengo prisa; tengo sueño</i> • ter. [...]</p>	<p>te•ner tenér [...] 6 Experimentar una [sensación física] o [necesidad]; en portugués, <i>estar com, sentir: tengo calor; tengo hambre; tengo dolor de cabeza; tengo prisa; tengo sueño.</i></p>
<p>po•ner ponér [...] 6 Empezar a mostrar; *adoptar: <i>el jefe puso mala cara cuando se lo dije.</i> • fazer [...]</p>	<p>po•ner ponér [...] 6 Empezar a mostrar; adoptar; en portugués, <i>fazer: el jefe puso mala cara cuando se lo dije.</i> 7 ponerse prnl. Pasar a un estado de ánimo; en portugués, <i>ficar: mi cuñado se pondrá muy contento al vernos llegar.</i> [...]</p>

Fonte: OxED2 (2012), Señas (2001), BNC (2007) e Corpes (2021)

Na versão adaptada do verbete *fold*, além de se explicitar os tipos de complemento do verbo, foi acrescentado um exemplo do BNC (2007) para ilustrar o complemento de “partes do corpo”. O mesmo procedimento foi feito na adaptação do verbete *go^l*, com a diferença de que, diante da possibilidade de amálgama entre a restrição de seleção e a coligação *go doing* para manifestar o significado em questão, esta última foi incluída entre parênteses. Por fim, manteve-se a marcação tipográfica em negrito que o OxED2 (2012) aplica às combinatórias léxicas presentes nos exemplos de *go^l*; no caso da coligação *go doing*, o sufixo *-ing* dos exemplos também recebeu marcação em negrito.

Nas adaptações do espanhol, enquanto em *tener* foi preciso apenas explicitar os complementos na acepção correspondente, em *poner* foi necessário incluir uma acepção inteira, pois o significado correspondente à restrição de seleção *ponerse + estado de ánimo* não estava registrado no Señas (2001). Neste último caso, recorreu-se ainda aos alinhamentos do Corpes (2021) para fornecer um exemplo (“mi cuñado se pondrá muy contento al vernos llegar”).

8.2.3 Propostas para a descrição de fraseologias (MEL’ČUK, 2015)

Para descrever fraseologias, adotou-se a perspectiva de Mel’čuk (2015), pois foi a que forneceu mais subsídios teóricos para definir o fenômeno. De fundamental importância é o conceito de pivô semântico, já que, para que uma combinação possa ser considerada uma fraseologia sob a perspectiva em questão, seu pivô semântico (*i.e.* seu principal traço de significado) deve estar ausente da combinação. Sob a perspectiva da compreensão, há duas maneiras de se registrar uma fraseologia: como uma UL independente, sob a forma de uma palavra-entrada; ou como uma unidade dependente, subordinada ao verbete de um de seus constituintes. Nos dicionários de aprendizes de inglês, a primeira opção vem se tornando cada vez mais comum, enquanto a segunda é a mais adotada pelo Señas (2001). Por outro lado, sob a perspectiva da produção, ainda não há uma proposta de inclusão de fraseologias em dicionários semasiológicos de forma tal que o consulente possa chegar à unidade fraseológica sem conhecê-la previamente.

Frente a esse cenário, nossa proposta em relação às fraseologias diz respeito muito mais a uma metodologia de descrição com vistas à função de produção linguística. Além da inclusão sob o verbete de um dos constituintes ou como palavra-entrada na macroestrutura, propomos as fraseologias sejam incluídas também no verbete correspondente ao seu pivô semântico. Para ilustrar a proposta, tomaram-se as seguintes unidades fraseológicas: *to grow up* (EVP, 2015, A2) e *to do somebody’s best* (EVP, 2015, B1) do inglês, cujos pivôs semânticos são *adult* e

effort, respectivamente. Para o espanhol, o PCIC (2006) associa as fraseologias somente aos níveis B2 em diante. A perspectiva que adotamos na tese, no entanto, é a de que não há como passar por qualquer nível de aprendizagem de LE, por mais básico que seja, sem aprender combinatórias léxicas, independentemente do fenômeno específico de que se trate. Por essa razão, decidimos buscar fraseologias que tivessem como pivô semântico ULs atreladas aos níveis A2 e B1 do espanhol: *pegarse los ojos* e *mala pata*, que têm como pivô semântico as ULs *sueño* (A2) e *suerte* (B1), respectivamente.

Quadro 73 – Proposta de descrição secundária de fraseologias

Verbetes original	Versão adaptada
<p>grow /grəʊ/ verb (grew, grown) [...]</p> <p>to grow up phrasal verb A2 to become older or an adult: <i>She grew up in Madrid.</i></p> <p>adult¹ /'ædʌlt/ noun 1 A1 a person or animal that has finished growing and is not now a child.</p>	<p>grow /grəʊ/ verb (grew, grown) [...]</p> <p>to grow up phrasal verb A2 tornar-se mais velho ou adulto: <i>She grew up in Madrid.</i></p> <p>adult¹ /'ædʌlt/ noun 1 A1 uma pessoa ou animal que já terminou de crescer e não é mais uma criança ou filhote. A2 to grow up phrasal verb tornar-se mais velho ou adulto: <i>She grew up in Madrid.</i></p>
<p>best³ /best/ noun [...]</p> <p>4 do/try your best B1 to try very hard to do something: <i>I did my best to persuade him.</i> [...]</p> <p>effort /'efət/ noun 1 B1 an attempt to do something: <i>He was making an effort to be sociable.</i> 2 B1 [no plural] the energy that you need to do something: <i>I put a lot of effort into organizing the party.</i></p>	<p>best³ /best/ noun [...]</p> <p>4 do/try your best B1 tentar fazer algo com muito esforço: <i>I did my best to persuade him.</i> [...]</p> <p>effort /'efət/ noun 1 B1 uma tentativa de fazer algo: <i>He was making an effort to be sociable.</i> do/try your best B1 tentar fazer algo com muito esforço: <i>I did my best to persuade him.</i> 2 B1 [no plural] a energia necessária para fazer algo: <i>I put a lot of effort into organizing the party.</i></p>
<p>pe·gar peɣár [...]</p> <p>■ pegarse los ojos, fam. Tener mucho sueño: <i>estoy tan cansado que se me pegan los ojos.</i> • fecharem-se os olhos</p> <p>sue·ño suéño 1 m. Estado del que duerme: <i>durante el ~, el cuerpo repone las energías necesarias.</i> • sono 2 Deseo o necesidad de dormir: <i>ayer me acosté muy tarde y por eso hoy tengo tanto ~.</i> • sono [...]</p>	<p>pe·gar peɣár [...]</p> <p>■ pegarse los ojos, fam. Tener mucho sueño; en portugués, <i>fecharem-se os olhos: estoy tan cansado que se me pegan los ojos.</i></p> <p>sue·ño suéño 1 m. Estado del que duerme; en portugués, <i>sono: durante el ~, el cuerpo repone las energías necesarias.</i> 2 Deseo o necesidad de dormir; en portugués, <i>sono: ayer me acosté muy tarde y por eso hoy tengo tanto ~.</i> ■ pegarse los ojos, fam. Tener mucho sueño: <i>estoy tan cansado que se me pegan los ojos.</i> [...]</p>

<p>pata páta [...] ■ mala ~, <i>fig. fam.</i>, mala suerte: <i>tengo muy mala pata em los negócios.</i> ⇒ <i>pie.</i> • azar</p> <p>suerte suerte [...] 2 Situación o circunstancia buena o mala: <i>tuve la mala ~ de tropezar con el único obstáculo que había en el camino; ¡qué buena suerte! ¡me ha tocado el primer premio de la rifa!</i> • sorte [...]</p>	<p>pata páta [...] ■ mala ~, <i>fig. fam.</i>, mala suerte; en portugués, <i>azar: tengo muy mala pata em los negócios.</i> ⇒ <i>pie.</i></p> <p>suerte suerte [...] 2 Situación o circunstancia buena o mala: <i>tuve la mala ~ de tropezar con el único obstáculo que había en el camino; ¡qué buena suerte! ¡me ha tocado el primer premio de la rifa!</i> ■ mala pata, <i>fig. fam.</i>, mala suerte; en portugués, <i>azar: tengo muy mala pata em los negócios</i> [...]</p>
---	--

Fonte: CaED2 (2011), Señas (2001)

A fraseologia *to grow up*, que também se enquadra no fenômeno dos *phrasal verbs*, estava registrada inicialmente no verbete *grow*. Para fins de compreensão, a única alteração feita foi a adaptação da paráfrase explanatória no molde das propostas do capítulo 6. A parte mais inovadora da adaptação, no entanto, diz respeito à inclusão da fraseologia sob o seu pivô semântico (*adult*). Dessa forma, toda vez que o consulente desejar realizar uma atividade de produção linguística que envolva a UL *adult*, poderá chegar à fraseologia em questão e, se julgar conveniente, usá-la.

A fraseologia *to do somebody's best*, por sua vez, estava registrada no verbete *best*³. Da mesma forma que no caso anterior, as alterações foram a nível de paráfrase explanatória. Para fins de produção, no verbete *effort*, incluiu-se a fraseologia em questão junto da aceção correspondente ao pivô.

Na contraparte do espanhol, as fraseologias *pegarse los ojos* e *mala pata* estão registradas nos verbetes *pegar* e *pata* do Señas (2001), respectivamente. Além das alterações mínimas nos verbetes de origem (supressão do equivalente em português e destaque tipográfico na sílaba tônica da transcrição fonológica), ambas as fraseologias foram incluídas junto das aceções correspondentes ao seu pivô semântico.

8.2.4 Propostas para a descrição de colocações (MEL'ČUK, 2015)

Para a definição e descrição das colocações, tomamos o aporte teórico-metodológico de Mel'čuk (2015). Novamente, o conceito de pivô semântico é fundamental, pois sua presença na combinatória é o que diferencia as colocações das fraseologias.

O principal desafio da aplicação da TST é a adaptação das FLs para uma linguagem acessível ao consulente não linguista. Alonso Ramos (2006), por exemplo, propõe a elaboração de glosas, isto é, de expressões parafrásticas das FLs, a partir de um conjunto de quatro parâmetros. O primeiro e o quarto tratam da priorização da clareza da glosa em comparação com a sua precisão semântica (ALONSO RAMOS, 2006, p. 75-76). Segundo a autora, para que uma glosa seja facilmente compreensível para um público não linguista, é preciso investir na clareza da glosa em detrimento da sua precisão semântica. O segundo parâmetro diz respeito à correção idiomática da glosa. Conforme Alonso Ramos (2006, p. 75-76), a combinação entre a glosa e a base da colocação deve ser válida e usual da língua – em outras palavras, a glosa e a base devem formar uma colocação. O terceiro parâmetro é relativo às FLs verbais. De acordo com Alonso Ramos (2006, p. 76), as glosas de FLs verbais devem manter a diferenciação das funções sintáticas da base (sujeito, objeto ou adjunto). A autora ainda salienta que dificilmente haverá uma única glosa para cada FL; segundo ela, é mais sensato trabalhar com um conjunto aberto de glosas para cada FL, de modo que se possa acrescentar novas glosas conforme a necessidade (ALONSO RAMOS, 2006, p. 74).

Dentre as orientações de Alonso Ramos (2006), o parâmetro mais difícil de se obedecer é o da obrigatoriedade de se formar uma colocação com a glosa e a base, ao mesmo tempo que se mantém um certo grau de generalização semântica para que uma mesma glosa possa substituir uma dada FL em diversas situações. Frente às dificuldades que tivemos em seguir esse parâmetro, optamos por propor glosas que refletissem o valor semântico das FLs.

Quadro 74 – Proposta de adaptação das FLs

Função léxica	Valor semântico	Proposta de glosa
Epit ^{restr}	especificação	tipo de ~
Epit ^{qual}	reforço; redundância	+ ênfase
Magn	intensificação	+ intensidade
AntiMagn	atenuação	– intensidade
Plus, Minus	comparação de alta/baixa gradação	grau de ~
Ver	condição (culturalmente) esperada	traço positivo de ~
Bon	juízo de valor positivo	traço positivo de ~
Pos	avaliação positiva	traço positivo de ~
AntiVer	violação de condição (culturalmente) esperada	traço negativo de ~
AntiBon	juízo de valor negativo	traço negativo de ~
AntiPos	avaliação negativa	traço negativo de ~
LOC _{in}	localização física	local
LOC _{ad}	movimento para uma direção	movimento para ~
LOC _{ab}	movimento a partir de um local	movimento de ~
LOC ^{temp}	localização temporal	data

Centr	localização central ou parte mais importante	núcleo de ~
Copul	estado	estado de ~
Oper	suporte	verbo + ~
Real	realização	verbo + ~
AntiReal	realização	verbo + ~
Func	suporte	~ + verbo
Fact	realização	~ + verbo
Labor	suporte	verbo + ~
Labreal	realização	verbo + ~
Prepar	preparação de A via B para ação C	preparar ~
Obstr	dificuldade de funcionamento ou realização	impedir ~
Son	som (característico) de A	som de ~
Prox	acontecimento iminente de A	~ prestes a acontecer
Incep	início	começar ~
Cont	continuação	continuar ~
Stop	interrupção	parar ~
Degrad	degradação	~ estragar
Fin	finalização	terminar ~
Caus	causa	causar ~
Propt	causa	causar ~
Instr	meio	meio
Involv	abrangência	envolver

Fonte: Elaboração própria a partir de Mel'čuk (2015)

Para ilustrar a aplicação da TST sobre a descrição lexicográfica das colocações, foram selecionados os seguintes itens combinatórios: *critically ill* (EVP, 2015, A2), [*to feel/ to be taken*] *ill* (EVP, 2015, A2), [*to spend/earn/make/save*] *money* (EVP, 2015, A1), *pocket money* (EVP, 2015, B1); *tener sueño* (EVP, 2015, A2), *buena persona* (EVP, 2015, B1), [*demasiado, muy*] *tarde* (EVP, 2015, A1), [*llamar por, hablar por, contestar al*] *teléfono* (EVP, 2015, A2).

Quadro 75 – Proposta de descrição de colocações

Colocação	Verbetes original	Versão adaptada
Magn(<i>ill</i>) = <i>critically ill</i> Oper(<i>ill</i>) = <i>to feel ill, to be taken ill</i>	ill / iɪ/ adj A2 not feeling well, or suffering from a disease: <i>He was in bed, ill.</i>	ill / iɪ/ adj A2 não se sentir bem, ou estar sofrendo de uma doença: <i>He was in bed, ill.</i> + intensidade critically ill <i>He said his mother is critically ill</i> verbo + ill to feel ill <i>Do you feel ill?</i> to be taken ill <i>Sean's teacher was taken ill</i>
Epit ^{restr} (<i>money</i>) = <i>pocket money</i>	money / mʌni/ verb [no plural] A1 the coins or pieces of paper that are used for buying things: <i>How much money have you got?</i> •	money / mʌni/ verb [no plural] A1 moedas ou pedaços de papel usados para comprar coisas: <i>We have no money.</i>

<p>Real(money) = to spend money, to earn money, to make money, to save money</p>	<p>He spends all his money on clothes • The company's not making (=earning) any money at the moment.</p>	<p>tipo de money pocket money <i>I don't get pocket money, my parents just get me things I need.</i> verbo + money to spend money <i>He spends all his money on clothes</i> to earn money <i>She took three part time jobs to earn money</i> to get money <i>How much money have you got?</i> to make money <i>The company's not making any money at the moment.</i> to save money <i>It is always possible to save money on entertainment</i></p>
<p>Oper(sueño) = tener sueño</p>	<p>sue·ño suéño 1 m. Estado del que duerme: <i>durante el ~, el cuerpo repone las energías necesarias.</i> • sono 2 Deseo o necesidad de dormir: <i>ayer me acosté muy tarde y por eso hoy tengo tanto ~.</i> • sono [...]</p>	<p>sue·ño suéño 1 m. Estado del que duerme; en portugués, <i>sono</i>: <i>durante el ~, el cuerpo repone las energías necesarias.</i> 2 Deseo o necesidad de dormir; en portugués, <i>sono</i>: <i>ayer me acosté muy tarde y por eso hoy tengo tanto ~.</i> verbo + sueño tener sueño ■ pegarse los ojos, <i>fam.</i> Tener mucho sueño; en portugués, <i>fecharem-se os olhos</i>: <i>estoy tan cansado que se me pegan los ojos.</i> [...]</p>
<p>Bon(persona) = buena persona</p>	<p>per·so·na persóna 1 f. Individuo de la especie humana: <i>habría un centenar de personas; varias personas entraron en el banco.</i> • pessoa [...]</p>	<p>per·so·na persóna 1 f. Individuo de la especie humana; en portugués, <i>pessoa</i>: <i>habría un centenar de personas; varias personas entraron en el banco.</i> atributo positivo de persona buena persona <i>Es una buena persona, siempre dispuesta a hacer un favor.</i> [...]</p>
<p>Magn(tarde) = demasiado tarde, muy tarde</p>	<p>tar·de tárde [...] - 3 adv. t. A una hora avanzada del día o de la noche: <i>Ayer se acostó a las cuatro y hoy se há levantado muy ~; salimos a bailar y llegamos a casa ~.</i> • tarde [...]</p>	<p>tar·de tárde [...] - 3 adv. t. A una hora avanzada del día o de la noche; en portugués, <i>tarde</i>: <i>Salimos a bailar y llegamos a casa ~.</i> + intensidad demasiado tarde <i>Siempre se despierta demasiado tarde.</i> muy tarde <i>Ayer se acostó a las cuatro y hoy se há levantado muy ~.</i> [...]</p>

Real(<i>teléfono</i>) = <i>llamar por teléfono, hablar por teléfono, contestar al teléfono</i>	te·lé·fo·no teléfono [...] 2 m. Aparato que sirve para hablar a larga distancia: <i>¿puedo usar su teléfono, por favor?</i> • telephone [...]	te·lé·fo·no teléfono [...] 2 m. Objeto que sirve para hablar a gran distancia; en portugués, <i>telefone: ¿puedo usar su teléfono, por favor?</i> verbo + teléfono llamar por teléfono <i>No se decide a llamarla por teléfono.</i> hablar por teléfono <i>Estas cosas no son para hablarlas por teléfono.</i> contestar al teléfono <i>Mi madre no contesta al teléfono desde hace un par de días.</i>
--	---	---

Fonte: Elaboração própria com base nos dados do EVP (2015), PCIC (2006), CaED2 (2011), Señas (2001), BNC (2007) e Corpes (2021)

Em consonância com as decisões tomadas no capítulo 6, manteve-se a decisão de usar o português para as glosas dos dicionários de inglês e o espanhol para as glosas do Señas (2001). Os colocados os colocados foram alocados sob as glosas de FLs e receberam destaque em negrito para facilitar a sua localização. Além disso, junto de cada colocação foi incluído um exemplo extraído do BNC (2007) e do Corpes (2021) para ilustrar a combinatória.

Em geral, não houve dificuldade em aplicar a versão adaptada das FLs sobre os verbetes do CaED2 (2011) e do Señas (2001). Apesar do aparente êxito da aplicação da proposta, é esperado que nem todas as FLs estandar previstas pela TST possam dar conta do valor semântico das colocações do inglês e do espanhol. Para esses casos, resta ao lexicógrafo propor FLs não estandar e, por conseguinte, novas glosas. Além disso, estamos cientes da necessidade de testar a TST, tanto em sua versão original em FLs como em sua versão adaptada em glosas, junto a um conjunto mais amplo de colocações. Trata-se de uma etapa posterior à investigação à nível teórico proposta na presente oportunidade, assim como a testagem das estratégias de descrição lexicográfica de colocações e as demais combinatórias tratadas na tese junto do público-alvo das propostas.

CONCLUSÕES

Estabelecer um perfil de usuário para um dicionário pode parecer uma obviedade ou uma tarefa simples, mas não é. Essa é uma das primeiras conclusões a que se pôde chegar após tratarmos de um aspecto que deveria ser considerado como fundamental ao se planejar uma obra lexicográfica. É a partir do perfil de usuário que deveria emanar parte significativa das decisões a respeito do que incluir em um dicionário, por que fazê-lo e como. Uma vez que o dicionário não é para o lexicógrafo, mas para um determinado público, esse público deveria ser considerado tão primordial quanto a descrição da língua em uso.

Dentre as diversas classes de obras lexicográficas, como os dicionários gerais de língua, de uso, de dúvidas, de sinônimos, de fraseologias, de imagens, entre tantos outros, o usuário almejado pelos dicionários pedagógicos é o que se mostra menos difuso porque a classe de obras em questão está destinada a um contexto específico: o ensino-aprendizagem de línguas. No caso das outras classes de dicionários citadas, o único contexto de uso que pode ser mais claramente delimitado é o âmbito laboral em que estão inseridos revisores de texto, tradutores e escritores. Em função de ser norteado por um currículo, o ensino-aprendizagem de línguas permite que se estabeleça de maneira mais precisa as características, o conhecimento prévio e as necessidades linguísticas dos estudantes inseridos nesse contexto. Essa máxima é ainda mais verdadeira quando se trata do ensino-aprendizagem de línguas estrangeiras, pois há uma considerável diversidade de fontes de dados sobre o público-alvo desse contexto: escalas de níveis de proficiência, sendo a do QECRL (2001) a mais difundida; documentos que orientam o ensino-aprendizagem, como o QECRL (2001), o PCIC (2006) e o EP (2011); e corpora de aprendizes, como o CAES (2020) e o CLC (2019).

Embora seja mais fácil estabelecer o perfil de usuário de um dicionário de aprendizes porque o contexto em que esse usuário está inserido é facilmente identificável, no capítulo 1, constatou-se que, na prática, essa facilidade não culmina na compilação de obras sob medida para um público específico. Essa afirmação é válida especialmente para os dicionários de aprendizes de inglês e de espanhol. Para que pudessem estar em conformidade com as orientações da Metalexigrafia no que diz respeito à inclusão de informações discriminantes (*i.e.* relevantes para o usuário) e à apresentação didática da descrição linguística (*i.e.* de uma maneira fácil de compreender), os dicionários de aprendizes de inglês e de espanhol teriam de estabelecer um perfil de usuário mais específico. Somente a partir de um público-alvo bem delimitado é que seria possível determinar quais informações são relevantes para esse público e qual seria a maneira mais didática de apresentá-las.

A falta de avanços na Lexicografia de Aprendizes no que diz respeito ao perfil de usuário se deve, em parte, à relativa estagnação da (Meta)Lexicografia quando se trata de investigar o tema em questão. Conforme o capítulo 2, as pesquisas empíricas sobre o usuário são esparsas e, no que concerne à teoria, a área pôde avançar até a especificação de um conjunto de variáveis sobre o comportamento do usuário de dicionários de aprendizes e que permitem defini-lo. Por um lado, é preciso considerar que a Metalexigrafia não possui subsídios teórico-metodológicos suficientes para dimensionar e aprofundar tais variáveis a ponto de convertê-las em estratégias de descrição lexicográfica condizentes com um perfil de usuário. Nesse sentido, não se pode exigir da área o que ela naturalmente não pode fornecer. Por outro lado, é preciso reconhecer que a Metalexigrafia também é responsável pela sua estagnação porque se detém nas suas limitações, com poucas iniciativas pela busca de respostas em outras áreas do conhecimento. A tomada de iniciativa rumo à interdisciplinaridade deveria ser um movimento natural da área, já que a Metalexigrafia é uma ciência aplicada, isto é, que se ampara em conhecimentos teórico-metodológicos de outras áreas da Linguística, tais como a Linguística de *Corpus* e a Semântica, para proveito próprio.

É nesse cenário que a tese se insere, com o objetivo de explorar e aplicar princípios teórico-metodológicos para um aprimoramento do desenho de dicionários de aprendizes de língua inglesa e de língua espanhola destinados aos níveis A2 e B1 de aprendizagem. Nossa proposta de aprimoramento requeria que as variáveis que definem o perfil de usuário do qual a tese se ocupa fossem destrinchadas. As áreas que elencamos como primordiais para essa tarefa foram a Linguística Aplicada, a Ciência Cognitiva e a Linguística Contrastiva (ou Learner Corpus Research). Paralelamente à dimensão do perfil de usuário, também constatamos que era necessário atentar também para a descrição de um aspecto da língua em uso cuja descrição ainda é vacilante nos dicionários de aprendizes: as combinatórias léxicas.

No tocante ao perfil de usuário, as principais variáveis envolvidas na sua definição são a proficiência já atingida na língua estrangeira, as necessidades de aprendizagem para que se atinja o próximo nível de proficiência e a língua materna.

As contribuições da Linguística Aplicada dizem respeito à determinação do conhecimento prévio na língua estrangeira e do conhecimento que constitui a etapa de aprendizagem na qual o aprendiz está inserido. Conforme tratado nos capítulos 3 e 5, os documentos EP (2011) e PCIC (2006) provaram ser úteis na avaliação dos definidores das paráfrases explanatórias e notas de uso, bem como na determinação do conjunto léxico mínimo que precisa constar na macroestrutura de um dicionário de aprendizes de níveis A2 e B1 de aprendizagem. Embora não respondam completamente à questão do léxico discriminante para

um dicionário de aprendizes, ambos os documentos têm um papel fundamental na aproximação do dicionário em relação ao ensino-aprendizagem.

Quanto às projeções futuras sobre o uso do EP (2011) e do PCIC (2006) como fonte de dados para a Lexicografia, é fato que ainda há muito o que se aproveitar de ambos. À medida que forem atualizados – em especial o PCIC (2006) – será possível ter uma noção ainda mais clara sobre o conhecimento prévio e as necessidades de aprendizagem dos usuários de dicionários de aprendizes das duas tradições.

A aproximação entre dicionário e ensino-aprendizagem ainda requeria que se compreendesse qual é o papel do dicionário nesse contexto, segundo salientado no capítulo 3. À luz da Linguística Aplicada, constatou-se que o dicionário é uma ferramenta mediadora do processo de aprendizagem, e não uma ferramenta para aprendizagem de línguas propriamente tal. Isso significa que se deve investir em facilitar a compreensão das informações das obras ao invés de utilizá-las como recurso para praticar habilidades de compreensão leitora, por exemplo. Embora o uso do dicionário naturalmente requeira a compreensão leitora, o principal objetivo da obra não pode estar no fornecimento de oportunidades de desenvolvimento dessa competência, mas sim na apresentação didática das informações para que estas sejam compreendidas da maneira mais fácil e rápida possível.

As contribuições da Ciência Cognitiva e da Linguística Contrastiva, por sua vez, lançam luz sobre a importância de se considerar a língua materna do usuário nos dicionários de aprendizes para que estes sejam mais didáticos. Por um lado, a língua materna do usuário se mostrou fundamental na determinação da língua a ser usada em função metalinguística em paráfrases explanatórias e notas de uso, conforme exposto nos capítulos 6 e 7. Com base nas oportunidades de ocorrência de influências interlinguísticas facilitadoras do processamento, concluiu-se que, para o usuário brasileiro, os dois segmentos informativos serão mais didáticos se estiverem redigidos em português, no caso dos dicionários de aprendizes de inglês A2-B1; e em espanhol, no caso de uma adaptação do Señas (2001) aos níveis A2 e B1.

Por outro lado, constatou-se que a língua materna do usuário é, entre outros fatores, uma importante variável no que diz respeito aos erros e dúvidas dos aprendizes. Os erros e dúvidas formam parte das características do usuário e contam com um espaço próprio nos dicionários de aprendizes, que é o das notas de uso. Os dicionários tomados como ponto de partida para as nossas investigações de fato incluíam notas de uso que tratavam de dúvidas e aspectos errôneos, mas, como o seu perfil de usuário é genérico, faltava determinar um conjunto de erros e dúvidas próprios dos brasileiros quando estes estudam inglês e espanhol. Dentre as áreas às quais se recorreu para as pesquisas da tese, a que pôde fornecer um diagnóstico mais preciso sobre esses

aspectos foi a Linguística Contrastiva. A área se ampara em conjuntos de dados sobre grupos de aprendizes de línguas, sendo os *corpora* de aprendizes as ferramentas mais recorrentemente empregadas. As pesquisas com esse tipo de ferramenta apresentam mais entraves se comparadas àquelas que fazem uso de *corpora* gerais de língua. Em primeiro lugar, além de estarem disponíveis em menor número, os *corpora* de aprendizes têm uma extensão menor que os gerais e ainda podem ser de acesso restrito, como é o caso do CLC (2019). Em segundo lugar, há uma alta demanda de recursos humanos, financeiros e de tempo que torna a compilação de *corpora* de aprendizes bastante dispendiosa. Em terceiro lugar, a distribuição dos dados por língua materna não é ampla. Nesse sentido, é preciso considerar que os pesquisadores brasileiros são afortunados pela existência de dados de aprendizes brasileiros de inglês no CLC (2019) e de espanhol no CAES (2020).

Por todas as razões comentadas acima, a concessão de acesso aos dados do CLC (2019) e a gratuidade de acesso ao CAES (2020) tiveram muito valor para as nossas pesquisas para além da relevância linguística dos dados em si. No capítulo 7, com base no apanhado teórico da Linguística Contrastiva, foram testadas metodologias de coleta de dados do CLC (2019) e foram investigadas ferramentas complementares ao CAES (2020), como o Fórum português-espanhol (WR, 2020) e o NDiFAEP (2014). Quanto ao CLC (2019), a falta de critérios estatísticos para avaliar o nível de relevância da frequência de um erro para fins lexicográficos foi um entrave na seleção de conteúdo. O estabelecimento de critérios estatísticos para essa tarefa é, sem dúvidas, necessário para que se fundamente a seleção dos dados em pesquisas futuras. Quanto às fontes de dados de erros e dúvidas em espanhol, a extensão limitada do CAES (2020) e a maneira confusa com que estão manifestadas algumas das dúvidas do Fórum Português-Espanhol (WR, 2020) são dois desafios para o pesquisador. Ainda assim, foi possível coletar uma amostra considerável de dados para simular as propostas de adaptação das notas de uso do Señas (2001).

Apesar das dificuldades particulares de cada base de dados, ainda há muito o que se investigar e aproveitar do CLC (2019), do CAES (2020) e do Fórum Português-Espanhol (WR, 2020). Além da necessária determinação de critérios estatísticos, pesquisas futuras a respeito da aplicação lexicográfica dos dados dessas ferramentas podem analisar outros tipos de ocorrências, como os casos de *overuse* (uso excessivo) e *underuse* (uso abaixo do normal) de ULs do inglês e do espanhol. Para tanto, a medida da frequência por milhão é um expediente útil para comparar dados de *corpora* de diferentes extensões.

De maneira paralela às contribuições de outras áreas, as propostas da tese a respeito das paráfrases explanatórias e das notas de uso também se ampararam nos aportes da Metalexigrafia.

No tocante às paráfrases explanatórias, as fórmulas propostas por Farias (2013) se mostraram úteis na adaptação do padrão de redação desse segmento informativo, tanto na sua versão original em português como na versão que propusemos ao espanhol. De forma complementar, a proposta de Longo (2001) e Zavaglia (2016) a respeito da inclusão de uma paráfrase sinonímica após a paráfrase principal de dicionários bilíngues se mostrou útil especialmente nos casos em que a paráfrase principal permanecia opaca após as adaptações. À medida que houver continuidade nas pesquisas sobre o embasamento semântico na elaboração de técnicas parafrásticas mais específicas para os substantivos, será possível aprimorar ainda mais a redação de paráfrases explanatórias para essa classe de palavras.

No que concerne às contribuições da Metalexigrafia a respeito das notas de uso, empregou-se o pouco que havia em termos de redação desse segmento informativo para propor parâmetros de redação e constituição do *layout*. Além do embasamento na prática lexicográfica, parte significativa das propostas se ampara nas contribuições de Bugueño Miranda (2019f) e de Bugueño Miranda e Borba (2020) a respeito dos tipos de orientação normativa em dicionários.

Um conceito de especial importância para as notas de uso e os verbetes em geral, mas que não pudemos desenvolver na tese, é o grau de saliência das informações. Investigar o princípio da saliência é preciso para que se determine o limite de uso de recursos visuais (ícones, ilustrações e quadros) e tipográficos em um verbete, evitando sobrecarregá-lo ou mesmo anular a capacidade de localização de cada tipo de informação. Pesquisas futuras podem recorrer à área da Psicologia Cognitiva e do Design Visual para lançar luz a essa questão, particularmente por meio do Efeito de Saliência [Efeito Von Restorff], da Teoria do Código Dual [Dual-code Theory] e do Princípio da Região Comum [Law of Common Region].

Finalmente, além do perfil de usuário, outro objeto de interesse da tese foram as combinatórias léxicas. Trata-se de um tema amplo e com perspectivas em parte discordantes, mas que, na verdade, podem ser consideradas como complementares sob a ótica da Metalexigrafia. A importância de se descrever as combinatórias léxicas reside no fato de que são um fenômeno tão comum nas línguas que seria impossível associá-las a um ou outro nível de proficiência, pois são pertinentes a todos os níveis. Além disso, a arbitrariedade com que estão constituídas e as diferenças entre as línguas na seleção dos componentes de uma combinação (ou mesmo da combinação inteira) tornam-nas fatos linguísticos dos quais um dicionário de aprendizes não pode prescindir, mesmo em se tratando de níveis de aprendizagem

A2 e B1, ou ainda A1 e A2. No capítulo 8, foi possível sintetizar as principais propostas teóricas sobre o particular, identificar os tipos de combinatórias e determinar, a partir da sua definição conceitual, parâmetros para a sua identificação e descrição lexicográfica. Para dar continuidade aos estudos com base nas abordagens comentadas na tese, faz-se necessário testar a aplicação da descrição das combinatórias sobre uma amostra mais extensa de itens. A etapa posterior a esta é a de testagem do didatismo das propostas junto ao público-alvo.

Todas as propostas realizadas no decorrer da tese estão centradas no fornecimento de informações discriminantes e didáticas nos dicionários de aprendizes em geral, com especial atenção aos que se ocupam da descrição do inglês e do espanhol. O próximo passo das investigações, para além das projeções futuras comentadas a respeito de cada área, é a testagem da efetividade das propostas junto ao público-alvo delimitado no capítulo 2. Em um futuro ainda mais longínquo, havendo a oportunidade de se proceder com a adaptação de qualquer uma das obras tratadas, é interessante avaliar qual suporte, entre impresso e eletrônico, seria mais adequado para a veiculação do dicionário e, no caso do suporte eletrônico, para que tipo de dispositivo. Nesse sentido, dados de pesquisas como as empreendidas pelo Conselho Gestor da Internet no Brasil indicam que, em se tratando do meio eletrônico, o principal dispositivo que os brasileiros usaram em 2020 para acessar a internet (quando têm acesso a ela, naturalmente) foi o *smartphone* (CETIC.BR; NIC.BR, 2021a, p. 04). Segundo o órgão, 99% da população com dez anos ou mais que tem acesso à internet o faz, em alguma medida, pelo *smartphone*; e para 58% desse grupo de brasileiros, o acesso à internet ocorre exclusivamente pelo dispositivo em questão. Esse tipo de informação é de extrema relevância no momento de se decidir qual suporte tem maior alcance em uma dada população.

Um comentário final que é preciso fazer diz respeito ao tema das ilustrações. Inicialmente, o tratamento lexicográfico das ilustrações era um dos temas do nosso projeto de tese de doutorado. Chamou-nos a atenção o fato de que tanto os dicionários de aprendizes de inglês como o Señas (2001) fazem uso desse expediente em seus verbetes e/ou em outro componente do dicionário (*middle matter* e *back matter*), mas sem que houvesse muitos dados na teoria acerca de parâmetros de seleção e inclusão de ilustrações. No âmbito das pesquisas interdisciplinares, havíamos proposto a investigar teorias no âmbito da Psicologia Cognitiva, precisamente sobre a percepção visual (*bottom-up* e *top-down*) e a percepção de objetos (holística e por partes). Contudo, as pesquisas a respeito do perfil de usuário tomaram uma dimensão e tempo maiores do que o previsto inicialmente, em especial no que concerne ao tratamento das notas de uso e à inesperada possibilidade de acesso ao CLC (2019). Apesar das mudanças no direcionamento da tese, investigações sobre o estabelecimento de parâmetros para

a inclusão de ilustrações constituem mais uma das perspectivas futuras de desenvolvimento das pesquisas no tocante ao aprimoramento dos dicionários de aprendizes em geral, e de inglês e espanhol em particular.

REFERÊNCIAS BIBLIOGRÁFICAS

DICIONÁRIOS DE LÍNGUA, GRAMÁTICAS E OUTRAS OBRAS DE REFERÊNCIA LINGUÍSTICA

CaED2. MCINTOSH, Colin (senior editor). **Cambridge Essential English Dictionary**. 2nd ed. Cambridge: Cambridge University Press, 2011.

CALD4. CAMBRIDGE. **Cambridge Advanced Learner's Dictionary**. 4th ed. Cambridge: Cambridge University Press, 2013.

CCAD6. HAPERCOLLINS. **Collins Cobuild Advanced Dictionary**. Glasgow: HarperCollins, 2009.

CCAD9. HAPERCOLLINS. **Collins Cobuild Advanced Dictionary**. Glasgow: HarperCollins, 2018.

CCID4. HANDS, Penny; MACAULAY, Alison (senior editors). **Collins Cobuild Intermediate Learner's Dictionary**. Glasgow: HarperCollins, 2018.

CoED2. COMBLEY, Rosalind; WEDGEWORTH, Laura; WOODFORD, Kate. **Cobuild Essential English Dictionary**. For elementary and pre-intermediate learners of English. 2nd ed. Glasgow: HarperCollins, 2014.

DAVIES, Mark; PRETO-BAY, Ana Maria R.A frequency dictionary of Portuguese. Core vocabulary for learners. New York; London: Routledge, 2008.

DComP. BOSQUE, Ignacio (Dir.). **Diccionario Combinatorio Práctico del Español Contemporáneo**. Madrid: SM, 2011.

DiCE. ALONSO RAMOS, Margarita. **Diccionario de Colocaciones del Español**. Disponível em: <http://www.dicesp.com/paginas>. Acesso em: 13 set. 2021.

DiELE. ALVAR EZQUERRA, Manuel (ed.). **Diccionario Vox para la enseñanza de la lengua española**. Barcelona: Universidad Alcalá de Henares / Bibliograf, 1995.

DiSal. GUTIÉRREZ CUADRADO, Juan. (dir.). **Diccionario Salamanca de la Lengua Española**. Madrid: Santillana, 2006.

DRAEe. REAL ACADEMIA ESPAÑOLA. **Diccionario de la Real Academia Española**. 23^a edición. Madrid: Espasa, 2014. Disponível em: <http://dle.rae.es/?id=DgIqVCc>. Acesso em: 30 abr. 2017.

DUPC. BORBA, Francisco S. (Org.). **Dicionário UNESP do Português Contemporâneo**. São Paulo: UNESP, 2004.

Espasa-Calpe. **Diccionario de la Lengua Española para Estudiantes de Español**. Madrid: Espasa-Calpe, 2002.

FredIt. JUILLAND, Alphonse; TRAVERSA, Vincenzo. **Frequency dictionary of Italian words**. Berlin/Boston: De Gruyter Mouton, 1973.

GDLE. BOSQUE, Ignacio; DEMONTE, Violeta (dir.). **Gramática Descriptiva de la Lengua Española**. Madrid: Espasa-Calpe, 1999.

GramLE. REAL ACADEMIA ESPAÑOLA; ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA. **Nueva gramática de la lengua española**. Madrid: Espasa, 2009.

Hou. HOUAISS, Antônio. **Novo dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2009

LAS5. PEARSON. **Longman Active Study Dictionary**. 5th edition. Harlow: Pearson Education, 2010.

LDOCE6. PEARSON. **Longman Dictionary of Contemporary English**. 6th edition. Harlow: Pearson Education, 2014. Disponível em: <https://www.ldoceonline.com/>. Acesso em: 29 out. 2019.

LEstilo. GARCÍA DE LA CONCHA, Víctor (coord.). **Libro de Estilo de la Lengua Española**. Madrid: Espasa, 2018.

LWord2. MAYOR, Michael (editorial director). **Longman Wordwise Dictionary**. For pre-intermediate – intermediate learners. 2nd ed. Harlow: Pearson, 2008.

LWord2e. MAYOR, Michael (editorial diretor). **Longman Wordwise Dictionary**. For pre-intermediate – intermediate learners. 2nd ed. Harlow: Pearson, 2008. CD-ROM.

MEDAL2. RUNDELL, Michael. (Ed.). **Macmillan English Dictionary for Advanced Learners**. 2nd edition. London: Macmillan, 2007. Disponível em: <https://www.macmillandictionary.com/>. Acesso em: 05 out. 2021.

MGramLE. REAL ACADEMIA ESPAÑOLA; ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA. **Nueva gramática de la lengua española. Manual**. Madrid: Espasa, 2010.

MPLP. BIDERMAN, Maria Tereza Camargo; CARVALHO, Carmen Silvia. **Meu primeiro livro de palavras**. São Paulo: Ática, 2010.

MWALE. MERRIAM-WEBSTER. **Merriam-Webster's Advanced Learner's English Dictionary**. Springfield (MA): Merriam-Webster, Inc., 2008.

NDiFAEP. BUGUEÑO MIRANDA, Félix Valentín (dir.). **Novo Dicionário de Falsos Amigos Espanhol/Português**. Florianópolis: Bookess, 2014.

OALD8. TURNBULL, Joanna. (Ed.). **Oxford Advanced Learner's Dictionary**. 8th edition. Oxford: Oxford University Press, 2010.

OALD8e. TURNBULL, Joanna. (Ed.). **Oxford Advanced Learner's Dictionary**. 8th edition. Oxford: Oxford University Press, 2010. CD-ROM.

ODE. STEVENSON, Angus (Ed.). **Oxford Dictionary of English**. 3rd edition. Oxford: Oxford University Press, 2010.

OED. SIMPSON, John; WEINER, Edmund (Eds.). **Oxford English Dictionary**. Oxford: Oxford University Press, 2021. Disponível em: <https://oed.com/>. Acesso em: 16 abr. 2021.

OxColl. MCINTOSH, Colin; FRANCIS, Ben; POOLE, Richard. **Oxford Collocations Dictionary: For Students of English**. Oxford: Oxford University Press, 2009.

OxED2. WATERS, Alison; BULL, Victoria. **Oxford Essential Dictionary**. For elementary and pre-intermediate learners of English. 2nd ed. Oxford: Oxford University Press, 2012.

OxED2e. WATERS, Alison; BULL, Victoria. **Oxford Essential Dictionary**. For elementary and pre-intermediate learners of English. 2nd ed. Oxford: Oxford University Press, 2012. CD-ROM.

OxThes. LEA, Diana; BRADBURY, Jennifer; POOLE, Richard; WARREN, Helen. **Oxford Learner's Thesaurus: A dictionary of synonyms**. Oxford: Oxford University Press, 2008.

Redes. BOSQUE, Ignacio (Dir.). **Redes - Diccionario Combinatorio del Español Contemporáneo**. Madrid: SM, 2004.

Sac. SACCONI, Luiz Antonio. **Grande Dicionário Sacconi da língua portuguesa**. Comentado, crítico e enciclopédico. São Paulo: Nova geração, 2010.

Señas. UNIVERSIDAD ALCALÁ DE HENARES. **Señas**. Diccionario para la Enseñanza de la Lengua Española para Brasileños. São Paulo: Martins Fontes, 2001.

CORPORA, LISTAS DE PALAVRAS E OUTRAS FERRAMENTAS ELETRÔNICAS DE LINGUAGEM

BNC. OXFORD UNIVERSITY PRESS. **British National Corpus**. 2007. Disponível em: <http://corpus.byu.edu/bnc/>. Acesso em: 25 fev. 2017.

Br-ICLE. SARDINHA, T. B. **The Brazilian Portuguese Sub-corpus of ICLE**. São Paulo: Pontifícia Universidade Católica de São Paulo, 2002. Disponível em: <http://www2.lael.pucsp.br/corpora/bricle/>. Acesso em: 01 fev. 2017.

CAES. INSTITUTO CERVANTES. **Corpus de Aprendices de Español**. Versión 1.3. [S.l.]: Instituto Cervantes, 2020. Disponível em: <http://galvan.usc.es/caes>. Acesso em: 14 mai. 2020.

CAMBRIDGE ASSESSMENT. **Key and Key for Schools Vocabulary List**. From 2020. A2 Key; A2 Key for schools. Cambridge: University of Cambridge Local Examinations Syndicate, 2018a. Disponível em: <https://www.cambridgeenglish.org/Images/506886-a2-key-2020-vocabulary-list.pdf>. Acesso em: 19 set. 2019.

CAMBRIDGE ASSESSMENT. **Preliminary and Preliminary for Schools Vocabulary List**. From 2020. B1 Preliminary; B1 Preliminary for schools. Cambridge: University of Cambridge Local Examinations Syndicate, 2018b. Disponível em:

<https://www.cambridgeenglish.org/images/506887-b1-preliminary-2020-vocabulary-list.pdf>. Acesso em: 19 set. 2019.

CLC. CAMBRIDGE UNIVERSITY PRESS; CAMBRIDGE ENGLISH LANGUAGE ASSESSMENT. **Cambridge Learner Corpus**. Coded version. Brighton: Lexical Computing Limited, 2019. Disponível em: <https://cup.sketchengine.co.uk/rasp/#login?next=https%3A%2F%2Fcup.sketchengine.co.uk%2F>. Acesso em: 21 dez. 2020.

COCA. DAVIES, Mark. **Corpus of Contemporary American English**. 2020. Disponível em: <https://www.english-corpora.org/coca/>. Acesso em: 31 ago. 2021.

Corpes. REAL ACADEMIA ESPAÑOLA. **Corpus del Español del Siglo XXI**. Versão 0.93. 2021. Disponível em: <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>. Acesso em: 25 abr. 2021.

CP. DAVIES, Mark. **Corpus do Português**. News on the Web Corpus. 2012-2019. 2019. Disponível em: <https://www.corpusdoportugues.org/now/>. Acesso em: 18 jul. 2020.

EGP. CAMBRIDGE. **English Grammar Profile**. Cambridge: Cambridge University Press, 2015. Disponível em: <https://www.englishprofile.org/english-grammar-profile/egp-online>. Acesso em: 27 out. 2019.

EVP. CAMBRIDGE. **English Vocabulary Profile**. Cambridge: Cambridge University Press, 2015. Disponível em: <https://www.englishprofile.org/wordlists/evp>. Acesso em: 27 out. 2019.

ICLE. GRANGER, Sylviane. **International Corpus of Learner English**. Louvain-la-Neuve: Université Catholique de Louvain, 2002-2020. Disponível em: <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/icle.htm>. Acesso em: 14 abr. 2019.

NIM. GUASCH, Marc, BOADA, Roger, FERRÉ, Pilar, & SÁNCHEZ-CASAS, Rosa. NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. **Behavior Research Methods**, v. 45, n. 3, 765-771, 2013. Disponível em: <https://psico.fcep.urv.cat/utilitats/nim/index.php>. Acesso em: 28 set. 2020.

OXFORD. **Oxford 3000**. Oxford: Oxford University Press, 2019a. Disponível em: <https://www.oxfordlearnersdictionaries.com/wordlists/oxford3000-5000>. Acesso em: 27 out. 2019.

OXFORD. **Oxford Phrase List**. Oxford: Oxford University Press, 2019b. Disponível em: <https://www.oxfordlearnersdictionaries.com/wordlists/oxford-phrase-list>. Acesso em: 27 out. 2019.

RAE. REAL ACADEMIA ESPAÑOLA. Conozca (algo más) el CORPES: listados de frecuencias. **Real Academia Española**, 30 jun. 2021. Disponível em: <https://www.rae.es/noticia/conozca-algo-mas-el-corpes-listados-de-frecuencias>. Acesso em: 3 ago. 2021.

WR. KELLOGG, Michael. **WordReference Language Forums: Português-Español**. Weston: WordReference, 2020. Disponível em: <https://forum.wordreference.com/forums/portugu%C3%AAs-espa%C3%B1ol.55/>. Acesso em: 21 dez. 2020.

QUADRO EUROPEU COMUM DE REFERÊNCIA E DOCUMENTOS RELACIONADOS

BENSON, Phil. **Teaching and Researching Autonomy in Language Learning**. Harlow: Longman, 2001.

Breakthrough. TRIM, John L. M. **Breakthrough**. s/l, s/d (manuscrito não publicado oficialmente). Disponível em: <http://www.englishprofile.org/the-cefr/further-reading>. Acesso em: 20 mar. 2018.

CEFR. COUNCIL OF EUROPE. **Common European Framework of Reference for Languages: learning, teaching, assessment**. Cambridge: Cambridge University Press, 2001. Disponível em: <https://rm.coe.int/1680459f97>. Acesso em: 18 out. 2019.

CEFR-Co. COUNCIL OF EUROPE. **Common European Framework of Reference for Languages: learning, teaching, assessment. Companion Volume with New Descriptors**. Council of Europe, 2018. Disponível em: <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>. Acesso em: 18 out. 2019.

EP. UNIVERSITY OF CAMBRIDGE LOCAL EXAMINATIONS SYNDICATE. **English Profile**. Introducing the CEFR for English. Version 1.1. Cambridge: Cambridge University Press, 2011. Disponível em: <http://www.englishprofile.org/images/pdf/theenglishprofilebooklet.pdf>. Acesso em: 11 dez. 2017.

PCIC. INSTITUTO CERVANTES. **Plan Curricular del Instituto Cervantes**. 3 vol. Madrid: Biblioteca Nueva, 2006. Disponível em: https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/. Acesso em: 18 jul. 2020.

QECL. CONSELHO DA EUROPA. **Quadro Europeu Comum de Referência para as línguas: aprendizagem, ensino, avaliação**. Portugal: Edições ASA, 2001. Disponível em: <http://www.dge.mec.pt/quadro-europeu-comum-de-referencia-para-linguas>. Acesso em: 09 fev. 2016.

Threshold 1990. VAN ECK, Jan A.; TRIM, John L. M. **Threshold**. Cambridge: CUP, 1998.

Vantage 1990. VAN ECK, Jan A.; TRIM, John L. M. **Vantage**. Cambridge: CUP, 2001.

Waystage 1990. VAN ECK, Jan A.; TRIM, John L. M. **Waystage**. Cambridge: CUP, 1998.

DEMAIS REFERÊNCIAS

ALDERSON, J. Charles; FIGUERAS, Neus; KUIJPER, Henk; NOLD, Guenter; TAKALA, Sauli; TARDIEU, Claire. Analysing tests of Reading and listening in relation to the Common

European Framework of Reference: the experience of the Dutch CEFR Construct Project. **Language Assessment Quarterly**, [s.l.], v. 3, n. 1, p. 3-30, 2005.

ALENCAR, José de. **Senhora**. Jaraguá do Sul: Gráfica a Editora Avenida, 2005.

ALONSO RAMOS, Margarita. Glosas para las colocaciones en el Diccionario de Colocaciones del Español. In: ALONSO RAMOS, Margarita (coord.). **Diccionarios y fraseología**. Corunha: Universidade da Coruña, 2006. p. 59-88.

ANDERSEN, Birger Enevold; NIELSEN, Sandro. Ten Key Issues in Lexicography for the Future. In: BERGENHOLTZ, Henning; NIELSEN, Sandro; TARP, Sven (Eds.). **Lexicography at a Crossroads**. Dictionaries and Encyclopedias today, lexicographical tools tomorrow. Bern: Peter Lang, 2009. p. 355-365.

BARTSCH, Sabine. **Structural and Functional Properties of Collocations in English**. A corpus study of lexical and pragmatic constraints on lexical co-occurrence. Tübingen: Gunter Narr Verlag Tübingen, 2004.

BLACKBURN, Simon. **Dicionário Oxford de Filosofia**. Rio de Janeiro: Jorge Zahar Editor, 1997.

BLANCO, Cindy. 2021 Duolingo Language Report. **Duolingo Blog**, [s.l.], 06 dez. 2021. Disponível em: <https://blog.duolingo.com/2021-duolingo-language-report/>. Acesso em: 02 fev. 2022.

BOGAARDS, Paul. Uses and users of dictionaries. In: STERKENBURG, Piet van (Org.). **A Practical Guide to Lexicography**. Amsterdam: John Benjamins, 2003, p. 26–33.

BOGULSKI, Cari A.; BICE, Kinsey; KROLL, Judith. Bilingualism as a desirable difficulty: Advantages in word learning depend on regulation of the dominant language. **Bilingualism: Language and Cognition**, Cambridge, v. 22, n. 5, p. 1052-1067, 2019. DOI: <https://doi.org/10.1017/S1366728918000858>.

BOLZAN, Rosane Maria; DURÃO, Adja B. A. B. O trabalho com dicionários em sala de aula: relato de uma contribuição para a formação docente em lexicografia. **Moara**, Belém (PA), n. 36, p. 181-196, 2011. Disponível em: <http://dx.doi.org/10.18542/moara.v1i36.1110>. Acesso em: 03 out. 2019.

BORBA, Laura Campos de. Emprego de dicionários monolíngues de espanhol no ensino-aprendizagem de espanhol como língua estrangeira: níveis A1 e A2. In: 27ª SEMANA DE LETRAS – Shakespeare e Cervantes 400 anos depois, 2016, Caxias do Sul. **Anais da 27ª Semana de Letras. Shakespeare e Cervantes 400 anos depois**. Caxias do Sul: Universidade de Caxias do Sul, 2016. p. 323-336. Disponível em: <https://www.ucs.br/site/midia/arquivos/anais-27-semana-letras.pdf>. Acesso em: 27 nov. 2017.

BORBA, Laura Campos de. **Panorama da Lexicografia Hispânica**: subsídios para o professor de ELE. Saarbrücken: NEA - Novas Edições Acadêmicas, 2017a. 116p.

BORBA, Laura Campos de. **Desenvolvendo a autonomia no ensino-aprendizagem de espanhol como língua estrangeira**: o auxílio dos dicionários monolíngues de espanhol. 2017b. 163 f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras,

Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <https://lume.ufrgs.br/handle/10183/169039>. Acesso em: 18 jul. 2018.

BORBA, Laura Campos de. En búsqueda de elementos definitorios para el término competencia comunicativa. **Entrehojas**, Ontario (CAN), v. 7, n. 1, p. 01-21, 2017c. Disponível em: <https://ojs.lib.uwo.ca/index.php/entrehojas/article/view/6173>. Acesso em: 18 out. 2019.

BORBA, Laura Campos de. Princípios de uma fundamentação teórica para a inclusão de ilustrações em dicionários infantis à luz da Psicologia Cognitiva. In: XIII Semana de Extensão, Pesquisa e Pós-Graduação UNIRITTER (SEPesq), 2017, Porto Alegre. **Atas da XIII SEPesq UNIRITTER**. Porto Alegre: Uniritter, 2018a. p. 1-12.

BORBA, Laura Campos de. A Teoria Sentido-Texto e suas possibilidades de aplicação em dicionários de aprendizes do espanhol. **Revista da Anpoll**, Florianópolis, v. 1, n. 45, p. 156-172, 2018b. Disponível em: <https://doi.org/10.18309/anp.v1i45.1118>. Acesso em: 01 jul. 2019.

BORBA, Laura Campos de. Capacidades e limitações dos dicionários de aprendizes de espanhol como língua estrangeira. In: SOUSA, I. V. **Letras, Linguística e Artes: perspectivas críticas e teóricas**. Ponta Grossa: Atena Editora, 2019a. p. 220-235. Disponível em: <https://www.atenaeditora.com.br/arquivos/ebooks/letras-linguistica-e-artes-perspectivas-criticas-e-teoricas>. Acesso em: 19 jun. 2019.

BORBA, Laura Campos de. Panorama da Lexicografia Hispânica. In: BUGUEÑO MIRANDA, F. V.; BORBA, L. C. (orgs.). **Manual de (Meta)Lexicografia**. Goiânia: Editora Espaço Acadêmico, 2019b. Capítulo 8, p. 83-98. Disponível em: <http://www.ufrgs.br/metalexigrafia/manual-de-meta-lexicografia>. Acesso em: 14 mai. 2020.

BORBA, Laura Campos de. Produção textual no ensino-aprendizagem de espanhol como língua estrangeira: a contribuição dos dicionários de sinônimos. In: SOUSA, I. V. (Org.). **Letras, Linguísticas e Artes: perspectivas críticas e teóricas 2**. Ponta Grossa: Atena Editora, 2019c. p. 290-304. Disponível em: <https://www.atenaeditora.com.br/arquivos/ebooks/letras-linguistica-e-artes-perspectivas-criticas-e-teoricas-2>. Acesso em: 06 jan. 2021.

BORBA, Laura Campos de. Collins Cobuild Intermediate Learner's Dictionary: aliando o rigor da descrição linguística às funções de compreensão e produção. **Caletrosópio**, Ouro Preto, v. 8, n. 2, p. 290-297, 2020a. Disponível em: <https://periodicos.ufop.br:8082/pp/index.php/caletrosopio/article/view/4499>. Acesso em: 04 jan. 2020.

BORBA, Laura Campos de. Improving learner's dictionaries: a discussion based on an interdisciplinary approach. **Fórum Linguístico**, Florianópolis, v. 17, n. 3, p. 5146-5166, 2020b. Disponível em: <https://doi.org/10.5007/1984-8412.2020.70844>. Acesso em: 14 set. 2021.

BORBA, Laura Campos de; BUGUEÑO MIRANDA, Félix Valentín. Análise de cinco dicionários semasiológicos de língua espanhola: a correlação entre o *Front Matter* e a Macro e Microestrutura. **Extensio**, Florianópolis, v. 9, n. 4, p. 32-43, 2012. Disponível em: <https://doi.org/10.5007/1807-0221.2012v9n14p32>. Acesso em: 04 mai. 2019.

BORBA, Laura Campos de; BUGUEÑO MIRANDA, Félix Valentín. Verdade e mito dos dicionários de aprendizes do inglês e do alemão. **Domínios de Linguagem**, Uberlândia, v. 12, n. 4, p. 2165-2203, 2018. Disponível em: <https://doi.org/10.14393/DL36-v12n4a2018-10>. Acesso em: 04 mai. 2019.

BORBA, Laura Campos de; BUGUEÑO MIRANDA, Félix Valentín. Lexicografia pedagógica: existe? In: BUGUEÑO MIRANDA, F. V.; BORBA, L. C. (orgs.). **Manual de (Meta)Lexicografia**. Goiânia: Editora Espaço Acadêmico, 2019a. Capítulo 16, p. 140-144. Disponível em: <http://www.ufrgs.br/metalexigrafia/manual-de-meta-lexicografia>. Acesso em: 08 out. 2019.

BORBA, Laura Campos de; BUGUEÑO MIRANDA, Félix Valentín. Passado, presente e futuro dos dicionários de aprendizes. **Domínios de Linguagem**, Uberlândia, v. 13 n. 3, p. 1018-1040, 2019b. Disponível em: <https://doi.org/10.14393/DL39-v13n3a2019-9>. Acesso em: 08 out. 2019.

BOSQUE, Ignacio. Sobre el concepto de colocación y sus límites. **Lingüística Española Actual**, Madrid, v. 23, n. 1, p. 9-40, 2001. Disponível em: <https://sites.google.com/site/ignaciobosquemunoz/publicaciones-y-presentaciones/2001>. Acesso em: 24 nov. 2021.

BRIGGS, Asa; BURKE, Peter. **Uma história social da mídia: de Gutenberg à Internet**. 3ª edição. Rio de Janeiro: Zahar, 2016.

British Council. **Demandas de Aprendizagem de Inglês no Brasil**. Elaborado com exclusividade para o British Council pelo Instituto de Pesquisa Data Popular. São Paulo: British Council, 2014. Disponível em: <https://www.britishcouncil.org.br/pesquisas-infograficos>. Acesso em: 28 jun. 2018.

British Council. **O Ensino de Inglês na Educação Pública Brasileira**. Elaborado com exclusividade para o British Council pelo Instituto de Pesquisas Plano CDE. São Paulo: British Council, 2015. Disponível em: <https://www.britishcouncil.org.br/pesquisas-infograficos>. Acesso em: 28 jun. 2018.

BRYBAERT, Marc; DUYCK, Wouter. Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? **Bilingualism: Language and Cognition**, Cambridge, v. 13, n. 3, p. 359-371, 2010. DOI: <https://doi.org/10.1017/S1366728909990344>.

BUGUEÑO MIRANDA, Félix Valentín. Falsos amigos, falsos cognatos, heterossemânticos: uma simples escolha de designações? **Organon**, Porto Alegre, v. 16, n. 32-33, p. 183-192, 2002. Disponível em: <https://doi.org/10.22456/2238-8915.29793>. Acesso em: 13 nov. 2020.

BUGUEÑO MIRANDA, Félix Valentín. O que o professor deve saber sobre a nominata do dicionário de língua. **Revista Língua e Literatura**, Frederico Westphalen, v. 6/7, n. 10/11, p. 17-31, 2004/2005. Disponível em: <http://revistas.fw.uri.br/index.php/revistalinguaeliteratura/article/view/36>. Acesso em: 04 mai. 2019.

BUGUEÑO MIRANDA, Félix Valentín. Léxico e ensino: Señas (2000), um dicionário para aprendizes do espanhol? *In*: MARTINS, E.; CANO, W.; MORAES FILHO, W. (org.). **Léxico e morfofonologia**: perspectivas e análise. Uberlândia: EDUFU, 2006, p. 216-232.

BUGUEÑO MIRANDA, Félix Valentín. O dicionário como reflexo da língua. **Expressão**, Santa Maria (UFSM), v. 11, n. 1, p. 97-105, 2007a.

BUGUEÑO MIRANDA, Félix Valentín. O que é macroestrutura no dicionário de língua? *In*: ISQUERDO, A. N.; ALVES, I. M. (org.). **As ciências do léxico**: lexicologia, lexicografia e terminologia. São Paulo: Humanitas, 2007b. v. III, p. 261-272.

BUGUEÑO MIRANDA, Félix Valentín. Sobre a microestrutura em dicionários semasiológicos do alemão. **Contingentia**, Porto Alegre, v. 4, n. 2, p. 60-72, 2009a. Disponível em: <https://seer.ufrgs.br/contingentia/article/view/11414>. Acesso em: 04 mai. 2019.

BUGUEÑO MIRANDA, Félix Valentín. Para uma taxonomia de paráfrases explanatórias. **Alfa**, São Paulo, v. 53, n. 1, p. 243-260, 2009b. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/1686>. Acesso em: 04 mai. 2019.

BUGUEÑO MIRANDA, Félix Valentín. O dicionário bilíngue como problema linguístico e lexicográfico. *In*: HWANG, A.; NADIN, O. **Linguagens em interação III**: estudos do léxico. Maringá: Clichetec, 2010. p. 65-91.

BUGUEÑO MIRANDA, Félix Valentín. Balanço e perspectivas da lexicografia. **Cadernos de Tradução**, Florianópolis, (UFSC), v. 32/2, p. 15-37, 2013.

BUGUEÑO MIRANDA, Félix Valentín. A estruturação de um dicionário. *In*: BUGUEÑO MIRANDA, F. V.; BORBA, L. C. (orgs.). **Manual de (Meta)Lexicografia**. Goiânia: Editora Espaço Acadêmico, 2019a. Capítulo 3, p. 14-29. Disponível em: <http://www.ufrgs.br/metalexigrafia/manual-de-meta-lexicografia>. Acesso em: 14 mai. 2020.

BUGUEÑO MIRANDA, Félix Valentín. Sobre a classificação de obras lexicográficas. *In*: BUGUEÑO MIRANDA, F. V.; BORBA, L. C. (orgs.). **Manual de (Meta)Lexicografia**. Goiânia: Editora Espaço Acadêmico, 2019b. Capítulo 4, p. 30-36. Disponível em: <http://www.ufrgs.br/metalexigrafia/manual-de-meta-lexicografia>. Acesso em: 14 mai. 2020.

BUGUEÑO MIRANDA, Félix Valentín. Esboço de um panorama da lexicografia em língua inglesa. Com especial atenção ao estudante de inglês como língua estrangeira. *In*: BUGUEÑO MIRANDA, F. V.; BORBA, L. C. (orgs.). **Manual de (Meta)Lexicografia**. Goiânia: Editora Espaço Acadêmico, 2019c. Capítulo 7, p. 69-77. Disponível em: <http://www.ufrgs.br/metalexigrafia/manual-de-meta-lexicografia>. Acesso em: 14 mai. 2020.

BUGUEÑO MIRANDA, Félix Valentín. Princípios teóricos básicos do Manual. *In*: BUGUEÑO MIRANDA, F. V.; BORBA, L. C. (orgs.). **Manual de (Meta)Lexicografia**. Goiânia: Editora Espaço Acadêmico, 2019d. Capítulo 2, p. 11-12. Disponível em: <http://www.ufrgs.br/metalexigrafia/manual-de-meta-lexicografia>. Acesso em: 14 mai. 2020.

BUGUEÑO MIRANDA, Félix Valentín. Panorama da Lexicografia Brasileira. *In*: BUGUEÑO MIRANDA, F. V.; BORBA, L. C. (orgs.). **Manual de (Meta)Lexicografia**.

Goiânia: Editora Espaço Acadêmico, 2019e. Capítulo 9, p. 94-105. Disponível em: <http://www.ufrgs.br/metalexicografia/manual-de-meta-lexicografia>. Acesso em: 14 mai. 2020.

BUGUEÑO MIRANDA, Félix Valentín. El vocabulario de la orientación en el uso de la lengua en los diccionarios de lingüística. **Nueva Revista del Pacífico**, Valparaíso, n. 71, p. 50-79, 2019f. Disponível em: <http://www.nuevarevistadelpacifico.cl/index.php/NRP/article/view/152>. Acesso em: 04 dez. 2020.

BUGUEÑO MIRANDA, Félix Valentín; BORBA, Laura Campos de. As combinatórias léxicas e o ensino da língua espanhola: o quanto ajudam os dicionários bilíngues português-espanhol? **Hispanic Research Journal**, Londres, v. 20, n. 03, p. 214-238, 2019. Disponível em: <https://doi.org/10.1080/14682737.2019.1651572>. Acesso em: 05 out. 2019.

BUGUEÑO MIRANDA, Félix Valentín; BORBA, Laura Campos de. Orientación normativa en diccionarios monolingües de inglés como lengua extranjera. **Colombian Applied Linguistics Journal**, Bogotá, v. 22, n. 2, p. 157-168, 2020. Disponível em: <https://doi.org/10.14483/22487085.15528>. Acesso em: 30 abr. 2021.

BUGUEÑO MIRANDA, Félix Valentín; FARIAS, Virginia Sita. Informações discretas e discriminantes no artigo léxico. **Cadernos de Tradução**, Florianópolis, v. 2, n. 18, p. 115-135, 2006. Disponível em: <https://periodicos.ufsc.br/index.php/traducao/article/view/6944>. Acesso em: 07 mai. 2019.

BUGUEÑO MIRANDA, Félix Valentín; FARIAS, Virginia Sita. Proposta de um modelo de avaliação de dicionários escolares de Língua Portuguesa. In: SIMPÓSIO NACIONAL DE LETRAS E LINGÜÍSTICA, SIMPÓSIO INTERNACIONAL DE LETRAS E LINGÜÍSTICA, 2013, Uberlândia. **Anais do SILEL**. v. 3, n. 1. Uberlândia: UFU, 2018. p. 1-20. Disponível em: http://www.ileel.ufu.br/anaisdosilel/wp-content/uploads/2014/04/silel2013_1100.pdf. Acesso em: 19 jun. 2019.

BUGUEÑO MIRANDA, Félix Valentín; ZANATTA, Flávia. Procedimentos medioestruturais em dicionários semasiológicos de língua portuguesa. **Lusorama**, Frankfurt am Main, n. 83-84, p. 80-97, 2010.

CAN, Cem. A learner corpus-based study on verb errors of Turkish EFL learners. **Journal of Education and Training Studies**, Beaverton, v. 5, n. 9, p. 167-175, 2017. Disponível em: <https://doi.org/10.11114/jets.v5i9.2612>. Acesso em: 12 mar. 2021.

CETIC.BR; NIC.BR. **Resumo Executivo** - Pesquisa sobre o uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros - TIC Domicílios 2020. São Paulo: Comitê Gestor da Internet no Brasil, 2021a. Disponível em: <https://cgi.br/publicacao/resumo-executivo-pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2020/>. Acesso em: 19 dez. 2021.

CETIC.BR; NIC.BR. **Pesquisa sobre o uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros** - TIC Domicílios 2020. São Paulo: Comitê Gestor da Internet no Brasil, 2021b. Disponível em: <https://cgi.br/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2020/>. Acesso em: 19 dez. 2021.

CHI, Amy. Researching Pedagogical Lexicography. In: JACKSON, Howard (Ed.). **The Bloomsbury Companion to Lexicography**. London: Bloomsbury Publishing, 2015. p. 165-187.

CHOI, Joon Suh; CROSSLEY, Scott. **ARTE: Automatic Readability Tool for English**. Version 1.1. [s.l.]: NLP Tools for the Social Sciences 2020. Disponível em: www.linguisticanalysistools.org. Acesso em: 29 out. 2020.

COWIE, Anthony P. Phraseology. In: FONTENELLE, Thierry (org.). **Practical Lexicography: A Reader**. Oxford: OUP, 2008. Chapter 11, p. 163-168.

CREESE, Sharon; MCGILLIVRAY, Barbara; NESI, Hilary; RUNDELL, Michael; SULE, Katalin. Everything You Always Wanted to Know about Dictionaries (But Were Afraid to Ask): A Massive Open Online Course. In: ČIBIČ, Jaka; GORJANC, Vojko; KOSEM, Iztok; KREK, Simon (Orgs.). **Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts**. Ljubljana: Ljubljana University Press, 2018, p. 59–66. Disponível em: <https://euralex.org/publications/everything-you-always-wanted-to-know-about-dictionaries-but-were-afraid-to-ask-a-massive-open-online-course/>. Acesso em: 3 out. 2019.

CRYSTAL, David. **English as a Global Language**. 2nd edition. New York: Cambridge University Press, 2003.

CROSSLEY, Scott; SKALICKY, Stephen; DASCALU, Mihai. Moving beyond classic readability formulas: new methods and new models. **Journal of Research in Reading**, [s.l.], v. 42, n. 3-4, p. 541-561, 2019. DOI: 10.1111/1467-9817.12283.

CRYSTAL, David. **A Dictionary of Linguistics and Phonetics**. 6th edition. Oxford: Blackwell, 2008.

CUNHA, Claudio de Assis da. O léxico e as unidades lexicais: revisitando a teoria. **Guavira Letras**, Três Lagoas, v. 1, n. 30, p. 15-30, 2019. Disponível em: <http://websensors.net.br/seer/index.php/guavira/article/view/814>. Acesso em: 03 fev. 2022.

DE COCK, Sylvie; GRANGER, Sylviane. Computer Learner Corpora and Monolingual Learners' Dictionaries: The Perfect Match. **Lexicographica**, [s.l.], v. 20, p. 72-86, 2005.

DE GROOT, Annette M. B. Bilingual Memory. In: GROSJEAN, François; LI, Ping. **The Psycholinguistics of Bilingualism**. Oxford: Wiley-Blackwell, 2013. Capítulo 8, p. 154-172.

DE GROOT, Annette M. B.; VAN HELL, Janet G. The Learning of Foreign Language Vocabulary. In: KROLL, Judith; DE GROOT, Annette M. B. (eds.) **Handbook of Bilingualism**. Psycholinguistic Approaches. Oxford: Oxford University Press, 2005. p. 9-29.

DEL RÍO, Iria; MENDES, Amália. Error annotation in a Learner Corpus of Portuguese. In: 11TH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2018, Miyazaki. **Conference Papers**. Miyazaki: ELRA, 2018. p. 4116-4119. Disponível em: <http://www.lrec-conf.org/proceedings/lrec2018/summaries/1002.html>. Acesso em: 11 mar. 2021.

DÍAZ-CORRALEJO C. Joaquín. Aportaciones de la Didáctica de las Lenguas y las Culturas. In: SÁNCHEZ LOBATO, Jesús; SANTOS GARGALLO, Isabel (dir.). **Vademécum para la formación de profesores**. Enseñar español como segunda lengua (L2) / lengua extranjera (LE). Tomo 1. Madrid: SGEL, 2004. p. 243-257.

DiClavELE. MARTÍN PERIS, Ernesto (org.). **Diccionario de Términos Clave de ELE**. [S.l.]: SGEL, Instituto Cervantes, 2008. Disponível em http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/indice.htm. Acesso em: 18 out. 2019.

DILLI, Camila; SCHOFFEN, Juliana Roquele; SCHLATTER, Margarete. Parâmetros para a avaliação de produção escrita orientados pela noção de gênero do discurso. In: SCHOFFEN, Juliana Roquele et al. **Português como Língua Adicional**: reflexões para a prática docente. Porto Alegre: Editora Bem Brasil, 2012. Cap. Dd, p. çç.

DOMÍNGUEZ VÁZQUEZ, María José; VALCÁRCEL RIVEIRO, Carlos. Hábitos de uso de los diccionarios entre los estudiantes universitarios europeos: ¿nuevas tendencias? In: CÓRDOBA RODRÍGUEZ, Félix; GONZÁLEZ SEOANE, Ernesto Xosé; SÁNCHEZ-PALOMINO, María Dolores (Orgs.). **Lexicografía de las lenguas románicas**. Berlin: Walter de Gruyter, 2015, v. 2, p. 165–189.

DÖRNYEI, Zoltán. **The Psychology of Second Language Acquisition**. Oxford: Oxford University Press, 2009.

DUARTE FERREIRA, Gabriela. Testagem de critérios de avaliação de um vocabulário básico. **Entrepalavras**, Fortaleza, v. 7, n. 1, p. 160-174, 2017. Disponível em: <http://dx.doi.org/10.22168/2237-6321.7.7.1.160-174>. Acesso em: 30 ago. 2020.

DURÃO, Adja B. A. Barbieri. **Análisis de errores en la interlengua de brasileños aprendices de español y de españoles aprendices de portugués**. 2ª edição. Londrina: EDUEL, 2004.

DZIEMIANKO, Anna. User-Friendliness of Noun and Verb Coding Systems in Pedagogical Dictionaries of English: a Case of Polish Learners. **International Journal of Lexicography**, Oxford, v. 24, n. 1, p. 50–78, 2011.

DZIEMIANKO, Anna. An insight into the visual presentation of signposts in English learners' dictionaries online. **International Journal of Lexicography**, Oxford, v. 29, n. 4, p. 490-524, 2016. Disponível em: <https://doi.org/10.1093/ijl/ecv040>. Acesso em: 07 mai. 2019.

EDUCATION FIRST. **EF English Proficiency Index**. A Ranking of 100 countries and regions by English skills. Lucerne: Education First, 2019. Disponível em: <https://www.ef.com.br/epi/downloads/>. Acesso em: 02 abr. 2019.

ERES FERNÁNDEZ, Gretel. et al. **Materiais didáticos de espanhol**: entre a quantidade e a diversidade. 2012. 68f. Relatório final de pesquisa não financiada – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2012. Disponível em: <http://www4.fe.usp.br/wp-content/uploads/cepel/materiales-didacticos-de-espanol-informe.pdf>. Acesso em: 06 abr. 2017.

ERMAN, Britt; WARREN, Beatrice. The idiom principle and the open choice principle. **Text**, [s.l.], v. 20, n. 1, p. 29-62, 2000.

FARIAS, Virginia Sita. *Whole-sentence definition* versus definição por *genus proximum + differentiae specifica*: um contraste entre duas técnicas definitórias. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 17, n. 1, p. 73-100, 2009. Disponível em: <http://periodicos.letras.ufmg.br/index.php/relin/issue/view/199>. Acesso em: 07 fev. 2017.

FARIAS, Virginia Sita. Considerações preliminares sobre o pós-comentário na microestrutura de dicionários semasiológicos. **ReVEL**, Porto Alegre, v. 9, n. 17, 2011. Disponível em: <http://www.revel.inf.br/pt/edicoes/?id=21>. Acesso em: 18 jul. 2020.

FARIAS, Virginia Sita. **Sobre a definição lexicográfica e seus problemas**: fundamentos para uma teoria geral dos mecanismos explanatórios em dicionários semasiológicos. 2013. 398 f. Tese (Doutorado em Letras) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <http://hdl.handle.net/10183/90167>. Acesso em: 07 fev. 2017.

FENN, P. Error analysis and language training in university FL courses: an applied linguistic assessment with a focus on syntax. In: BÜRCEL, Ch.; SIEPMANN, D. (Hrsg). **Sprachwissenschaft und Fremdsprachenunterricht**: Spracherwerb und Sprachkompetenzen im Fokus. Band 13. Baltmannsweiler: Schneider Verlag Hohengehren, 2014. p. 79-105.

FERNÁNDEZ, Sonsoles. **Interlengua y análisis de errores en el aprendizaje del español como lengua extranjera**. Madrid: Edelsa, 1997.

FERNÁNDEZ DE LA TORRE, M^a. Dolores. Uso de los diccionarios y posibilidades pedagógicas en la enseñanza de segundas lenguas: el caso de la filología inglesa. In: AYALA CASTRO, Marta (coord.). **Diccionarios y enseñanza**. Alcalá de Henares: Publicaciones de la Universidad de Alcalá, 2001. p. 275-286.

FERREIRA CABRERA, Anita; ELEJALDE GÓMEZ, Jéssica. Análisis de errores recurrentes en el Corpus de Aprendices de Español como Lengua Extranjera, CAELE. **Revista Brasileira de Linguística Aplicada**, Belo Horizonte, v. 17, n. 3, p. 509-537, 2017. Disponível em: <https://doi.org/10.1590/1984-6398201710927>. Acesso em: 03 mar. 2021.

FILLMORE, Charles. Frame Semantics. In: THE LINGUISTIC Society of Korea (Ed.). **Linguistics in the morning calm**. Seoul: Hanshin, 1982. p. 111-138.

FIRTH, John R. A synopsis of linguistic theory, 1930-1955. In: FIRTH, John R. **Studies in Linguistic Analysis**. Oxford: Basil Blackwell, 1962. p. 1-32.

FORNARI, Michelle. Kühn. Conceção e desenho do Front Matter do dicionário de falsos amigos espanhol-português. **Voz das Letras**, Concórdia, n. 9, s.p., 2008.

GALISSON, Robert; COSTE, Daniel (orgs.). **Dicionário de Didática das Línguas**. Tradução coordenada por Fernanda Irene Fonseca. Coimbra: Livraria Almedina, 1983.

GAO, Jun; LIU, Haitao. Valency and English learners' thesauri. **International Journal of Lexicography**, Oxford, v. 32, n. 3, p. 326-361, 2019. Disponível em: <https://doi.org/10.1093/ijl/icy025>. Acesso em: 07 mai. 2020.

GIBBONS, Pauline. **Scaffolding language, scaffolding learning**: Teaching second language learners in the midstream classroom. Second Edition. Portsmouth, New Hampshire: Heinemann, 2015.

GILQUIN, Gaëtanelle; GRANGER, Sylviane. Learner language. In: BIBER, Douglas; REPPEN, Randi (eds.). **The Cambridge Handbook of English Corpus Linguistics**. Cambridge: Cambridge University Press, 2015. p 418-435.

GOLESTANI, Narly; ALARIO, F-Xavier; MERIAUX, Sébastien; LE BIHAN, Denis; DEHAENE, Stanislas; PALLIER, Christophe. Syntax production in bilinguals. **Neuropsychologia**, [s.l.], v. 44, n. 7, p. 1029-1040, 2006.

GRANGER, Sylviane. Error-tagged Learner Corpora and CALL: a promising synergy. **CALICO Journal**, Sheffield, v. 20, n. 3, 465-480, 2003. Disponível em: <https://journals.equinoxpub.com/CALICO/article/view/23244>. Acesso em: 12 mar. 2021.

GRANGER, Sylviane. Contrastive interlanguage analysis: a reappraisal. **International Journal of Learner Corpus Research**, [s.l.], v. 1, n. 1, p. 7-24, 2015a.

GRANGER, Sylviane. The contribution of learner corpora to reference and instructional materials design. In: GRANGER, Sylviane; GILQUIN, Gaëtanelle; MEUNIER, Fanny (eds.). **The Cambridge Handbook of Learner Corpus Research**. Cambridge: Cambridge University Press, 2015b. p. 486-510.

GRANGER, Sylviane. Learner corpora in foreign language education. In: THORNE, Steven; MAY, Stephen (eds.). **Language, Education and technology**. Encyclopedia of Language and Education. Cham: Springer, 2017.

GRANGER, Sylviane. Has Lexicography Reaped the Full Benefit of the (Learner) Corpus Revolution?. In: XVIII EURALEX INTERNATIONAL CONGRESS, 2018, Ljubljana. **Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts**. Ljubljana: Ljubljana University Press, Faculty of Arts, 2018. p. 17-24. Disponível em: <https://euralex.org/publications/has-lexicography-reaped-the-full-benefit-of-the-learner-corpus-revolution/>. Acesso em: 12 mar. 2021.

GRANGER, Sylviane. Have Learner Corpus Research and Second Language Acquisition finally met? In: LE BRUYN, Bert; PAQUOT, Magali (Eds.). **Learner Corpus Research Meets Second Language Acquisition**. Cambridge: Cambridge University Press, 2021. p. 243-257.

GRANGER, Sylviane; GILQUIN, Gaëtanelle; MEUNIER, Fanny. Introduction: learner corpus research – past, present and future. In: GRANGER, Sylviane; GILQUIN, Gaëtanelle; MEUNIER, Fanny (eds.). **The Cambridge Handbook of Learner Corpus Research**. Cambridge: Cambridge University Press, 2015. p. 1-6.

GREEN, David. Mental control of the bilingual lexico-semantic system. **Bilingualism: Language and Cognition**, [s.l.], v. 1, n. 3, p. 67-81, 1998.

GREEN, David; ABUTALEBI, Jubin. Language control in bilinguals: The adaptive control hypothesis. **Journal of Cognitive Psychology**, [s.l.], v. 25, n. 5, p. 515-530, 2013. DOI: <https://doi.org/10.1080/20445911.2013.796377>.

GROSJEAN, François. Bilingualism: a short introduction. In: GROSJEAN, François; LI, Ping. **The Psycholinguistics of Biligualism**. Oxford: Wiley-Blackwell, 2013. Capítulo 1, p. 5-25.

GYLLSTAD, H. Grammatical Collocation. In: CHAPELLE, Carol A. (Org.). **The Encyclopedia of Applied Linguistics**. New Jersey: Wiley-Blackwell, 2013. p. 01-06.

HALASA, Najwa H.; AL-MANASEER, Majeda. The use of the first language in second language learning reconsidered. **College Student Journal**, Jordan, v. 46, n. 01, p. 71-81, 2012.

HALL, Graham. Method, methods and methodology. In: HALL, Graham (ed.). **The Routledge Handbook of English Language Teaching**. Abingdon: Routledge, 2016, p. 209-223.

HARTMANN, Reinhard R. K.; JAMES, Gregory. **Dictionary of Lexicography**. Londres: Routledge, 2001.

HAUSMANN, Franz Josef. Apprendre le vocabulaire, c'est apprendre les collocations. Comment enseigner et apprendre les combinaisons lexicales?. In: HAAG, Elke (ed.). **Collocations, phraséologie, lexicographie**. Études 1977-2007 et Bibliographie. Aachen: Shaker Verlag, 2007. p. 49-61.

HEUBERGER, Reinhard. Learners' dictionaries: History and development; current issues. In: Durkin, Philip (Ed.). **The Oxford Handbook of Lexicography**. Oxford: Oxford University Press, 2016. p. 25-43

HOEY, Michael. **Lexical Priming: A new theory of words and language**. London/New York: Routledge, 2005.

HÖFFLING, Camila. **Traçando um perfil de usuários de dicionários – estudantes de Letras em Habilitação em Língua Inglesa: um novo olhar sobre dicionários para aprendizes e a formação de um usuário autônomo**. 2006. 375 f. Tese (Doutorado em Letras) – Programa de Pós-Graduação em Letras, Universidade Estadual Paulista Júlio de Mesquita Filho, Araraquara, 2006. Disponível em: <https://repositorio.unesp.br/handle/11449/103619>. Acesso em: 02 fev. 2022.

HOWKINS, John; FILIPOVIC, Luna. **Criterial Features in L2 English**. Specifying the Reference Levels of the Common European Framework. Cambridge: Cambridge University Press, 2012. (English Profile Studies Series 1)

Instituto Cervantes. **El Español: una lengua viva**. Informe 2018. Madrid: Instituto Cervantes, 2018. Disponível em: https://cvc.cervantes.es/lengua/espanol_lengua_viva/. Acesso em: 28 jun. 2018.

JACKSON, Howard. **Lexicography**. An Introduction. London: Routledge, 2002.
JARVIS, Scott; PAVLENKO, Aneta. **Crosslinguistic Influence in Language and Cognition**. Kindle edition. New York; London: Routledge, 2008.

JARVIS, Scott. Crosslinguistic Influence and Multilingualism. *In*: CHAPPELLE, Carol A. (Org.). **The Encyclopedia of Applied Linguistics**. Oxford: Wiley–Blackwell, 2013, p. 01-08.

JUDD, Elliot L.; TAN, Lihua; WALBERG, Herbert J. **Teaching Additional Languages**. Brussels: International Academy of Education; Geneva: International Bureau of Education, 2001. Disponível em: http://www.ibe.unesco.org/fileadmin/user_upload/archive/Publications/educationalpractices_eriespdf/prac06e.pdf. Acesso em: 1 abr. 2020.

KASPARY, Cíntia Voos. Considerações sobre o uso do dicionário em dois manuais de ensino-aprendizagem de francês língua estrangeira. **Travessias interativas**, São Cristóvão (SE), v. 7, n. 14, p. 295-314, 2017. Disponível em: <https://seer.ufs.br/index.php/Travessias/article/view/9137/7160>. Acesso em: 28 out. 2019.

KELDER, Marceline. **The Revised Hierarchical Model tested**: Connections of Form and Meaning for L2 Nouns and Verbs. 2018. 58 f. Dissertação (Master in Language and Communication Coaching) – Radboud University Nijmegen, Nijmegen, Holanda, 2018. Disponível em: <https://theses.uibn.ru.nl/handle/123456789/5963>. Acesso em: 22 jul. 2020.

KIVISTÖ-DE SOUZA, Hanna. Brazilian EFL learners' awareness about L2 phones: Is mall pronounced as 'mal'? **Trabalhos em Linguística Aplicada**, Campinas, n. 56.1, p. 235-258, 2017. Disponível em: <http://dx.doi.org/10.1590/010318135071200021>. Acesso em: 07 mai. 2019.

KLOTZ, Michael; HERBST, Thomas. **English Dictionaries**. A Linguistic Introduction. Berlin: Erich Schmidt Verlag, 2016.

KOSEM, Iztok; LEW, Robert; MÜLLER-SPITZER, Carolin; *et al.* The Image of the Monolingual Dictionary Across Europe. Results of the European Survey of Dictionary use and Culture. **International Journal of Lexicography**, v. 32, n. 1, p. 92–114, 2018.
KOSCH, Inge. Expectation Levels in Dictionary Consultation and Compilation. **Lexikos**, Stellenbosch, v. 23, p. 201-208, 2013. Disponível em: <https://doi.org/10.5788/23-1-1211>. Acesso em: 18 jul. 2020.

KROLL, Judith; BIALYSTOK, Ellen. Understanding the consequences of bilingualism for language processing and cognition. **Journal of Cognitive Psychology**, s.l., v. 25, n. 5, p. 497-514, 2013. Disponível em: <https://doi.org/10.1080/20445911.2013.799170>. Acesso em: 09 mar. 2020.

KROLL, Judith; DUSSIAS, Paola E.; BICE, Kinsey; PERROTTI, Lauren. Bilingualism, Mind, and Brain. **Annual Review of Linguistics**, s.l., v. 1, n. 1, p. 377-394, 2015. Disponível em: <https://doi.org/10.1146/annurev-linguist-030514-124937>. Acesso em: 09 mar. 2020.

KROLL, Judith; STEWART, Erika. Category Interference in translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations. **Journal of Memory and Language**, s.l., v. 33, n. 2, p. 149-174, 1994. DOI: <https://doi.org/10.1006/jmla.1994.1008>

KROLL, Judith F.; VAN HELL, Janet G.; TOKOWICZ, Natasha; GREEN, David W. The Revised Hierarchical Model: A Critical Review and Assessment. **Bilingualism: Language and Cognition**, Cambridge, v. 13, n. 3, p. 373-381, 2010. Disponível em: <https://doi:10.1017/S136672891000009X>. Acesso em: 01 jul. 2019.

KROLL, Judith; VAN HELL, Janet G.; TOKOWICZ, Natasha; GREEN, David W. The Revised Hierarchical Model: A critical review and assessment. **Bilingualism: Language and Cognition**, s.l., v. 13, n. 3, p. 373-381, 2010. Disponível em: <https://doi.org/10.1017/S136672891000009X>. Acesso em: 09 mar. 2020.

LANTOLF, James; THORNE, Steven. Sociocultural theory and Second language learning. In: VAN PATTEN, B.; WILLIAMS, J. (eds.). **Theories in Second Language Acquisition**. New Jersey: Lawrence Erlbaum, 2007. p. 197-220.

LARSEN-FREEMAN, Diane. Another step to be taken – Rethinking the end point of the interlanguage continuum. In: HAN, Zhaohong; TARONE, Elaine. **Interlanguage. Forty years later**. Amsterdam / Philadelphia: John Benjamins, 2014. p. 203-220.

LEECH, Geoffrey; NESI, Hilary. 'Moving towards perfection: the learners' (electronic) dictionary of the future'. In: **The Perfect Learners' Dictionary (?)**. Tübingen: Max Niemeyer Verlag, 1999, p. 295–306.

LEECH, Geoffrey; SVARTVIK, Jan. **A communicative grammar of English**. 3rd ed. Harlow: Pearson Education, 2002.

LEGALLOIS, D. La colligation: autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique?. **Corpus**, Nice, n. 11, p. 31-54, 2012. Disponível em: <https://corpus.revues.org/2183>. Acesso em: 22 fev. 2021.

LEHECKA, T. Collocation and Colligation. In: ÖSTMAN, J.; VERSCHUEREN, J. **Handbook of Pragmatics**. Amsterdam: John Benjamins, 2015.

LEW, Robert. User studies: Opportunities and limitations. In: AKASU, Kaoru; UCHIDA, Satoru (Orgs.). **ASIALEX2011 Proceedings. Lexicography: Theoretical and practical perspectives**. Kyoto: Asian Association for Lexicography, 2011, p. 07–16. Disponível em: <https://www.asialex.org/pdf/Asialex-Proceedings-2011.pdf>. Acesso em: 5 out. 2019.

LI, Ping. Successive Language Acquisition. In: GROSJEAN, François; LI, Ping. **The Psycholinguistics of Bilingualism**. Oxford: Wiley-Blackwell, 2013. Capítulo 7, p. 132-152.

LITTLEWOOD, William. Second Language Learning. In: DAVIES, Alan; ELDER, Catherine. **The Handbook of Applied Linguistics**. Oxford: Blackwell Publishing, 2004, p. 501-524.

LLORIÁN, Susana. Claves de una revisión de los *Niveles de Referencia para el Español*, basada em metodología de corpus. **MarcoELE**, Valencia, n. 25, p. 1-34, 2017. Disponível em: <https://marcoele.com/revision-de-los-nre-basada-en-corpus/>. Acesso em: 02 fev. 2021.

LONGO, Beatriz Nunes de Oliveira. Uma proposta de definição para dicionários bilíngues. **Estudos Linguísticos**, São Paulo, v. 29, p. 286-291, 2000. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos>. Acesso em: 20 fev. 2021.

LU, Huaguo; WEI, Xiangqing. Structuring Polysemy in English Learners' Dictionaries: A Prototype Theory-Based Model. **International Journal of Lexicography**, Oxford, v. 32, n. 1, p. 20-37, 2019. Disponível em: <https://doi.org/10.1093/ijl/ecy018>. Acesso em: 07 mai. 2019.

MA, Fengyang; CHEN, Peiyao; GUO, Taomei; KROLL, Judith. When late second language learners access the meaning of L2 words: Using ERPs to investigate the role of the L1 translation equivalent. **Journal of Neurolinguistics**, s.l., v. 41, p. 50-69, 2017. DOI: <http://dx.doi.org/10.1016/j.jneuroling.2016.09.006>

MACDONALD, Penny. "We all make mistakes!". Analysing an Error-coded Corpus of Spanish University Students' Written English. **Complutense Journal of English Studies**, Madrid, v. 26, p. 103-129, 2016. Disponível em: <http://dx.doi.org/10.5209/CJES.53273>. Acesso em: 09 mar. 2021.

MALAVAR, Irania. Diminutivos en el habla de aprendices de español. **MarcoELE**, Valencia, n. 29, jul-dez, 2019. Disponível em: https://marcoele.com/descargas/29/malaver-diminutivos_aprendices.pdf. Acesso em: 03 mar. 2021.

MARTINEZ, Ron. A framework for the inclusion of multi-word expressions in ELT. **ELT Journal**, London, v. 67, n. 2, p. 184-198, 2013.

MARTINS, Analigia. Relatório de idiomas Duolingo 2020: Brasil. **Duolingo Blog**, [s.l.], 15 dez. 2020. Disponível em: <https://blog.duolingo.com/brazil-language-report-2020/>. Acesso em: 02 fev. 2022.

MATTES, Marlene Gonçalves; BUGUEÑO MIRANDA, Félix Valentín. Por que um dicionário é um texto? **Gragoatá**, Niterói, v. 20, n. 38, p. 91-110, 2015. Disponível em: <https://doi.org/10.22409/gragoata.v20i38.33302>. Acesso em: 13 nov. 2019.

MCNAMARA, Tim. Language Testing. In: DAVIES, Alan; ELDER, Catherine. **The Handbook of Applied Linguistics**. Oxford: Blackwell Publishing, 2004, p. 763-783.

MEADE, Gabriela; MIDGLEY, Katherine; HOLCOMB, Phillip. An ERP Investigation of L2-L1 Translation Priming in Adult Learners. **Frontiers in Psychology**, Lausanne, v. 9, p. 1-10, 2018. Disponível em: <https://doi.org/10.3389/fpsyg.2018.00986>. Acesso em: 22 jul. 2020.

MEC AFIRMA que o Ciência sem Fronteiras terá 5 mil bolsistas na pós-graduação. **Portal Ministério da Educação**, 02 abr. 2017. Disponível em: <http://portal.mec.gov.br/ultimas-noticias/212-educacao-superior-1690610854/46981-mec-afirma-que-o-ciencia-sem-fronteiras-tera-5-mil-bolsistas-na-pos-graduacao>. Acesso em: 17 nov. 2017.

MELČUK, Igor. **Semantics**: from meaning to text. Vol. 3. Philadelphia: John Benjamins, 2015.

MILLIOT, Jim. The world's 52 Largest Book Publishers, 2016. **Publishers Weekly**, USA, 26 ago. 2016. Disponível em: <http://www.publishersweekly.com/pw/by-topic/international/international-book-news/article/71268-the-world-s-52-largest-book-publishers-2016.html>. Acesso em: 08 jan. 2018.

MONTEIRO, Rafael Augusto; SALES SANTOS, Roney Lira de; PARDO, Thiago Alexandre S. **FakeCheck**. Universidade de São Paulo, 2018. Disponível em: <http://nilc-fakenews.herokuapp.com/>. Acesso em: 04 mai. 2019.

MOREIRA, Glauber Lima. Dificuldades de compreensão leitora en el artículo lexicográfico del léxico del turismo en los diccionarios Señas y Clave. **Trama**, Marechal Cândido Rondon, v. 13, n. 30, p. 249-271, 2017. Disponível em: <http://e-revista.unioeste.br/index.php/trama/article/view/15763>. Acesso em: 10 nov. 2020.

MORTON, Tom. Content and language integrated learning. *In*: HALL, Graham (ed.). **The Routledge Handbook of English Language Teaching**. Abingdon: Routledge, 2016, p. 252-264.

MUÑOZ MARTÍN, Ricardo. Looking Toward the Future of Cognitive Translation Studies. *In*: SCHWIETER, John W.; FERREIRA, Aline. (eds.). **The Handbook of Translation and Cognition**. Hoboken: Wiley-Blackwell, 2017. p. 555-572.

NASCIMENTO, Francisco I. Lexicografia Multimodal: como as crianças percebem os recursos visuais dos dicionários infantis. **Discursividades**, Campina Grande, v. 3, n. 2, p. 87-103, 2018. Disponível em: revista.uepb.edu.br/index.php/REDIS/REVISTA/article/view/4951. Acesso em: 6 ago. 2020.

NAIR, Vishnu K.K.; BIEDERMANN, Britta; NICKELS, Lyndsey. Consequences of late bilingualism for novel word learning: Evidence from Tamil–English bilingual speakers. **International Journal of Bilingualism**, [s.l.], v. 20, n. 4, p. 473-487, 2016. DOI: <https://doi.org/10.1177%2F1367006914567005>.

NEGUERUELA-AZAROLA, Eduardo; GARCÍA, Próspero. Sociocultural theory and the language classroom. *In*: HALL, Graham (org.). **The Routledge Handbook of English Language Teaching**. Abingdon: Routledge, 2016. P. 295-309.

NESI, Hilary. Dictionary Use. *In*: CHAPELLE, Carol A. (Org.). **The Encyclopedia of Applied Linguistics**. Oxford: Wiley–Blackwell, 2013. p. 01-05.

NESI, Hilary. Researching Users and Uses of Dictionaries. *In*: JACKSON, Howard (Org.). **The Bloomsbury Companion to Lexicography**. London: Bloomsbury, 2015a. p. 62–74.

NESI, Hilary. Thirty years of user studies– and what we still need to find out. *In*: LI, Lan; MCKEOWN, Jamie; LIU, Liming (Orgs.). **Words, Dictionaries and Corpora: Innovations in reference science. Proceedings of ASIALEX 2015**. Hung Hom (Hong Kong): Hong Kong Polytechnic University, 2015b, p. 01-08. Disponível em: <http://asialex.org/pdf/Asialex-Proceedings-2015.pdf>. Acesso em: 19 set. 2019.

NESI, Hilary. The Demands of Users and the Publishing World: Printed or Online, Free or Paid For? *In: DURKIN, Philip (Org.). The Oxford Handbook of Lexicography*. Oxford: Oxford University Press, 2016. p. 579–589.

NESI, Hilary; BOONMOH, Atpat. A close look at the use of pocket electronic dictionaries for receptive and productive purposes. *In: FITZPATRICK, Tess; BARFIELD, Andy (Orgs.). Lexical Processing in Second Language Learners. Papers and Perspectives in Honour of Paul Meara*. Clevedon (UK): Multilingual Matters, 2009. p. 67–81.

NOGUEIRA, Marcia Castelo Branco. **Ouvindo a voz do (pré)adolescente brasileiro da geração digital sobre o livro didático de inglês desenvolvido no Brasil**. 2007. 182f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007. Disponível em: <https://doi.org/10.17771/PUCRio.acad.10439>. Acesso em: 30 jun. 2018.

NUNAN, David. The Impact of English as a Global language on Educational Policies and Practices the Asia-Pacific Region. **TESOL Quarterly**, New York, v. 37, n. 4, p. 589-613, 2003.

ODLIN, Terence. Crosslinguistic influence in Second Language Acquisition. *In: CHAPPELLE, Carol A. (Org.). The Encyclopedia of Applied Linguistics*. Oxford: Wiley–Blackwell, 2013, p. 01-06.

O'DONNELL CHRISTOFFERSEN, Katherine. A contrastive analysis of dar 'give' in English and Brazilian Portuguese: semantic-syntactic relationships and implications for L2 instruction. **Brazilian English Language Teaching Journal**, v. 7, n. 1, p. 26-43, 2016. Disponível em: <http://dx.doi.org/10.15448/2178-3640.2016.1.23491>. Acesso em: 07 mai. 2019.

PALFREYMAN, David M.; BENSON Philip. Autonomy and Its Role in English Language Learning: Practice and Research. *In: GAO, Xuesong. (eds) Second Handbook of English Language Teaching*. Vol. 2. Cham: Springer, 2019. Capítulo 36, p. 661-681.

PARADIS, Michel. Key concepts, framework, and clarifications. *In: PARADIS, Michel. Declarative and Procedural Determinants of Second Languages*. Amsterdam/Philadelphia: John Benjamins, 2009. Capítulo 1, p. 1-35.

PARODI, Giovanni. Corpus de aprendices de espanhol (CAES). **Journal of Spanish Language Teaching**, [s.l.], vol. 2, n. 2, p. 194-200, 2015.

PEREIRA, Renato Rodrigues; NADIN, Odair Luiz. Dicionário enquanto gênero textual: por uma proposta de categorização. **Acta Scientiarum**, Maringá, v. 41, n. 1, s.p., 2019. Disponível em: <https://doi.org/10.4025/actascilangcult.v41i1.43835>. Acesso em: 03 ev. 2022.

PLANTA AFRICANA conhecida como “milagre da natureza”, é capaz de ‘Lavar’ o excesso de açúcar do sangue e combater diversas doenças. **Portal Ciência e Saúde**, 06 abr. 2019. Disponível em: <https://portal.ciencia-saude.com/plantamilenar-2/>. Acesso em: 04 mai. 2019.

POARCH, Gregory J., VAN HELL, Janet G., & KROLL, Judith F. Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? **Bilingualism: Language and Cognition**, Cambridge, v. 18, n. 3, p. 357-371, 2015. DOI: <https://doi.org/10.1017/S1366728914000558>.

PONTES, Antônio Luciano. O verbal e o não-verbal em dicionários didáticos: um enfoque multimodal. In: ARAÚJO, J. C.; BIASE-RODRIGUES, B.; DIEB, M. (org.). **Seminários linguísticos: discurso, análise linguística, ensino e pesquisa**. Mossoró: Edições UERN, 2010a.

PONTES, Antônio Luciano. Multimodalidade em dicionários escolares. In: ISQUERDO, A. N.; BARROS, Lídia Almeida. (org.). **As ciências do léxico: Lexicologia, Lexicografia, Terminologia**. Mato Grosso do Sul: UFMS, 2010b.

PONTES, Antônio Luciano. Articulação verbo-visual em verbetes lexicográficos de dicionários escolares brasileiros. In: PONTES, Antônio Luciano. et al. (orgs.). **Perspectivas em Lexicografia e terminologia**. Fortaleza: EdUECE, 2018. cap. 1, p. 16-50. Disponível em: <portal-archipelagus.azurewebsites.net/farol/eduece/ebook/perspectivas-em-lexicografia-e-terminologia/525727/>. Acesso em: 28 ago. 2020.

RANALLI, Jim; NURMUKHAMEDOV, Ulugbek. Learner Dictionaries. In: CHAPELLE, Carol A. (Ed.). **The Encyclopedia of Applied Linguistics**. New Jersey: Wiley-Blackwell, 2014. p. 01-06.

REOLON J., Carolina. **Mecanismos de otimização para um Learner's Dictionary do inglês para aprendizes brasileiros**. 2013. 224 f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <http://hdl.handle.net/10183/78161>. Acesso em: 18 out. 2019.

RICHARDS, Jack C.; SCHMIDT, Richard. **Dictionary of Language Teaching & Applied Linguistics**. 4th edition. Harlow: Pearson Education, 2010.

RINGBOM, Håkan. **Cross-linguistic Similarity in Foreign Language Learning**. Clevedon; Buffalo; Toronto: Multilingual Matters, 2006. Ch. 2, p. 5-9. Disponível em: <https://books.google.com.br/books?id=oeEM1tK1fy8C&lpg=PP1&dq=ringbom&hl=pt-BR&pg=PT30#v=onepage&q=ringbom&f=false>. Acesso em: 09 set. 2020.

RUNDELL, Michael. The dictionary of the future. In: GRANGER, Sylviane (Ed.), **Optimizing the Role of Language in Technology-Enhanced Learning**. Louvain-la-Neuve: Kaleidoscope, 2007. p. 49-51. Disponível em: <https://cdn.uclouvain.be/public/Exports%20readdir/adri/documents/GRANGER-SYLVIANE-2007.pdf>. Acesso em: 19 fev. 2021

RUNDELL, Michael. Recent Trends in English Pedagogical Lexicography. In: FONTENELLE, Thierry. **Practical Lexicography: a reader**. New York: Oxford University Press, 2008.

SÁNCHEZ RUFAT, Anna. Apuntes sobre las combinaciones léxicas y el concepto de colocación. **Anuario de Estudios Filológicos**, Cáceres, n. 33, p. 291-306, 2010. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=3401843>. Acesso em: 09 nov. 2021.

SÁNCHEZ RUFAT, Anna. Estrategias para la Enseñanza de Secuencias Formulaicas en el Aula de Español como Lengua Extranjera. In: NÍKLEVA, Dimitrika G. (ed.). **Necesidades y tendencias en la formación del profesorado de español como lengua extranjera**. Berna: Peter Lang, 2018. p. 257-282.

SANTOS, Thaísa Maria Rocha. Estudo da relação verbo-visual no dicionário Aurelino. In: PONTES, A. L. et al. (orgs.). **Perspectivas em Lexicografia e terminologia**. Fortaleza: EdUECE, 2018. cap. 1, p. 287-301. Disponível em: portal-archipelagus.azurewebsites.net/farol/eduece/ebook/perspectivas-em-lexicografia-e-terminologia/525727/. Acesso em: 28 ago. 2020.

SANTOS, Thaísa Maria Rocha; CORIOLANO, Raoni Reinaldo; PONTES, Antônio Luciano. O uso de fotografias em verbetes de dicionários infantis. **Leitura**, Maceió, n. 63, p. 138-153, 2019. Disponível em: [10.28998/2317-9945.2019n63p138-153](https://doi.org/10.28998/2317-9945.2019n63p138-153). Acesso em: 28 ago. 2020.

SELINKER, Larry. Interlanguage 40 years on. Three themes from here. In: HAN, Zhaohong; TARONE, Elaine. **Interlanguage. Forty years later**. Amsterdam / Philadelphia: John Benjamins, 2014. p. 221-246.

SIEPMANN, Dirk. Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects. **International Journal of Lexicography**, Oxford, v. 18, n. 4, p. 409-443, 2005.

SINCLAIR, John. **Corpus, Concordance, Collocation**. Oxford: OUP, 1991.

SINCLAIR, John. In praise of the dictionary. In: WILLIAMS, Geoffrey; VESSIER, Sandra (orgs.). **Proceedings of the 11th EURALEX International Congress**. Lorient: Université de Bretagne-Sud, 2004. p. 1-12. Disponível em: <https://euralex.org/publications/in-praise-of-the-dictionary/>. Acesso em: 04 nov. 2021.

SOLER MONTES, Carlos. La variación gramatical y el aprendizaje de los tiempos verbales del pasado en el contexto del español como lengua extranjera. In: AINCIBURU, María Cecilia (ed.). **La adquisición del sistema verbal del español**. Peter Lang: Frankfurt am Main, 2017. p. 235-266. Disponível em: https://www.research.ed.ac.uk/portal/files/43626856/Soler_MontesASVE2017LaVariacionGramatical.pdf. Acesso em: 03 mar. 2021.

STEYER, Kathrin. Patterns. Phraseology in a state of flux. **International Journal of Lexicography**, v. 28, n. 3, p. 279-298, 2015.

TARP, Sven. **Lexicography in the Borderland between Knowledge and Non-Knowledge, General Lexicographical Theory with Particular Focus on Learner's Lexicography**. Berlin, Boston: De Gruyter, 2008.

TARP, Sven. Pedagogical Lexicography: Towards a New and Strict Typology Corresponding to the Present State-of-the-Art. **Lexikos**, Stellenbosch, v. 21, p. 217-231, 2011. Disponível em: <https://doi.org/10.5788/21-1-44>. Acesso em: 10 set. 2019.

TEER, Jacqueline Vaccaro. A lematização de expressões idiomáticas em dicionários para aprendizes. **Domínios de Linguagem**, Uberlândia, v. 12, n. 4, p. 2363-2383, 2018. Disponível em: <https://doi.org/10.14393/DL36-v12n4a2018-16>. Acesso em: 10 nov. 2020.

THE HISTORY of Cobuild. A new generation of dictionaries for learners of English. **Collins**, [s.d.]. Disponível em: <https://collins.co.uk/pages/elt-cobuild-reference-the-history-of-cobuild>. Acesso em: 04 nov. 2021.

THE OXFORD 3000. **Oxford Learners Dictionaries website**, 2019. Disponível <https://www.oxfordlearnersdictionaries.com/about/oxford3000>. Acesso em: 14 fev. 2020.

THORNE, Steven; TASKER, Thomas. Sociocultural and cultural-historical theories of language development. *In*: SIMPSON, James. (Ed.). **The Routledge Handbook of Applied Linguistics**. London: Routledge, 2011. p. 487-500.

TLFi. CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE; UNIVERSITÉ DE LORRAINE. **Trésor de la langue Française informatisé**. [s.l.]: ATILF/Université de Lorraine, 1997. Disponível em: <http://www.atilf.fr/tlfi>. Acesso em: 14 set. 2020.

TRAPPES-LOMAX, Hugh. Discourse Analysis. *In*: DAVIES, Alan; ELDER, Catherine. **The Handbook of Applied Linguistics**. Oxford: Blackwell Publishing, 2004, p. 133-164.

TRAXLER, Matthew J. Bilingual language processing. *In*: TRAXLER, Matthew J. **Introduction to Psycholinguistics: Understanding Language Science**. Oxford: Wiley-Blackwell, 2012. Capítulo 11, p. 415-446.

VAN DEN BRANDEN, Kris. Task-based language teaching. *In*: HALL, Graham (ed.). **The Routledge Handbook of English Language Teaching**. Abingdon: Routledge, 2016, p. 238-251.

VAN HELL, Janet G.; TANNER, Darren. Second Language Proficiency and Cross-Language Lexical Activation. **Language Learning**, Michigan, Ann Arbor, v. 62, n. 2, p. 148-171, 2012. DOI: <https://doi.org/10.1111/j.1467-9922.2012.00710.x>.

VAN ORDEN, Guy C. A rows is a rose: spelling, sound, and reading. **Memory & Cognition**, [s.l.], v. 15, n. 3, p. 181-198, 1987.

VASILJEVIC, Zorana. Teaching Collocations in a Second Language: Why, What and How?. **ELTA Journal**, Belgrado, v. 2, n. 2, p. 48-73, 2014. Disponível em: <http://eltajournal.org.rs/elta-journal-volume-2-no-2-dec-2014/>. Acesso em: 15 mar. 2019.

VERLINDE, Serge; BINON, Jean. Pedagogical Lexicography Revisited. *In*: BERGENHOLTZ, Henning; NIELSEN, Sandro; TARP, Sven. (Eds.). **Lexicography at a Crossroads**. Dictionaries and Encyclopedias today, lexicographical tools tomorrow. Bern: Peter Lang, 2009. p. 69-89.

VOLKOVYSKAYA, Evgenia. **Lexico-semantic processing in adult monolingual Russian and bilingual Russian (L1) - English (L2) Speakers**. 2017. 203 f. Tese (Doutorado) – Middlesex University, London, Reino Unido, 2017. Disponível em: <http://eprints.mdx.ac.uk/22968/>. Acesso em: 22 jul. 2020.

VYGOTSKY, Lev S. **A formação social da mente**. 4ª ed. São Paulo: Martins Fontes, 1991.

WALTER, Elizabeth; WOODFORD, Kate. **Common Errors in English**. Your essential guide to accurate English. Glasgow: HarperCollins, 2015.

WARTENBURGER, Isabell; HEEKEREN, Hauke; ABUTALEBI, Jubin; CAPPA, Stefano; VILLRINGER, Arno; PERANI, Daniela. Early setting of grammatical processing in the bilingual brain. **Neuron**, [s.l.], v. 37, p. 159-170, 2003.

WB. COLLINS. **WordBanks**. [S.l.]: Collins, 2005. Disponível em: <https://wordbanks.harpercollins.co.uk/>. Acesso em: 19 jun. 2019.

WEIR, Cyril. Limitations of the Common European Framework for developing comparable examinations and tests. **Language Testing**, [s.l.] v. 22, n. 3, p. 281-300, 2005.

WELKER, Herbert Andreas. **Panorama Geral da Lexicografia Pedagógica**. Brasília: Thesaurus, 2008. Disponível em: <https://hawelk.wixsite.com/hawelker/livros-publicados>. Acesso em: 11 set. 2019.

WELKER, Herbert Andreas. **Dictionary Use: A general survey of empirical studies**. Brasília: Author's Edition, 2010. Disponível em: <https://hawelk.wixsite.com/hawelker/livros-publicados>. Acesso em: 5 out. 2019.

WHALEY, Lindsay J. **Introduction to Typology: The Unity and Diversity of Language**. Thousand Oaks; London; New Delhi: Sage, 1997.

WOLFER, Sascha; KOSEM, Iztok; LEW, Robert; *et al.* Web-based Exploration of Results From a Large European Survey on Dictionary Use and Culture: ESDexplorer. **Lexikos**, Stellenbosch, v. 28, p. 440–447, 2018. Disponível em: <https://doi.org/10.5788/28-1-1473>. Acesso em: 05 out. 2019.

WRAY, Alison. Formulaic language. **Language Teaching**, [s.l.]. v. 46, n. 3, p. 316-334, 2013.

WU, Yan Jing; THIERRY, Guillaume. Unconscious translation during incidental foreign language processing. **Neuroimage**, [s.l.], v. 59, n. 4, p. 3468-3473, 2012. DOI: <https://doi.org/10.1016/j.neuroimage.2011.11.049>.

WU, Yan; CRISTINO, Filipe; LEEK, Charles; THIERRY, Guillaume. Non-selective lexical access in bilinguals is spontaneous and independent of input monitoring: Evidence from eye tracking. **Cognition**, [s.l.], v. 129, n. 2, p. 418-425, 2013. DOI: <https://doi.org/10.1016/j.cognition.2013.08.005>

YIP, Virginia. Simultaneous Language Acquisition. In: GROSJEAN, François; LI, Ping. **The Psycholinguistics of Bilingualism**. Oxford: Wiley-Blackwell, 2013. Capítulo 6, p. 109-131.

ZACARIAS, Regiani Aparecida Santos. Researching Brazilian students' needs and proposing lexicographical solutions for Portuguese-English learner's dictionaries. In: FERREIRA, A. A. G. D.; FARGETTI, C. M.; MURAKAWA, C. A. A. (org.). **Variedades do léxico**. Araraquara: Letraria, 2015. *E-book*. Disponível em: <https://www.lettraria.net/variedades-do-lexico/>. Acesso em: 07 mai. 2019.

ZANATTA, Flávia. **A normatividade e seu reflexo em dicionários semasiológicos de Língua Portuguesa**. 2010. 270f. Dissertação (Mestrado em Letras) – Programa de Pós-

Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010. Disponível em: <http://hdl.handle.net/10183/25429>. Acesso em: 24 out. 2017.

ZAVAGLIA, Adriana. A elaboração de um verbete-modelo para a construção de um dicionário temático bilíngüe de cores francês-português. **Estudos Linguísticos**, São Paulo, v. 29, p. 843-848, 2000. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos>. Acesso em: 20 fev. 2021.

ZAVAGLIA, Claudia. Unidades paremiológicas em dicionários bilíngües: em defesa da definição. In: SILVA, Suzete (org.). **Fraseologia & CIA**. vol. 2. Campinas: Pontes Editores, 2016 [impressão: 2020]. Capítulo 3.

ZAVAGLIA, Claudia; ZAVAGLIA, Adriana. A elaboração de um dicionário trilíngüe temático de cromônimos italiano-português-francês/francês-português-italiano: reflexões e considerações. **Linguística**, Madrid/São Paulo, v. 12, p. 235-247, 2000a. Disponível em: https://www.mundoalfal.org/es/pt_vol12. Acesso em: 20 fev. 2021.

ZAVAGLIA, Claudia; ZAVAGLIA, Adriana. A elaboração de um dicionário trilíngüe temático de cromônimos italiano-português-francês/francês-português-italiano: reflexões e considerações. **Cadernos do CNLF**, São Paulo, v. 4, n. 10, [s.p.], 2000b. Disponível em: <http://filologia.org.br/>. Acesso em: 20 fev. 2021.

ZHAO, Gang. Making a User-friendly Bilingual Dictionary for Chinese Translators: on the Revision of A New Century Chinese-English Dictionary. **International Journal of Lexicography**, Oxford, v. 29, n. 4, p. 452-489, 2015. DOI: <https://doi.org/10.1093/ijl/ecv031>.

ANEXO 1

CaED2 (2011)	
Verbetes de unidades léxicas A2	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
appointment (ac.) a time you have arranged to see someone	B1 (present perfect simple [recent past]: <i>have arranged</i>); A2 (subordinated clause with <i>to</i> [purpose])
landscape (ac.) the appearance of an area of land, especially in the countryside	B1 (focusing adverb <i>especially</i>); A1 (prepositions <i>of, in</i> + noun phrase)
library (ac.) a place with a lot of books that you can read or borrow	A2 (that-clause); A1 (conjunction <i>or</i>)
glad (ac.) happy about something	A1 (phrase)
pleasant (ac.) good or enjoyable	A1 (conjunction <i>or</i>)
remember (ac.2) 2 remember to do something (A2) to not forget to do something	A1 (clause)
rent (ac.1) to pay money to live in a building that someone else owns	A2 (subordinated clause with <i>to</i> [purpose]; that-clause)
rent (ac.2) to pay money to use something for a short time	A2 (subordinated clause with <i>to</i> [purpose]); A1 (preposition <i>for</i> + noun phrase)
rent (n.) Common mistake: rent and hire In British English you rent something for a long time. [...] In British English you hire something for a short time. [...] In American English the word rent is used in both situations.	A2 (passive: <i>is used</i>); A1 (preposition <i>for</i> + noun phrase)
already (ac.1) before now, or before a time in the past	A2 (preposition <i>before</i>); A1 (conjunction <i>or</i> ; preposition <i>in</i> + noun phrase)
across (ac.1) from one side to the other	A1 (preposition <i>from... to</i> + noun phrase)
across (ac.2) on the opposite side of	B2 (stranded preposition: <i>of</i>); A1 (preposition <i>on</i> + noun phrase)
among (ac.2) in a particular group	A1 (preposition <i>in</i> + noun phrase)

Verbetes de unidades léxicas B1	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
edge (ac.) the part around something that is furthest from the centre	B1 (<i>furthest</i>); A2 (that-clause)
flavour (ac.) the taste of a type of food or drink	A1 (conjunction <i>or</i> ; preposition <i>of</i> + noun phrase)
label ¹ (ac.1) a piece of paper or material that is attached to something and gives you information about it	A2 (that-clause; passive); A1 (conjunctions <i>or</i> , <i>and</i>)
label ¹ (ac.2) a company that makes things for sale, mainly clothes	B1 (focusing adverb <i>mainly</i>); A2 (that-clause)
gorgeous (ac.) very beautiful or pleasant	A1 (conjunction <i>or</i>)
likely (ac.1) expected	C1 (past participle as adjective: <i>expected</i>)
rent (ac.3) (also rent out) to allow someone to pay you money to live in your building	A2 (subordinated clause with <i>to</i> [purpose])
bury (ac.1) to put a dead body into the ground	A1 (preposition <i>into</i> + noun phrase)
remind (ac.1) to make someone remember something, or remember to do something	A1 (conjunction <i>or</i>)
remind (n.) Common mistake: remind or remember ? If you remember a fact or something from the past, you keep it in your mind, or bring it back into your mind. [...] When you remind someone to do something, you make them remember it. [...]	B2 (them [gender neutral]); A2 (conditional clause; when-clause); A1 (conjunction <i>or</i> ; prepositions <i>in</i> , <i>into</i> + noun phrase)
remind (phr. v.) remind someone of something/someone phrasal verb. (B1) to make someone think of something or someone else	A1 (conjunction <i>or</i>)
already (ac.2) used to say that something has happened earlier than you expected	A2 (that-clause; present perfect tense: <i>has happened</i> ; comparative adjective: <i>earlier than</i>)
pretty ¹ (ac.) quite, but not very	A2 (degree adverb <i>quite</i>); A1 (conjunction <i>but</i>)

among (ac.1) <i>in the middle of something</i>	<i>A2 (in the middle of)</i>
towards (ac.1) <i>in the direction of someone or something</i>	<i>B1 (in the direction of);</i> <i>A1 (conjunction or)</i>

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxEED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 2

OxED2 (2012)	
Verbetes de unidades léxicas A2	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
appointment (ac.) an arrangement to see somebody at a particular time	A2 (subordinated clause with <i>to</i> [purpose]); A1 (preposition <i>at</i> + noun phrase)
landscape (ac.) everything you can see in an area of land	A1 (preposition <i>in</i> + noun phrase)
library (ac.1) a room or building where you go to borrow or read books	B1 (relative where-clause); A2 (subordinated clause with <i>to</i> [purpose]); A1 (preposition <i>with</i> ; conjunction <i>or</i>)
library (n.) WHICH WORD? Be careful! You cannot buy books from a library. The place you buy books is called a bookshop or a bookstore	A2 (passive: <i>is called</i>); A1 (conjunction <i>or</i>)
glad (ac.1) happy about something	A1 (phrase)
glad (n.) WORD BUILDING You are usually glad or pleased about a particular event or situation: [...] You use happy to describe a state of mind [...] or before a noun	A2 (adverb of frequency <i>usually</i> ; subordinated clause with <i>to</i> [purpose]); A1 (conjunction <i>or</i> ; preposition <i>about</i> + noun phrase)
pleasant (ac.) nice, enjoyable or friendly	A1 (conjunction <i>or</i>)
rent (ac.1) to pay to live in a place or to use something that belongs to another person	A2 (subordinated clause with <i>to</i> [purpose]; that-clause); A1 (conjunction <i>or</i>)
already (ac.1) before now or before then	A2 (preposition <i>before</i>); A1 (conjunction <i>or</i>)
already (n.) WHICH WORD? Already or yet? Yet means the same as already, but you only use yet in negative sentences and questions	B1 (adverb of restriction <i>only</i> [nowhere else]); A1 (conjunctions <i>but</i> , <i>and</i> ; preposition <i>in</i> + noun phrase)
across (ac.1) from one side to the other side of something	A1 (prepositions <i>from... to</i> , <i>of</i> + noun phrase)
across (ac.2) on the other side of something	A1 (prepositions <i>on</i> , <i>of</i> + noun phrase)

<p>across (n.)</p> <p>WHICH WORD?</p> <p>Across or over?</p> <p>We can use across or over to mean ‘on or to the other side’. [...] We usually use over to talk about crossing something high. [...] With ‘room’ we usually use across</p>	<p>B1 (verb + gerund: <i>talk about crossing</i>; <i>we</i> [all people]);</p> <p>A2 (subordinated clause with <i>to</i> [purpose]; adverb of frequency <i>usually</i>);</p> <p>A1 (prepositions <i>on, to, with</i> + noun phrase; conjunction <i>or</i>)</p>
<p>among (ac.2)</p> <p>in a particular group of people or things</p>	<p>A1 (prepositions <i>in, of</i> + noun phrase; conjunction <i>or</i>)</p>
<p>among (n.)</p> <p>WHICH WORD?</p> <p>Among or between?</p> <p>We use among when we are talking about more than two people or things: [...] If there are only two people or things, we use between</p>	<p>B1 (<i>we</i> [all people]);</p> <p>A2 (when-clause; comparative phrase <i>more than</i>; conditional clause; adverb of restriction <i>only</i>);</p> <p>A1 (conjunction <i>or</i>)</p>
Verbetes de unidades léxicas B1	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
<p>edge (ac.1)</p> <p>the part along the end or side of something</p>	<p>B1 (preposition <i>along</i> + noun phrase);</p> <p>A1 (conjunction <i>or</i>)</p>
<p>flavour (ac.)</p> <p>the taste of food</p>	<p>A1 (preposition <i>of</i> + noun phrase)</p>
<p>label¹ (ac.)</p> <p>a piece of paper or material on something that tells you about it</p>	<p>A2 (that-clause);</p> <p>A1 (conjunction <i>or</i>)</p>
<p>gorgeous (ac.)</p> <p>very good or attractive</p>	<p>A1 (conjunction <i>or</i>)</p>
<p>likely (ac.)</p> <p>If something is likely, it will probably happen</p>	<p>B1 (<i>will</i> to express predictions);</p> <p>A2 (conditional clause)</p>
<p>rent (ac.2)</p> <p>to let somebody live in a place or use something that belongs to you, if they pay you</p>	<p>B2 (they [gender neutral]);</p> <p>A2 (that-clause; conditional clause);</p> <p>A1 (conjunction <i>or</i>)</p>
<p>bury (ac.1)</p> <p>to put a dead body in the ground</p>	<p>A1 (preposition <i>in</i> + noun phrase)</p>
<p>remind (ac.1)</p> <p>to help somebody remember something that they must do</p>	<p>B2 (they [gender neutral]);</p> <p>A2 (that-clause)</p>
<p>remind (ac.2)</p> <p>to make somebody remember somebody or something</p>	<p>A1 (conjunction <i>or</i>)</p>
<p>pretty¹ (ac.)</p> <p>quite; fairly</p>	<p>A2 (degree adverb <i>quite</i>)</p>

among (ac.1) <i>in the middle of a group of people or things</i>	<i>A2 (in the middle of);</i> <i>A1 (preposition of + noun phrase; conjunction or)</i>
towards (ac.1) <i>in the direction of somebody or something</i>	<i>B1 (in the direction of);</i> <i>A1 (conjunction or)</i>

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxEED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 3

CoED2 (2014)	
Verbetes de unidades léxicas A2	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
appointment (ac.) an arrangement to see someone at a particular time	A2 (subordinated clause with <i>to</i> [purpose]); A1 (preposition <i>at</i> + noun phrase)
landscape (ac.) everything you can see when you look across an area of land	A2 (when-clause)
library (ac.) a place where books, newspapers, DVDs and music are kept for people to use or borrow	B1 (relative where-clause); A2 (passive: <i>are kept</i> ; subordinated clause with <i>to</i> [purpose]); A1 (conjunctions <i>and</i> , <i>or</i>)
glad (ac.) happy and pleased about something	A1 (conjunction <i>and</i>)
pleasant (ac.1) enjoyable or attractive	A1 (conjunction <i>or</i>)
pleasant (ac.2) nice and friendly	A1 (conjunction <i>and</i>)
remember (ac.3) to not forget to do something	A1 (clause)
rent (ac.1) to pay the owner of something in order to be able to use it yourself	B1 (conjunction <i>in order to</i> [purpose])
already (ac.1) used for showing that one thing happened before another thing	B1 (verb + gerund: <i>used for showing</i>); A2 (that-clause; adverb of time <i>before</i>)
already (n.) LANGUAGE HELP already or yet? In British English, you usually use already and yet with the perfect tenses. [...] In American English, people often use a past tense	A2 (adverbs of frequency <i>usually</i> , <i>often</i>) A1 (conjunction <i>and</i> ; preposition <i>in</i> + noun phrase)
across (ac.1) from one side of something to the other side of it	A1 (prepositions <i>from... to</i> , <i>of</i> + noun phrase)
across (ac.2) going from one side of something to the other side of	B2 (subject in gerund: <i>going</i>); A1 (preposition + noun phrase)
among (ac.2) included in a particular group	A2 (passive); A1 (preposition + noun phrase)

<p>among (ac.3) used for saying that most people in a group have a particular opinion or feeling</p>	<p>B1 (verb + gerund: <i>used for saying</i>; quantifying determiner <i>most</i>); A2 (that-clause); A1 (preposition <i>in</i> + noun phrase; conjunction <i>or</i>)</p>
<p>among (n.) LANGUAGE HELP among or between? If there are more than two people or things, use among. If there are only two people or things, use between</p>	<p>A2 (conditional clause; comparative phrase <i>more than</i>; adverb of restriction <i>only</i>)</p>
Verbetes de unidades léxicas B1	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
<p>edge (ac.) the part of something that is farthest from the middle</p>	<p>A2 (that-clause; superlative with one syllable adjective <i>farthest</i>)</p>
<p>flavour (ac.) the taste of a food or drink</p>	<p>A1 (conjunction <i>or</i>; preposition <i>of</i> + noun phrase)</p>
<p>label (ac.) a piece of paper or plastic that is attached to an object to give information about it</p>	<p>A2 (that-clause; passive; subordinated clause with <i>to</i> [purpose]); A1 (conjunction <i>or</i>)</p>
<p>gorgeous (ac.) very pleasant or attractive</p>	<p>A1 (conjunction <i>or</i>)</p>
<p>likely (ac.2) used for saying that a person or thing will probably do something</p>	<p>B1 (<i>will</i> to express predictions); A2 (that-clause); A1 (conjunction <i>or</i>)</p>
<p>rent (ac.2) (also rent something out) to let someone have and use something in exchange for money</p>	<p>B1 (complex preposition: <i>in exchange for</i>); A1 (conjunction <i>and</i>)</p>
<p>bury (ac.2) to put the body of a dead person into a grave and cover it with earth</p>	<p>A1 (prepositions <i>of</i>, <i>into</i>, <i>with</i> + noun phrase; conjunction <i>and</i>)</p>
<p>remind (ac.1) to say something that makes someone think about a fact or an event that they already know about</p>	<p>B2 (stranded preposition: <i>about</i>; they [gender neutral]); A2 (that-clause); A1 (conjunction <i>or</i>)</p>
<p>remind (ac.2) to say something that makes someone remember to do something</p>	<p>A2 (that-clause)</p>
<p>remind (ac.3) to be similar to another person or thing and make someone think about them</p>	<p>B2 (them [gender neutral]); A1 (conjunctions <i>or</i>, <i>and</i>)</p>
<p>already (ac.2) used for showing that a situation exists now or that it started earlier than expected</p>	<p>B1 (verb + gerund: <i>used for showing</i>); A2 (that-clause; comparative adjective: <i>earlier than</i>); A1 (conjunction <i>or</i>)</p>

<p>pretty (ac.2) used before an adjective or adverb to mean 'fairly'</p>	<p>B1 (preposition <i>before</i> [in order]); A2 (subordinated clause with <i>to</i> [purpose]); A1 (conjunction <i>or</i>)</p>
<p>among (ac.1) in the middle of a group of people or things</p>	<p>A2 (<i>in the middle of</i>); A1 (preposition <i>of</i> + noun phrase; conjunction <i>or</i>)</p>
<p>towards (ac.1) in the direction of something or someone</p>	<p>B1 (<i>in the direction of</i>); A1 (conjunction <i>or</i>)</p>

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 4

LWord2 (2008)	
Verbetes de unidades léxicas A2	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
appointment (ac.1) a meeting that has been arranged for a particular time and place	B1 (present perfect simple [recent past]: <i>has been arranged for</i>) A2 (that-clause; passive)
landscape (ac.) a large area of land that you can see	A2 (that-clause)
library (ac.) a room or building containing books that you can borrow, and computers that you can use to get information	A2 (that-clause; subordinated clause with <i>to</i> [purpose]) A1 (conjunctions <i>and, or</i>)
library (n. 1) SPELLING This word is often spelled wrongly. The correct spelling is: library .	A2 (passive; adverb of frequency <i>often</i>)
library (n. 2) WORD CHOICE library or bookshop? A library is a place where you can borrow books for a short time. A bookshop or bookstore is a place where you can buy books.	B1 (relative where-clause) A1 (conjunction <i>or</i>)
glad (ac.) if you are glad, you are pleased and happy about something	A2 (conditional clause) A1 (conjunction <i>and</i>)
pleasant (ac.1) enjoyable or nice	A1 (conjunction <i>or</i>)
pleasant (ac.2) polite and friendly	A1 (conjunction <i>and</i>)
remember (ac.2) if you remember to do something, you do not forget to do it	A2 (conditional clause)
rent (ac.1) to pay money to live in a place or to use something such as a car	A2 (subordinated clause with <i>to</i> [purpose]; <i>such as</i>) A1 (conjunction <i>or</i>)
already (ac.1) before	A2 (adverb of time <i>before</i>)
already (n.) WORD CHOICE already or yet?	A2 (adverb of frequency <i>usually</i>) A1 (conjunction <i>and</i> ; prepositions <i>in, at</i> + noun phrase)

<p>You use already in positive statements: [...]</p> <p>You usually use yet in negative sentences and questions: [...]</p> <p>Yet usually comes at the end of a sentence.</p>	
<p>across (ac.1)</p> <p>from one side of something to the other</p>	A1 (prepositions <i>from... to; of + noun phrase</i>)
<p>across (ac.2)</p> <p>on the other side of something</p>	A1 (prepositions <i>on, of + noun phrase</i>)
<p>among (ac.1)</p> <p>used to say which group of people is affected by something</p>	A2 (passive) A1 (determiner <i>which</i>)
Verbetes de unidades léxicas B1	
Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)	
Classificação dos definidores conforme EVP (2015)	Classificação dos padrões gramaticais conforme EGP (2015)
<p>edge (ac.1)</p> <p>the part of something that is furthest from the middle</p>	B1 (superlative adjective <i>furthest</i>) A2 (that-clause)
<p>flavour (ac.1)</p> <p>the taste that a food or drink has</p>	A2 (that-clause) A1 (conjunction <i>or</i>)
<p>label¹ (ac.1)</p> <p>a piece of paper or cloth with information on it that is fixed to an object</p>	A2 (that-clause; passive) A1 (conjunction <i>or</i>)
<p>label¹ (ac.2)</p> <p>a record company</p>	A1 (phrase)
<p>gorgeous (ac.)</p> <p>very beautiful or pleasant</p>	A1 (conjunction <i>or</i>)
<p>likely (ac.1)</p> <p>if something is likely to happen, it will probably happen</p>	B1 (<i>will</i> to express predictions) A2 (conditional clause)
<p>rent (ac.2)</p> <p>also rent out</p> <p>if you rent a house, room etc to someone, you allow them to use it, and they pay you money</p>	B2 (they, them [gender neutral]); A2 (conditional clause) A1 (conjunction <i>and</i>)
<p>bury (ac.1)</p> <p>to put a dead body into a GRAVE (=a hole that has been dug in the ground for this purpose)</p>	B2 (passive form of present perfect simple: <i>has been dug</i>) A2 (that-clause) A1 (prepositions <i>into, in, for + noun phrase</i>)
<p>remind (ac.1)</p> <p>to make someone remember something that they must do</p>	B2 (they [gender neutral]); A2 (that-clause)
<p>remind (ac.2)</p> <p>remind someone of someone</p>	B2 (they, them [gender neutral]) A2 (conditional clause) A1 (conjunction <i>because</i>)

if someone reminds you of another person, they make you think of that person because they are similar to them	
already (ac.2) now, rather than later	B1 (comparative clause with <i>rather than</i>)
pretty ¹ (ac.) fairly or quite	A2 (degree adverb <i>quite</i>) A1 (conjunction <i>or</i>)
among (ac.2) surrounded by things	A1 (phrase)
towards (ac.1) in the direction of someone or something	B1 (<i>in the direction of</i>) A1 (conjunction <i>or</i>)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), O_xED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 5

Tipos de definidores	Incidência de paráfrases e notas de uso em cada dicionário			
	CaED2	OxED2	CoED2	LWord2
Un. léxicas A2 com definidores A1 somente	<i>glad</i> (ac.) <i>remember</i> (ac. 2)	<i>glad</i> (ac.)	<i>remember</i> (ac. 3)	-
Un. léxicas A2 com definidores A2	<i>library</i> (ac.) <i>rent</i> (ac. 2)	<i>library</i> (n.) <i>library</i> (ac.) <i>rent</i> (ac. 1) <i>already</i> (ac.)	<i>library</i> (ac.) <i>glad</i> (ac.) <i>pleasant</i> (ac. 2)	<i>library</i> (n. 2) <i>glad</i> (ac.) <i>pleasant</i> (ac. 2) <i>remember</i> (ac. 2) <i>rent</i> (ac. 1) <i>already</i> (ac. 1)
Un. léxicas A2 com definidores B1+	<i>appointment</i> (ac.) <i>landscape</i> (ac.) <i>pleasant</i> (ac.) <i>rent</i> (ac. 1) <i>rent</i> (n.) <i>already</i> (ac. 1) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 2)	<i>appointment</i> (ac.1) <i>landscape</i> (ac.) <i>glad</i> (n.) <i>pleasant</i> (ac.) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>across</i> (n.) <i>among</i> (ac. 2) <i>among</i> (n.)	<i>appointment</i> (ac.1) <i>landscape</i> (ac. 1) <i>pleasant</i> (ac. 1) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>across</i> (ac. 2) <i>among</i> (ac. 2) <i>among</i> (ac. 3) <i>among</i> (n.)	<i>appointment</i> (ac. 1) <i>landscape</i> (ac.) <i>library</i> (ac.) <i>library</i> (n. 1) <i>pleasant</i> (ac. 1) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 1)
Un. léxicas B1 com definidores A1 e/ou A2	<i>gorgeous</i> (ac.) <i>pretty</i> ^l (ac.) <i>among</i> (ac. 1)	<i>edge</i> (ac. 1) <i>gorgeous</i> (ac.) <i>among</i> (ac. 1)	<i>edge</i> (ac. 1) <i>gorgeous</i> (ac.) <i>among</i> (ac. 1)	<i>label</i> ^l (ac. 2) <i>gorgeous</i> (ac.)
Un. léxicas B1 com definidores B1	<i>edge</i> (ac.) <i>flavour</i> (ac.) <i>label</i> ^l (ac. 2) <i>rent</i> (ac. 3) <i>bury</i> (ac. 1) <i>remind</i> (ac. 1) <i>remind</i> (n.) <i>remind</i> (ph.v.) <i>towards</i> (ac. 1)	<i>flavour</i> (ac.) <i>likely</i> (ac.) <i>rent</i> (ac. 2) <i>bury</i> (ac. 1) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>pretty</i> ^l (ac.) <i>towards</i> (ac. 1)	<i>flavour</i> (ac.) <i>label</i> (ac.) <i>rent</i> (ac. 2) <i>bury</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>remind</i> (ac. 2) <i>remind</i> (ac. 3) <i>towards</i> (ac. 1)	<i>edge</i> (ac. 1) <i>flavour</i> (ac.) <i>likely</i> (ac. 1) <i>rent</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>already</i> (ac. 2) <i>pretty</i> ^l (ac.) <i>among</i> (ac. 2) <i>towards</i> (ac. 1)
Un. léxicas B1 com definidores B2+	<i>label</i> ^l (ac. 1) <i>already</i> (ac. 2) <i>likely</i> (ac. 1)	<i>label</i> ^l (ac.)	<i>likely</i> (ac. 2) <i>already</i> (ac. 2) <i>pretty</i> (ac. 2)	<i>label</i> ^l (ac. 1) <i>bury</i> (ac. 1)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 6

Tipos de padrões gramat.	Incidência de paráfrases e notas de uso em cada dicionário			
	CaED2	OxED2	CoED2	LWord2
Un. léxicas A2 com padrões gramaticais A1	<i>glad</i> (ac.) <i>pleasant</i> (ac.) <i>remember</i> (ac. 2) <i>across</i> (ac. 1) <i>among</i> (ac. 2)	<i>landscape</i> (ac.) <i>glad</i> (ac.) <i>pleasant</i> (ac.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 2)	<i>glad</i> (ac.) <i>pleasant</i> (ac. 2) <i>pleasant</i> (ac. 1) <i>remember</i> (ac. 3) <i>across</i> (ac. 1)	<i>pleasant</i> (ac. 1) <i>pleasant</i> (ac. 2) <i>across</i> (ac. 1) <i>across</i> (ac. 2)
Un. léxicas A2 com padrões gramaticais A2	<i>library</i> (ac.) <i>rent</i> (ac. 1) <i>rent</i> (ac. 2) <i>rent</i> (n.) <i>already</i> (ac. 1)	<i>appointment</i> (ac.1) <i>library</i> (n.) <i>glad</i> (n.) <i>rent</i> (ac. 1) <i>already</i> (ac.) <i>among</i> (n.)	<i>appointment</i> (ac.1) <i>landscape</i> (ac. 1) <i>already</i> (n.) <i>among</i> (ac. 2) <i>among</i> (n.)	<i>landscape</i> (ac.) <i>library</i> (ac.) <i>library</i> (n. 1) <i>glad</i> (ac.) <i>remember</i> (ac. 2) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>already</i> (n.) <i>among</i> (ac. 1)
Un. léxicas A2 com padrões gramaticais B1+	<i>appointment</i> (ac.) <i>landscape</i> (ac.) <i>across</i> (ac. 2)	<i>library</i> (ac.) <i>already</i> (n.) <i>across</i> (n.)	<i>library</i> (ac.) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 3)	<i>library</i> (n. 2) <i>appointment</i> (ac. 1)
Un. léxicas B1 com padrões gramaticais A1 e/ou A2	<i>flavour</i> (ac.) <i>label^l</i> (ac. 1) <i>gorgeous</i> (ac.) <i>rent</i> (ac. 3) <i>bury</i> (ac. 1) <i>remind</i> (ac. 1) <i>remind</i> (ph.v.) <i>already</i> (ac. 2) <i>pretty^l</i> (ac.) <i>among</i> (ac. 1)	<i>flavour</i> (ac.) <i>label^l</i> (ac.) <i>gorgeous</i> (ac.) <i>likely</i> (ac.) <i>bury</i> (ac. 1) <i>remind</i> (ac. 2) <i>pretty^l</i> (ac.) <i>among</i> (ac. 1)	<i>edge</i> (ac. 1) <i>flavour</i> (ac.) <i>label</i> (ac.) <i>gorgeous</i> (ac.) <i>likely</i> (ac. 2) <i>bury</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>among</i> (ac. 1)	<i>flavour</i> (ac.) <i>label^l</i> (ac. 1) <i>label^l</i> (ac. 2) <i>gorgeous</i> (ac.) <i>likely</i> (ac. 1) <i>pretty^l</i> (ac.) <i>among</i> (ac. 2)
Un. léxicas B1 com padrões gramaticais B1	<i>edge</i> (ac.) <i>label^l</i> (ac. 2) <i>towards</i> (ac. 1)	<i>edge</i> (ac. 1) <i>remind</i> (ac. 1) <i>towards</i> (ac. 1)	<i>rent</i> (ac. 2) <i>already</i> (ac. 2) <i>remind</i> (ac. 3) <i>pretty</i> (ac. 2) <i>towards</i> (ac. 1)	<i>edge</i> (ac. 1) <i>rent</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>already</i> (ac. 2) <i>towards</i> (ac. 1)
Un. léxicas B1 com padrões gramaticais B2+	<i>likely</i> (ac. 1) <i>remind</i> (n.)	<i>rent</i> (ac. 2)	-	<i>bury</i> (ac. 1)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 7

Tipos de elementos	Incidência de paráfrases e notas de uso em cada dicionário			
	CaED2	OxED2	CoED2	LWord2
Un. léxicas A2 definidas por elementos A1	<i>glad</i> (ac.) <i>remember</i> (ac. 2)	<i>glad</i> (ac.)	<i>remember</i> (ac. 3)	-
Un. léxicas A2 definidas por elementos A2	<i>library</i> (ac.) <i>rent</i> (ac. 2)	<i>library</i> (n.) <i>rent</i> (ac. 1) <i>already</i> (ac.)	<i>glad</i> (ac.) <i>pleasant</i> (ac. 2)	<i>glad</i> (ac.) <i>pleasant</i> (ac. 2) <i>remember</i> (ac. 2) <i>rent</i> (ac. 1) <i>already</i> (ac. 1)
Un. léxicas A2 definidas por elementos B1+	<i>appointment</i> (ac.) <i>landscape</i> (ac.) <i>pleasant</i> (ac.) <i>rent</i> (ac. 1) <i>rent</i> (n.) <i>already</i> (ac. 1) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 2)	<i>appointment</i> (ac.1) <i>landscape</i> (ac.) <i>library</i> (ac.) <i>glad</i> (n.) <i>pleasant</i> (ac.) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>across</i> (n.) <i>among</i> (ac. 2) <i>among</i> (n.)	<i>appointment</i> (ac.1) <i>landscape</i> (ac. 1) <i>library</i> (ac.) <i>pleasant</i> (ac. 1) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>across</i> (ac. 2) <i>among</i> (ac. 2) <i>among</i> (ac. 3) <i>among</i> (n.)	<i>appointment</i> (ac. 1) <i>landscape</i> (ac.) <i>library</i> (ac.) <i>library</i> (n. 1) <i>library</i> (n. 2) <i>pleasant</i> (ac. 1) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 1)
Un. léxicas B1 definidas por elementos A1 e/ou A2	<i>gorgeous</i> (ac.) <i>pretty</i> ^l (ac.) <i>among</i> (ac. 1)	<i>gorgeous</i> (ac.) <i>likely</i> (ac.) <i>among</i> (ac. 1)	<i>edge</i> (ac. 1) <i>gorgeous</i> (ac.) <i>among</i> (ac. 1)	<i>label</i> ^l (ac. 2) <i>gorgeous</i> (ac.) <i>likely</i> (ac. 1)
Un. léxicas B1 definidas por elementos B1	<i>edge</i> (ac.) <i>flavour</i> (ac.) <i>label</i> ^l (ac. 2) <i>rent</i> (ac. 3) <i>bury</i> (ac. 1) <i>remind</i> (ac. 1) <i>remind</i> (ph.v.) <i>towards</i> (ac. 1)	<i>edge</i> (ac. 1) <i>flavour</i> (ac.) <i>bury</i> (ac. 1) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>pretty</i> ^l (ac.) <i>towards</i> (ac. 1)	<i>flavour</i> (ac.) <i>label</i> (ac.) <i>rent</i> (ac. 2) <i>bury</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>remind</i> (ac. 3) <i>towards</i> (ac. 1)	<i>edge</i> (ac. 1) <i>flavour</i> (ac.) <i>rent</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>already</i> (ac. 2) <i>pretty</i> ^l (ac.) <i>among</i> (ac. 2) <i>towards</i> (ac. 1)
Un. léxicas B1 definidas por elementos B2+	<i>label</i> ^l (ac. 1) <i>likely</i> (ac. 1) <i>remind</i> (n.) <i>already</i> (ac. 2)	<i>label</i> ^l (ac.) <i>rent</i> (ac. 2)	<i>likely</i> (ac. 2) <i>already</i> (ac. 2) <i>pretty</i> (ac. 2)	<i>label</i> ^l (ac. 1) <i>bury</i> (ac. 1)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 8

Tipo de romanismo	Romanismo do inglês	Símile ortográfico em português	S.O.	Equivalente em português
1	beautiful	-	-	bonito
1	because	-	-	porque
1	money	moeda	0,48	dinheiro
1	pay	-	-	pagar
1	people	-	-	pessoas
1	place	-	-	lugar
1	very	-	-	muito
1	around	-	-	ao redor
1	pleasant	-	-	agradável
1	pleased	-	-	satisfeito; contente
1	probably	provável	0,46	provavelmente
1	type	tipo	0,38	tipo
1	allow	-	-	permitir
1	exchange	-	-	troca
1	expected	esperar	0,38	esperado
1	mind	-	-	mente
1	statement	-	-	declaração
1	surrounded	-	-	rodeado
1	taste	-	-	sabor
Total de romanismos tipo 1: 19				
2	person	personagem	0,65	pessoa
2	large	largo	0,70	grande
2	arranged	arranjar	0,65	agendado
2	enjoyable	enjoativo	0,52	agradável
2	past	pasta	0,73	passado
2	material	material	1,00	tecido
2	particular	particular	1,00	determinado
2	record	recorde	0,77	gravadora [record company]
Total de romanismos tipo 2: 8				
3	car	carro	0,59	carro
3	computer	computador	0,80	computador
3	group	grupo	0,62	grupo
3	music	música	0,75	música
3	paper	papel	0,73	papel
3	part	parte	0,73	parte
3	question	questão	0,64	pergunta
3	adjective	adjetivo	0,67	adjetivo
3	adverb	advérbio	0,72	advérbio
3	area	área	1,00	área
3	attractive	atrativo	0,67	atrativo
3	centre	centro	0,72	centro
3	company	companhia	0,67	companhia
3	correct	correto	0,69	correto
3	describe	descrever	0,53	descrever
3	fact	fato	0,53	fato
3	included	incluir	0,63	incluído
3	information	informação	0,61	informação
3	negative	negativo	0,75	negativo

3	opinion	opinião	0,70	opinião
3	piece	peça	0,57	pedaço
3	plastic	plástico	0,78	plástico
3	usually	usualmente	0,60	normalmente
3	containing	contendo	0,51	contendo
3	direction	direção	0,62	direção
3	especially	especial	0,76	especialmente
3	event	evento	0,78	acontecimento
3	exist	existe (verbo flex.)	0,75	existe (verbo flex.)
3	object	objeto	0,64	objeto
3	opposite	oposto	0,61	oposto
3	appearance	aparência	0,56	aparência
3	similar	similar	1,00	parecido
3	situation	situação	0,59	situação
3	fixed	fixado	0,56	fixado
3	affected	afetar	0,54	afetado
3	use	usa (verbo flex.)	0,59	usa (verbo flex.)
3	used	usar	0,51	usado
3	positive	positivo	0,77	afirmativo
Total de romanismos tipo 3: 38				
4	remember	rememorar	0,77	lembrar
4	sentence	sentença	0,75	oração
2	polite	polido	0,63	educado
4	arrangement	arranjo	0,59	combinado
4	attached	atachado	0,65	anexado
4	cover	cobrir	0,57	cobrir
4	purpose	propósito	0,51	fim [for this purpose]
Total de romanismos tipo 4: 7				
Total absoluto de romanismos: 72				

Fonte: Elaboração da autora a partir dos dados do CaED2 (2011), OxE2 (2012), CoED2 (2014), LWord2 (2008) e da ferramenta NIM (2013)

Legenda: S.O. = índice de similaridade ortográfica; ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do EVP (2015)

ANEXO 9

Tipos de elementos	Incidência de paráfrases e notas de uso em cada dicionário			
	CaED2	OxED2	CoED2	LWord2
Un. léxicas A2 definidas por elementos A1	<i>glad</i> (ac.) <i>remember</i> (ac. 2)	<i>glad</i> (ac.)	<i>remember</i> (ac. 3)	-
Un. léxicas A2 definidas por elementos A2	<i>library</i> (ac.) <i>rent</i> (ac. 2) <i>rent</i> (n.)	<i>library</i> (n.) <i>rent</i> (ac. 1) <i>already</i> (ac.) <i>already</i> (n.) <i>among</i> (n.)	<i>glad</i> (ac.) <i>pleasant</i> (ac. 2) <i>among</i> (n.)	<i>library</i> (ac.) <i>glad</i> (ac.) <i>pleasant</i> (ac. 2) <i>remember</i> (ac. 2) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>among</i> (ac. 1)
Un. léxicas A2 definidas por elementos B1+	<i>appointment</i> (ac.) <i>landscape</i> (ac.) <i>pleasant</i> (ac.) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 2)	<i>appointment</i> (ac.1) <i>landscape</i> (ac.) <i>library</i> (ac.) <i>glad</i> (n.) <i>pleasant</i> (ac.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>across</i> (n.) <i>among</i> (ac. 2)	<i>appointment</i> (ac.1) <i>landscape</i> (ac. 1) <i>library</i> (ac.) <i>pleasant</i> (ac. 1) <i>rent</i> (ac. 1) <i>already</i> (ac. 1) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2) <i>among</i> (ac. 2) <i>among</i> (ac. 3)	<i>appointment</i> (ac. 1) <i>landscape</i> (ac.) <i>library</i> (n. 1) <i>library</i> (n. 2) <i>pleasant</i> (ac. 1) <i>already</i> (n.) <i>across</i> (ac. 1) <i>across</i> (ac. 2)
Un. léxicas B1 definidas por elementos A1 e/ou A2	<i>gorgeous</i> (ac.) <i>pretty</i> ^l (ac.) <i>among</i> (ac. 1) <i>towards</i> (ac. 1)	<i>gorgeous</i> (ac.) <i>likely</i> (ac.) <i>among</i> (ac. 1) <i>towards</i> (ac. 1)	<i>edge</i> (ac. 1) <i>gorgeous</i> (ac.) <i>likely</i> (ac. 2) <i>among</i> (ac. 1) <i>towards</i> (ac. 1)	<i>label</i> ^l (ac. 2) <i>gorgeous</i> (ac.) <i>likely</i> (ac. 1) <i>towards</i> (ac. 1)
Un. léxicas B1 definidas por elementos B1	<i>edge</i> (ac.) <i>flavour</i> (ac.) <i>label</i> ^l (ac. 2) <i>rent</i> (ac. 3) <i>bury</i> (ac. 1) <i>remind</i> (ac. 1) <i>remind</i> (ph.v.) <i>already</i> (ac. 2)	<i>edge</i> (ac. 1) <i>flavour</i> (ac.) <i>bury</i> (ac. 1) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>pretty</i> ^l (ac.)	<i>flavour</i> (ac.) <i>label</i> (ac.) <i>rent</i> (ac. 2) <i>bury</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>remind</i> (ac. 3) <i>already</i> (ac. 2) <i>pretty</i> (ac. 2)	<i>edge</i> (ac. 1) <i>flavour</i> (ac.) <i>rent</i> (ac. 2) <i>remind</i> (ac. 1) <i>remind</i> (ac. 2) <i>already</i> (ac. 2) <i>pretty</i> ^l (ac.) <i>among</i> (ac. 2)
Un. Léxicas B1 definidas por elementos B2+	<i>label</i> ^l (ac. 1) <i>likely</i> (ac. 1) <i>remind</i> (n.)	<i>label</i> ^l (ac.) <i>rent</i> (ac. 2)	-	<i>label</i> ^l (ac. 1) <i>bury</i> (ac. 1)

Fonte: Elaboração da autora com base nos dados do CaED2 (2011), OxED2 (2012), CoED2 (2014), LWord2 (2008), EVP (2015) e EGP (2015)

ANEXO 10

CaED2 – ULs A2		
Verbetes	Paráfrase explanatória ou nota de uso	CAREC
<i>glad</i>	happy about something.	-0.279
<i>remember, 2</i>	to not forget to do something.	0.321
<i>library</i>	a place with a lot of books that you can read or borrow.	-0.077
<i>rent, 2</i>	to pay money to use something for a short time.	0.014
<i>rent, n.</i>	In British English you rent something for a long time. In British English you hire something for a short time. In American English the word rent is used in both situations.	0.326
<i>across, 1</i>	from one side to the other.	-0.272
<i>across, 2</i>	on the opposite side of.	-0.165
<i>pleasant</i>	good or enjoyable.	-0.025
<i>landscape</i>	the appearance of an area of land, especially in the countryside.	0.003
<i>rent, 1</i>	to pay money to live in a building that someone else owns.	0.03
<i>already, 1</i>	before now, or before a time in the past.	0.105
<i>among, 2</i>	in a particular group.	0.259
<i>appointment</i>	a time you have arranged to see someone.	0.343
CaED2 – ULs B1		
Verbetes	Paráfrase explanatória ou nota de uso	CAREC
<i>gorgeous</i>	very beautiful or pleasant.	-0.071
<i>towards</i>	in the direction of someone or something.	0.194
<i>pretty</i>	quite, but not very.	0.41
<i>among, 1</i>	in the middle of something.	0.455
<i>already, 2</i>	used to say that something has happened earlier than you expected.	-0.101
<i>flavour</i>	the taste of a type of food or drink.	-0.086
<i>bury, 1</i>	to put a dead body into the ground.	-0.063
<i>label, 2</i>	a company that makes things for sale, mainly clothes.	-0.028
<i>rent, 3</i>	to allow someone to pay you money to live in your building.	0.044
<i>remind, p.v.</i>	to make someone think of something or someone else.	0.293
<i>edge</i>	the part around something that is furthest from the centre.	0.369
<i>remind, 1</i>	to make someone remember something, or remember to do something.	0.4
<i>likely</i>	expected.	---
<i>remind, n.</i>	If you remember a fact or something from the past, you keep it in your mind, or bring it back into your mind. When you remind someone to do something, you make them remember it.	0.073
<i>label, 1</i>	a piece of paper or material that is attached to something and gives you information about it.	0.085

OxED2 – ULs A2		
Verbetes	Paráfrase explanatória ou nota de uso	CAREC
<i>glad</i>	happy about something.	
<i>among, n.</i>	We use among when we are talking about more than two people or things. If there are only two people or things, we use between .	-0.096
<i>rent, 1</i>	to pay to live in a place or to use something that belongs to another person.	-0.015
<i>already</i>	before now or before then.	0.024
<i>library, n.</i>	Be careful! You cannot buy books from a library . The place you buy books is called a bookshop or a bookstore .	0.098
<i>already, n.</i>	Yet means the same as already , but you only use yet in negative sentences and questions.	0.28
<i>across, 1</i>	from one side to the other side of something.	-0.235
<i>pleasant</i>	nice, enjoyable or friendly.	-0.194
<i>landscape</i>	everything you can see in an area of land.	-0.101
<i>across, 2</i>	on the other side of something.	-0.014
<i>across, n.</i>	We can use across or over to mean ‘on or to the other side’. We usually use over to talk about crossing something high. With ‘room’ we usually use across .	0.005
<i>glad, n.</i>	You are usually glad or pleased about a particular event or situation. You use happy to describe a state of mind or before a noun.	0.01
<i>library</i>	a room or building where you go to borrow or read books.	0.045
<i>among, 2</i>	in a particular group of people or things.	0.057
<i>appointment</i>	an arrangement to see somebody at a particular time.	0.169
OxED2 – ULs B1		
Verbetes	Paráfrase explanatória ou nota de uso	CAREC
<i>among, 1</i>	in the middle of a group of people or things.	-0.255
<i>gorgeous</i>	very good or attractive.	0.069
<i>likely</i>	If something is likely, it will probably happen.	0.166
<i>towards</i>	in the direction of somebody or something.	0.209
<i>bury</i>	to put a dead body in the ground.	-0.125
<i>flavour</i>	the taste of food.	-0.089
<i>remind, 1</i>	to help somebody remember something that they must do.	0.205
<i>edge</i>	the part along the end or side of something.	0.276
<i>remind, 2</i>	to make somebody remember somebody or something.	0.446
<i>pretty</i>	quite; fairly.	1.23
<i>label</i>	a piece of paper or material on something that tells you about it.	-0.042
<i>rent, 2</i>	to let somebody live in a place or use something that belongs to you, if they pay you.	0.082

CoED2 – ULs A2		
Verbetes	Paráfrase explicatória ou nota de uso	CAREC
<i>remember, 3</i>	to not forget to do something.	
<i>glad</i>	happy and pleased about something.	-0.202
<i>pleasant, 2</i>	nice and friendly.	-0.046
<i>among, n.</i>	If there are more than two people or things, use among . If there are only two people or things, use between .	0.07
<i>landscape</i>	everything you can see when you look across an area of land.	-0.286
<i>rent, 1</i>	to pay the owner of something in order to be able to use it yourself.	-0.169
<i>across, 1</i>	from one side of something to the other side of it.	-0.161
<i>across, 2</i>	going from one side of something to the other side of.	-0.131
<i>among, 2</i>	included in a particular group.	-0.002
<i>among, 3</i>	used for saying that most people in a group have a particular opinion or feeling.	0.003
<i>library</i>	a place where books, newspapers, DVDs and music are kept for people to use or borrow.	0.044
<i>already, n.</i>	In British English, you usually use already and yet with the perfect tenses. In American English, people often use a past tense.	0.097
<i>already, 1</i>	used for showing that one thing happened before another thing.	0.119
<i>appointment</i>	an arrangement to see someone at a particular time.	0.227
<i>pleasant, 1</i>	enjoyable or attractive.	0.363
CoED2 – ULs B1		
Verbetes	Paráfrase explicatória ou nota de uso	CAREC
<i>among, 1</i>	in the middle of a group of people or things.	-0.255
<i>likely, 2</i>	used for saying that a person or thing will probably do something.	0.012
<i>gorgeous</i>	very pleasant or attractive.	0.042
<i>towards, 1</i>	in the direction of something or someone.	0.194
<i>edge</i>	the part of something that is farthest from the middle.	0.238
<i>bury, 2</i>	to put the body of a dead person into a grave and cover it with earth.	-0.119
<i>rent, 2</i>	to let someone have and use something in exchange for money.	0.058
<i>flavour</i>	the taste of a food or drink.	0.079
<i>label</i>	a piece of paper or plastic that is attached to an object to give information about it.	0.11
<i>already, 2</i>	used for showing that a situation exists now or that it started earlier than expected.	0.155
<i>pretty, 2</i>	used before an adjective or adverb to mean 'fairly'.	0.162
<i>remind, 3</i>	to be similar to another person or thing and make someone think about them.	0.193
<i>remind, 1</i>	to say something that makes someone think about a fact or an event that they already know about.	0.194
<i>remind, 2</i>	to say something that makes someone remember to do something.	0.246

LWord2 – ULs A2		
Verbetes	Paráfrase explanatória ou nota de uso	CAREC
<i>already, 1</i>	before.	---
<i>among, 1</i>	used to say which group of people is affected by something.	-0.122
<i>pleasant, 2</i>	polite and friendly.	-0.092
<i>glad</i>	if you are glad, you are pleased and happy about something.	-0.069
<i>rent, 1</i>	to pay money to live in a place or to use something such as a car.	-0.022
<i>library</i>	a room or building containing books that you can borrow, and computers that you can use to get information.	0.092
<i>remember, 2</i>	if you remember to do something, you do not forget to do it.	0.257
LWord2 – ULs B1		
Verbetes	Paráfrase explanatória ou nota de uso	CAREC
<i>gorgeous</i>	very beautiful or pleasant.	-0.071
<i>likely, 1</i>	if something is likely to happen, it will probably happen.	0.046
<i>towards</i>	in the direction of someone or something.	0.194
<i>label, 2</i>	a record company.	0.25
<i>already, 2</i>	now, rather than later.	-0.031
<i>rent, 2</i>	if you rent a house, room etc to someone, you allow them to use it, and they pay you money.	-0.02
<i>remind, 2</i>	if someone reminds you of another person, they make you think of that person because they are similar to them.	0.016
<i>flavour</i>	the taste that a food or drink has.	0.221
<i>edge</i>	the part of something that is furthest from the middle.	0.222
<i>remind, 1</i>	to make someone remember something that they must do.	0.348
<i>among, 2</i>	surrounded by things.	0.493
<i>pretty</i>	fairly or quite.	0.588
<i>bury</i>	to put a dead body into a GRAVE (a hole that has been dug in the ground for this purpose).	0.02
<i>label, 1</i>	a piece of paper or cloth with information on it that is fixed to an object.	0.125

ANEXO 11

SIMILARIDADE ORTOGRÁFICA DAS ULs DA AMOSTRA			
Item da amostra	Símile em português	Similaridade ortográfica	Equivalente correto em português
bajar	-	-	descer
invitar	-	-	convidar
aburrido	-	-	entediante; tedioso
ancho	-	-	largo
demasiado	-	-	demais
chiste	-	-	piada
contraseña	-	-	senha
huelga	-	-	greve
rebaja	-	-	desconto, oferta
enfadar	-	-	aborrecer
mareado	-	-	enjoadado
roto	-	-	quebrado
anoche	-	-	ontem à noite
billete [dinero]	bilhete	0,86	nota; cédula
cita [reunión]	citar (verbo)	0,73	consulta; encontro
copa [bebida]	copa	1,00	trago
dirección [ubicación]	direção	0,60	endereço
gimnasio	ginásio	0,89	academia
propina	propina	1,00	gorjeta
contestar [responder]	contestar	1,00	responder
reparar [examinar]	repassar	0,93	revisar [texto]; repassar [conteúdo]
mayor [edad]	maior	0,78	mais velho
todavía [vigente]	todavia	1,00	ainda
regla [sangre]	regra	0,78	menstruação
tapa [alimento]	tapa	1,00	porção
ligar [relación]	ligar	1,00	paquerar
saber [sabor]	saber	1,00	ter gosto de
fatal [algo muy malo]	fatal	1,00	muito ruim; horrível
ligero [alimento]	ligeiro	0,87	leve
vago [perezoso]	vago	1,00	vagabundo

Nota: Foram omitidos os acentos (´, ^) e diacríticos (~ do português, ¨ do espanhol). O ç do português foi substituído por c. A justificativa está em 5.2.2.

ANEXO 12

Verbetes de ULs A2	
bi·lle·te 1 <i>m.</i>	Papel que imprime y emite un banco y que representa una cantidad de la moneda de un país: [...]
ci·ta 1 <i>f.</i>	Fijación del día, la hora y el lugar para encontrarse dos o más personas: [...]
co·pa [...] 2 <i>f.</i>	Bebida alcohólica: [...]
di·rec·ción [...] 3	Nombre de la población y la calle y número de la casa donde vive una persona: [...]
gim·na·sio <i>m.</i>	Instalación o establecimiento que dispone de aparatos para hacer ejercicio físico y practicar ciertos deportes: [...]
pro·pi·na <i>f.</i>	Cantidad de dinero que se da voluntariamente para mostrar *satisfacción por el buen trato o el servicio recibido: [...]
ba·jar 1 <i>intr.-prnl.</i>	[de un sitio; a un sitio] Ir de un lugar alto a otro que está más bajo: [...]
con·tes·tar 1 <i>tr.</i>	[algo; a alguien, a algo] Dar una respuesta a lo que se pregunta, dice o escribe: [...]
in·vi·tar 1 <i>tr.</i>	[a alguien] Pedir a una persona que esté en una celebración, comida u otra cosa: [...]
re·pa·sar 2.	[algo] Examinar o volver a mirar: [...]
a·bu·rri·do, da 1 <i>adj.</i>	Que no tiene interés; que aburre: [...] ☒ Se usa con el verbo <i>ser</i> [...]
an·cho 1 <i>adj.</i>	Que tiene una extensión de lado a lado excesiva o mayor de lo normal: [...]
ma·yor 1 <i>adj.</i>	(ser vivo) Que tiene más edad: [...]
de·ma·sia·do, da 1 <i>adv. c.</i>	Superando lo justo o lo necesario; excesivamente: [...]
to·da·ví·a 1 <i>adv. t.</i>	Hasta un momento determinado; hasta ahora: [...]
Verbetes de ULs B1	
chis·te 1 <i>m.</i>	Historia corta o dibujo que hace reír: [...]
con·tra·se·ña <i>f.</i>	Expresión o señal que se dan unas personas a otras para entenderse entre sí, para ser reconocidas o para que les sea permitida una cosa determinada: [...]
huel·ga <i>f.</i>	Paro del trabajo que hacen los *obreros de común acuerdo, generalmente como protesta o para *imponer ciertas condiciones a los dueños de la empresa o al gobierno: [...]
re·ba·ja <i>f.</i>	Disminución o reducción, especialmente en una cantidad o un precio: [...]
re·gla [...] 5	Sangre que todos los meses expulsan naturalmente las mujeres y las hembras de ciertos animales: [...]
ta·pa [...] 2	Alimento ligero que se sirve como acompañamiento de una bebida: [...]
en·fa·dar 1 <i>tr.-prnl.</i>	[a alguien] Causar disgusto o enfado: [...]
li·gar 6 <i>intr.-prnl. fam.</i>	[a/con alguien] Establecer relaciones amorosas o sexuales: [...]
sa·ber 4 <i>intr.</i>	Tener sabor: [...]
fa·tal [...] 1 <i>adj.</i>	Que es muy malo; que no se puede soportar: [...]
li·ge·ro, ra [...] 3	(alimento) Que se puede *digerir fácilmente: [...]
ma·re·a·do, da <i>adj.</i>	Que experimenta una sensación desagradable en la cabeza y en el estómago, generalmente a causa del movimiento de un vehículo o de una nave: [...]
ro·to, ta [...] 2	Que no sirve, no funciona o no se puede usar: [...] ☒ Es el participio de <i>romper</i> . [...]
va·go, ga 1 <i>adj.-s.</i>	Que no gusta del trabajo: [...]
a·no·che <i>adv. t.</i>	En la noche de ayer: [...]

Fonte: Elaboração da autora com base nos dados do Señas (2001) e do PCIC (2006)
 Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do PCIC (2006)

ANEXO 13

Tipo de cognato	Palavra do espanhol	Símile ortográfico em português	S.O.	Equivalente em português
Tipo 1: Similaridade baixa ou nula	y	--	--	e
	un	um	0,44	um
	la	a	0,45	a
	del	--	--	do
	o	ou	0,43	ou
	en	em	0,44	em
	el	--	--	o
	hacer	--	--	fazer
	hace	--	--	faz
	hacen	--	--	fazem
	bajo	--	--	baixo
	tiene	--	--	tem
	escribe	--	--	escreve
	paro	--	--	paralisação
	al	--	--	ao
	muy	--	--	muito
	ayer	--	--	ontem
	u	--	--	ou
	dice	--	--	diz
	volver	--	--	voltar
	dibujo	--	--	desenho
	cantidad	quantidade	0,47	quantidade
	poblacion	--	--	população
	buen	--	--	bom
	hembras	--	--	fêmeas
	lo	o	0,45	o
	aburre	--	--	entediado; cansar
obreros	--	--	operários	
enfado	--	--	aborrecer	
señal	--	--	sinal	
fijación	--	--	fixação	
reír	--	--	rir	
Total de cognatos tipo 1: 32				
Tipo 2: Falsos amigos	dos	dos	1,00	dois
	calle	calar	0,57	rua
	no	no	1,00	não
	deportes	deportar	0,68	esportes
	mirar	mirar	1,00	olhar
	hasta	haste	0,70	até
	personas	personagem	0,69	pessoas
	persona	personagem	0,71	pessoa
	ligero	ligeiro	0,87	leve
	esté	este	0,66	esteja
	aparatos	aparados	0,88	aparelhos
nave	nave	1,00	navio	
Total de cognatos tipo 2: 12				
Tipo 3: Cognatos	de	de	1,00	de
	que	que	1,00	que
	una	uma	0,61	uma

Tipo 3: Cognatos	banco	banco	1,00	banco
	país	país	1,00	país
	hora	hora	1,00	hora
	lugar	lugar	1,00	lugar
	más [cantidad]	mais	0,73	mais
	más [grado]	mais	0,73	mais
	para [utilidad]	para	1,00	para
	nombre	nome	0,78	nome
	número	número	1,00	número
	casa	casa	1,00	casa
	vive	vive	1,00	vive
	ejercicio	exercício	0,84	exercício
	dinero	dinheiro	0,76	dinheiro
	ir	ir	1,00	ir
	lugar	lugar	1,00	lugar
	alto	alto	1,00	alto
	está	está	1,00	está
	comida	comida	1,00	refeição
	con	com	0,59	com
	mayor	maior	0,78	maior
	es	es	1,00	é
	momento	momento	1,00	momento
	ahora	agora	0,78	agora
	corta	curta	0,78	curta
	unas	umas	0,72	umas
	ser	ser	1,00	ser
	trabajo	trabalho	0,83	trabalho
	los	os	0,58	os
	empresa	empresa	1,00	empresa
	precio	preço	0,86	preço
	todos	todos	1,00	todos
	meses	meses	1,00	meses
	mujeres	mulheres	0,83	mulheres
	animales	animais	0,82	animais
	malo	mal	0,73	mal
	soportar	suportar	0,90	suportar
	noche	noite	0,66	noite
	papel	papel	1,00	papel
	moneda	moeda	0,86	moeda
	para [modif. sust.]	para	1,00	para
practicar	praticar	0,94	praticar	
a	a	1,00	a	
dar	dar	1,00	dar	
dan	dão	0,59	dão	
sí	si	1,00	si	
respuesta	resposta	0,84	resposta	
otro	outro	0,80	outro	
pregunta	pergunta	0,86	pergunta	
otra	outra	0,80	outra	
da	da	1,00	dá	
normal	normal	1,00	normal	
necesario	necessario	0,94	necessário	
otras	outras	0,84	outras	

Tipo 3: Cognatos	como	como	1,00	como
	gobierno	governo	0,82	governo
	bebida	bebida	1,00	bebida
	por	por	1,00	por
	nada	nada	1,00	nada
	puede	pode	0,75	pode
	cabeza	cabeça	0,82	cabeça
	estómago	estômago	1,00	estômago
	gusta	gosta	0,78	gosta
	donde	onde	0,66	onde
	encontrarse	encontrar	0,77	encontrarem-se
	se	se	1,00	se
	interes	interesse	0,73	interesse
	usa	usa	1,00	usa
	lado	lado	1,00	lado
	historia	história	1,00	história
	expresión	expressão	0,63	expressão
	entenderse	entender	0,76	entenderem-se
	les	lhes	0,73	lhes
	sea	seja	0,73	seja
	sangre	sangue	0,82	sangue
	naturalmente	naturalmente	1,00	naturalmente
	alimento	alimento	1,00	alimento
	relaciones	relações	0,83	relações
	sexuales	sexuais	0,86	sexuais
	tener	ter	0,86	ter
	sabor	sabor	1,00	sabor
	sirve	serve	0,78	serve
	funciona	funciona	1,00	funciona
	usar	usar	1,00	usar
	velocidad	velocidade	0,81	velocidade
	romper	romper	1,00	quebrar
	representa	representa	1,00	representa
	bebida	bebida	1,00	bebida
	alcohólica	alcoólica	1,00	alcoólica
	instalacion	instalacao	0,70	instalação
	establecimiento	estabelecimento	0,90	estabelecimento
	dispone	dispoe	0,88	dispõe
	satisfaccion	satisfacao	0,67	satisfação
	servicio	servico	0,93	serviço
	pedir	pedir	1,00	pedir
	celebracion	celebracao	0,70	celebração
examinar	examinar	1,00	examinar	
extensión	extensao	0,65	extensão	
excesiva	excessiva	0,93	excessiva	
justo	justo	1,00	justo	
excesivamente	excessivamente	0,93	excessivamente	
determinado	determinado	1,00	determinado	
permitida	permitida	1,00	permitida	
determinada	determinada	1,00	determinada	
generalmente	geralmente	0,97	geralmente	
protesta	protesto	0,74	protesto	
condiciones	condições	0,87	condições	

Tipo 3: Cognatos	disminución	diminuição	0,58	diminuição
	reducción	redução	0,62	redução
	especialmente	especialmente	1,00	especialmente
	sirve	serve	0,78	serve
	establecer	estabelecer	0,91	estabelecer
	adecuada	adequada	0,88	adequada
	digerir	digerir	1,00	digerir
	fácilmente	facilmente	1,00	facilmente
	experimenta	experimenta	1,00	experimenta
	sensación	sensação	0,65	sensação
	desagradable	desagradável	0,74	desagradável
	movimiento	movimento	0,95	movimento
	vehículo	veículo	0,89	veículo
	imprime	imprime	1,00	imprime
	emite	emite	1,00	emite
	físico	físico	1,00	físico
	ciertos	certos	0,88	certos
	voluntariamente	voluntariamente	1,00	voluntariamente
	mostrar	mostrar	1,00	mostrar
	trato	trato	1,00	trato
	recibido	recebido	0,88	recebido
	verbo	verbo	1,00	verbo
	superando	superando	1,00	superando
	entre	entre	1,00	entre
	cosa	coisa	0,80	coisa
	común	comum	0,68	comum
	acuerdo	acordo	0,79	acordo
	imponer	impor	0,82	impor
	ciertas	certas	0,88	certas
	dueños	donos	0,57	donos
	expulsan	expulsam	0,75	expulsam
acompañamiento	acompanhamento	0,85	acompanhamento	
causar	causar	1,00	causar	
disgusto	desgosto	0,73	desgosto	
amorosas	amorosas	1,00	amorosas	
manera	maneira	0,87	maneira	
causa	causa	1,00	causa	
participio	participio	1,00	participio	
Total de cognatos tipo 3: 147				
Total absoluto de cognatos: 191				

Fonte: Elaboração da autora com base nos dados do Señas (2001) e do PCIC (2006)
 Legenda: ■ = A1; ■ = A2; ■ = B1; ■ = B2 ou C1 ou C2; ■ = ausente do PCIC (2006)

ANEXO 14

Ranking de tipos de erros de brasileiros adultos A1-B1 no CLC (2019)

Tipo de erro	Nº absoluto	Comparação com erros
Total	38.809	46% dos erros BR A1-C2 18+
1. RP: Substituir pontuação	4.198	10,8% dos erros BR A1-B1 18+
2. S: Ortografia	2.566	6,6% dos erros BR A1-B1 18+
3. RT: Substituir preposição	2.559	6,6% dos erros BR A1-B1 18+
4. TV: Tempo verbal incorreto	2.417	6,2% dos erros BR A1-B1 18+
5. RV: Substituir verbo	2.341	6,0% dos erros BR A1-B1 18+
6. MP: Ausência de pontuação	1.890	4,8% dos erros BR A1-B1 18+
7. FV: Forma nominal do verbo incorreta	1.686	4,3% dos erros BR A1-B1 18+
8. UP: Pontuação desnecessária	1.341	3,4% dos erros BR A1-B1 18+
9. MA: Ausência de elemento anafórico	1.241	3,2% dos erros BR A1-B1 18+
10. MD: Ausência de determinante	1.222	3,1% dos erros BR A1-B1 18+
11. RN: Substituir substantivo	1.144	2,9% dos erros BR A1-B1 18+
12. MT: Ausência de preposição	1.076	2,7% dos erros BR A1-B1 18+
13. W: Erro de posição sintática	968	2,5% dos erros BR A1-B1 18+
14. R: Substituir recurso linguístico	916	2,3% dos erros BR A1-B1 18+
15. UD: Determinante desnecessário	881	2,2% dos erros BR A1-B1 18+
16. UT: Preposição desnecessária	839	2,1% dos erros BR A1-B1 18+
17. AGV: Erro de concordância verbal	696	1,8% dos erros BR A1-B1 18+
18. MV: Ausência de verbo	550	1,4% dos erros BR A1-B1 18+
19. RY: Substituir advérbio	544	1,4% dos erros BR A1-B1 18+
20. RJ: Substituir adjetivo	532	1,3% dos erros BR A1-B1 18+
21. FN: Forma nominal incorreta	522	1,3% dos erros BR A1-B1 18+
22. SX: Ortografia confusa	512	1,3% dos erros BR A1-B1 18+
23. RA: Substituir anáfora	500	1,2% dos erros BR A1-B1 18+
24. M: Ausência de recurso linguístico	434	1,1% dos erros BR A1-B1 18+
25. RD: Substituir determinante	412	1% dos erros BR A1-B1 18+
26. DN: Erro de derivação nominal	394	1% dos erros BR A1-B1 18+
27. DJ: Erro de derivação adjetival	394	1% dos erros BR A1-B1 18+
28. L: Registro inapropriado	369	0,9% dos erros BR A1-B1 18+
29. MC: Ausência de nexos oracionais	359	0,9% dos erros BR A1-B1 18+
30. AGN: Erro de concordância nominal	320	0,8% dos erros BR A1-B1 18+
31. SA: Ortografia americana	309	0,8% dos erros BR A1-B1 18+
32. MN: Ausência de substantivo	291	0,7% dos erros BR A1-B1 18+
33. UV: Verbo desnecessário	285	0,7% dos erros BR A1-B1 18+
34. IV: Flexão verbal incorreta	270	0,7% dos erros BR A1-B1 18+
35. UA: Anáfora desnecessária	256	0,6% dos erros BR A1-B1 18+
36. DY: Erro de derivação adverbial	242	0,6% dos erros BR A1-B1 18+
37. U: Recurso linguístico desnecessário	241	0,6% dos erros BR A1-B1 18+
38. UY: Advérbio desnecessário	206	0,5% dos erros BR A1-B1 18+
39. MY: Ausência de advérbio	191	0,5% dos erros BR A1-B1 18+
40. IN: Flexão nominal incorreta	158	0,4% dos erros BR A1-B1 18+
41. RC: Substituir nexos oracionais	144	0,4% dos erros BR A1-B1 18+
42. UN: Substantivo desnecessário	141	0,3% dos erros BR A1-B1 18+
43. CE: Erro complexo	141	0,3% dos erros BR A1-B1 18+
44. IJ: Flexão adjetival incorreta	137	0,3% dos erros BR A1-B1 18+
45. FD: Forma incorreta do determinante	131	0,3% dos erros BR A1-B1 18+
46. RQ: Substituir quantificador	124	0,3% dos erros BR A1-B1 18+

47. UC: Nexoracional desnecessário	123	0,3% dos erros BR A1-B1 18+
48. MQ: Ausência de quantificador	122	0,3% dos erros BR A1-B1 18+
49. CN: Erro de contabilidade nominal	119	0,3% dos erros BR A1-B1 18+
50. ID: Flexão do determinante incorreta	111	0,3% dos erros BR A1-B1 18+
51. AGA: Erro de concordância anafórica	108	0,3% dos erros BR A1-B1 18+
52. DV: Erro de derivação verbal	101	0,2% dos erros BR A1-B1 18+
53. AGD: Erro de concordância do determ.	93	0,2% dos erros BR A1-B1 18+
54. AS: Erro na estrutura argumental	87	0,2% dos erros BR A1-B1 18+
55. DA: Erro de derivação anafórica	77	0,2% dos erros BR A1-B1 18+
56. DD: Erro de derivação do determinante	74	0,2% dos erros BR A1-B1 18+
57. MJ: Ausência de adjetivo	71	0,2% dos erros BR A1-B1 18+
58. FJ: Forma adjetival incorreta	63	0,1% dos erros BR A1-B1 18+
59. IQ: Flexão incorreta do quantificador	45	0,1% dos erros BR A1-B1 18+
60. X: Forma negativa incorreta	44	0,1% dos erros BR A1-B1 18+
61. AGQ: Erro de concordância do quantific.	42	0,1% dos erros BR A1-B1 18+
62. CQ: Quantificador incorreto	41	0,1% dos erros BR A1-B1 18+
63. UJ: Adjetivo desnecessário	38	0,1% dos erros BR A1-B1 18+
64. FQ: Forma incorreta do quantificador	36	0,1% dos erros BR A1-B1 18+
65. DT: Erro de derivação preposicional	32	0,1% dos erros BR A1-B1 18+
66. AG: Erro de concordância	28	0,07% dos erros BR A1-B1 18+
67. UQ: Quantificador desnecessário	24	0,06% dos erros BR A1-B1 18+
68. DI: Erro de derivação do determinante	24	0,06% dos erros BR A1-B1 18+
69. DC: Erro de derivação do nexoracional	16	0,03% dos erros BR A1-B1 18+
70. FY: Forma adverbial incorreta	15	0,03% dos erros BR A1-B1 18+
71. DQ: Erro de derivação do quantificador	10	0,02% dos erros BR A1-B1 18+
72. CL: Erro colocacional ou tautológico	7	0,02% dos erros BR A1-B1 18+
73. FA: Forma anafórica incorreta	6	0,01% dos erros BR A1-B1 18+
74. IA: Flexão anafórica incorreta	5	0,01% dos erros BR A1-B1 18+
75. IY: Flexão adverbial incorreta	1	0,002% dos erros BR A1-B1 18+
Erros não catalogados ou não existentes	156	0,4% dos erros BR A1-B1 18+