



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

MAGALI SANCHES DURAN

PARÂMETROS PARA A ELABORAÇÃO DE DICIONÁRIOS
BILÍNGÜES DE APOIO À CODIFICAÇÃO ESCRITA EM
LÍNGUAS ESTRANGEIRAS

MAGALI SANCHES DURAN

PARÂMETROS PARA A ELABORAÇÃO DE DICIONÁRIOS BILÍNGÜES DE APOIO À CODIFICAÇÃO ESCRITA EM LÍNGUAS ESTRANGEIRAS

Tese apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista, *Campus* de São José do Rio Preto, para obtenção do título de Doutor em Estudos Lingüísticos (Área de Concentração: Análise Lingüística)

BANCA EXAMINADORA

Prof^a. Dr^a. Claudia Maria Xatara
Professor Livre-Docente
UNESP – São José do Rio Preto
Orientador

Prof. Dr. John Robert Schmitz
Professor Titular
UNICAMP

Prof. Dr. Bento Carlos Dias da Silva
Professor Assistente Doutor
UNESP – Araraquara

Prof^a. Dr^a. Claudia Zavaglia
Professor Assistente Doutor
UNESP – São José do Rio Preto

Prof. Dr. Douglas Altamiro Consolo
Professor Livre-Docente
UNESP – São José do Rio Preto

São José do Rio Preto, 16 de dezembro de 2008

Duran, Magali Sanches.

Parâmetros para elaboração de dicionários bilíngües de apoio à codificação escrita em línguas estrangeiras / Magali Sanches Duran. - São José do Rio Preto : [s.n.], 2008.

167 f. : il ; 30 cm.

Orientador: Claudia Maria Xatara

Tese (doutorado) - Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas

1. Lexicografia. 2. Lexicografia pedagógica. 3. Dicionários bilíngües. 4. Escrita. 5. Língua estrangeira - Escrita. I. Xatara, Claudia Maria. II. Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU - 81'374

Ficha catalográfica elaborada pela Biblioteca do IBILCE
Campus de São José do Rio Preto - UNESP

Dedicatória

Dentre os pacifistas espalhados pelo mundo, famosos ou anônimos, estão aqueles que trabalham para superar as barreiras lingüísticas e melhorar a comunicação entre os povos. A todos que compartilham dessa motivação, dedico esta tese.

Agradecimentos

Muitos colaboraram, direta ou indiretamente, para a realização desta pesquisa e, a todos, dirijo meus agradecimentos.

Quero dizer um “muito obrigada” especial a:

Kleber Sartorio, companheiro, ouvinte atento, inspirador de muitas idéias e provedor de soluções para problemas que apareceram nas diversas fases da pesquisa. Sem você tudo teria sido diferente e mais difícil;

Claudia Xatara, minha orientadora, por ter acreditado em meu projeto e ter me dado total liberdade na condução da pesquisa;

Herbert Welker, **Philippe Humblé**, e **Jean Binon**, pela disposição em trocar idéias sobre assuntos lexicográficos de interesse em comum. Vocês se tornaram amigos por quem nutro grande admiração;

Batia Laufer, **Hilary Nesi** e **Sylvie De Cock**, por terem gentilmente enviado alguns dos trabalhos essenciais para a revisão bibliográfica;

À **CAPES**, duplamente: por ter concedido a Bolsa que permitiu minha dedicação exclusiva ao projeto e por disponibilizar os Periódicos CAPES, recurso maravilhoso para acessarmos o discurso da comunidade científica internacional;

E, finalmente, a meus pais, **Tonico e Nadir**, que são meu “chão”, e a minhas filhas, **Nina e Taís**, que mantêm meu coração aquecido.

Resumo. O objetivo deste trabalho é apresentar parâmetros que possam ser utilizados para orientar melhorias e inovações lexicográficas, a fim de atender as dificuldades dos brasileiros na codificação escrita em línguas estrangeiras. Revisam-se criticamente várias pesquisas, em busca de subsídios que contribuam para diagnosticar as dificuldades de escrever em língua estrangeira, dentre elas os trabalhos produzidos por teóricos da Lexicografia, as pesquisas sobre o uso do dicionário e estudos sobre *corpus* de aprendizes. Realiza-se, também, uma pesquisa empírica sobre o uso dos dicionários na codificação a fim de complementar, com dados primários, as informações obtidas nas demais fontes. Uma vez conhecidas as necessidades dos usuários de dicionários, analisam-se as obras lexicográficas que expressam o propósito de auxiliar a produção em língua estrangeira. Nessa análise, verifica-se em que medida as necessidades identificadas são atendidas e como são atendidas. Chega-se, assim, à conclusão de que algumas das dificuldades solucionáveis por meio da consulta ao dicionário já são satisfatoriamente atendidas pelas obras disponíveis, embora seja oportuno otimizar o acesso dos usuários às informações. Há necessidades, porém, que não são atendidas pelos dicionários ou são atendidas apenas parcialmente. A principal delas é a busca de equivalentes para itens lexicais do português. Há carência de dicionários bilíngües na direção português→línguas estrangeiras que possuam, principalmente, a grande maioria dos sentidos e as colocações mais recorrentes de cada item lexical do português, além de ampla exemplificação do uso dos equivalentes em línguas estrangeiras. Para elaborar esses dicionários, o primeiro passo é definir uma nomenclatura que tome por base o léxico ativo do português do Brasil, incluindo lexias simples e complexas, com suas respectivas acepções. A partir dessa etapa, então, é sugerida uma proposta de se apresentar equivalentes em línguas estrangeiras, por meio de exemplos que ilustrem seu emprego. Para empreender todo o processo de elaboração desse tipo de dicionário, propõe-se o reaproveitamento de várias classes de trabalhos já realizados como: construção de redes semânticas, levantamento de lexias complexas, construção de *corpus* de aprendizes brasileiros, construção de *corpora* paralelos de português com outros idiomas etc.

Palavras-chave: dicionários bilíngües. dicionários para codificação. dicionários de produção. escrita em língua estrangeira.

Abstract. This work aims at presenting useful parameters for lexicographical improvements and innovations that meet encoding difficulties of Brazilian dictionary users. A critical review of lexicographers' works, and researches on second language writing, productive dictionary use, and learner's corpus, contributes to diagnose the problem of foreign language encoding. Also, an empirical research about dictionary use in encoding activities completes with primary data the information obtained from secondary sources. After identifying what dictionary users need, this work presents an analysis of five dictionaries which expressly mention their purpose to help users to write in a foreign language. Such analysis tries to measure to what extent, and how, users' needs are satisfied. The conclusion is that reference works are able to solve some of learners' difficulties, although it is desirable to improve the ways users retrieve information from them. However, there are some needs not met by dictionaries or only partially met. The major of them is the search for equivalents in foreign languages that correspond to Portuguese lexical items. There is a lack of bilingual dictionaries Portuguese-foreign language that present, specially, every meaning and recurrent collocations for Portuguese lexical items, besides examples of equivalents use in foreign language. To compile such dictionaries, the first step is to define a word-list based on Brazilian Portuguese active vocabulary, including simple and complex lexical units with all their respective senses. Then, it is suggested to find equivalents in foreign language with examples that illustrate pedagogically their use. For this work, results from other projects may be reused, such as semantic networks, collocation extraction, Brazilian learners corpus, and parallel corpora of Brazilian Portuguese and foreign languages.

Key-Words: bilingual dictionaries. encoding dictionaries. dictionaries for production. active dictionaries. second language writing.

SUMÁRIO

INTRODUÇÃO	10
1. A ESCRITA EM LÍNGUA ESTRANGEIRA	19
1.1. O processo de produção da escrita em língua estrangeira	22
1.2. O uso da língua materna na escrita em língua estrangeira.....	25
1.3. Recursos tecnológicos na escrita em língua estrangeira	27
1.4. Plágio em língua estrangeira.....	29
1.5. Gêneros textuais e escrita em língua estrangeira.....	29
1.6. Avaliação da escrita em língua estrangeira	30
1.7. Língua materna, nível de proficiência e a escrita em língua estrangeira.....	34
1.8. Conclusões sobre o relacionamento do dicionário com a escrita em língua estrangeira	35
2. DICIONÁRIOS PARA CODIFICAÇÃO: OPINIÕES TEÓRICAS	37
2.1. A inadequação dos dicionários tradicionais para a atividade de codificação em língua estrangeira.....	37
2.2. A nomenclatura de um dicionário de apoio à codificação.....	46
2.3. A mídia para um dicionário de apoio à codificação.....	47
2.4. Informações lexicográficas necessárias para a codificação em língua estrangeira	48
3. NECESSIDADES DO USUÁRIO DE DICIONÁRIOS NA CODIFICAÇÃO EM LÍNGUA ESTRANGEIRA	54
3.1. Pesquisas sobre o uso do dicionário na codificação	55
3.1.1. Ard (1982)	58
3.1.2. Krings (1986 <i>apud</i> WELKER, 2006).....	59
3.1.3. Nesi (1987)	60
3.1.4. Nesi (1994)	62
3.1.5. Bogaards (1996)	63
3.1.6. Atkins e Varantola (1997)	64
3.1.7. Harvey e Yuill (1997).....	66
3.1.8. Christianson (1997)	69
3.1.9. Winkler (1998)	71
3.1.10. Komuro e Yamada (2000)	72
3.1.11. Frankerberg-Garcia (2005).....	74

3.1.12.	East (2006).....	75
3.1.13.	Laufer e Levitzky-Aviad (2006).....	76
3.1.14.	Gomes (2006).....	77
3.1.15.	Conclusões.....	78
3.2.	Os estudos de <i>corpus</i> de aprendizes	81
3.3.	Conclusões gerais do capítulo.....	88
4.	COLETA DE DADOS PRIMÁRIOS SOBRE O USO DO DICIONÁRIO NA CODIFICAÇÃO EM LÍNGUA ESTRANGEIRA.....	90
4.1.	Metodologia.....	90
4.2.	Estudo-piloto.....	94
4.3.	Definição do texto para a atividade e dos sujeitos de pesquisa.....	96
4.4.	Análise dos resultados	100
4.5.	Conclusões.....	108
5.	A ANÁLISE DE SOLUÇÕES	111
5.1.	Longman Language Activator (1993).....	111
5.1.1.	A seleção da nomenclatura	113
5.1.2.	A metodologia	113
5.1.3.	A microestrutura	115
5.2.	Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde (2000)	117
5.3.	Bilingual Dictionary Plus (2006).....	120
5.4.	Dicionário de Aprendizagem Português-Espanhol (2007)	121
5.5.	Longman Dicionário Escolar (2002)	124
5.6.	Conclusões.....	126
6.	FERRAMENTAS DE SUPORTE PARA CODIFICAÇÃO EM LÍNGUAS ESTRANGEIRAS	129
6.1.	Recursos pré-escrita	132
6.1.1.	Características do dicionário bilíngüe para apoio à codificação	135
6.1.2.	Desenvolvimento dos dicionários bilíngües para codificação escrita em línguas estrangeiras .	142
6.2.	Recursos pós-escrita	146
6.2.1.	Considerações adicionais	147
	CONSIDERAÇÕES FINAIS.....	152

REFERÊNCIAS BIBLIOGRÁFICAS.....	157
ANEXOS	167

LISTA DE SIGLAS

DAFLES	Dictionnaire d’Aprentissage du Français Langue Étrangère
DAPE	Dicionário de Aprendizagem Português Espanhol
DB	Dicionário Bilíngüe
DB+	Bilingual Dictionary Plus
DM	Dicionário Monolíngüe
DPE	Dicionário de Aprendizagem Português Espanhol
L1	Língua materna
L2	Língua estrangeira ou segunda língua, indiferentemente
LLA	Longman Language Activator

LISTA DE TABELAS

Tabela 1 - Sujeitos de Pesquisa	97
Tabela 2 - Relação entre número de buscas e de consultas por usuário	104
Tabela 3 - Consultas por participante/idioma, distribuídas por tipo de recurso utilizado.....	105
Tabela 4 - Motivos de consulta aos dicionários, discriminando participantes e idiomas	106

LISTA DE TELAS

Tela 1 - Desambigüização de sentidos	137
Tela 2 - Exemplos bilíngües e notas de uso.....	138
Tela 3 - Extensão de exemplos em L2.....	140

LISTA DE FIGURAS

Figura 1 - Verbetes de dicionário bilíngüe tradicional.....	12
Figura 2 - Verbetes de dicionário bilíngüe pedagógico	13
Figura 3 - Fatores que influenciam a escrita em L2	35
Figura 4 - Perspectivas de análise dos dados.....	99
Figura 5 - Resultado de consulta ao DAFLES (1).....	118
Figura 6 - Resultado de consulta ao DAFLES (2).....	119
Figura 7 - Modelo de verbete do DB+.....	121
Figura 8 - Recursos de um ambiente para escrita em L2.....	131
Figura 9 - Esquema de links	141

INTRODUÇÃO

Escrever em língua estrangeira (L2) é uma tarefa que apresenta uma série de dificuldades além daquelas já enfrentadas na tarefa de escrever em língua materna (L1). Isso significa que, além de exigir o mesmo esforço cognitivo desde a geração de idéias até sua expressão sob forma de texto, a escrita em L2 ainda exige que se reúnam estratégias para encontrar, escolher e empregar itens lexicais que não são suficientemente familiares para serem acessados, escolhidos e empregados com a mesma facilidade de itens lexicais da L1. Mas é um desafio que precisa ser superado, pois um grande volume de textos precisa ser exportado diariamente no mundo globalizado. O uso de dicionários é uma constante na tentativa de superar as dificuldades de escrever em L2, o que coloca em foco as discussões acerca da adequação das obras lexicográficas para atender as necessidades dos usuários nesse tipo de situação.

A demanda por dicionários de apoio à codificação em L2 (não só escrita, mas também oral) não é uma novidade. Dicionários bilíngües (DBs) tornam-se necessários sempre que uma sociedade deseja comunicar conteúdos para outra sociedade que não compartilha o mesmo idioma. É o que já ocorria, por exemplo, no contexto da colonização. Quem exportava discursos (nesse exemplo, o colonizador) necessitava desse tipo de dicionário.

No Brasil, esse era o caso do dicionário português-tupi, *Vocabulário na Língua Brasileira*, de 1621, “elaborado para que os missionários pudessem pregar em tupi” (NUNES, 1996, p. 91). Os jesuítas precisavam aprender o tupi não apenas para entenderem os índios, mas principalmente para se fazerem entender por eles. Embora um dicionário português-tupi pudesse, teoricamente, ser direcionado à sociedade falante do tupi para decodificação do português, não era esse, obviamente, o propósito do dicionário em questão. Observando um verbete desse dicionário citado por Nunes (1996), percebem-se informações sobre o

comportamento pragmático da língua tupi, ou seja, relevantes para um estrangeiro, mas inconscientemente assimiladas por qualquer falante nativo:

Pancada, pelo sinal dela que fica na carne ou lugar aonde se deu. – Bora, 1, Porá, ou com a partícula guera que faz pretérito. Ut. Itapora, ou, Itapopuera, i. pancada, ou sinal onde me firou a pedra et sic de coeteris: e se foi dada por mão de alguém, ora seja pancada ou pedrada. Moaçabora. (NUNES, 1996, p. 101)

Partindo-se do item *pancada*, em português, são fornecidas informações para codificar a mesma idéia em tupi: se a intenção for se referir ao efeito da pancada em quem a recebeu, diz-se *borá*; se a intenção for se referir à pancada de uma pedra, diz-se *itapora* ou *itapopuera*; se a intenção for se referir a uma pancada dada pela mão de alguém, diz-se *moaçabora*.

Esse dicionário apresenta o português na situação de língua do colonizador (estrangeiro) em relação ao tupi, língua do colonizado. No entanto, nos séculos que se seguiram, o português tornou-se a língua do Brasil colonizado, deixando de ser estrangeira em terras brasileiras.

Oficialmente, o Brasil deixou de ser colônia há 200 anos, mas o processo de descolonização cultural ainda está em curso. A história da Lexicografia bilíngüe brasileira é testemunha desse processo. Os primeiros DBs em que uma das línguas é o português do Brasil começaram a ser produzidos em meados do século XX, para atender a demanda criada pela importação de textos literários de outros países. Na segunda metade do mesmo século, os DBs começaram a refletir também a preocupação em atender as necessidades ligadas à transferência de tecnologia industrial. A ênfase, pois, dos DBs, era atender as atividades de recepção de L2 e não de produção em L2.

Sendo assim, até recentemente não havia no Brasil nenhum dicionário que apresentasse situação similar à encontrada no dicionário português-tupi, ou seja, situação em que o português é a língua conhecida de partida, língua de um conteúdo, e a L2 é a língua na

qual se deseja transmitir esse conteúdo e sobre a qual informações de léxico, de gramática e de uso são apresentadas na microestrutura do dicionário.

Só uma forte demanda pela exportação de conteúdos brasileiros em L2 poderia deflagrar a produção de dicionários específicos para as necessidades de codificação. Tal condição começou a aparecer no final do século XX, quando o Brasil passou a aumentar sua participação no comércio, na política e na cultura internacionais.

Essa situação parece ser a responsável por alguns dos avanços na Lexicografia brasileira na virada do século. A evolução mais sensível é a observada nos dicionários escolares inglês-português / português-inglês, talvez porque, a partir dos anos 80, a procura pelo aprendizado de idiomas estrangeiros, notadamente o inglês, tenha aumentado muito. Esses DBs são conhecidos por DBs pedagógicos e, embora suas seções L1-L2 não atendam completamente as necessidades de quem escreve em L2, atendem mais do que os DBs tradicionais.

Compare-se o conteúdo da entrada “ter” em um dicionário português-inglês de 1960 com a mesma entrada num dicionário pedagógico de 2002 (um dos mais recentes do mercado):

Figura 1 - Verbetes de dicionário bilíngüe tradicional

**DICIONÁRIO PORTUGUÊS-INGLÊS THE PORTUGUESE-ENGLISH
DICTIONARY, organized and compiled by Hygino Aliandro.**

ter, vt. To have; to possess, to own, to hold. Tenho um livro, I have a book (I've got a book). – a pouca vergonha de, to have the impudence, the cheek to do. – que fazer, to have something to do. - ciúmes, to be jealous. – frio, to be cold. – fome, to be hungry. – sede, to be thirsty; - calor, to be warm. – medo, to be afraid; - acanhamento, to be shy. – vergonha, to be ashamed. – razão, to be right. Não – razão, to be wrong. – a bondade, to be so kind as. – de, to have to, to be to. Ir – a, to lead, to end in. Que tem você? What is the matter with you? Você tem horas? Have you the time on you? Ele não tinha nenhum dinheiro consigo, he hadn't any money on him. Se eu o tivesse sabido, had I known it, if I had known it. – falta, to be in want. – raiva, to be angry. – para dar e vender, to have enough and to spare of something. Ter em muito, to have a good opinion of someone. (Coisa: to value or to set store by, or to attach importance to something). (ALIANDRO, 1960)

Figura 2 - Verbetes de dicionário bilíngüe pedagógico

LONGMAN DICIONÁRIO ESCOLAR INGLÊS-PORTUGUÊS PORTUGUÊS-INGLÊS PARA ESTUDANTES BRASILEIROS

ter verbo

to have é a tradução na maioria dos contextos:

Eles têm muito dinheiro. They have a lot of money. *Na semana passada tivemos duas provas.* We had two tests last week. *Ela teve uma menina.* She had a little girl.

No presente, **to have got** é mais freqüente do que **to have** no inglês britânico falado:

Tenho três irmãs. I have three sisters./I've got three sisters.

Note que em inglês tem-se que usar o artigo indefinido nos seguintes contextos:

Ele tem carro/celular/computador. He has a car/ a cellphone/ a computer.

Não se usa o auxiliar **to do** para formar frases negativas e interrogativas com **to have got** :

Não tenho bicicleta. I don't have a bike./I haven't got a bike. *Você tem patins?* Have you got rollerskates?/Do you have rollerskates?

EXCEÇÕES

IDADE

Tenho quatorze anos. I'm fourteen years old. *Quantos anos tem o João?* How old is João?

SENSAÇÕES, SENTIMENTOS

Tenho muita sede. I'm very thirsty. *Ele tem medo de trovoadas.* He's afraid of thunder. *Ela tem carinho por mim.* She's fond of me.

MEDIDAS

Tem vários metros de comprimento. It's several meters long. *Quanto tem de largura?* How wide is it?

EXPRESSÕES

não tem de quê you're welcome **ter a ver** *Uma coisa não tem nada a ver com a outra.* One thing has nothing to do with the other. *Isso não tem nada a ver com você.* That has nothing to do with you.

ter que verbo [auxiliar] 1 (para expressar obrigação, necessidade) to have to *Tivemos que ajudá-lo.* We had to help him. ► No presente usa-se também **to have got to**, embora mais freqüentemente no inglês britânico *Temos que estar lá às oito.* We have to be there at eight o'clock./We've got to be there at eight o'clock. ► Não se usa o auxiliar **to do** para formar frases negativas ou interrogativas com **to have got to** *Não tenho que fazê-lo hoje.* I haven't got to do it today./I don't have to do it today. *Você tem que sair?* Have you got to go out? / Do you have to go out? Para fazer recomendações também se usam os modais **must**, **should** e **ought to** *Você tem que ouvir este CD.* You must listen to this CD. / You've got to listen to this CD. *Você tem que ir visitá-la.* You should go and see her. / You ought to go and see her. Para fazer conjecturas ou expressar certeza, usa-se o modal **must** *Isso tem que estar aqui.* It must be here.

• **ter feito verbo [auxiliar] 1** (pretérito perfeito composto) **eu tenho/você tem etc. feito** I/you etc. have been doing *Você tem assistido os jogos?* Have you been watching the games? **2** (mais-que-perfeito) **eu/ele etc. tinha feito** I/he etc. had done *Ele já tinha visto o filme.* He had already seen the movie. **3** (futuro do pretérito composto) **eu/ela etc. teria feito** I/she etc. would have done *Eu teria ido se pudesse.* I would have gone if I could.

(BUENO DE ABREU *et al.*, 2002)

Observam-se mudanças no conteúdo e na forma de exibição. Apesar de o DB mais antigo ter um rico conteúdo, apresenta os equivalentes em seqüência, com poucas pistas para que o aprendiz se decida por um deles ou possa empregá-los adequadamente. Já no DB pedagógico, inseriu-se uma metalinguagem na língua do aprendiz, informando questões relevantes para brasileiros que precisam codificar em L2, levantadas pela análise de *corpus* de aprendizes.

Embora a gama de sentidos apresentada no DB pedagógico seja mais extensa e denote o cuidado com o aspecto contrastivo entre o uso do português do Brasil e o inglês, há informações no DB mais antigo que não foram contempladas no DB pedagógico, como, por exemplo, os equivalentes *to own* e *to hold* e os equivalentes para: *ter a pouca vergonha de / ter a bondade / o que você tem? / você tem horas? / ter falta / ter para dar e vender / ter em muito* (alta conta). Isso indica que o DB pedagógico, embora seja inovador em diversos sentidos, não é completo e, portanto, não torna dispensáveis as obras lexicográficas mais tradicionais. As características desse dicionário serão estudadas detalhadamente no item 5.5.

Portanto, a transição de país-objeto-de-colonização para país-sujeito-internacional se dá na medida em que os discursos brasileiros são comunicados para outras nações e isso precisa ser feito em outras línguas, pois se o fossem em língua portuguesa, a probabilidade de alcançar seus destinatários seria muito menor.

Contudo, há uma grande diferença entre a comunicação corriqueira e a comunicação eficiente de sutilezas do pensamento em L2. Quanto mais elaborado e complexo o conteúdo a ser expresso, maior é o desconforto ao tentar expressá-lo em L2. A fluência necessária para situações cotidianas, em L2, muitas vezes é insuficiente para garantir a comunicação de um discurso de cunho mais intelectual, ou seja, idéias sofisticadas expressas em L1 podem soar infantilizadas se codificadas em L2 de uso cotidiano.

Embora pareça haver consenso sobre o fato de que escrever em L2 é uma tarefa que apresenta todas as dificuldades de escrever em L1 e mais algumas, ainda pouco se sabe sobre quais dessas dificuldades poderiam ser superadas com um aparato lexicográfico adequado e menos ainda sobre como deveriam ser os dicionários para apoio à produção de textos em L2.

Esse tema vem crescendo nas discussões entre teóricos da Lexicografia. No Brasil, contudo, o tema foi pouco abordado até o presente momento. Talvez o trabalho mais significativo a respeito seja a tese do Prof. Philippe Humblé (1997), da UFSC, intitulada “A

new model for a foreign learner's dictionary". Já recentemente, na UNB, Denise Gomes (2006) realizou uma pesquisa empírica sobre o uso do DB na redação em L2 (alemão). Como seqüência a essas iniciativas, propõe-se, neste trabalho, fazer uma descrição do estado da arte no que diz respeito a dicionários para apoio à codificação escrita em L2 e acrescentar novas contribuições sob forma de parâmetros para o desenvolvimento de obras desse tipo.

Para entender essa proposta, é fundamental que se perceba que um dicionário pode ser adequado para a finalidade de decodificar uma L2, mas inadequado para a finalidade de codificar em L2. Vários estudos já foram publicados falando da necessidade do desenvolvimento de dicionários específicos para suporte à "produção", à "redação" e à "codificação" em L2.

Pode-se considerar que um texto escrito em L2 seja resultado de uma dessas três situações:

1. Falante de L1 produz diretamente um texto em L2 (o texto em L2 é um texto original)
2. Falante de L1 traduz para L2 um texto produzido por ele mesmo em L1 (o autor do texto em L2 é o mesmo do texto em L1, mas o texto em L2 não é o original e sim uma tradução)
3. Falante de L1 traduz para L2 um texto produzido por outro falante de L1 (o autor em L2 é um tradutor do texto de outro autor: o texto em L2 é uma tradução).

Na codificação em L2, a liberdade do redator diminui à medida que aumentam as restrições impostas por elementos aos quais ele deve algum grau de fidelidade, ou seja, a liberdade é maior na primeira das três situações citadas, diminui na segunda e é muito pequena na terceira situação. Por exemplo, ao se pedir que um aprendiz escreva algo em L2, ele pode escolher dizer o que seu repertório lexical lhe permita, descartando as idéias que suas limitações de conhecimento da L2 não lhe permitam expressar. Nessa situação, é possível evitar o uso do dicionário.

Já no caso de uma tradução de L1 para L2, se o tradutor é também o autor do texto de origem, poderá sentir alguma liberdade em promover alterações na forma de expressar em L2 as idéias do texto de origem. Mas se o texto de origem for de autoria de outra pessoa, a liberdade de expressão diminui drasticamente e nessa situação é ainda mais provável que o tradutor recorra ao uso do dicionário.

Para a Lexicografia, no entanto, não são relevantes essas distinções, pois em qualquer das situações citadas, se o usuário busca auxílio no dicionário é porque precisa de informações para codificar em L2. Mas, como deve ser uma obra lexicográfica de suporte à codificação? Para poder responder a essa questão, um longo caminho de investigação foi percorrido. O assunto mostrou-se mais abrangente do que se acreditava a princípio e novas interfaces com outras linhas de pesquisa apareceram no decorrer desse percurso, como a Lingüística de *Corpus*, a Lingüística Aplicada, a Lingüística Computacional, a Análise do Discurso e a Psicolingüística. Essa interdisciplinaridade exigiu esforços para amalgamar os conhecimentos e para fazer uma delimitação do tema.

Apesar de arriscado, o resultado dessa empreitada foi gratificante, uma vez que se acredita que as fronteiras entre as ciências não sejam rígidas e que o diálogo com outras áreas reforce os conceitos da própria especialidade em que se atua.

O resultado do percurso de pesquisa foi organizado em seis capítulos.

No capítulo 1, com o objetivo de tecer um pano de fundo para as discussões em torno do assunto em foco, apresenta-se uma revisão sobre os diversos aspectos da escrita em L2 que vêm sendo estudados atualmente. Nesse capítulo, procura-se expor toda a complexidade da atividade de escrever em L2, atividade para a qual o uso do dicionário pode contribuir apenas parcialmente.

Os quatro capítulos seguintes constituem a fase diagnóstica. Nessa fase procurou-se conhecer o problema e coletar subsídios para soluções. No capítulo 2 comentam-se os

trabalhos na Lexicografia acerca da necessidade de dicionários para apoio à codificação em L2. Esse capítulo mostra que, embora a idéia não seja inovadora, a melhor forma para sua implementação ainda não é um consenso.

Na seqüência, o capítulo 3 exhibe uma revisão crítica dos trabalhos produzidos em dois caminhos diferentes, mas que levam à descoberta das necessidades dos usuários em situação de codificação escrita em L2. Um desses caminhos é a pesquisa do uso do dicionário em atividades de redação em L2 e de tradução para L2. O outro caminho é a análise de *corpus* de aprendizes de L2.

No capítulo 4, descreve-se a pesquisa que se realizou, sobre o uso do dicionário na codificação, com a finalidade de explorar empiricamente a situação-base do problema em foco, já que os demais dados levantados são de fontes secundárias.

Finalizando a fase diagnóstica, são analisados, no capítulo 5, os dicionários que apresentam como proposta suprir a lacuna, apontada nos capítulos anteriores, relativa a obras que atendam as necessidades de codificação. Pretende-se, com isso, observar pontos fortes que possam ser imitados e pontos fracos que devam ser evitados em novos projetos de mesma natureza.

No capítulo 6, os resultados da fase diagnóstica são sintetizados na elaboração de propostas. Traçam-se linhas gerais para reunião de vários recursos em um mesmo ambiente computacional para escrita em L2, um dos quais um DB específico para busca de equivalentes a partir da L1. Em seqüência, foca-se a descrição de parâmetros específicos para a elaboração desse tipo de DB. E, finalmente, apresenta-se um capítulo com as conclusões.

É oportuno discorrer sobre a terminologia que será usada neste trabalho. Encontra-se, freqüentemente, o termo *produção* para designar tanto o processo quanto o produto das manifestações orais e escritas em L2. O processo de produção, porém, envolve operações cognitivas que ainda estão sendo investigadas por lingüistas e psicólogos, conforme será

discutido no capítulo 1. Para fins de designar a seleção lexical, que é a parte do processo de produção em que se buscam as palavras adequadas para exprimir determinado conteúdo em L2, será adotado o termo *codificação* e para designar o produto desse processo, será adotado o termo *redação* ou *tradução para L2*. O termo *produção* será empregado sempre num sentido mais abrangente, compreendendo todas as operações cognitivas envolvidas na criação de um texto.

Quanto ao tipo de dicionário em foco nesta pesquisa, alguns o chamam de *dicionário de codificação*, *dicionário de produção* ou *dicionário ativo*. O termo *produção*, conforme já explicado, é muito abrangente, e o termo *ativo* contrapõe-se a *passivo*, dois atributos que não parecem ser adequados para designar o dicionário em si. Assim, optou-se por chamá-lo, no âmbito desta tese, de *dicionário para codificação*, com o cuidado de especificar que se focaliza a *codificação escrita em L2*, uma vez que a codificação oral é uma situação diferente e exigiria outro tipo de pesquisa.

Outra opção terminológica é a utilização das abreviaturas *L1*, para a língua materna e *L2*, para língua estrangeira. Contudo, nos contextos em que se trabalha a distinção entre a L2 falada em ambiente de L2 e a L2 falada fora do ambiente de L2, decidiu-se utilizar duas formas por extenso e em itálico para designar, respectivamente, os dois conceitos: *segunda língua* e *língua estrangeira*. Assim, *segunda língua* é, por exemplo, o francês falado por brasileiros na França e *língua estrangeira* é o francês falado por brasileiros no Brasil.

1. A ESCRITA EM LÍNGUA ESTRANGEIRA

O objetivo desta tese é chegar à definição de uma proposta de solução para apoio aos brasileiros que escrevem em L2 e toda boa solução pressupõe uma boa compreensão do problema. De início, pensou-se em estudar as dificuldades de escrever em L2 apenas por meio de uma pesquisa bibliográfica restrita ao paradigma da Lexicografia, o que será apresentado nos capítulos 2 e 3. Os estudos lexicográficos sobre a relação dos dicionários com a escrita em L2 não pareceram, porém, ser suficientes para um diagnóstico do problema. Sendo assim, decidiu-se fazer incursões em outras áreas da Linguística onde quer que a questão da escrita em L2 fosse tratada. Essa linha de raciocínio mostra-se coerente com as tendências da Lexicografia, que conquistou avanços nos últimos anos exatamente por tentar conhecer mais sobre seus usuários e suas respectivas necessidades. Embora grande parte das leituras feitas não seja objeto de estudo nesta pesquisa, o conjunto forneceu um pano de fundo para tratar o foco do trabalho.

Essa fase do percurso revelou-se bastante complexa, pois a produção científica a respeito é extensa e os aspectos investigados já estão constituindo novas especializações. O assunto – escrita em L2 – gerou vários tópicos de discussão específica na literatura (cf. SILVA & BRICE, 2004):

- Processo de redação em L2;
- A avaliação da produção escrita em L2, seus diversos métodos e respectivas vantagens e limitações;
- Comparação entre redação em L1 e L2;
- Levantamento das variáveis que afetam a redação em L2;

- Comparação entre redação em *segunda língua* (num ambiente da língua, em que o usuário é estrangeiro) e redação em *língua estrangeira* (num ambiente em que o usuário é nativo e a língua é estrangeira);
- Questões de ideologia na escrita em L2;
- Aspectos de construção de voz e identidade por meio da redação em L2;
- Influência de computadores e tecnologia na escrita em L2;
- Plágio em L2;
- Gramática e vocabulário na escrita em L2;
- Métodos para ensino de redação em L2;
- Estudos do texto escrito em L2 (gênero, foco, contexto, nível).

A bibliografia sobre esses diferentes tópicos é abundante e seria impossível ler todo o material a que se tem acesso. Optou-se, assim, por uma leitura exploratória de resumos, procurando construir um panorama sobre o assunto. Nesse processo foram identificados vários trabalhos que mereceram uma leitura completa e cuidadosa, pois traziam subsídios relevantes para o propósito desta tese e que serão aqui abordados.

Nos estudos sobre a escrita em L2 encontram-se argumentos que reforçam a importância de desenvolver um aparato lexicográfico para suporte à escrita em L2 no Brasil, principalmente para a escrita em inglês.

Conforme dizem Silva e Brice (2004) “Devido, em parte, à globalização e à subsequente necessidade de comunicar via computador, a escrita em segunda língua tornou-se um foco importante, se não o dominante, dos trabalhos nos estudos de segunda língua”¹² (p. 70). Desde 1992, uma revista internacional dedica-se a publicar estudos sobre o assunto: *Journal of Second Language Writing*³. Grande parte desses estudos trata do inglês. Aliás, os

¹ Due partly to globalization and the subsequent need to communicate via computer, second language writing has become an important if not dominant focus of work in second language studies.

² Esta e as demais citações de autores estrangeiros são traduções minhas.

³ <http://www.jslw.org/>

estudos sobre a escrita que tratam do inglês como *língua estrangeira* tendem, na opinião desses autores, a superar os estudos do mesmo tipo com o inglês no papel de *segunda língua*.

Flowerdew (1999) constata que o inglês tornou-se a língua internacional de divulgação científica e pesquisadores do mundo todo se vêem compelidos a escrever nesse idioma. O autor aponta pontos positivos e negativos nesse fato. O inglês proporciona, por exemplo, a rápida disseminação do conhecimento humano e promove o entendimento entre as culturas. Por outro lado, autores não-nativos ficam em desvantagem se comparados aos nativos de língua inglesa no momento da seleção para publicação. Assim, o mundo fica privado das descobertas que deixam de ser divulgadas, em especial aquelas dos países em desenvolvimento, que têm menos recursos para apoiar autores no desafio de escrever em inglês. Além disso, segundo Flowerdew, há de se considerar o prejuízo cultural para as línguas que são preteridas como meio de divulgação científica; a carga ideológica que acompanha a língua (no caso, o inglês) e o risco de a homogeneização lingüística eliminar diversidades culturais.

Flowerdew comenta que, nos países de primeiro mundo cuja língua não é o inglês, um grande aparato lingüístico e tecnológico apóia os autores na tarefa de escrever em inglês. Mas essa não é a realidade dos países em desenvolvimento. O que o autor observou pode ser aplicado ao Brasil, onde o desenvolvimento de recursos para apoiar os autores de trabalhos científicos na escrita em inglês poderia contribuir para o aumento da divulgação de suas pesquisas internacionalmente. Da mesma forma, seria fundamental desenvolver o mesmo tipo de aparato para os brasileiros escreverem em outros idiomas estrangeiros de grande expressão internacional, como o espanhol e o francês. Isso não significa que línguas de menor expressão deveriam ser preteridas, mas apenas que existem prioridades a serem observadas no atendimento das demandas.

As pesquisas específicas sobre a escrita em L2 fornecem muitos subsídios para o propósito desta tese e serão comentados a seguir por tópicos. Grande parte dessas pesquisas utilizou a técnica do protocolo verbal durante o processo de escrita, na tentativa de entender como ele acontece. Contudo, essa técnica é sempre utilizada durante redações em ambiente acadêmico. Para investigar outros aspectos, como as características do texto escrito em L2, alguns estudos começaram a utilizar um *corpus* de documentos originais escritos em L2: correspondências, e-mails, publicações etc.

1.1. O processo de produção da escrita em língua estrangeira

Conhecer o processo de escrita em L2 ainda é um desafio para os pesquisadores. Acredita-se que os processos de produção oral e escrita em L2 sejam em grande parte semelhantes aos processos de produção oral e escrita em L1. O processo de produção escrita, por sua vez, “herdou” parte dos resultados da pesquisa sobre a produção da fala. Segundo Johnson e Johnson (1998),

Não há razão para acreditar que, em linhas gerais, os processos de escrita sejam substancialmente diferentes dos processos da fala. Mas como a escrita não é, de forma alguma, a fala colocada no papel, a produção de língua escrita envolve regras, convenções e procedimentos que diferem daqueles encontrados na fala.⁴ (JOHNSON & JOHNSON, 1998, p. 263)

Willem Levelt (1989, *apud* LEVELT *et al.*, 1999) propôs um modelo de produção da linguagem constituído por três grandes componentes: o primeiro é o *Conceitualizador* e trata da geração da mensagem num estágio pré-verbal; o segundo é o *Formulador* e trata da transformação da mensagem pré-verbal em mensagem verbal; o terceiro é o *Articulador*, que trata da transformação da mensagem verbal mental em manifestação física.

⁴ There is no reason to believe that, in broad outline, writing processes are substantially different from those of speech. But because writing is emphatically not ‘speech written down’, the production of written language will involve rules, conventions and procedures which differ from those found in speech.

O modelo de Levelt tem grande aceitabilidade tanto no campo das ciências da cognição quanto em outras áreas, como a Linguística Computacional e a Inteligência Artificial. Isso talvez se deva ao fato de Levelt ter considerado todas as contribuições de seus predecessores para o assunto e ter revisado periodicamente o modelo, incorporando evidências das pesquisas sobre processamento da linguagem.

Johnson e Johnson (1998) fazem uma breve revisão das pesquisas sobre o assunto e afirmam que o modelo de Levelt foi corroborado por diversos estudos. O acesso ou seleção lexical, que ocorre no componente *Formulador* do modelo, é o processo cognitivo de maior relevância para fins lexicográficos. Os estudos sugerem que esse estágio ocorra em dois sub-estágios: primeiro é feita a seleção do lema e depois a seleção do lexema (as flexões, portanto, ocorrem nesse segundo momento). O modelo de Garret (1975, *apud* JOHNSON & JOHNSON, 1998, p. 262) para a seleção lexical, por exemplo, já distinguia a seleção de palavras plenas da seleção de palavras funcionais, pois as evidências mostram que essas duas classes de palavras jamais são confundidas em falhas da língua: “o prato de pato selvagem” pode ser equivocadamente trocado por “o pato de prato selvagem”, mas nunca por “prato pato de o selvagem”. É como se as palavras funcionais fossem estruturais e deixassem lacunas para serem preenchidas pelas palavras plenas.

Fortkamp, Prebianca e Pires (2005) fazem uma revisão teórica sobre as hipóteses levantadas para explicar a seleção lexical, ou seja, como se dá a passagem da mensagem pré-verbal para mensagem verbal. No caso da produção linguística de bilíngües, tudo indica que o acesso lexical seja independente de língua. Isso significa que todos os itens lexicais de L1 e L2 que atendessem às propriedades semânticas da mensagem pré-verbal seriam ativados simultaneamente e, só então, o cérebro selecionaria um desses itens na língua em que a mensagem precisasse ser codificada.

Falantes não-nativos, contudo, não têm como os bilíngües uma “paridade” entre nível de conhecimento de L1 e L2. Dessa forma, pode-se levantar a hipótese de que, quando se pretende produzir uma mensagem em L2, ocorram situações em que apenas itens lexicais em L1 sejam ativados. Outra possibilidade decorrente da lacuna em L2 seria a ativação de itens lexicais em L2 que apresentassem alguns dos traços semânticos da mensagem pré-verbal (sinônimos, hiperônimos, hipônimos etc.). Em ambas as situações, abrem-se oportunidades para consultas ao dicionário, bilíngüe, no primeiro caso, e monolíngüe, no segundo caso. O dicionário, assim, funcionaria como um complementador do léxico mental.

Segundo Clahsen e Felsen (2006), os estudos sobre o processamento lingüístico de L2 por não-nativos são mais recentes que os de L1. Os autores revisam a literatura a respeito e concluem que as diferenças entre processamento por nativos e não-nativos são menores do que se costuma acreditar. Não há evidências conclusivas, contudo, sobre o quanto a L1 interfere no processamento da L2. Ao que parece, o “caminho” percorrido pelo falante não-nativo para processar uma determinada língua é mais longo que o caminho percorrido pelo nativo (provavelmente pela participação da L1), o que explicaria o tempo de resposta maior observado em não-nativos. O modelo de três componentes de Levelt aplica-se bem à produção em L2, porém, tendo por parâmetro tudo que já foi investigado sobre a seleção lexical em L1, ainda há muitos aspectos a serem investigados sobre a seleção lexical em L2.

Acredita-se que a seleção lexical em L2 compartilhe de um conjunto de dificuldades com a seleção lexical em L1, porém são as dificuldades típicas da seleção lexical em L2 que se tem em mente no presente estudo. Se em L1 normalmente se possui um grande estoque de conhecimento da língua em uso (léxico mental), em L2 esse estoque costuma ser relativamente menor e insuficiente para orientar as escolhas lexicais.

Como a seleção lexical ocorre quando o sujeito já tem idéia do que deseja ou precisa expressar e o uso do dicionário acontece especificamente nesse momento, o dicionário não

contribui para o processo de produção como um todo, conforme já explicado no capítulo introdutório, quando se justificou a escolha do termo *dicionário para apoio à codificação*. Esse tipo de dicionário não teria o propósito de auxiliar a geração de idéias, mas apenas de traduzi-las em palavras.

1.2. O uso da língua materna na escrita em língua estrangeira

Woodall (2002) refutou o resultado de pesquisas que dizem que o processo de produção escrita em L2 é basicamente igual ao de L1. Segundo ele, existe uma diferença: quando se redige em L1, pensa-se apenas em L1, mas quando se redige em L2, pode-se pensar em L1 e em L2. O mais comum é haver uma alternância de pensamento entre L1 e L2. O autor pesquisou essas alternâncias e concluiu que elas são relativamente frequentes e têm diferentes funções. A L1 é mais usada quando o nível de proficiência em L2 é menor ou a tarefa é mais difícil. O uso da L1 parece ter influência positiva quando a distância entre a L2 e a L1 é menor (no caso, inglês-espanhol) e negativa quando essa distância é maior (no caso, inglês-japonês). Embora o autor indique a necessidade de novas pesquisas sobre o assunto, buscando verificar a influência de fatores individuais na alternância de línguas, sua conclusão geral é de que “...L1 é um importante recurso para quem escreve em L2”⁵ (WOODALL, 2002, p. 28). Esse argumento pode ser aplicado para defender a idéia de DBs e não apenas DMs para suporte à atividade escrita em L2.

Uzawa (1996), por sua vez, comparou redações em L1 com redações em L2 dos mesmos sujeitos de pesquisa, bem como redações em L1 e L2 com traduções na direção L1-L2. Ele verificou que o tipo de atenção na atividade de redação, tanto em L1 quanto em L2 é similar. Na tradução L1-L2, contudo, constatou que há muito mais atenção ao uso da língua. Por isso, a qualidade das traduções foi bem superior à qualidade das redações. Ele explica o

⁵ L1 is an important resource for L2 writers.

fato dizendo que a redação exige atenção na geração de idéias, na planificação do texto e no uso da língua. Já na tradução L1-L2, o texto já está pronto e a atenção pode ser concentrada no uso da língua. O autor também observou que na redação os alunos utilizaram menos do que sabiam:

De maneira geral, na tarefa de tradução deste estudo, os participantes ficaram liberados das atividades cognitivas de gerar e organizar idéias e puderam se concentrar nas atividades lingüísticas, estendendo seus níveis lingüísticos para o “nível $i + 1$ ” com o auxílio de seus dicionários, ao passo que na tarefa de redigir em L2 eles só usaram o vocabulário e a sintaxe que eles conheciam bem; parece que escreveram redações em inglês no nível “ $i - 1$ ”, usando somente palavras, expressões e sintaxe prontamente acessíveis a eles.⁶ (UZAWA, 1996, p. 288)

O autor cita estudos anteriores (KOBAYASHI, 1992 e TUDOR, 1987, *apud* UZAWA, 1996) que compararam redações feitas diretamente em L2 com redações feitas em L1 e depois traduzidas para L2. A conclusão desses estudos foi de que o preparo em L1 garantiu melhor qualidade às redações.

Uzawa (1996) também cita estudos em que os alunos ousaram demais no uso da língua e geraram textos incompreensíveis. Por isso, conclui que os exercícios de tradução L1-L2 ajudam mais o aprendizado de L2 do que os exercícios de redação, pois promovem a hipótese de *output* compreensível de Swain (1985, *apud* UZAWA, 1996), ou seja, forcem gradualmente o aprendiz a elevar seu nível de proficiência em L2.

A questão do nível da língua nas redações também é abordada no estudo de Wong (2005), o qual observou que os alunos ousam menos em redações quando a representação mental que fazem de sua audiência é a de um avaliador e a representação mental do objetivo da redação é uma avaliação. Ousar menos significa menor probabilidade de errar.

⁶ Generally speaking, in the translation task in this study, the participants were freed from the cognitive activities of generating and organizing ideas, and they were able to concentrate on linguistic activities, stretching their linguistic levels to the “ $i + 1$ level” with the help of their dictionaries, whereas in the L2 writing task, they just used vocabulary and syntax that they knew well; it seems that they wrote English compositions at the “ $i - 1$ ” level, using only words, expressions, and syntax readily accessible to them.

O assunto também é abordado por Ard (1982) que chama atenção para o fato de que quem usa mais o dicionário como um andaime para um nível mais avançado de uso da língua pode cometer mais erros, mas também pode aprender mais.

1.3. Recursos tecnológicos na escrita em língua estrangeira

Dentro do conjunto de recursos que podem auxiliar a produção escrita em L2, destaca-se o trabalho de Yoon e Hirvela (2004), que investigaram as atitudes dos estudantes de L2 em relação ao uso de *corpus* como ferramenta de auxílio à escrita em L2. Os autores salientam que além dos *corpora* gerais de língua, os *corpora* de domínios específicos e gêneros textuais específicos também são muito valiosos para a escrita. Os estudantes pesquisados receberam instrução sobre como utilizar concordanciadores (programas para consulta de *corpora*) durante um curso de escrita em L2.

Esse estudo é de interesse direto para este trabalho, pois os estudantes compararam o uso do *corpus* com o uso do dicionário. O uso de *corpus* foi considerado valioso para descobrir como combinar palavras em L2, pois a consulta traz diversos exemplos de uso da palavra-chave, o que aumentou a confiança dos sujeitos de pesquisa. Na opinião dos pesquisados, em comparação com o dicionário, o uso do *corpus* demanda mais tempo para análise do resultado da consulta. *Corpus* e dicionário foram considerados ferramentas complementares: “... um dicionário é útil para a aquisição do significado de palavras, mas um *corpus* é mais útil para aprender como e onde empregar palavras em contexto⁷” (YOON E HIRVELA, 2004, p. 277).

Dos 22 sujeitos da pesquisa de Yoon e Hirvela (2004), apenas um disse utilizar exclusivamente dicionário impresso e dois disseram utilizar o impresso e o eletrônico. Os outros dezenove participantes declararam utilizar apenas dicionários eletrônicos.

⁷ ... a dictionary is useful for acquiring the meaning of words, but a corpus is more useful for learning how and where to put words in context.

Já no que diz respeito à mídia da escrita em L2, Li (2006) comparou a escrita à mão com a escrita em computador. Seus resultados corroboraram estudos anteriores ao apontar que os processadores de texto ajudaram os autores a revisar mais. Segundo Li:

As capacidades do processador de texto, tais como corrigir ortografia e gramática, mover blocos, deletar blocos, formatar, consultar *thesaurus* e armazenar informações facilitam a geração, a revisão e a edição de texto. Conseqüentemente, os alunos têm demonstrado reorganizar textos escritos sem muito esforço se comparado com a escrita à mão; escrever no computador libera os alunos de ter que reescrever suas redações, o que, por sua vez, diminui sua resistência às revisões.⁸ (LI, 2006, p. 6)

Ainda segundo o autor, os computadores remodelam o processo de escrita e de revisão das pessoas que adaptaram suas habilidades de escrita para essa mídia. Ele defende que não se pode forçar alguém a escrever em uma mídia à qual não está adaptado. Assim, não vê sentido em forçar quem escreve no computador a escrever à mão (e vice-versa), pois isso poderia influir no resultado do processo.

Podem-se aproveitar essas considerações para refletir sobre a mídia de um dicionário. Quando se pretende fornecer aparato lexicográfico para usuários que escrevem no computador, parece que os dicionários eletrônicos são mais adequados, pois a mídia impressa poderia afetar negativamente o desempenho desses usuários. No caso da pesquisa de Yoon e Hirvela (2004), já comentada, mais de 90% dos sujeitos pesquisados declararam utilizar apenas dicionários eletrônicos. Embora essa pesquisa não tenha representatividade estatística, a questão da mídia deve ser cuidadosamente considerada no desenvolvimento de propostas de soluções lexicográficas. Da mesma forma como na escrita em papel a disposição para revisões é menor, pois é mais trabalhosa, pode-se levantar a hipótese de que a disposição para consultas a dicionários impressos é menor que para consultas a dicionários eletrônicos.

⁸ The capacity of Word processing – such as spell – and Grammar-checking, block moving, block deleting, formatting, thesaurus, and storage of information – facilitate text generating, revising and editing. Consequently, student writers have been shown to rearrange written texts without much effort as compared to pen-based writing; writing on computers frees students from recopying their essays, which in turn cases their resistance to revising.

1.4. Plágio em língua estrangeira

Um dos motivos que, segundo Yoon e Hirvela (2004), justificam a segurança que a consulta a *corpus* proporciona àqueles que escrevem em L2 é o fato de se poder encontrar o item lexical procurado, sendo utilizado exatamente no sentido em que se pretende, além de se poder observar quais os itens lexicais coocorrem mais freqüentemente nessa situação. Isso, porém, pode ter conseqüências negativas. Ao que parece, na tentativa de reproduzir uma linguagem “natural” em seus próprios discursos, os estrangeiros apropriam-se de seqüências cada vez maiores dos discursos dos falantes nativos, o que pode ser muitas vezes classificado como plágio. Existem estudos que procuram definir a questão do plágio em L2 (PECORARI, 2003; ABASI *et al.*, 2006). Esses estudos indicam que, muitas vezes, não existe má fé da parte de quem escreve. Essa é uma questão a ser levada em conta no momento da elaboração dos exemplos de um dicionário. Pode-se refletir, por exemplo, sobre quais seqüências lexicais já pertencem à língua e quais seqüências são de “propriedade privada” (constituem marca de autoria).

1.5. Gêneros textuais e escrita em língua estrangeira

Também dentro do assunto da produção escrita em L2, há muitos pesquisadores dedicados ao estudo dos gêneros textuais com o objetivo de explicitar suas características e possibilitar a aquisição de L2 para fins específicos, em especial para fins acadêmicos. Uma revisão de 60 estudos empíricos sobre como as pessoas aprendem os gêneros é feita por Tardy (2006). A autora diz que

Tanto autores de L1 quanto de L2 parecem adquirir conhecimento [de gêneros] implicitamente por meio da exposição a textos e também fazer uso explícito de textos-modelo para adquirir

conhecimento sobre a estrutura do gênero — contudo parece que esta estratégia ocorre com mais regularidade nos estudos de autores de L2.⁹ (TARDY, 2006, p. 95)

Outra oportunidade para o aprendizado de gêneros, segundo Tardy, são as interações orais entre colegas do mesmo domínio. Porém, a autora observa que os estrangeiros se beneficiam menos do aprendizado dos gêneros nessas situações, pois a falta de habilidade lingüística muitas vezes os mantém os isolados nessas interações.

1.6. Avaliação da escrita em língua estrangeira

Avaliar a produção escrita é uma tarefa difícil que requer critérios estabelecidos previamente. A análise e a avaliação de produções escritas dependem do estabelecimento de critérios com relação ao que são erros e ao que são inadequações. Ezer e Sivan (2005) defendem que aspectos como lógica, retórica, coesão e coerência, percepção da audiência, voz e pensamento crítico devam ser considerados nessa tarefa.

Uma questão é o texto de qualidade e outra é o texto sem erros. Como afirma Ard (1982), a ausência de erros não implica maior qualidade do conteúdo. Ard observa também que aprendizes que ousam menos podem cometer menos erros. Os aprendizes que usam dicionários, contudo, ousam mais, empregando itens lexicais que não fazem parte de seu léxico mental em L2, e isso pode levá-los a cometer mais erros do que se não usassem o dicionário.

Engber (1995) investiga a relação entre riqueza lexical e as notas atribuídas às redações de alunos. Em sua revisão teórica, ela cita outros trabalhos nos quais se constata que a noção de boa redação está relacionada à diversidade lexical. Linnarud (1986, *apud* ENGBER, 1995), por exemplo, verificou que não-nativos variam menos o vocabulário,

⁹ Both L1 and L2 writers seem to build knowledge implicitly through exposure to texts and also make explicit use of model texts to build knowledge of generic structure—though this strategy seems to come up with greater regularity in studies of L2 writers.

utilizam menos expressões idiomáticas e colocações e suas palavras mais freqüentes não são as mesmas dos falantes nativos.

Engber (1995) menciona ainda pesquisas sobre redação em L2 onde foram medidos: 1) a quantidade de itens lexicais diferentes sobre o total de itens lexicais; 2) a quantidade de itens lexicais pouco freqüentes sobre total de itens lexicais (medida de sofisticação) e 3) o percentual de cada classe gramatical nas redações.

Em sua própria pesquisa, a autora trabalhou com 66 alunos de nível intermediário em ambiente multilíngüe, constituído por aprendizes de inglês de diversas nacionalidades. Dez professores treinados deram nota para as 66 redações resultantes e as notas foram comparadas com as medidas quantitativas de riqueza lexical. O resultado da pesquisa de Engber mostra grande correlação entre a nota atribuída às redações e a diversidade lexical combinada com ausência de erros lexicais. Em suma, isso significa que, se houvesse duas redações sem erros, aquela que apresentasse maior diversidade lexical receberia uma nota maior. A autora confessa a dificuldade de determinar o que são erros e salienta que resultados diferentes talvez pudessem ser obtidos se o tema da redação, o nível de proficiência em L2 e a idade dos alunos fossem outros.

Escrever bem em L2, assim como em L1, é uma questão controversa. O que é escrever bem? Erros de natureza morfossintática são mais facilmente identificáveis, até por programas de computador. Mas há textos que são *instintivamente* ou *intuitivamente* classificados como ruins ou inadequados, sem que se consiga explicar a razão. No que tange o julgamento de redações é preciso estar atento para o risco da interferência de preconceitos de várias origens.

Rajagopalan (2006) alerta para a questão do que se pretende ao almejar um desempenho semelhante ao do “falante nativo”, apagando as marcas da própria cultura no discurso que se produz em L2.

O conceito de “naturalidade” também merece uma reflexão dentro dessa ótica. Num idioma como o inglês, por exemplo, que já possui mais falantes não-nativos do que nativos, o que deveria ser considerado “natural”?

Esse tipo de preocupação poderia levar a outro extremo, que seria aceitar tudo que não pudesse ser gramaticalmente reprovado. Ao tratar de redações em L1, Oliveira (2006) diz que muitas imprecisões são percebidas pelos professores, mas muitas vezes não são apontadas porque os professores não conseguem justificar o que os leva a rejeitar determinadas construções e temem estar impondo seu idioleto aos alunos.

É preciso se deter, portanto, na busca de características concretas para a avaliação da qualidade das redações, a fim de separar o que é justificável como bom uso da língua daquilo que é gerado por motivações pessoais e culturais.

A redação em L1, aliás, já apresenta muitas dificuldades e conhecê-las pode ajudar a entender parte das dificuldades enfrentadas durante a redação em L2.

Oliveira (2006) procedeu à avaliação de redações de vestibulandos e propôs nove categorias de “impropriedade vocabular” ou “inexatidão lexical”. Segundo o autor:

O professor às vezes, sendo um usuário maduro da língua escrita, intui como inadequada determinada combinação de palavras empregadas pelo aluno, mas tem dificuldade para persuadi-lo de que se trata de impropriedade, porque lhe falta uma teoria sobre **propriedade vocabular**. (OLIVEIRA, 2006, p. 50, grifo do autor)

Trata-se, portanto, de um esforço no sentido de teorizar o que era só intuição. São nove as impropriedades vocabulares levantadas pelo autor:

1. Substituição de palavras em expressões idiomáticas. Ex.: “costurando arestas” ao invés de “aparando as arestas”;
2. Criação, ainda que por processos válidos de formação de palavras, de neologismos comunicacionalmente inadequados: “Existem vários tipos de erros, com pequenas, médias e grandes conseqüências para o **errante** e para a sociedade”¹⁰;
3. Incompatibilidade entre a palavra escolhida e o tema ou o gênero do texto: em notícia de jornal: “O gorila fugiu da jaula. O **primata**, no entanto já foi capturado pelos

¹⁰ O autor deixou de esclarecer, contudo, que se trata de um neologismo semântico, pois o item lexical “errante” já existe com outro sentido.

seguranças do zoológico” [segundo o autor, as palavras *animal* ou *fera* seriam mais adequadas à situação];

4. Conflito entre a natureza pejorativa ou meliorativa das escolhas lexicais e a orientação argumentativa do texto. Seria o caso, por exemplo, de um texto que argumentasse contra a ditadura militar e a favor dos exilados e se referisse a estes como “subversivos”. É uma impropriedade que se percebe no nível do texto e não da frase;
5. Incompatibilidade semântica entre vocábulos no interior do sintagma ou da oração: “Não **possuíamos** nenhum representante nesse Colégio Eleitoral.” (*tínhamos* seria adequado); “Pesquisas são realizadas para a **descoberta** de novas máquinas.” (*desenvolvimento/ a criação/ a invenção* seriam adequados);
6. Inexatidão com redundância: “O **emprego da verba** pública tem que ser **investido** em saúde”. Nesse caso, ou omite-se “o emprego” ou omite-se “investido”: “A verba pública tem que ser investida em saúde” / “O emprego da verba pública tem que ser em saúde” (é diferente do pleonasma, pois no pleonasma há redundância, mas não inexatidão);
7. Emprego de vocabulário informal em situações formais e vice-versa (registro inadequado). Está relacionado ao excesso de oralidade na escrita: “A **moleza** da Justiça tem uma explicação, que seria o número excessivo de processos para cada juiz julgar” (nesse caso, seria mais adequado utilizar a palavra *morosidade*);
8. Atração paronímica: confusão entre parônimos em virtude da semelhança fônica entre eles: “O pensamento **frui** livremente” (emprego de *fruir* por *fluir*). “A igreja deve não só **enveredar esforços** para a legalização do jogo, como também fiscalizar o dinheiro que nele for jogado” (emprego de *enveredar* por *envidar*);
9. Vaguidade decorrente do emprego, em contextos inadequados, de palavras com baixa densidade semântica: “...leis que tenham por **fator** principal a criação...” (a palavra *objetivo* seria mais precisa). (OLIVEIRA, 2006, adaptação minha)

Todas essas inadequações vocabulares são passíveis de ocorrer na redação em L2. É o que mostram os estudos baseados em *corpora* de aprendizes (DE COCK & GRANGER, 2005). Conhecê-las pode dar subsídios para que o lexicógrafo contemple, nos dicionários de apoio à codificação, informações suficientes para evitá-las.

Tono (2003), Dagneaux, Denness e Granger (1998), Hasan *et al.* (2004) também na linha de pesquisa que utiliza *corpus* de aprendizes, mostram como uma tipologia de erros pode sustentar a etiquetagem desse tipo de *corpus* e possibilitar a análise automática de erros, abrindo a possibilidade de pesquisar amostras cada vez maiores de aprendizes de L2.

Finalmente, Blattner (1999) mostra que há vários métodos para se avaliar a escrita e cada um se presta a mensurar aspectos diferentes. As formas de avaliação têm pontos fortes e fracos. A autora defende que tanto professores quanto alunos devem inteirar-se dos critérios utilizados para avaliar a escrita, pois isso os torna aptos a ter um posicionamento crítico em

relação aos métodos de avaliação. Além disso, segundo a autora, conhecer as formas de avaliação pode aumentar a autocrítica do sujeito que redige.

A conscientização sobre todos esses aspectos da avaliação da produção escrita contribuiu para que se desenvolvesse maior espírito crítico face às pesquisas revisadas nesta tese, pois a maioria dos estudos sobre escrita em L2 contempla uma avaliação em algum momento.

1.7. Língua materna, nível de proficiência e a escrita em língua estrangeira

Sasaki (2000) revisa a literatura sobre estudos empíricos da escrita em L2 e observa pontos de convergência dos resultados comprovados:

- aprendizes menos habilitados planejam menos e revisam mais;
- a habilidade para a escrita em L1 transfere-se para a escrita em L2;
- deve haver um “talento” para escrita, independente da proficiência em L2 e que afeta a redação em L2;
- o uso mental da L1 pode melhorar a qualidade da redação em L2,
- a qualidade das redações parece mais associada a estratégias de escrita do que à proficiência em L2.

Resultados de pesquisas empíricas indicam, contudo, que a competência escrita em L2 é influenciada pelo nível de conhecimento de L2. Wang e Wen (2004) analisaram o processo de escrita em L2 e descobriram que a proficiência lingüística pode afetar os textos escritos, as estratégias de escrita e a participação da L1 nessa atividade. Utilizando protocolos verbais, Liu (2004) percebeu que os aprendizes de L2 com maior habilidade na escrita em L2 utilizavam mais a L2 enquanto “pensavam em voz alta” do que os menos experientes na mesma situação (estes utilizavam mais a L1). Na mesma linha, Haan e Esch (2005) fizeram uma pesquisa empírica cujos resultados indicam que aprendizes de L2 mais maduros

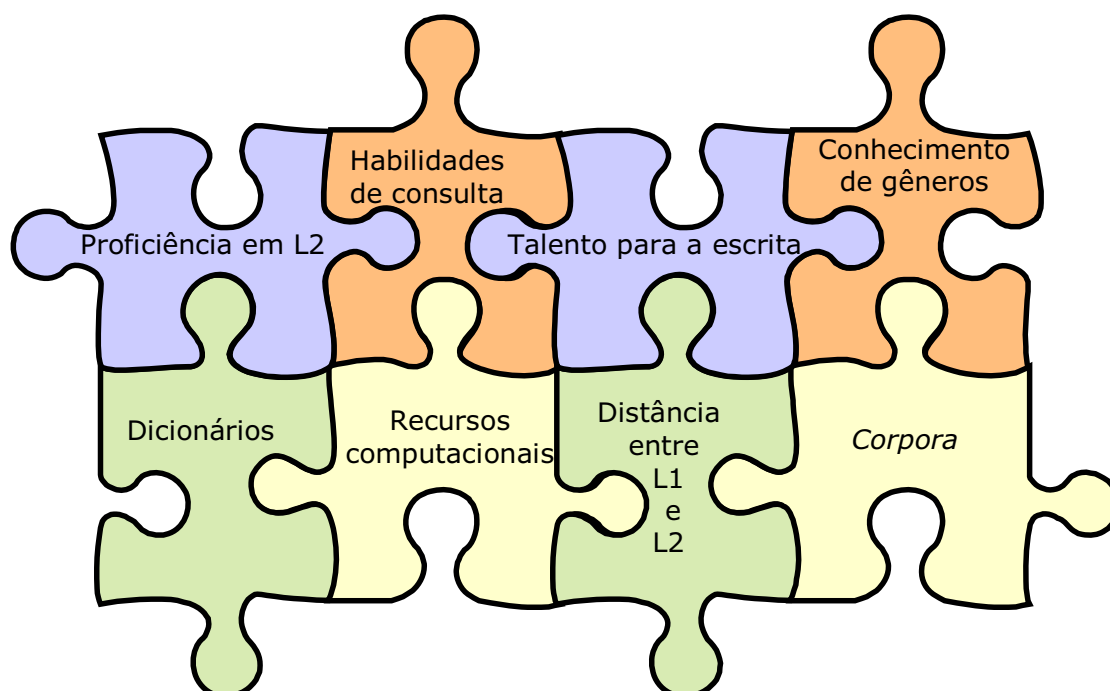
(avançados) produzem itens lexicais e sentenças mais longas e, acima de tudo, são mais fluentes na produção textual.

Parece haver, portanto, uma decrescente participação da L1 no processo de produção textual em L2 à medida que aumenta a proficiência lingüística em L2.

1.8. Conclusões sobre o relacionamento do dicionário com a escrita em língua estrangeira

A principal conclusão a que se chegou após as leituras comentadas neste capítulo é de que uma solução lexicográfica para apoio à codificação escrita em L2, por melhor que seja, não será capaz de suprir todas as necessidades dos usuários para desenvolver uma escrita em L2 de qualidade. Isso significa que o dicionário é uma ferramenta, um acessório nesse processo, o que, de forma alguma, diminui sua importância. A seguir são relacionados fatores que influenciam a escrita em L2:

Figura 3 - Fatores que influenciam a escrita em L2



Observe-se que alguns desses fatores dependem do usuário (as quatro “peças” superiores), outros não (as quatro “peças” inferiores).

Mesmo que o desenvolvimento da proficiência lingüística diminua as dificuldades de escrita em L2, há sempre dificuldades que persistem e as obras de referência devem oferecer informações para que os indivíduos possam saná-las. Portanto, como obras de referência, os dicionários são uma ferramenta de suporte essencial para a escrita em L2.

Há, no entanto, dois problemas impedindo que os dicionários maximizem sua parcela de contribuição para os indivíduos que escrevem em L2: um deles é a falta de habilidade dos usuários para lidar com dicionários e o outro é a falta de adequação dos dicionários para as necessidades de codificação. A solução para o primeiro problema é adotar o ensino do uso do dicionário. Já a solução para o segundo problema exige que se descubra como modelar uma obra lexicográfica que seja realmente útil durante as atividades de codificação. Quanto mais fácil de usar for esse tipo de solução, aliás, menos complexo se tornará o ensino do uso do dicionário. Esse é precisamente o desafio desta pesquisa.

Iniciando a fase diagnóstica do problema, apresenta-se, no capítulo 2, uma revisão crítica dos trabalhos na área da Lexicografia que apontam a inadequação dos dicionários tradicionais para as necessidades de codificação e sugerem possíveis soluções.

2. DICIONÁRIOS PARA CODIFICAÇÃO: OPINIÕES TEÓRICAS

Dentre os trabalhos produzidos por lexicógrafos ou teóricos da Lexicografia, há vários que tratam do apoio lexicográfico à atividade de codificação em L2. Esses trabalhos podem ser agrupados segundo sua natureza:

- I. Artigos de opinião, normalmente escritos por pessoas que conhecem a Lexicografia e o Ensino de L2;
- II. Relatos de pesquisas empíricas
 - sobre o uso do dicionário nas atividades de codificação em L2;
 - sobre a análise de *corpora* de textos em L2 produzidos por aprendizes.
- III. Dicionários elaborados para a finalidade de apoio à codificação (relativamente raros).

Esses três tipos de trabalho integram o processo de evolução da Lexicografia – artigos de opinião discutindo a idéia; pesquisas empíricas investigando a idéia; projetos lexicográficos implementando a idéia. Embora esses momentos não obedeçam a uma seqüência rígida, decidiu-se adotar essa divisão para conferir maior clareza à exposição. Assim, trabalhos do tipo I serão discutidos neste capítulo; trabalhos do tipo II serão discutidos nos capítulos 3 e 4 e trabalhos do tipo III serão analisados no capítulo 5.

2.1. A inadequação dos dicionários tradicionais para a atividade de codificação em língua estrangeira

Desde as primeiras décadas do século XX já se percebia que os dicionários tradicionais não atendiam adequadamente os usuários empenhados em codificar em L2. Essa percepção de “falta” sugeria, portanto, a necessidade de desenvolver dicionários para apoio à codificação em L2.

Shcherba (1940) manifesta consciência do problema que é codificar em L2: “De fato, raramente alguém precisa ser lembrado da palavra mais apropriada para expressar um pensamento em sua própria língua (...), mas quando se usa uma língua estrangeira, isso acontece o tempo todo.”¹¹ (SHCHERBA, 1940, p. 336).

O autor também declara que os DBs (que ele chama de dicionários de tradução) não são adequados para a codificação e que o ideal seria recorrer a um DM de L2, o que, porém, exigiria um nível de conhecimento muito avançado dessa língua:

...dicionários de tradução continuam sendo nosso *malum necessarium*, até que um novo tipo de dicionário seja criado. Devemos tentar aliviar suas falhas por meio de vários paliativos que poderiam levar, finalmente, de forma indireta, à criação do tipo de dicionário de língua estrangeira que me parece ser ideal.¹² (SHCHERBA, 1940, p. 341)

O autor fala do que seriam esses “paliativos”: “as falhas dos dicionários de tradução poderiam ser parcialmente eliminadas, acrescentando-se várias notas e exemplos.”¹³ (SHCHERBA, 1940, p. 341).

Embora critique os DBs e defenda o uso de DM de L2, o autor sugere soluções lexicográficas que incluem a L1 ao lado da L2. Uma dessas soluções é a bilinguização (embora ele não utilize esse termo): “Em 1936 eu já indicava o que considero uma solução radical para o problema, no prefácio de meu Dicionário russo-francês: poder-se-ia criar dicionários de definição da língua estrangeira na língua nativa dos alunos.”¹⁴ (SHCHERBA, 1940, p. 341). Ao detalhar essa solução, ele não exclui os equivalentes. Mas sua “solução ideal” é outra: “...ainda precisamos de dicionários que possibilitariam à pessoa que conheça as

¹¹ Indeed, one rarely needs to be reminded of the most appropriate word to express a thought in one’s native language (...), but when using a foreign language this happens at every turn.

¹² ... translating dictionaries remain our *malum necessarium*, until a new type of dictionary is created. We must try to alleviate their faults by various palliatives that ultimately might lead – in a roundabout way – to the creation of the type of foreign dictionary I perceive to be ideal.

¹³ The faults of translating dictionaries could be partially eliminated by adding various notes and examples.

¹⁴ Already in 1936, I indicated what I consider a radical solution to the problem in the preface of my Russian-French Dictionary: one could create foreign defining dictionaries in the students’ native language.

bases gramaticais de uma dada língua em seu aspecto ativo (...) traduzir textos de não-ficção para a língua estrangeira sem cometer erros grosseiros.¹⁵” (SHCHERBA, 1940, p. 342).

O autor diz que esse tipo de dicionário na direção russo-L2 seria dirigido aos russos e não serviria para fornecer aos não-russos a compreensão dos significados das palavras russas, mas sim para fornecer “aos russos as indicações exatas de como traduzir palavras russas em vários contextos de forma a serem compreendidos e evitarem parecer ridículos¹⁶” (SHCHERBA, 1940, p. 342).

Cowie (1999), por sua vez, relata que A.S. Hornby e Harold Palmer, na segunda metade dos anos 30, chegaram à conclusão que os DMs tradicionais, desenvolvidos para falantes nativos, eram tão inadequados para os aprendizes estrangeiros codificarem quanto os DBs: “Palmer produziu um dicionário para codificação pioneiro. (...) O dicionário de Palmer (...) deixou sua marca no Dicionário Idiomático e Sintático de Hornby (...) e ajudou a construir seu caráter fortemente ‘produtivo’.¹⁷” (COWIE, 1999, p. 13).

Apesar de serem contemporâneos e terem detectado o mesmo tipo de deficiência, Hornby e Palmer, diferentemente de Shcherba, decidiram produzir DMs – e não DBs - mais adequados aos aprendizes de idiomas estrangeiros.

Por que um mesmo problema levou a soluções diferentes? Acredita-se que isso se deva ao ponto de vista desses autores. Shcherba era um lexicógrafo russo preocupado em fornecer soluções de língua francesa para russos (lexicógrafo nativo da L1), ao passo que Hornby e Palmer eram ingleses preocupados em fornecer soluções lexicográficas de língua inglesa para estrangeiros (lexicógrafos nativos da L2).

¹⁵ ... dictionaries are still needed that would enable the person who knows the basics of a given language’s grammar in its active aspect (...) to translate nonfictional texts into the foreign language without making gross mistakes.

¹⁶ ... provide Russians with exact indications of how to translate Russian words in various contexts in order to be understood and to avoid appearing ridiculous.

¹⁷ Palmer produced a pioneering encoding dictionary.(...) Palmer dictionary (...) left its mark on Hornby’s Idiomatic and Syntactic English Dictionary (...) and helped to shape its strongly “productive” character

Ao contrário do que se vê em Shcherba (1940), a idéia de desenvolver dicionários exclusivamente para apoio à codificação não estava explícita nos dicionários de Hornby e Palmer, mas, segundo Cowie, “...não há muita dúvida de que, tanto filosoficamente quanto na prática, Palmer e Hornby favoreceram a função de codificação. No nível teórico, o compromisso com a função de codificação era virtualmente total.”¹⁸ (COWIE, 1999, p. 175).

A Lexicografia monolíngüe voltada para aprendizes de L2 desenvolveu-se fortemente, ao passo que as idéias de Shcherba sobre DBs mais adequados à codificação foram tardiamente divulgadas no ocidente (a tradução de seu trabalho para o inglês é de 1995).

Nos anos 80, no entanto, diversas pesquisas apontaram que, apesar da evidente superioridade das informações contidas nos DMs de aprendizes, os DBs continuavam a ser mais utilizados. Isso coincidiu com a época em que a participação do léxico e da L1 no ensino de L2 sofreu uma revalorização. Essas condições motivaram a iniciativa de bilingualizar um DM de aprendizes, o *Oxford student's dictionary of current English*, de A. S. Hornby, inicialmente para o hebraico e depois para mais de 30 idiomas. O sucesso dessa iniciativa, contudo, satisfaz principalmente as necessidades de decodificação.

A comparação das vantagens e desvantagens entre DBs e DMs para aprendizes está presente em grande parte dos trabalhos lexicográficos, incluindo aqueles que mencionam especificamente as necessidades lexicográficas para a codificação (v. DURAN, 2004, p. 54-58 para argumentos favoráveis e desfavoráveis com relação a DBs e DMs).

Os DBs tradicionais são particularmente criticados, pois não distinguem os equivalentes de tradução e não informam quando e como esses equivalentes podem ser empregados. Se na decodificação o usuário pode “testar” os diversos equivalentes apresentados para inferir qual é mais apropriado para o contexto que motivou a consulta, na codificação não há como fazer o mesmo. Um só equivalente pode dar a falsa idéia de que

¹⁸ ...the can be little doubt that both philosophically and in practice Palmer and Hornby favoured encoding function. At theoretical level, commitment to the encoding function was virtually total.

constitui um sinônimo perfeito da palavra-entrada, mas vários equivalentes enumerados sem distinção também podem, igualmente, induzir a uma má escolha lexical.

Por outro lado, os DMs dão informação sobre o emprego e sobre os diversos sentidos dos itens lexicais, mas para acessar essas informações o usuário tem que ter uma palavra-chave em L2 para iniciar a busca, o que nem sempre ocorre.

No que diz respeito ao tipo de dicionário mais adequado para a codificação, existe uma predileção dos teóricos pelos DMs. A seguir apresentam-se duas citações para ilustrar essa questão, mas são inúmeras as opiniões favoráveis aos DMs, tanto na Lexicografia quanto no Ensino de línguas.

Shcherba (1940) afirmava que “...qualquer pedagogo sério aconselha os estudantes a descartarem os dicionários de tradução tão logo possível e mudar para um dicionário de definições da língua estrangeira”¹⁹ (SHCHERBA, 1940, p. 341).

Cinquenta anos depois, a mesma afirmação parece repetir-se nas palavras de McCarthy (1990), evidenciando um lugar-comum nas crenças sobre dicionários:

Como podemos lidar com o problema de encorajar aprendizes a usar seus dicionários para codificar? Preferencialmente, gostaríamos que os estudantes utilizassem um bom dicionário de aprendizes ou um dicionário de combinações para descobrir padrões gramaticais apropriados, colocações e informações sobre o registro de palavras sobre as quais eles têm dúvida ou que só conhecem parcialmente ou usar um *thesaurus* ou algum tipo de léxico para acessar novos campos lexicais quando têm significados a expressar, mas carecem de palavras para expressá-los.²⁰ (MCCARTHY, 1990, p. 143)

Essa idéia pró DMs é marcada principalmente na cultura inglesa e na americana. Aqui se abre um parêntese para discutir essa questão. Talvez o que sustente tal postura não seja a pura lógica lexicográfica, mas sim a lógica de mercado. Os DMs de inglês para aprendizes representam atualmente um dos segmentos mais lucrativos das editoras de dicionários e os

¹⁹ ... any true pedagogue advises students to discard translating dictionaries as soon as possible and switch to the defining dictionary of the foreign language.

²⁰ How can one get round the problem of encouraging learners to use their dictionaries for encoding? Ideally, we would want students to use a good learner's dictionary, or a combinatory dictionary, to find out about appropriate grammatical patterns, collocations, and a registerial information for words they are unsure of or only partially know, or to use a thesaurus or lexicon-type work for access to new lexical fields when they have meanings to express but lack the words to express them.

países de língua inglesa monopolizam sua produção. Um mesmo dicionário é comercializado em vários países do mundo, ou seja, o mercado de DMs para aprendizes é global. No entanto, a produção de DBs que incluam informações relevantes para os aprendizes de cada nacionalidade, além de exigir a participação de inteligência estrangeira (lexicógrafos falantes da L1 que fará par com o inglês), poderá ameaçar a participação dos DMs de aprendizes no mercado, já que poderão substituí-los em muitas situações. DMs para aprendizes permitem um ganho de escala (conseqüentemente, grandes investimentos justificam-se diante da expectativa de fortes vendas). Já os DBs feitos para atender necessidades de uma nacionalidade específica não têm o mesmo ganho de escala e, portanto, ameaçam os lucros das grandes editoras de DMs para aprendizes. Em resumo, bons DBs viriam para dividir o mercado. Isso já aconteceu no caso de DBs para decodificação, quando a Kernerman Dictionaries, de Israel, iniciou o já mencionado processo de bilinguização do *Oxford Student's Dictionary of Current English*, de A. S. Hornby, gerando os dicionários conhecidos como *Password*.

Em 1995, Laufer levantou a necessidade de dicionários que fizessem para a codificação o que o *Password* fez para a decodificação. Um protótipo de sua proposta (que será discutido no capítulo 5) foi desenvolvido e testado dez anos depois por Laufer e Levitzky-Aviad (2006), que declaram:

...no processo de formular um pensamento em L2, o aprendiz freqüentemente será interrompido pela falta de uma palavra (...) o que provavelmente virá à cabeça é o equivalente em L1 da palavra procurada, ao invés de um sinônimo, antônimo ou outra palavra semanticamente relacionada em L2. É assim, já que palavras em nossa língua dominante são acessadas mais facilmente do que palavras em línguas menos familiares para nós.²¹ (LAUFER & LEVITZKY-AVIAD, 2006, p. 4)

Nessa situação, as autoras acreditam que o usuário precisaria de opções de tradução para o item lexical da L1 que veio à mente, com suas respectivas especificações gramaticais,

²¹ ... in the process of formulating the thought in the foreign language, the learner will often get stuck for a word. What would probably come to mind is the L1 word that is needed rather than a synonym, antonym, or any other semantically related word in L2.

semânticas e pragmáticas, além de exemplos de uso. Hoje, segundo as autoras, o usuário só consegue reunir essas informações utilizando simultaneamente um DB e um DM, o que é demorado e pouco prático. O índice remissivo L1-L2 contido em alguns dicionários bilingualizados não é, na opinião das autoras, adequado às necessidades de produção, pois muitas palavras em L1 têm mais de um sentido e mais de um equivalente em L2.

Outro autor que aborda a questão é Rundell (1999a), que admite a possibilidade de elaborar DBs para a codificação. Porém, o próprio Rundell (1999b) volta a defender os DMs para aprendizes como melhor solução, mesmo reconhecendo toda a limitação que as entradas exclusivamente em L2 representam para quem está codificando.

Conclui-se que há duas questões centrais para abordar o problema dos dicionários para codificação. A primeira delas é decidir se há necessidade ou vantagem em separar as funções de codificar e de decodificar para fins de projeto lexicográfico. A outra questão é saber se é possível escolher apenas um tipo de entrada, em L1 ou em L2, isto é, optar por um DB L1-L2 ou por um DM de L2.

Não se encontra consenso entre os teóricos com relação à necessidade ou vantagem de desenvolver dicionários específicos para a codificação. Humblé (1997) defende a idéia de desdobrar o DB segundo a função, ou seja, um DB para apoio à decodificação e um DB para apoio à codificação.

Rundell (1999a), no entanto, chama atenção para o fato de que a distinção entre produção e recepção nem sempre é simples e algumas situações podem exigir um misto de habilidades de recepção e produção. Contudo, o autor acredita que essa distinção “continua útil tanto para usuários de dicionários quanto para dicionaristas”²² (RUNDELL, 1999b, p. 36).

²² remains a useful one both for dictionary users and for dictionary makers.

O mesmo autor, contudo, pergunta: “É possível produzir um dicionário bilíngüe de aprendiz que sirva às duas finalidades – decodificar e codificar? Para mim a resposta seria ‘sim’”²³ (RUNDELL, 1999b, p. 105).

Varantola (2003), por sua vez, parece acreditar que um mesmo dicionário possa atender às duas funções e que a função de decodificação, além de exigir menos do que a função de codificação, pode ser satisfeita com parte das informações fornecidas para a codificação:

Produzir textos de alta qualidade ou traduções em L2 deve ser um dos testes mais rigorosos para a utilidade de um dicionário. Se um dicionário pode satisfazer as necessidades do usuário nesses contextos, ele também pode satisfazê-las em situações que demandam menos, como usar um dicionário para compreensão de texto. Essas situações, obviamente, representam os extremos opostos na escala de uso ativo *versus* passivo do dicionário, ou na escala de necessidades de codificação *versus* decodificação do usuário.²⁴ (VARANTOLA, 2003, p. 228)

Já Binon e Verlinde (2001) e Bogaards e Hannay (2004) defendem soluções que atendam as duas funções, codificação e decodificação, porém não simultaneamente.

Essa questão é fundamental para esta pesquisa e a opinião que se formou após muita reflexão é de que, embora uma base de conhecimentos lexicais possa ser utilizada para fornecer conteúdos lexicográficos de auxílio à codificação e à decodificação, o formato de exibição desses conteúdos deve ser específico para cada finalidade, ou seja, um só formato de exibição não satisfaria necessidades de codificação e de decodificação.

Quanto à segunda questão levantada, sobre a possibilidade de escolher apenas uma forma de entrada, em L1 ou L2, talvez ainda não haja consenso simplesmente porque, na codificação, não seja possível prescindir nem das características do DB nem das características do DM. Na codificação, ora se consulta um item em L1 (quando se busca um

²³ Is it possible to produce a bilingual learner’s dictionary which serves both purposes – decoding as well as encoding? To me the answer would be ‘yes’.

²⁴ Producing high-quality texts or translations in one’s L2 must be one of the most demanding tests for the usability of a dictionary. If a dictionary can satisfy the user’s needs in these contexts it can also satisfy them in less demanding situations, such as using a dictionary for text comprehension. These situations obviously represent the opposite ends on the scale of active vs. passive dictionary use, or on the scale of user’s encoding vs. decoding needs.

equivalente), ora se consulta um item em L2 (quando a dúvida é com relação à adequação da palavra escolhida ou quanto ao seu emprego).

Essas discussões sugerem que os usuários necessitam de muitos tipos de informações, além da possibilidade de entradas tanto em L1 quanto em L2. Essa parece ser também a conclusão de Béjoint (2000), que amplia o espectro dos dicionários para apoio à codificação, como se vê nos três excertos a seguir:

...o período assistiu ao aumento na produção de dicionários para **codificação**, os quais parecem estar progredindo em várias frentes ao mesmo tempo. O tipo de dicionário significado-texto, o dicionário de colocações, os dicionários organizados tematicamente, todos são variações sobre o mesmo tema da produção linguística...²⁵ (BÉJOINT, 2000, p. vii, grifo meu)

O termo *dicionário de uso* é usado como um termo genérico para todos os dicionários que são especificamente projetados para facilitar ou melhorar o **uso produtivo** da língua. Eles incluem *thesauri*, dicionários de sinônimos, dicionários de pronúncia etc.²⁶ (BÉJOINT, 2000, p. 53, grifo meu)

O que a nova lexicografia para **codificação** precisa é de um novo conceito de dicionários, algo na linha do Robert analógico, com uma rede de referências cruzadas que permita múltiplas entradas para os dados.²⁷ (BÉJOINT, 2000, p. 85, grifo meu)

Portanto, uma solução lexicográfica para a codificação talvez tenha que partir da destruição dos estereótipos de DM e de DB, para ter oportunidade de atender as necessidades dos usuários.

Contudo, além de se obter melhor compreensão do problema com a leitura dos trabalhos teóricos que abordam a questão da codificação em L2, reuniram-se também as opiniões que constituem subsídios relevantes para esta tese.

²⁵ ...the period has seen an augmentation in the production of dictionaries for encoding, which seem to be progressing on several fronts at the same time. The Meaning → Text type dictionary, the dictionary of collocations, the thematically organized dictionary, are all variations on the same theme of linguistic production...

²⁶ The term *dictionary of usage* is used here as an encompassing term for all the dictionaries that are specifically designed to facilitate or improve the productive use of the language. They include thesauri, dictionaries of synonyms, dictionaries of spelling, etc.

²⁷ What the new lexicography for encoding needs is a new concept of dictionaries, something along the lines of the Robert analogies, with a network of cross-references that allow multiple entries into the data.

Esses subsídios serão agrupados em três blocos: o primeiro sobre as questões relativas à nomenclatura, o segundo sobre a questão da mídia e o terceiro sobre as informações lexicográficas julgadas necessárias para a codificação.

2.2. A nomenclatura de um dicionário de apoio à codificação

Shcherba (1940) fornece inspiração para a questão da escolha da nomenclatura de dicionários de apoio à codificação em L2 quando diferencia dicionários passivos e ativos (embora apresente isso no contexto dos DMs para nativos). Segundo o autor, os dicionários passivos, feitos para auxiliar a compreensão, deveriam conter todo o léxico passivo, incluindo todos os usos verificados na literatura da língua, mesmo os usos obsoletos. Os dicionários ativos, ao contrário, deveriam conter apenas o léxico em uso no momento de sua elaboração, ou seja, excluir todos os usos obsoletos.

Essa diferenciação parece coincidir com aquela feita por Béjoint “o melhor dicionário para decodificação é aquele que contém o maior número de entradas” e “o melhor dicionário para codificação é aquele que fornece a orientação mais detalhada sobre sintaxe e colocação”²⁸ (BÉJOINT, 1982, *apud* HUMBLÉ, 1997 p. 48).

Acredita-se que a idéia de diferenciar a nomenclatura segundo a finalidade do dicionário seja defensável, isto é, definir como entradas de um dicionário de apoio à codificação apenas o léxico ativo da língua, excluindo, portanto, itens lexicais pouco frequentes e usos obsoletos dos itens lexicais selecionados. Isso evitaria a necessidade de notas avisando que determinados itens lexicais ou usos estejam ultrapassados.

Schmitz (2001), por sua vez, apresenta um argumento que revela a importância de que os dicionários para codificação sejam desenvolvidos por brasileiros e para brasileiros. Segundo o autor, DBs feitos a partir do ponto de vista dos brasileiros poderiam apresentar

²⁸ ‘the best dictionary for *decoding* is the one that contains the largest number of entries’ and ‘the best dictionary for *encoding* is one that provides the most detailed guidance on syntax and collocation’

expressões típicas do português do Brasil, como *jogo de cintura* e *quebrar o galho*. Assim, a nomenclatura do dicionário deveria ser selecionada de forma a garantir a inclusão de unidades léxicas que permitissem a expressão de nossa cultura, como *caju*, *pitanga*, *umbanda*, *jangada*, *favela*, *sem-terra*, etc. Nas palavras do autor: “desta forma o dicionário bilíngüe será desestrangeirizado e tornar-se-á uma obra nacional, verdadeiramente brasileira e não um instrumento de dominação cultural” (SCHMITZ, 2001, p. 169).

Com base nessas duas opiniões, portanto, um dicionário para apoio à codificação deveria ter sua nomenclatura selecionada a partir do ponto de vista da L1 e incluir apenas os itens lexicais pertencentes ao vocabulário ativo, excluindo itens lexicais e usos obsoletos.

2.3. A mídia para um dicionário de apoio à codificação

Ao defender a idéia de desdobrar o dicionário bilíngüe segundo a função: DBs para decodificação e DBs para codificação, Humblé (1996) diz que a mídia eletrônica torna isso factível, mas, segundo ele,

Apesar de seus benefícios óbvios, a lexicografia tem feito pouco uso da tecnologia computacional porque ninguém tem uma idéia precisa do que tem que ser feito. Resolver essa questão implica pesquisar de que tipo de dicionário os aprendizes necessitam para codificar.²⁹ (HUMBLÉ, 1997, p. 45)

Varantola (2003) acredita que a mídia eletrônica seja a melhor opção e explica por que:

A lexicografia impressa evoluiu e aperfeiçoou seus resultados ao longo de séculos e produziu conquistas culturais e lingüísticas notáveis, ao passo que a mídia eletrônica nos dá uma nova liberdade, liberando-nos da camisa-de-força do alfabeto e expandindo nossas estratégias de consulta além dos limites da busca por palavras-entrada.³⁰ (VARANTOLA, 2003, p. 229)

²⁹ In spite of its obvious benefits, lexicography has as yet made little use of computer technology, because no one has a precise idea of what has to be done. Solving this question implies research as to the kind of dictionary learners need for encoding.

³⁰ Print lexicography has evolved and perfected its results over centuries and produced remarkable cultural and linguistic achievements, whereas the electronic médium gives us new freedom by liberating us from the straightjacket of the alphabet, and by expanding our look-up strategies beyond the limits of headword searches.

Béjoint (2000) também demonstra uma expectativa positiva com relação à mídia eletrônica:

Não existe ainda um dicionário “perfeito” para codificação e pode nunca vir a existir. Qualquer dicionário que tente ser exaustivo, como o *Dictionnaire explicatif et combinatoire* de Mel’cuk’s, está fadado a ser tão extenso e complicado que será impossível de usar – exceto, talvez, **se a informação for eventualmente colocada no computador**, o que facilitará o processo de busca (Knowles 1990: 1665). Obviamente um dicionário para codificação terá uma ordem semântica e não alfabética, ou os usuários serão incapazes de encontrar uma palavra ou expressão quando eles têm apenas o conceito.³¹ (BÉJOINT, 2000, p. 85, grifo meu)

Parece haver, portanto, consenso entre os teóricos da Lexicografia sobre as vantagens de se desenvolver novas soluções lexicográficas em mídia eletrônica e essa evidência será retomada no capítulo 6, na definição das propostas deste trabalho.

2.4. Informações lexicográficas necessárias para a codificação em língua estrangeira

Foram reunidas neste item diversas opiniões sobre o tipo de informação que os teóricos da Lexicografia julgam necessário nos dicionários para apoio à codificação. Essas opiniões serão avaliadas e aproveitadas na fase de proposta de solução deste trabalho.

O elemento mais valorizado para a codificação pela maioria dos teóricos da Lexicografia é o exemplo.

Segundo Shcherba (1940), dicionários que contêm muitos exemplos costumam ser considerados bons. O autor esclarece que os exemplos podem aparecer de duas formas: a primeira é dar muitos exemplos e deixar por conta do usuário inferir os significados; a segunda é explicar os significados ou fornecer equivalentes e dar exemplos apenas para ilustrar as definições. Para um dicionário onde se possa verificar se determinado uso existe na língua, é importante incluir uma série de exemplos, porém “não é necessário, por exemplo,

³¹ There is not yet a ‘perfect’ dictionary for encoding, and there may never be. Any dictionary attempting to be exhaustive, like Mel’cuk’s *Dictionnaire explicatif et combinatoire* (DEC), is bound to be so huge and complicated that will be impossible to use – except, perhaps, if the information is eventually fed into a computer that will simplify the search process (Knowles 1990: 1665). Clearly, a dictionary for encoding will have a semantic, not an alphabetical arrangement, or users will be unable to find a word or phrase when they just have the concept.

incluir casos repetitivos ou casos muito similares em contexto”³² (SHCHERBA, 1940, p. 329).

A opinião de Shcherba parece não ser compartilhada por Jacobsen *et al.* (1991):

Os chamados dicionários ativos, que preferimos chamar de dicionários de codificação (...) porque eles vão da língua do usuário para a língua estrangeira, não necessitam de exemplificação dos lemas e sub-lemas de L2, pois se pressupõe que o usuário competente seja capaz de dar ele mesmo esses exemplos.³³ (JACOBSEN *et al.*, 1991, p. 2786)

Mas essa é uma opinião isolada, pois existe grande coincidência de opiniões acerca de que, com base exemplos, seja possível inferir uma série de informações relevantes para a codificação.

Laufer (1992) e Katzaros (2004) apontam a multifuncionalidade dos exemplos em um dicionário. No entanto, salientam que, em muitos casos, a eficácia dessa estratégia depende da capacidade de inferência do usuário e dos exemplos em si.

Por sua vez, Worsch (1999) diz que “a idéia de mostrar a valência dos verbos por meio de frases-modelo ao invés dos complicados códigos de gramática é uma conquista importante orientada a usuário...”³⁴ (p. 106). Essa estratégia, portanto, significa que os exemplos foram escolhidos ou elaborados especificamente para evidenciar alguns aspectos do uso.

Humblé (2003) defende que o exemplo seja o elemento mais importante dentro dos dicionários para apoio à codificação e fala sobre sua própria experiência de definição de um projeto desse tipo:

O fato de o dicionário ter uma finalidade bem definida, a de ajudar para a produção da língua, fez com que cortássemos tudo o que julgávamos não ser necessário para esse processo. Assim chegamos à conclusão que um dicionário que constasse só de exemplos seria o mais adequado. (HUMBLÉ, 2003, p. 27)

³² there is no need, for example, to include repetitive cases or cases that are very similar in context....

³³ So-called active dictionaries, which we prefer to call encoding dictionaries (...) as they go from the language of the user to the foreign language, do not need exemplification of SL lemmas and sub-lemmas, since the competent dictionary user may be supposed to be able to furnish such examples himself.

³⁴ The idea of showing the valency of verbs by giving sentence patterns instead of complicated grammar codes is an important user-oriented achievement....

Concorda-se que seja válida a opção de fornecer exemplos para que o usuário possa extrair deles as informações de que precisa, pois se evita, assim, uma série de outras formas mais complexas para transmitir informações sobre uso, registro, valência, significado etc. A vantagem dessa opção, aliás, é evitar a tendência apontada por Béjoint: “...tudo no dicionário que toma tempo por ser difícil de localizar ou difícil de interpretar – como a **informação para codificação** – tenderá a ser menos usado que as informações mais facilmente acessíveis³⁵” (BÉJOINT, 2000, p. 152, grifo meu).

Além dos exemplos, as notas de uso também são apontadas como uma solução para auxiliar na codificação. Shcherba (1940) já preconizava o tipo de nota que hoje se encontram nos modernos DMs para aprendizes de L2 que utilizam evidências de *corpus* de aprendizes. São exemplos de erros comuns dos aprendizes, o que ele chamou de material lingüístico negativo: “esse material é, na verdade, marcado com uma etiqueta especial: ‘não é assim que as pessoas falam’”³⁶ (SHCHERBA, 1940, p. 331).

Rundell e Granger (2007) dão exemplo do tipo de nota que estão elaborando e que parece coincidir com o que Shcherba imaginava:

information

Remember you cannot make **information** plural. Do not say ‘informations’.

*Could you send me some information about your courses?
We've been able to find out several pieces of information.
For more information please contact our office.
~~For more informations please contact our office.~~*

Outro elemento lexicográfico apontado como essencial para a codificação é a informação sintática. Nas palavras de Chan e Taylor (2001):

A informação sintática é mais importante para fins de codificação. Se aprendizes estrangeiros são capazes de usar uma palavra de forma aceitável é uma questão que depende de seu conhecimento sobre as funções sintáticas dessa palavra, incluindo exigências de

³⁵ Everything in a dictionary that takes time because it is difficult to locate or difficult to interpret – like the information on encoding – will tend to be used less than the more easily accessible information.

³⁶ This material is in effect marked with a special label: “that’s not how people talk”.

valência de um adjetivo (ex: se um adjetivo exige uma sentença iniciada por *to* + infinitivo como em *He is willing to be honest*, ou uma sentença iniciada por *that*, como em *It is nice that you could come.*). (Herbst 1984) Também têm que ter consciência do complemento de um substantivo (...) e da transitividade de um verbo...³⁷ (CHAN & TAYLOR, 2001, p. 170)

No que diz respeito ao comportamento combinatório dos itens lexicais, há uma extensa literatura e a inclusão desse tipo de informação nos dicionários é constantemente mencionada.

Xatara (2001), por exemplo, afirma que:

Sejam idiomáticas ou não, as combinações sintático-semânticas freqüentes e fixas, isto é, cristalizadas pelo uso em uma língua, deveriam constar num DLB [Dicionário de Língua Bilíngüe] e ser especificadas, sempre que possível, com traduções também freqüentes e cristalizadas. (XATARA, 2001, p. 184)

A autora cita vários exemplos, dentre os quais, *bourré comme une vache*, que não significa apenas “muito bêbado”, mas “bêbado como um gambá”. Segundo ela, ao se recuperar o sentido metafórico das expressões, evita-se o empobrecimento do texto na língua-meta. Com respeito especificamente à codificação, Xatara diz que:

Seja como for, se um DLB ainda quisesse de fato **favorecer a construção de enunciados na língua estrangeira**, ele deveria especificar sistemática e não esporadicamente as combinações freqüentes, mesmo no caso de simetria contrastiva, como por exemplo *sommeil lourd* – “sono pesado”. (XATARA, 2001, p. 185, grifo meu)

Berber Sardinha (2004) destaca que os padrões detectados pela Lingüística de *Corpus* deveriam constar dos DBs, pois “a quebra de padrões entre uma língua e outra pode trazer implicações relacionadas à fidedignidade, aceitabilidade e legibilidade do texto traduzido ou vertido” (p. 238).

O autor explica que os padrões revelados a partir da análise de *corpus* podem ser classificados em três tipos:

³⁷ Syntactic information is most important for encoding purposes. Whether foreign learners can use a word acceptably depends on their knowledge of its syntactic functions, including the valency requirements of an adjective (e. g. whether the adjective takes a to-infinitive clause, as in *He is willing to be honest*, or a *that*-clause, as in *It is nice that you could come.*) (Herbst 1984) They must also be aware of the complementation of a noun (...) and transitivity of a verb ...

- Colocação: associação entre itens lexicais ou entre o léxico e campos semânticos.
- Coligação: associação entre itens lexicais e gramaticais.
- Prosódia semântica: associação entre itens lexicais e conotação (negativa, positiva ou neutra). (BERBER SARDINHA, 2004, p. 40)

Ele considera que a “informação relativa à conotação, derivada da exploração de *corpora* eletrônicos, deveria constar de glossários e dicionários, especialmente os de produção, de que dependem os tradutores” (BERBER SARDINHA, 2004, p. 250).

O valor das colocações e coligações para a codificação é reconhecido por Siepmann (2005 e 2006). O autor critica uma prática comum nos dicionários: introduzir as colocações por meio dos exemplos de uso de lexias simples. Para ele, as colocações são lexias complexas e toda lexia complexa deveria ter o tratamento de uma entrada independente, com sua própria definição, seus sinônimos e seus exemplos de uso (afirmações com as quais se concorda plenamente). Após discorrer sobre as vantagens que a organização onomasiológica tem sobre a semasiológica para fins de registrar colocações em um dicionário, Siepmann (2006) apresenta uma proposta de dicionário temático bilíngüe de colocações. As colocações seriam extraídas de *corpora* paralelos temáticos. Como sua proposta visa a atender também as necessidades da codificação oral, Siepmann sugere que os *corpora* contenham material típico da Web, como fóruns, *e-mails* e salas de bate-papo, o que compensaria a falta de amostras temáticas de língua falada.

Outros tipos de informações são apontados por Béjoint ao comentar características do *Advanced Learner's Dictionary* (cuja primeira edição é de 1963): “indicação sistemática de outras informações necessárias para codificação: se substantivos são contáveis ou incontáveis, qual o plural dos substantivos, os comparativos dos adjetivos, as formas verbais irregulares etc.”³⁸ (BÉJOINT, 2000, p. 83).

Em resumo, segundo esses teóricos, um dicionário para codificação deveria ter:

³⁸ ...the systematic indication of other information necessary for encoding: whether nouns are countable or uncountable, what are the plural of nouns, the comparatives of adjectives, irregular verbal forms, etc.

- muitos exemplos, atestando diversos usos da palavra-entrada;
- informações gramaticais sobre flexões da palavra-entrada (plural, feminino, conjugação verbal etc);
- notas alertando sobre erros comuns dos aprendizes;
- seqüências lexicais recorrentes (expressões idiomáticas e colocações);
- informações sobre prosódia semântica;
- estrutura sintática por meio de frases-modelo;
- contraste entre L1 e L2;
- valência dos substantivos, adjetivos e verbos.

Acredita-se que todos esses tipos de informações sejam realmente relevantes para quem codifica em L2. Contudo, reunir e exibir essas informações em um mesmo dicionário é um grande desafio para os lexicógrafos, principalmente se houver o compromisso de conferir um caráter pedagógico à obra.

Na revisão crítica apresentada neste capítulo procurou-se reunir as contribuições dos trabalhos teóricos da Lexicografia sobre dicionários para codificação. No próximo capítulo será feita uma revisão crítica das pesquisas empíricas que fornecem subsídios para o propósito desta tese.

3. NECESSIDADES DO USUÁRIO DE DICIONÁRIOS NA CODIFICAÇÃO EM LÍNGUA ESTRANGEIRA

No capítulo 1 concluiu-se que o uso de dicionários é um dos fatores que podem contribuir para a escrita em L2 e, no capítulo 2, foram abordados trabalhos de teóricos da Lexicografia que apregoam a necessidade de desenvolver dicionários que contemplem as necessidades dos usuários para a finalidade de codificação em L2. Neste capítulo serão discutidas duas formas de descobrir quais são essas necessidades.

No item 3.1 serão abordadas as pesquisas sobre o uso do dicionário em atividades de codificação. Essas pesquisas têm diferentes objetivos, dentre os quais: comparar dicionários; validar dicionários; levantar estratégias de consulta; avaliar a influência de diferentes variáveis no uso do dicionário. Todas elas, porém, têm algo em comum: investigam o processo do uso do dicionário na codificação e avaliam o efeito do uso do dicionário no texto codificado em L2 (redação ou tradução L1-L2). Seus resultados podem fornecer evidências que apontem como deve ser e o que deve conter uma solução lexicográfica para apoio à codificação em L2, a fim de satisfazer as necessidades dos usuários.

Já no item 3.2 será discutida a alternativa que tem rendido maiores contribuições aos lexicógrafos que buscam atender as necessidades de codificação em L2: a análise de erros em *corpus* de aprendizes. Esses erros podem ter sido cometidos com ou sem a consulta a dicionários. Cabe ao lexicógrafo identificar oportunidades de prover informações nos dicionários que possam evitar esses erros. Nem todos os erros, porém, são sanáveis por meio da consulta ao dicionário. Mesmo aqueles que o são, dependem da decisão do usuário de fazer a consulta e da habilidade do usuário em encontrar e interpretar corretamente a informação.

3.1. Pesquisas sobre o uso do dicionário na codificação

As pesquisas sobre o uso do dicionário começaram a ser feitas há pouco mais de vinte anos e estão no cerne do que hoje se conhece como Lexicografia Pedagógica.

Apesar de a Lexicografia Pedagógica ter ganhado nome e notoriedade recentemente, dicionários com preocupações pedagógicas são muito mais antigos, como diz Béjoint:

Os “verdadeiros” dicionários apareceram quando as sociedades começaram a ter relações comerciais ou culturais com comunidades que usavam línguas diferentes e precisavam de traduções. A **natureza pedagógica** daqueles primeiros dicionários bilíngües está quase sempre evidente em seus títulos: *The introductory to Wryte and to Pronounce French* (Barclay, 1521), *Dictionary of the French and English Tongues: Brief Directions for such as desire to learne the French Tongue* (Cotgrave, 1611), etc.³⁹ (BÉJOINT, 2000, p. 93, grifo meu)

Cowie (1999) salienta os avanços no aspecto pedagógico dos dicionários desenvolvidos por Palmer, Hornby e West na primeira metade do século XX, antes mesmo que qualquer pesquisa sobre o uso do dicionário tivesse sido feita ou, pelo menos, formalmente relatada:

Se a demanda por dicionários de aprendizes e a forma como eles evoluíram, particularmente nos últimos vinte anos, foi cada vez mais determinada pelas necessidades dos usuários, isto é, pelo que os usuários exigem – não era esse o caso no início. Nos anos de 1930 e 1940, dificilmente se esperaria que os potenciais usuários de um dicionário de inglês para aprendizes estrangeiros conseguissem expressar suas necessidades, pois não contavam com modelos existentes e nem mesmo com um reconhecimento manifesto sobre o papel dos dicionários no aprendizado de línguas.⁴⁰ (COWIE, 1999, p. 1)

³⁹ The ‘true’ dictionaries appeared when societies began to have commercial or cultural relations with communities that used different languages, and to need translations. The pedagogical nature of those early bilingual dictionaries is often clearly stated in their titles: *The introductory to Wryte ant to Pronounce French* (Barclay, 1521), *Dictionary of the French and English Tongues: Brief Directions for such as desire to learne the French Tongue* (Cotgrave, 1611), etc.

⁴⁰ If demand for learners’ dictionaries, and the way in which they have evolved over the past twenty years in particular, have been increasingly user-driven-determined, that is, by what users require, or are thought to require – this was not the case at the beginning. In the 1930s and 1940s, potential users of the EFL dictionary could hardly be expected to articulate their needs in the absence of existing models, or of any informed understanding of the role of dictionaries in language learning.

Segundo Cowie, os primeiros DMs para estrangeiros foram iniciativas bem sucedidas graças ao conhecimento que seus empreendedores adquiriram ensinando inglês no exterior (Palmer e Hornby no Japão e West na Índia).

Essa experiência, de confecção de dicionários por professores que conheciam as dificuldades de seus alunos, levou os lexicógrafos a perceberem que uma boa solução pressupõe o conhecimento do problema, ou seja, é preciso conhecer as dificuldades dos usuários para conceber soluções lexicográficas capazes de superá-las. Portanto, quando a *Dictionary Conference* de Bloomington, de 1960, estabeleceu que “dicionários deveriam ser planejados tendo em mente um determinado conjunto de usuários e suas necessidades específicas”⁴¹ (HOUSEHOLDER & SAPORTA, 1962, *apud* BOGAARDS, 1999, p. 32), não estava introduzindo uma inovação de pensamento, ou seja, “dicionários sempre foram moldados pelas necessidades de seus usuários, ou pelo menos pelo que os lexicógrafos conheciam ou imaginavam sobre aquelas necessidades...”⁴² (BÉJOINT, 2000, p. 169).

A resolução de Bloomington expressou o reconhecimento dos lexicógrafos sobre a necessidade de conhecer melhor os usuários dos dicionários e suas dificuldades, ao invés de ficar apenas “imaginando”. Talvez não coincidentemente, nas décadas que se seguiram, essa resolução foi repetida muitas vezes, acompanhando o surgimento e a crescente multiplicação de pesquisas sobre o uso dos dicionários (v. BOGAARDS, 1999).

Enquanto a Lexicografia tradicionalmente envidou esforços para fazer um inventário cada vez mais completo do léxico das línguas, a Lexicografia Pedagógica tirou o foco da quantidade de informações e passou a focar as necessidades dos usuários. A qualidade de um dicionário começou, então, a ser avaliada não mais com base na quantidade de informações, mas principalmente com base na sua adequação às necessidades dos usuários.

⁴¹ Dictionaries should be designed with a special set of users in mind and for their specific needs.”

⁴² Dictionaries have always been shaped by the needs of their users, or at least by what lexicographers knew, or imagined, of those needs...

Humblé (1997) fez uma análise mais completa dessa mudança de paradigma e, retrocedendo na história, concluiu que, antes da formação dos estados nacionais europeus, os dicionários tinham propósitos práticos de auxiliar a comunicação. Foi a partir do momento em que unificação lingüística passou a ser coadjuvante no processo de unificação de nações, que os dicionários começaram a ser tratados como repositórios de tesouros e a ter como missão dar estabilidade às línguas, evitando, assim, que fossem corrompidas. Só no século XX é que a funcionalidade dos dicionários voltou a ser valorizada, ou seja, as informações neles contidas deveriam ter um propósito prático de auxílio à comunicação.

Foi para atender esse propósito que as pesquisas sobre o uso de dicionários começaram a ser feitas há pouco mais de vinte anos. No cerne dessas pesquisas estão contemplados tanto o aspecto do público-alvo (quem usa) quanto o aspecto da função (para que se usa), dois aspectos determinantes dos projetos lexicográficos.

As pesquisas empíricas sobre o uso do dicionário foram incluídas na extensa bibliografia crítica produzida por Dolezal e McCreary (1999) a respeito dos trabalhos em Lexicografia Pedagógica. Uma nova e completa bibliografia crítica foi produzida por Welker (2006), focando especificamente os 220 estudos empíricos sobre o uso do dicionário publicados até então. Esses trabalhos compreensivos fornecem um panorama para os pesquisadores compararem resultados e métodos desse tipo de pesquisa, consolidando um conhecimento básico para novas iniciativas.

Também relevante é o trabalho teórico de Wiegand (1998, *apud* HARTMANN, 1999), que discute as diversas possibilidades metodológicas de se realizar pesquisas empíricas sobre o uso do dicionário. Infelizmente, até o momento, o trabalho foi publicado apenas em alemão.

Como o problema em foco nesta pesquisa é a escolha lexical na codificação escrita em L2, serão reunidas e comentadas aqui as pesquisas que investigaram o uso de dicionários durante atividades que resultaram num produto escrito em L2. Somente por isso e apenas

dentro do escopo deste trabalho, não parece ser relevante tratar distintamente as pesquisas que investigaram o uso do dicionário na redação em L2 e na tradução L1-L2, pois ambas as situações exigem um uso semelhante do dicionário, ou seja, procuram-se formas adequadas para expressar determinado conteúdo em L2.

Tanto o processo de uso do dicionário quanto o produto do uso foram analisados na maioria das pesquisas aqui comentadas. Para investigar o processo de uso do dicionário, que envolve uma série de operações cognitivas, é comum a adoção de técnicas de pesquisa introspectiva, que tentam explicitar comportamentos não observáveis externamente e que revelam aspectos dos quais os próprios usuários talvez não tenham consciência.

Para verificar o efeito do uso do dicionário no texto produzido, observam-se os erros cometidos apesar da consulta ao dicionário, mas não todos os erros. Assim, é preciso que durante o processo de uso sejam registrados os itens lexicais consultados, para que depois sejam identificados no texto produzido. Assim, o efeito do uso do dicionário não é avaliado sobre o texto todo, mas em trechos específicos onde os itens lexicais consultados são empregados.

A seguir serão discutidos trabalhos dessa natureza, procurando-se ressaltar as evidências que representam subsídios para o propósito desta tese.

3.1.1. Ard (1982)

Ard investigou, por meio do protocolo verbal, a diferença entre o uso e o não-uso do DB em redações escritas em inglês por estrangeiros nos Estados Unidos. Ele relatou dois experimentos: no primeiro solicitou que os aprendizes fizessem uma redação e sublinhassem as palavras que procuraram no dicionário. No segundo, solicitou que os aprendizes fizessem uma redação e expressassem em voz alta tudo o que estavam pensando durante a atividade. O

número de sujeitos dos dois experimentos não fica claro no artigo e o autor escolheu dois sujeitos de cada experimento para ilustrar suas constatações.

No primeiro experimento, Ard percebeu que, embora os sujeitos japoneses tivessem mais habilidade escrita que os sujeitos espanhóis, eles cometeram mais erros, o que o autor atribuiu ao “grau de intertradutibilidade” do par de línguas em questão. No caso do inglês-espanhol, a intertradutibilidade é maior do que no caso do inglês-japonês. Esse experimento levou Ard a observar que os tipos de erros estão associados a diferentes grupos de L1.

Essa constatação de Ard sugere, portanto, que há necessidade de o dicionário contemplar informações que permitam ao usuário evitar os tipos de erros mais frequentemente cometidos pelo grupo da L1 do qual faz parte. Estudos de tipos de erros que identifiquem os sujeitos por L1 são, portanto, um importante subsídio para a Lexicografia.

Outra importante hipótese levantada por Ard é a de que o DB não ajuda tanto a eliminar erros, mas sim a aumentar as habilidades de expressão em L2.

As constatações de Ard foram corroboradas pelas pesquisas sobre a escrita em L2 já mencionadas no capítulo 1: a) a distância entre um par de línguas é uma variável que afeta a escrita em L2; b) o dicionário representa um auxílio para que o aprendiz tente empregar um léxico mais rico do que já possui, embora nem sempre o resultado seja isento de erros.

3.1.2. Krings (1986 *apud* WELKER, 2006)

Krings⁴³ utilizou protocolo verbal para descobrir o que se passava na mente dos tradutores durante atividades de tradução L1-L2 e L2-L1. Os sujeitos da pesquisa eram estudantes de tradução e não tradutores experientes. No que diz respeito à tradução L1-L2, o autor observou que os dicionários foram utilizados em 64% das dificuldades detectadas pelo protocolo verbal. Esse dado mostra que nem todos os problemas enfrentados durante a escrita

⁴³ O artigo só foi publicado em alemão.

em L2 deflagram uma consulta ao dicionário, ou seja, consultar o dicionário é uma das estratégias dos usuários para tentar solucionar suas dificuldades.

Os sujeitos da pesquisa de Krings fizeram, ao todo, 260 consultas: 224 no DB e 36 no DM e uma mesma dificuldade muitas vezes gerou mais de uma consulta ao dicionário. O principal motivo de consulta aos DBs foi a busca de equivalentes, na seção L1-L2, seguido da busca de confirmação dos “candidatos” a equivalentes na seção L2-L1. O DM foi mais utilizado para esclarecer dúvidas de significado ou para observar restrições de uso de um possível equivalente.

3.1.3. Nesi (1987)

A autora adotou uma abordagem diferente das demais pesquisas: analisou se o uso do dicionário poderia ter evitado os erros constatados em redações de aprendizes. Partindo dos erros cometidos por estrangeiros na redação em inglês, ela consultou três DMs para aprendizes de inglês e avaliou se as informações disponíveis poderiam ter evitado os erros.

Sua conclusão é de que os dicionários utilizados na pesquisa priorizam o uso receptivo da linguagem, ou seja, a compreensão do significado. Quando o propósito é codificar em L2, as informações exibidas pelos três dicionários podem confundir os usuários. Esse estudo mostra como a análise de erros comuns entre aprendizes de uma mesma L1 e cultura de origem pode fornecer subsídios para a decisão sobre o que incluir em um dicionário para dar suporte à codificação em L2.

Embora não mencionasse naquela época o uso de *corpus* de aprendizes, Nesi renunciou o uso dessa fonte de dados para melhoria dos dicionários para aprendizes. A autora apresentou uma série de sugestões para aperfeiçoamentos das obras lexicográficas, com vistas a suprir as necessidades dos usuários na função de codificação em L2, todas elas

muito relevantes para o presente trabalho. Essas sugestões são apresentadas a seguir de forma resumida (adaptação minha):

- Listar os significados das palavras polissêmicas por ordem de frequência e alertar para significados pouco comuns cujo emprego é restrito a determinadas situações;
- Criar referências cruzadas entre dois itens lexicais frequentemente confundidos pelos usuários, esclarecendo o uso de um e de outro (isso depende da disponibilidade de resultados de pesquisas de erros comuns a determinados grupos de aprendizes);
- Criar e utilizar etiquetas que aprimorem informações sobre marcas de uso (registro) e, sempre que possível, criar referências cruzadas para itens lexicais sinônimos com marcas de uso diferentes;
- Inserir notas de uso nos itens lexicais cujo emprego seja frequentemente objeto de erro (pesquisas sobre erros comuns também são essenciais para isso);
- Especificar nas entradas as restrições de seleção e de colocação dos itens lexicais;
- Evitar definições baseadas em um só item lexical, pois isso estimula o aprendiz a acreditar em falsas equivalências entre itens lexicais;
- Marcar claramente os itens lexicais que possuem conotação negativa;
- Exibir, em entradas separadas, itens lexicais derivados de outros itens lexicais (por exemplo, adjetivos formados pelo verbo no particípio) e ressaltar sempre que seu significado divergir do significado do item lexical de origem;
- Fornecer informações culturais que possam melhorar a compreensão de um item lexical ou remeter o usuário a trabalhos que as contenham.

3.1.4. Nesi (1994)

Nessa pesquisa a autora propôs a dois grupos de estudantes universitários, um em Portugal (51 sujeitos) e outro na Malásia (44 sujeitos), a produção de frases em inglês utilizando, em cada uma, duas palavras dadas. O experimento teve por objetivo avaliar se existe influência da língua e da cultura de origem no uso produtivo do dicionário de inglês para aprendizes. Os dados obtidos confirmaram a influência, pois os estudantes portugueses, cuja língua apresenta maior proximidade com o inglês do que a língua malaia, fizeram menos consultas, consumiram menos tempo em cada consulta e obtiveram melhores notas nas frases do que os estudantes malaios. Essa conclusão fica mais reforçada pelo fato de um teste preliminar ter demonstrado que o vocabulário de inglês dos estudantes malaios era maior que o vocabulário de inglês dos estudantes portugueses.

Normalmente, a expectativa é de que, quanto maior o vocabulário do aprendiz, menor o número de consultas ao dicionário e maior a facilidade de produzir frases aceitáveis. A quebra dessa expectativa não somente mostra a influência de fatores culturais na produção em L2, mas também sugere que o conhecimento de um grande número de itens lexicais não garante a capacidade de empregar esses mesmos itens adequadamente.

Nesi observou, porém, que a influência da L1 nem sempre é positiva no uso do dicionário, pois quando estudantes reconhecem um item lexical da L1 como cognato de um item lexical da L2, tendem a ignorar as diferenças apresentadas no texto do respectivo verbete, podendo incorrer em erros.

Os resultados de Nesi evidenciam que falantes de diferentes LMs apresentam diferenças com relação ao uso de dicionários e corroboram o que Ard (1982) chamou de “grau de intertradutibilidade” entre duas línguas.

No entanto, o tipo de atividade escolhido para essa pesquisa de Nesi não parece ser adequado para investigar o uso do dicionário na codificação em L2. Atividades do mesmo

tipo foram relatadas em Laufer (1992), Laufer (1993), Nesi (1996) e Laufer e Hadar (1997), ou seja, o efeito do uso do dicionário na “produção” em L2 foi medido por meio da elaboração de frases que empregassem determinadas palavras em L2. Em todas essas experiências, as palavras fornecidas foram testadas em grupos de controle para garantir que não fossem conhecidas dos sujeitos da pesquisa. Os dicionários, nesse caso, foram utilizados para entender o significado dessas palavras. Esse tipo de atividade mede, portanto, por meio das frases produzidas, que tipo de dicionário ou que parte da microestrutura dos dicionários contribui mais para a compreensão de palavras desconhecidas (decodificação). Conclui-se que, nesse tipo de pesquisa, verificou-se o uso receptivo do dicionário e não o produtivo, como era a proposta.

3.1.5. Bogaards (1996)

O autor avaliou três DMs de aprendizes de língua inglesa, lançados na época da pesquisa e, para a atividade de codificação, comparou-os com o *Longman Language Activator* (LLA). A maior dificuldade de uso dos DMs na codificação é iniciar a consulta quando não se tem uma palavra-chave. Os dicionários analisados deveriam, assim, organizar os itens lexicais de forma a permitir que o acesso aos itens mais complexos fosse possível a partir de itens mais frequentes, semanticamente relacionados.

Para fazer sua avaliação, Bogaards simulou o uso desses dicionários em uma atividade de redação em L2, colocando-se na posição de um aprendiz com um vocabulário de 2000 itens. Assim, ao invés de recorrer a seu próprio léxico mental para escrever em L2, o autor empregou apenas os itens lexicais constantes da lista do vocabulário esperado para um aprendiz de nível intermediário. Quando precisava expressar uma idéia que não podia ser adequadamente expressa por um dos 2000 itens lexicais da lista, Bogaards tentava utilizar, como chave de consulta aos dicionários, os itens da lista que tivessem alguma relação

semântica com o item lexical procurado (hiperônimos, merônimos, sinônimos, antônimos etc.).

O autor observou que, no caso dos três DMs de aprendizes, o sucesso das consultas foi em torno de 50%, pois a partir do vocabulário básico nem sempre foi possível ter acesso aos itens lexicais procurados. O LLA, contudo, levou a um resultado inferior ao obtido com cada um dos DMs de aprendizes, o que chamou atenção do autor, já que os DMs analisados, ao contrário do LLA, não são específicos para a codificação. Esse fato será explicado no item 5.1, dedicado à análise das características do LLA.

O estudo de Bogaards salientou as deficiências de organização e de conteúdo dos dicionários utilizados, apontando caminhos para aperfeiçoá-los. Segundo o autor, os DMs de aprendizes, notadamente de língua inglesa, têm contemplado, a cada nova edição, mais aspectos relevantes para as necessidades de codificação.

3.1.6. Atkins e Varantola (1997)

As autoras adaptaram a técnica utilizada por Ard (1982) e Krings (1986), transformando o protocolo verbal em um protocolo escrito semi-estruturado. Os 103 sujeitos de pesquisa foram instruídos a formar pares e a escolherem uma atividade de tradução L1-L2 ou L2-L1. O fato de a tradução L1-L2 ter sido escolhida por 86 dos 103 sujeitos dessa pesquisa justifica sua inclusão neste capítulo.

Na pesquisa, um dos integrantes do par fazia tradução (o tradutor), enquanto o outro anotava em um formulário tudo o que o tradutor consultava no dicionário. O tradutor expressava oralmente tudo o que estava fazendo para que seu par não tivesse que inferir nada, apenas registrar. A planilha fornecida para o registro já previa a possibilidade de uma dúvida gerar mais de uma consulta e esse é um diferencial que não se encontra em outras pesquisas do gênero. Esse procedimento dispensou a transcrição dos protocolos verbais, fase mais

trabalhosa do emprego dessa técnica, possibilitando sua aplicação a um grande número de usuários. Além disso, o resultado dos registros já apresentava uma estrutura, o que deve ter facilitado a tabulação dos dados. Dessa forma, as autoras investigaram o processo de uso do dicionário, mas não o efeito do uso do dicionário nos textos produzidos. Contudo, por meio de um breve questionário ao final de cada busca, levantaram se os usuários estavam satisfeitos ou insatisfeitos com os dicionários consultados.

As pesquisadoras colocaram vários DBs e DMs à disposição dos participantes (todos impressos) e não determinaram quais deveriam ser consultados. Os resultados mostram que a taxa de satisfação dos usuários com os dicionários foi maior na tradução L2-L1 (72%) do que na tradução L1-L2 (58%). O DB foi utilizado em 70% das consultas e o DM em 30%. Além disso, 64% das consultas aos DBs foram consideradas bem sucedidas contra apenas 48% das consultas aos DMs.

O principal motivo de consulta na tradução L1-L2 foi a busca de equivalentes, seguido da tentativa de confirmar a adequação dos equivalentes escolhidos. Quanto às colocações, apenas os tradutores com maior proficiência na L2 procuraram informações dessa natureza no dicionário. Também foram os estudantes mais experientes que consultaram mais os dicionários, principalmente para confirmar a adequação de um equivalente escolhido para a tradução.

Das 1000 consultas realizadas, 40% não tiveram resultado satisfatório na opinião dos usuários, fato que as autoras atribuem não somente à falta de informações nos dicionários, mas também à falta de habilidade de consulta dos usuários. Segundo a pesquisa, foram os usuários mais experientes que demonstraram menor satisfação com o resultado das consultas, o que é interpretado, por Atkins e Varantola, também como sinal de que esses usuários possuem expectativas muito altas e pouco realistas em relação ao conteúdo dos dicionários.

3.1.7. Harvey e Yuill (1997)

Harvey e Yuill realizaram uma pesquisa com 211 sujeitos de várias nacionalidades, freqüentadores de cursos de inglês para estrangeiros (presumivelmente na Inglaterra, em função do vínculo institucional dos autores com escolas inglesas). Não houve restrição com relação à área de estudos ou de atuação profissional dos sujeitos, mas a pesquisa fez a identificação dessas variáveis e observou-se grande diversidade desses atributos. Foram fornecidos quatro temas para dissertação. Não foram fixados limites nem de tempo nem de extensão do texto (embora os pesquisadores tenham sugerido de 45 a 60 minutos).

Os autores utilizaram uma técnica introspectiva para obter informações sobre o uso do dicionário durante uma atividade de escrita em inglês. Forneceram três tipos de formulários pré-formatados que os sujeitos deveriam preencher antes e depois de cada consulta ao dicionário. O tipo A apresentava oito alternativas para que o sujeito assinalasse o motivo ou motivos que o levaram a consultar o dicionário e deveria ser preenchido antes da consulta. Depois de cada consulta os sujeitos eram solicitados a preencher os formulários B e C. O formulário B apresentava um fluxograma em que o sujeito assinalava se julgava a consulta bem-sucedida ou não. Seguindo o caminho de cada resposta (sim ou não), alternativas eram apresentadas para que o sujeito identificasse as razões das consultas mal-sucedidas ou as razões das consultas que, embora bem-sucedidas, apresentaram dificuldade, bem como indicasse qual a parte do verbete se mostrou mais útil. O formulário C indagava se o sujeito encontrara informações adicionais durante cada consulta e apresentava alternativas para que identificasse a natureza dessas informações e em que parte do verbete as encontrara.

Essa forma de pesquisa é da mesma natureza do protocolo verbal, ou seja, busca descobrir o raciocínio dos sujeitos de pesquisa, porém tem um intervalo um pouco maior entre o ato e o relato. Enquanto o protocolo verbal é livre e dá oportunidade de expressão de fatos não previstos pelo pesquisador, a técnica utilizada pelos autores é estruturada e limita a

investigação dos comportamentos apenas dentro das alternativas previstas nos formulários. A vantagem do segundo sobre o primeiro é o fato de não exigir transcrição (protocolo verbal) e ser de fácil tabulação. Esses fatores são importantes quando se pretende pesquisar uma amostra com grande número de usuários, caso em que a utilização do protocolo verbal seria praticamente inviável.

Harvey e Yuill permitiram o uso apenas de um dicionário, o *Collins Cobuild English Language Dictionary*. Portanto, ao se analisar as conclusões do estudo, é preciso ter em mente que os sujeitos não tinham a alternativa de consultar DBs nem outros tipos de DMs. Outro fator importante é que os sujeitos da pesquisa estavam inseridos em um contexto de língua inglesa, o que permite distingui-los como aprendizes de *segunda língua* e não de *língua estrangeira*.

A principal razão de consultas ao dicionário, por exemplo, foi a ortografia (24,4%). Os sujeitos procuravam a forma de escrever palavras que sabiam falar ou, pelo menos, que já haviam ouvido. Daí decorre, inclusive, a sugestão dos autores para os lexicógrafos: dar como chave de consulta a pronúncia de uma palavra: “guiando o usuário, via sons da fala, à ortografia correta de uma palavra”⁴⁴ (HARVEY & YUILL, 1997, p. 26). Não especificaram, porém, como isso seria feito, se eletronicamente por *softwares* de reconhecimento de voz ou se pela transcrição fonética com o uso do Alfabeto Fonético Internacional. Em ambientes em que o inglês é *língua estrangeira*, porém, dificilmente essa razão encabeçaria a lista das consultas mais freqüentes ao dicionário, pois nesses ambientes as oportunidades de contato com a L2 escrita superam as oportunidades de contato com a L2 falada. Seguindo esse raciocínio, seria menos provável que o conhecimento oral de um item lexical precedesse o conhecimento de sua forma escrita.

O segundo motivo de consulta ao dicionário apontado pela pesquisa é a busca do significado (18,2%), que se subdivide em três categorias: verificar se não se trata de um falso

⁴⁴ (...) guiding the reader via speech sounds to the correct spelling of a word.

cognato; verificar a polissemia do item lexical (quando o usuário conhece um dos sentidos e deseja saber se o item cobre os mesmos campos semânticos de seu equivalente na L1) e verificar a distinção de dois itens lexicais cujo sentido seja muito próximo e que os autores chamaram de “palavras confundíveis”.

Em terceiro lugar encontra-se a consulta para verificar a existência de um item lexical “intuído” pelo usuário. Nessa pesquisa, se o usuário não conhecesse uma palavra em inglês que exprimisse pelo menos aproximadamente a idéia desejada, ficaria sem a opção de iniciar a busca a partir da L1, já que o único dicionário disponível na pesquisa era monolíngüe.

A conclusão dos autores é de que “ao escrever, os aprendizes necessitam de dicionários que lhes forneçam, acima de tudo, informações sobre ortografia e significado. Todas as demais categorias de informações, incluindo informações sintáticas, colocacionais e sobre sinônimos são demandadas em um menor grau⁴⁵” (HARVEY & YUILL, 1997, p. 271). O fato de essa conclusão discordar das conclusões das demais pesquisas sobre o uso do dicionário na escrita em L2 permite concluir algo relevante para o objetivo desta tese: as características desejáveis em um dicionário concebido para auxiliar a codificação em ambiente de *segunda língua* podem (ou devem) ser diferentes das características desejáveis em um dicionário concebido para auxiliar a codificação em ambiente de *língua estrangeira*. No caso, a necessidade de uma entrada oral para obter a forma escrita parece ser uma necessidade típica de ambiente de *segunda língua*, pois a imersão nesse ambiente possibilita a aquisição oral antes da escrita. Esse fato sugere ainda que a distinção de registro oral/escrito seria muito relevante em ambiente de *segunda língua*, pois muito provavelmente nessa situação os aprendizes tendem a apresentar excesso de oralidade na escrita.

⁴⁵ Learners when writing require dictionaries to provide them above all with information on spelling and meaning. All other categories of information, including synonyms, syntactic, and collocational, are required to a lesser degree.

Em razão dessas evidências, é importante enfatizar que se almeja uma proposta de solução lexicográfica para apoiar brasileiros que escrevem em L2 em ambiente de L1. Essa proposta não seria a mais adequada, por exemplo, para brasileiros que residam no exterior.

3.1.8. Christianson (1997)

Christianson investigou o uso do dicionário na produção escrita por duas turmas de alunos japoneses que aprendiam inglês como L2 em um curso de ciências da computação no Japão, num total de 51 alunos. Primeiramente, ele solicitou aos alunos que, durante um semestre, anotassem todas as palavras que foram procuradas no dicionário e depois utilizadas em alguma atividade de escrita em L2 em sala de aula. De posse da lista de palavras procuradas e dos textos produzidos, o autor analisou quais foram empregadas adequadamente. As palavras empregadas inadequadamente foram agrupadas por tipo de erro (foram estipulados sete tipos de erros). Depois foram entrevistados os alunos que tiveram os melhores e os piores desempenhos, para tentar descobrir o tipo de raciocínio que tiveram durante as consultas das palavras que foram erroneamente empregadas.

Christianson relatou detalhadamente as condições do experimento. De um *corpus* de 41.024 palavras, apenas 1,6% foram consultadas nos dicionários e, das consultadas, 42% foram usadas incorretamente. A escolha dos dicionários foi livre, mas o autor observou que os alunos com pior desempenho utilizaram o mesmo dicionário que o aluno que teve o melhor desempenho. Isso o levou a sugerir que, “não importa qual dicionário é usado, seu valor depende mais do usuário do que do próprio dicionário”⁴⁶ (CHRISTIANSON, 1997, p. 34).

Ao focar os erros cometidos apesar do uso do dicionário, Christianson se perguntou: “Por que usar um dicionário se não para garantir que o que você está escrevendo está correto?”⁴⁷ (CHRISTIANSON, 1997, p. 35). Ele, aliás, toma como pressuposto a afirmação

⁴⁶ no matter which dictionary is used, its value depends on the user more than the dictionary itself...

⁴⁷ Why use a dictionary at all if not to ensure that what you are writing is correct?

de Hartmann de que “o principal objetivo de um dicionário é evitar ou pelo menos reduzir conflitos de comunicação que possam surgir do déficit lexical⁴⁸” (HARTMANN, 1987, *apud* CHRISTIANSON, 1997, p. 23). E essa é uma das questões discutidas por Christianson: o uso do dicionário provoca erros ou os evita? O autor conclui que, talvez sem o uso do dicionário, o número de erros seria maior.

A mesma questão foi discutida por Ard (1982), conforme já exposto no item 3.1.1. Para ele, o uso do dicionário permite ao sujeito expressar idéias que não poderiam ser expressas apenas com o emprego de seu léxico mental. Assim, Ard considera o dicionário importante para estimular o desenvolvimento da expressão escrita, a despeito do risco de erros nesse processo. Segundo esse autor, se o sujeito não se arrisca a usar as palavras que encontrou no dicionário, também não progride em sua competência escrita. Por isso, a visão de Ard acerca do objetivo do dicionário é diferente da visão de Hartmann e de Christianson.

Christianson complementou o levantamento de dados entrevistando os quatro alunos com melhor desempenho nas consultas a dicionários e os quatro com pior desempenho. Diz o autor: “para cada palavra [errada] eu comecei uma série de perguntas a fim de determinar qual tinha sido o processo de raciocínio do aluno e sua estratégia de consulta (Ard’s protocol)⁴⁹” (CHRISTIANSON, 1997, p. 28). É curioso observar que Christianson, apesar de ter feito as entrevistas praticamente um ano após os alunos terem consultado as palavras nos dicionários, acreditou estar aplicando o mesmo método utilizado por Ard (1982). Ao que parece, para ele não há diferença entre explicitar o raciocínio enquanto ele está ocorrendo ou após um ano. Teoricamente, no entanto, quanto maior o tempo decorrido entre a ação e a aplicação da técnica introspectiva, menor a confiabilidade dos resultados.

⁴⁸ The main purpose of a dictionary is to prevent or at least reduce communication conflicts which may arise from lexical deficit.

⁴⁹ For each word, I began a series of questions aimed at determining what the student’s thought process and look-up strategy (Ard’s ‘protocol’) had been.

Dentre os resultados de Christianson, três são particularmente relevantes para esta tese:

1. Não foi encontrada correlação entre o nível de competência em L2 e a habilidade de consulta a dicionários;
2. Os alunos mal-sucedidos nas consultas não têm o hábito de ler os exemplos e as informações gramaticais;
3. Alguns dos alunos bem-sucedidos nas consultas utilizaram DBs L2-L1 ou apenas DMs, o que levou o autor a questionar a tipologia que separa dicionários para produção (L1-L2, ativos ou para codificação) dos dicionários para recepção (L2-L1, passivos ou para decodificação).

Essa última questão coloca em dúvida um dos pressupostos do presente trabalho: o de que dicionários L1-L2 são as ferramentas lexicográficas mais adequadas para apoio à codificação em L2. Embora não se pretenda ignorar essa questão, acredita-se que bons estrategistas consigam chegar a bons resultados mesmo não contando com o melhor aparato, ou seja, enxergam oportunidades de extrair do dicionário as informações necessárias por meio de várias consultas a itens de alguma forma relacionados ao item procurado. Porém, o objetivo deste trabalho não é verificar como os usuários de dicionários poderiam ser treinados como bons estrategistas, mas sim apontar como dicionários mais adequados poderiam tornar o sucesso nas consultas algo possível para a maioria dos usuários e não apenas para os mais habilidosos. Resta descobrir, portanto, se o uso dos DBs L2-L1 na codificação tem sustentação lógica ou se é um subterfúgio para contornar as deficiências dos DBs L1-L2.

3.1.9. Winkler (1998)

O uso do dicionário na codificação em L2 também foi investigado por Winkler (1998) por meio do protocolo verbal. A autora estava interessada em comparar o uso das versões

impressa e em CD-ROM do *Oxford Advanced Learner's Dictionary* durante uma atividade de redação. Os sujeitos de pesquisa foram estudantes de inglês, na Inglaterra, de diversas nacionalidades. Seus resultados são particularmente relevantes para a questão do ensino do uso do dicionário, pois as habilidades requeridas para consultas bem sucedidas em dicionários eletrônicos e impressos são em grande parte diferentes. Se os usuários não desenvolverem habilidades de consulta, poderão não tirar proveito das inovações que estão se tornando disponíveis nos dicionários eletrônicos.

A autora observou que os alunos tiveram dificuldade em lidar com abreviações, códigos e símbolos. Sendo assim, eles evitavam todos os códigos, principalmente os que forneciam informações gramaticais, e tentavam inferir essas informações por meio dos exemplos. Por isso, a possibilidade de decifrar essas convenções lexicográficas com um duplo clique sobre o símbolo ou abreviação, na versão CD-ROM do dicionário, agradou os usuários.

Como na pesquisa de Harvey e Yuill (1997), a pesquisa de Winkler também revelou ser interessante possibilitar entradas pela pronúncia, o que, como já foi comentado, é relevante em ambiente de *segunda língua*, quando o usuário sabe como falar, mas não como escrever determinadas palavras. Por ter trabalhado unicamente com um DM, Winkler chegou à conclusão de que, para apoio à codificação em L2:

...uma forma híbrida, no formato de um dicionário bilingualizado, também pode ser desejável, já que esse dicionário auxiliaria os estudantes a encontrar palavras que eles não conhecem e, portanto, não podem procurar em um dicionário monolíngue para aprendizes de inglês.⁵⁰ (WINKLER, 1998, p. 10)

3.1.10. Komuro e Yamada (2000)

Komuro e Yamada dizem que “o tipo de dicionário associado à recepção parece ter tido precedência, em demanda, em uso, e no valor e atenção dados por usuários, professores,

⁵⁰ ...a hybrid form, in the shape of a bilingualised dictionary may be desirable as well since such a dictionary would assist students in finding words they do not know and therefore cannot look up in a monolingual English learners' dictionary.

pesquisadores e editores, sobre os dicionários associados à produção⁵¹” (KOMURO & YAMADA, 2000, p. 1). No sentido de colaborar para a diminuição dessa defasagem, os autores pesquisaram o uso de dicionários na codificação por meio de um questionário respondido por 167 japoneses.

Os resultados dessa pesquisa revelaram que os dicionários L1-L2 são usados para encontrar equivalentes e os dicionários L2-L1 para confirmar a adequação de um “candidato” a equivalente, o que surpreende os autores, pois DBs L2-L1 são comumente associados à recepção e não à produção. Os autores relatam também as estratégias utilizadas pelos sujeitos da pesquisa:

“Primeiro eles tentam expressar uma (dada) idéia com o vocabulário que eles possuem e quando eles não têm certeza sobre o uso da palavra que têm em mente, procuram-na no DB inglês-japonês para confirmar. Quando não conseguem pensar em nenhuma palavra para começar, eles pegam um DB japonês-inglês para encontrar um equivalente de tradução para sua idéia. (...) Quando a palavra em japonês que o usuário cogitou no início para expressar sua idéia não está listada como palavra-entrada, ele/ela tem que tentar uma outra palavra sinônima, já que a nomenclatura em um dicionário japonês-inglês é essencialmente limitada.⁵² (KOMURO & YAMADA, 2000, p. 6)

Os resultados de Komuro e Yamada corroboram os resultados de Christianson sobre o uso de DBs L2-L1 na codificação. Esse fato, portanto, merece atenção nesta pesquisa, pois talvez sugira que não existe um tipo de dicionário ideal para codificação (DB L1-L2 ou DB L2-L1 ou DM). É provável que os usuários precisem de um material lexicográfico mais completo que aquele fornecido isoladamente por diferentes tipos de dicionários e, principalmente, necessitem da possibilidade de acessar esse material via L1 ou via L2, dependendo da estratégia que julgarem mais conveniente no momento da consulta.

⁵¹ ...the type of dictionary associated with reception seems to have had precedence in demand, use, the value and attention attached and paid by users, teachers, researchers, and publishers over that associated with production.

⁵² They first try to express a (given) idea within the vocabulary they possess and when they are not sure about the usage of a word in mind, they look it up in their EJD for confirmation. When they cannot think of any words to start with, they reach out for a JED to find a translation equivalent for their idea. (...) When the Japanese word which a user first thought of to express her/his idea is not listed as a headword, s/he has to try another synonymous entry since the headword list in a JED is essentially limited.

3.1.11. Frankenberg-Garcia (2005)

A autora pesquisou não propriamente o uso do dicionário, mas o uso combinado de recursos disponíveis para solucionar dúvidas de escrita em L2, incluindo recursos disponíveis na Web. Os sujeitos de pesquisa foram dezesseis alunos do quarto ano do curso de tradução de inglês em uma atividade de tradução português-inglês. Os alunos preencheram uma planilha indicando o tipo de dúvida que motivou cada busca, as obras consultadas, os resultados e sua percepção sobre os recursos utilizados. Evidenciou-se o uso combinado de dicionários, de *corpora* de L2, banco de termos e outros. Os recursos mais utilizados foram: o *Eurodicauton*⁵³ (38%), seguido dos DBS (17%) e dos DMs de aprendizes (11%), de dicionários de colocações (10%), dois *corpora* de inglês (9%), de ferramentas de busca *on-line* (6%) e outros em menor escala. Apesar de ter sido o recurso mais consultado, o *Eurodicauton* recebeu uma avaliação de sua utilidade pior que os demais recursos. Com relação aos motivos de buscas, os três maiores foram: achar um equivalente (58%), confirmar um equivalente (18%) e encontrar a colocação adequada (16%). A autora concluiu que mais recursos não significam necessariamente melhores resultados e que os alunos precisam ser instruídos sobre como lidar concomitantemente com tantas opções de consulta.

Há de se observar que o gênero de texto escolhido para tradução influenciou o uso dos recursos lexicográficos, colocando o banco de termos em primeiro lugar em número de consultas. Isso demonstra que o peso da variável *gênero textual* deve ser considerado na análise dos resultados de pesquisas desse tipo.

O estudo de Frankenberg-Garcia abriu o foco da pesquisa sobre uso de dicionários, revelando que outras obras de referência concorrem ou se combinam com o uso do dicionário no apoio à codificação em L2. Além disso, seus resultados podem apresentar maior aplicabilidade ao contexto brasileiro, pois foi feita com falantes de português (em Portugal), o

⁵³ Base terminológica multilíngüe da União Européia, substituída em 2007 pela IATE (*Inter Active Terminology for Europe*) disponível no endereço: <http://iate.europa.eu/>.

que implica possível coincidência de tipos de dificuldade em relação à L2 e, teoricamente, compartilhamento dos mesmos recursos para consulta.

3.1.12. East (2006)

Interessado em investigar o impacto que o uso do DB traria se fosse permitido nos testes de proficiência em L2, East (2006) comparou redações escritas pelos mesmos 47 sujeitos com e sem o uso do DB. Sua pesquisa utilizou métodos quantitativos, diferentemente da maioria das pesquisas comentadas neste capítulo. O vocabulário das redações foi comparado com duas listas: a de um vocabulário básico e a de um vocabulário mais avançado. Os resultados obtidos mostram que o uso do dicionário contribuiu para a sofisticação lexical (os usuários utilizaram menos palavras do vocabulário básico e mais palavras que estavam fora até do vocabulário mais avançado). Além disso, o aumento da sofisticação foi maior nas redações dos alunos de nível intermediário, o que diminuiu a diferença entre alunos intermediários e avançados.

No entanto, das palavras consultadas nos dicionários, 50% apresentaram erros. Observados isoladamente, contudo, alunos de nível avançado obtiveram maior sucesso no uso do dicionário. Esse resultado diverge daquele obtido por Christianson (1997) e mostra correlação entre nível de proficiência e habilidade de consulta a dicionários.

Ao contrário do que era de se esperar, as notas atribuídas às redações com e sem uso do dicionário não apresentaram diferenças relevantes estatisticamente, “sugerindo que o dicionário não estava necessariamente agregando valor aos candidatos do teste em termos de contribuição positiva à qualidade de sua redação⁵⁴” (EAST, 2006, p. 15). O autor lança a hipótese de que os erros anularam o ganho de sofisticação lexical proporcionado pelo uso do dicionário. Ele acredita que o ensino do uso do dicionário poderia aumentar a proporção de

⁵⁴ ...suggesting that the dictionary was not necessarily adding any value to test takers in terms of positive contribution to the accuracy of their writing.

consultas bem sucedidas ao dicionário. Diferentemente dos demais autores aqui mencionados, East não faz nenhuma alusão à qualidade dos DBs utilizados.

3.1.13. Laufer e Levitzky-Aviad (2006)

Laufer e Levitzky-Aviad testaram um protótipo do DB para apoio à produção concebido por Laufer (1995). No experimento, 75 alunos israelenses traduziram 36 frases para o inglês, em quatro blocos de nove frases, cada bloco utilizando um tipo de dicionário. Um dos dicionários é o protótipo do DB para apoio à produção, *Bilingual dictionary plus*, que será analisado no capítulo 5. Os dicionários foram comparados pelo número de traduções (L1-L2) corretas obtidas após o uso e pela nota a eles atribuída pelos alunos. Nos dois quesitos o dicionário testado mostrou-se superior.

Os DBs L1-L2 raramente diferenciam os possíveis equivalentes em L2 para o item lexical de L1, bem como não fornecem informações sobre o uso de cada equivalente. O protótipo de dicionário testado, no entanto, contém: 1) traduções L1-L2; 2) informações em L2 para cada equivalente em L2 (definição e exemplos de uso); 3) itens lexicais semanticamente relacionados a cada equivalente e 4) significados adicionais em L1 de cada equivalente em L2. Desses quatro tipos de informação, apenas os dois primeiros mostraram-se úteis no experimento realizado pelas pesquisadoras. Além disso, Laufer e Levitzky-Aviad (2006) também testaram separadamente um protótipo impresso e um protótipo eletrônico do BD+ e, embora os resultados tenham sido igualmente satisfatórios em ambos os casos, os sujeitos da pesquisa manifestaram preferência pela versão eletrônica.

Os resultados dessa experiência devem ser considerados com cuidado, pois as autoras do protótipo de dicionário testado são as mesmas que determinaram as características do teste, o que pode, não intencionalmente, ter favorecido para que as características positivas do dicionário testado fossem ressaltadas. Não se sabe, também, se os sujeitos da pesquisa tinham

algum envolvimento do tipo professor-aluno com as pesquisadoras e se sabiam que o dicionário testado era de autoria delas, o que poderia afetar as opiniões expressas no teste. Além disso, é preciso reconhecer que a atividade de tradução de frases descontextualizadas não reproduz uma situação de codificação em L2 tão bem quanto um texto completo.

3.1.14. Gomes (2006)

A autora realizou a primeira pesquisa de que se tem conhecimento sobre o uso do dicionário na codificação em L2 no Brasil. No caso, os 80 sujeitos da pesquisa foram brasileiros aprendizes de alemão. Gomes utilizou protocolos escritos pelos próprios pesquisados durante uma atividade de redação em alemão e analisou tanto o processo do uso do dicionário na codificação em si quanto o efeito do uso do dicionário na redação resultante.

A análise do efeito do uso foi muito produtiva, pois, de posse das palavras consultadas, a autora verificou se as soluções adotadas foram satisfatórias, parcialmente satisfatórias ou insatisfatórias. Além disso, analisando os itens dos dicionários que foram consultados, ela pôde inferir se as falhas cometidas poderiam ser atribuídas à falta de habilidade dos usuários em usar os dicionários ou às deficiências dos dicionários, o que foi uma inovação no universo desse tipo de pesquisa. Gomes levantou estatísticas dos motivos de consulta, dos motivos de falhas dos alunos e dos motivos de falhas do dicionário, além das quantidades de consultas ao dicionário. Suas conclusões fornecem importantes subsídios para projetos dicionarísticos, além de corroborarem o que a maioria dos estudos desse tipo apregoa: há necessidade de instrução sobre como usar o dicionário tanto quanto de aperfeiçoamento dos DBs.

Dentre os motivos de falha do dicionário levantados por Gomes (2006) estão: a falta de equivalentes, a falta de elementos diferenciadores dos equivalentes, a falta de expressões e colocações e a falta de exemplos de uso. Uma das falhas mais observadas, no entanto, seria a

mais simples de sanar: a falta da forma plural de cada entrada. Além disso, Gomes observou incoerência nos DBs, pois não dão tratamento equânime aos itens lexicais. Em alguns verbetes encontram-se notas de uso, informação sobre as preposições que acompanham substantivos, verbos e adjetivos e informações sobre a regência verbal. Não existe critério aparente que explique a inclusão dessas informações para apenas alguns itens lexicais e não para todos.

Dentre os motivos de falha dos alunos, Gomes trabalha com a hipótese de que estejam relacionados à falta de conhecimentos lingüísticos tanto em L2 quanto em L1. Além disso, os alunos demonstraram dificuldade em compreender algumas informações do dicionário e tendência a ignorar parte dessas informações.

Gomes observou, ainda, o tipo de erro relacionado à ausência de decisão de consulta a dicionário:

...muitas vezes, ao utilizarmos o idioma estrangeiro não nos damos conta de que podemos estar diante de uma expressão construída de maneira diferente no nosso idioma. Dessa forma, nem sequer temos consciência de que deveríamos procurá-la no dicionário. (GOMES, 2006, p. 109)

3.1.15. Conclusões

Além das pesquisas aqui analisadas, Welker (2006) cita outras sobre o uso do dicionário em atividades de codificação em L2. Segundo o autor, a tradução L1-L2 também foi a circunstância de uso das pesquisas de Hatherall (1984), Wiegand (1985), Bogaards (1991), Varantola (1998), Lantolf, Labarca e den Tuinder (1985).

A redação em L2, por sua vez, também foi a circunstância de uso das pesquisas de Yokoyama (1994) e Momoi (1998). Não se teve acesso a essas pesquisas, mas pelo que foi possível depreender das resenhas apresentadas por Welker, seus resultados não contrariam os dados já apresentados na presente discussão.

Os resultados das pesquisas apresentadas devem ser comparados com ressalvas, pois as mesmas não foram feitas para serem comparadas, ou seja, não seguiram o mesmo método e o mesmo controle de variáveis.

Mesmo assim, a análise desse conjunto de pesquisas permitiu uma série de conclusões:

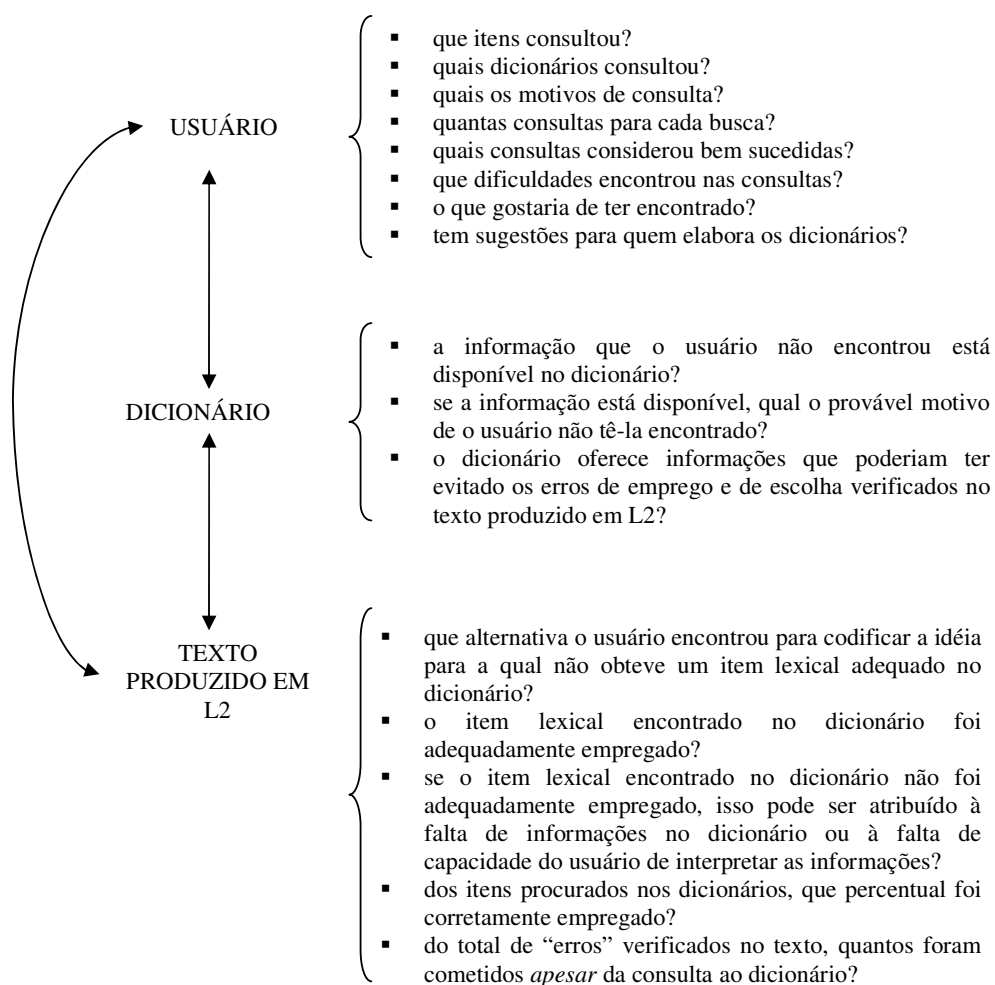
- Na atividade de codificação, a busca de equivalentes é o maior motivo de consulta a dicionários;
- A busca de equivalentes é, na grande maioria das vezes, iniciada por uma consulta à seção L1-L2 de um DB e, por esse motivo, os DBs são mais consultados nas atividades de codificação que os DMs;
- Frequentemente uma mesma busca gera uma segunda consulta com o objetivo de conferir a adequação do equivalente escolhido. Essa confirmação costuma ser feita ou na seção L2-L1 do DB ou no DM de L2;
- Às vezes a busca de confirmação de um equivalente é o motivo da primeira consulta, pois o equivalente foi inferido pelo usuário e não obtido por meio de outra consulta. Nesse caso, o primeiro dicionário consultado é um DB L2-L1 ou um DM;
- As consultas para confirmação evidenciam que as informações contidas nos DBs L1-L2 tradicionais não dão segurança ao usuário. Informações contidas na seção L2-L1 do DB e no DM de L2 parecem suprir parcialmente essa carência, o que indica que nem sempre é preciso produzir novos tipos de informações lexicográficas, mas simplesmente realocar as já existentes ou reuni-las em um só lugar;
- As pesquisas divergem no que diz respeito à correlação entre o nível de proficiência em L2 e o uso bem-sucedido de dicionários;

- Há divergências também quanto à correlação entre o nível de conhecimento lingüístico de L2 e a decisão de consultar o dicionário. Em alguns casos, usuários mais experientes consultaram mais o dicionário que usuários menos experientes;
- Há diferenças também nas conclusões sobre o tipo de dicionário mais útil à codificação. Essas divergências se justificam, pois nesse conjunto de pesquisas não há uniformidade de método nem controle de variáveis que permitam uma comparação dos resultados;
- Constituem variáveis na pesquisa sobre o uso do dicionário: o tipo de atividade proposta, o gênero textual do texto a ser codificado, os dicionários utilizados, a nacionalidade dos sujeitos da pesquisa, o ambiente em que o sujeito da pesquisa se encontra (de *segunda língua* ou de *língua estrangeira*) e o perfil dos sujeitos da pesquisa (idade, sexo, área de formação, nível de proficiência, domínio de outras L2 etc.);
- Falantes de diferentes L1 têm diferentes tipos de dificuldade em relação a uma mesma L2 e, por isso, têm diferentes níveis de aproveitamento das informações contidas nos DMs para aprendizes estrangeiros.

Adicionalmente, a revisão de pesquisas sobre o uso do dicionário na codificação permitiu que se levantassem pontos em comum entre elas, apresentados a seguir, que poderão contribuir para novas iniciativas no gênero.

Existem três fontes de dados que podem ser combinadas para se avaliar o uso do dicionário na codificação: o usuário, os dicionários consultados e o texto produzido em L2. Confrontar dados coletados nessas três fontes pode fornecer informações úteis tanto para o ensino do uso do dicionário quanto para o aperfeiçoamento e desenvolvimento de dicionários.

A seguir são relacionadas uma série de perguntas que podem orientar as pesquisas sobre o uso do dicionário na codificação em L2.



3.2. Os estudos de *corpus* de aprendizes

Corpus de aprendizes é uma abordagem muito utilizada em pesquisas relacionadas ao ensino e à aquisição de L2. A idéia de conhecer o aprendiz de uma L2 por meio da análise de um *corpus* eletrônico de suas produções foi em grande parte impulsionada pelo desenvolvimento da Lingüística de *Corpus* aplicada aos Estudos da Tradução.

Quando os Estudos da Tradução começaram a utilizar *corpora* automatizados, comparavam-se textos originais com suas respectivas traduções, uma análise que focava as relações bilíngües no que é conhecido como *corpus* paralelo. Esse tipo de *corpus* continua sendo utilizado para várias questões de pesquisa.

Os estudos de Toury (1978) sugeriram a existência de normas no texto traduzido, diferentes das normas observadas nas línguas de origem e de chegada. Era necessário, no entanto, confirmar a existência dessas normas por meio de estudos em larga escala. Baker (1993 e 1995), diante desse desafio, teve a idéia de montar um grande *corpus* monolíngüe de textos traduzidos e de textos originais, a fim de que pudessem ser comparados. Surgiu então uma nova proposta de abordagem para estudos da tradução: a do *corpus* comparável.

Mas a idéia de Baker não provocou avanços apenas nos Estudos da Tradução. Da mesma forma que as traduções apresentam certa norma dentro de uma língua, textos produzidos por não-nativos também podem evidenciar tendências e foi essa idéia que impulsionou os estudos baseados em *corpus* de aprendizes computadorizado.

Granger (1996) citou a proposta de Baker (1993) ao propor uma metodologia para estudo da interlíngua similar à proposta por Baker para estudo do texto traduzido. Enquanto na proposta de Baker contrastam-se traduções produzidas em uma língua com textos produzidos originalmente nessa mesma língua, na proposta de Granger contrastam-se textos produzidos por estrangeiros e por nativos em uma mesma língua. O que varia, portanto, nas duas propostas, não é a língua dos textos, mas a condição do autor:

tradutor	X	autor original,	(nos estudos da tradução);
estrangeiro	X	nativo,	(nos estudos da interlíngua).

Há mais de uma década, a proposta de Granger vem orientando pesquisas e ganhando maior sofisticação⁵⁵. A dificuldade inicial de digitalizar textos está sendo suplantada com a tendência de os professores solicitarem redações em formato eletrônico, acelerando a coleta de dados. Alguns problemas persistem e o principal deles é garantir que os sujeitos sejam comparáveis. Tono (2003) ressalta que a seleção dos sujeitos com base em ano escolar ou idade não garante que eles sejam comparáveis em termos de proficiência na língua. Por isso, o

⁵⁵ Lista com mais de 300 referências bibliográficas sobre *corpus* de aprendizes pode ser obtida em <http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/cecl.html>.

autor afirma que o ideal seria que os aprendizes que produziram material para o *corpus* fossem submetidos a um teste padrão de proficiência na língua. No entanto, ele mesmo reconhece que esse tipo de exigência poderia restringir o número de colaborações para o *corpus*. Assim, sua proposta é aplicar um teste rápido, de cerca de cinco minutos, a cada produção de material para o *corpus*. Testes desse tipo, segundo ele, já estão em fase de teste e seriam fundamentais para tornar os dados do *corpus* verdadeiramente comparáveis.

Normalmente se usa o critério idade ou série escolar para selecionar os sujeitos, mas isso não garante que eles tenham um mesmo grau de proficiência em L2. O ideal seria aplicar um teste padrão para medir a proficiência dos sujeitos, mas isso ainda é pouco exequível (v. TONO, 2003).

O primeiro projeto de *corpus* de aprendizes de âmbito internacional é o ICLE (*International Corpus of Learners of English*), constituído de vários *subcorpora* de redações de aprendizes de inglês de diferentes línguas maternas (v. GRANGER, 1996). O *subcorpus* do ICLE em português do Brasil está sendo construído sob direção de Berber Sardinha (2001). O ICLE permite comparar as variedades nativa e não-nativa de uma mesma língua, bem como a variedade não-nativa de diferentes origens (aprendizes com diferentes línguas maternas). Com isso, torna-se possível identificar características típicas de textos produzidos em inglês por estrangeiros em geral (independentemente de suas respectivas L1) e características típicas de textos produzidos em inglês por estrangeiros falantes de uma determinada L1. Outro projeto significativo, restrito aos aprendizes brasileiros, mas com várias línguas além do inglês, é o *corpus* de aprendizes do projeto COMET da USP (TAGNIN, 2003).

Os estudos baseados em *corpus* de aprendizes são uma oportunidade para se conhecer as dificuldades daqueles que escrevem em L2 e são, conseqüentemente, prováveis usuários de dicionários. Considerando as dificuldades identificadas pela análise das redações de aprendizes de L2, por exemplo, pode-se cogitar que tipo de necessidade de ação ou de

informação elas representam e, assim, desenvolverem-se recursos que supram tais necessidades.

Existe já um grande número de estudos de caso baseados em *corpus* de aprendizes reportando resultados que inspiram decisões de projetos lexicográficos. Conforme cita Granger (2004), o sobre-uso de vocabulário de alta frequência, o uso recursivo de um número limitado de expressões fixas, a influência do estilo da língua falada na escrita e a mistura de estilos formal e informal parecem ser características comuns a todos os aprendizes. Contudo, as causas dessas evidências ainda estão sendo pesquisadas, pois, como justifica a autora, a interlíngua avançada é o resultado de uma interação muito complexa de fatores, dentre eles o tipo de instrução e a interferência da L1.

Segundo Tono (2003), a estruturação adequada do *corpus* de aprendizes é essencial para dar confiabilidade aos resultados. Diversas variáveis precisam ser contempladas na identificação de cada componente do *corpus* para diminuir o risco de atribuir equivocadamente as causas dos comportamentos observados.

O *corpus* de aprendizes possibilita diversos tipos de análise. Pode-se comparar:

- a produção de aprendizes com a produção de nativos, para descobrir diferenças como sobre-uso, sub-uso, erros típicos de estrangeiros etc.;
- a produção de aprendizes de línguas maternas diferentes, para identificar a interferência da L1 na interlíngua;
- a produção de aprendizes em ambiente de L2 com a produção de aprendizes em ambiente de L1, para ver a influência do ambiente no aprendizado;
- a produção de aprendizes utilizando dicionário com a produção sem o uso do dicionário, para ver o quanto a ferramenta agrega de qualidade ao resultado;
- a produção de aprendizes de diferentes idades ou expostos a diferentes métodos etc.

As decisões dos lexicógrafos sobre o conteúdo e a forma de apresentação dos dicionários poderiam se beneficiar desses subsídios. O problema da mistura de estilos, por exemplo, poderia inspirar projetos lexicográficos que apresentassem exemplos de um mesmo conteúdo semântico expresso em diversos estilos. Uma vez identificadas as dificuldades sanáveis por meio da consulta a obras lexicográficas, como aquelas relacionadas a significados, colocações, grafia, pronúncia, marcas de uso etc., seria verificado se os dicionários disponíveis no mercado contêm as informações adequadas. Quando se detectasse a necessidade de agregar mais informações aos dicionários, seria preciso definir como elas poderiam ser obtidas e como deveriam ser exibidas para otimizar seu aproveitamento.

No entanto, nem todas as dificuldades expressas na produção dos aprendizes de L2 podem ser resolvidas pela Lexicografia. Existem dificuldades que só podem ser superadas com o ensino e outras que nem o ensino consegue superar, como é o caso de tendências que não constituem propriamente “erros”. Seria o caso, por exemplo, de tendências no texto do aprendiz semelhantes às apontadas pelos estudos da tradução (BAKER, 1996), como a simplificação, a explicitação, a normalização e a estabilização.

Evidências de *corpus* de aprendizes já são aproveitadas para a elaboração de DMs para falantes estrangeiros. No entanto, os DMs abordam especificamente os problemas comuns a várias interlínguas de uma língua, e não os problemas de transferência ou influência de L1. Esses problemas deveriam integrar as obras bilíngües.

De Cock e Granger (2005) dizem que os primeiros dicionários que utilizaram dados de *corpus* de aprendizes foram o *Longman Language Activator* e o *Longman Essential Activator*, os dois primeiros dicionários desenhados especialmente para a finalidade de apoio à codificação em L2.

Segundo as autoras, o principal interesse comercial de se manter um *corpus* de aprendizes é subsidiar o desenvolvimento de DMs para aprendizes. No entanto, elas mesmas

dizem que “... para dicionários bilingualizados, apenas dados de um grupo nacional específico de aprendizes pode ser usado⁵⁶” (DE COCK & GRANGER, 2005, p. 5).

Um *corpus* de aprendizes tem que conter grande volume de texto para ter representatividade; deve-se descartar a coleta oportunista e fazer a coleta segundo critérios controlados (idade, sexo, nível de proficiência, L1 etc.). As principais variáveis que influenciam a linguagem do aprendiz devem ser identificáveis, pois são críticas para a interpretação dos dados. Essas variáveis, segundo De Cock e Granger, têm que poder ser utilizadas como critério de busca pelos lexicógrafos e, por isso, as autoras concluem que “é necessário não apenas um *corpus* de L2, mas uma base de dados de L2 e um conjunto de ferramentas projetadas para auxiliar lexicógrafos...”⁵⁷ (DE COCK & GRANGER, 2005, p. 5).

Como o custo de se desenvolver, manter e etiquetar um *corpus* de aprendizes é alto, as autoras dizem que é recomendável constituir um grande *corpus* desse tipo que atenda a muitas aplicações (uso lexicográfico, desenvolvimento de material didático, estudos de interlínguas etc.).

Para uso lexicográfico, é preciso que os erros do *corpus* de aprendizes sejam localizados e as autoras descrevem quatro formas de fazer esse trabalho:

- 1) Busca baseada em *corpus*: procuram-se no *corpus* os tipos de erros já conhecidos;
- 2) Busca dirigida por *corpus*: utiliza-se o *corpus* sem hipóteses preliminares; compara-se, por exemplo, listas de frequência de *corpus* de aprendizes com listas de frequência de *corpus* de falantes nativos e as diferenças são investigadas;

⁵⁶ ...for bilingualised dictionaries, only data from one specific national learner group can be used...”

⁵⁷ ...what is needed is not simply an L2 corpus, but an L2 database and workbench designed to help lexicographers...

- 3) Análise das concordâncias: listam-se as concordâncias para todas as palavras e analisam-se todas até achar erros, o que seria perfeito se não consumisse tanto tempo;
- 4) Notação de erros: todos os erros do *corpus* recebem uma etiqueta apropriada; o método é árduo mas, segundo as autoras, compensa em vista dos benefícios; porém essa estratégia exige critério e definição de uma taxonomia de erros, pois a atividade está sujeita à influência de alto grau de subjetividade.

Se os erros estiverem “etiquetados”, torna-se possível levantar estatísticas dos erros mais recorrentes, o que pode orientar a decisão de quais devam ser contemplados pelas informações lexicográficas.

Com relação aos tipos de erros, De Cock e Granger afirmam que, “como todos os processadores de texto incluem corretores ortográficos, erros de ortografia poderiam ser vistos como sendo apenas de relevância marginal”⁵⁸ (p. 8). Há erros, porém, que os corretores ortográficos não detectam, como os erros de escolha lexical e os erros léxico-gramaticais. As autoras dizem ainda que o *corpus* de aprendizes mostra muitos erros passíveis de serem transformados em material lexicográfico, mas os dicionários impressos, com sua limitação de espaço, aproveitam apenas uma pequena proporção dessas evidências. Elas acreditam que o desenvolvimento da Lexicografia eletrônica vai mudar esse cenário e fazer uso intensivo dos subsídios fornecidos pelos *corpora* de aprendizes. Tudo indica que, uma vez que incorporar evidências de *corpus* de aprendizes no dicionário é um trabalho árduo e contínuo, os dicionários eletrônicos realmente são mais adequados para fazê-lo, pois permitem atualizar versões com maior facilidade.

De Cock e Granger também analisaram as notas de uso de dois dicionários que declaram ter acrescentado evidências de *corpus* de aprendizes. Elas verificaram que as notas

⁵⁸ As word processors all include automatic spell-checkers, orthographic errors could be regarded as of only marginal relevance.

de um dicionário não coincidem com as notas do outro. Além disso, muitas delas falam de erros que as autoras não identificaram no *corpus* de aprendizes internacional que gerenciam (o ICLE), o que coloca em dúvida a origem dessas notas.

Quanto ao perfil do profissional para esse trabalho, as autoras consideram que

...a análise de *corpora* de aprendizes sempre exigirá não apenas habilidades lexicográficas avançadas, mas também sólida experiência de ensino. Como não é fácil encontrar pessoas que tenham ambas habilidades, a exploração lexicográfica de *corpora* de aprendizes terá que ser geralmente apoiada em uma intensa colaboração entre lexicógrafos treinados e professores de língua estrangeira experientes.”⁵⁹ (DE COCK & GRANGER, 2005, p. 14)

Rundell e Granger (2007) relatam um projeto de parceria entre a universidade belga que sedia o ICLE e uma editora de dicionários (a Macmillan), o qual resultou na produção de cerca de 100 notas de uso para o DM para aprendizes de inglês da editora. Essas notas focaram os erros mais abrangentes e recorrentes, mostrando exemplos do uso incorreto retirados do ICLE, esclarecendo as causas dos erros e apresentando alternativas corretas. Esse relato tem o valor de orientação sobre como transformar evidências de *corpus* de aprendizes em notas de uso.

3.3. Conclusões gerais do capítulo

Neste capítulo foram reunidas duas frentes de pesquisa diferentes (pesquisas empíricas sobre o uso do dicionário na codificação e análise de *corpus* de aprendizes) que atendem uma mesma finalidade lexicográfica: descobrir quais são as necessidades dos usuários. As pesquisas sobre o uso do dicionário em atividades de codificação geraram evidências que permitem concluir que, nessa situação, o usuário precisa ter a possibilidade de iniciar a consulta tanto via L1 quanto via L2.

⁵⁹ the analysis of learner corpora will always require not only expert lexicographical skills but also solid teaching experience. As it is no easy matter to find people who have both of these skills, lexicographical exploitation of learner corpora will generally need to rely on close collaboration between trained lexicographers and experienced ELT teachers.

Tanto as pesquisas sobre o uso do dicionário quanto as pesquisas de *corpus* de aprendizes indicam a necessidade de se desenvolver uma solução específica de apoio à codificação em L2 para cada nacionalidade ou, pelo menos, para cada grupo de falantes de uma mesma L1.

Enquanto as pesquisas sobre uso do dicionário na codificação apontam os tipos de informações necessários para essa finalidade, a análise de *corpus* de aprendizes fornece subsídios para decidir que informações devam ser incluídas para cada item lexical. O envolvimento de lexicógrafos nos projetos de construção de *corpora* de aprendizes brasileiros, portanto, poderia maximizar o aproveitamento dos *corpora* resultantes para fins lexicográficos.

No próximo capítulo será relatada a pesquisa sobre o uso do dicionário empreendida dentro do escopo desta tese para explorar empiricamente a situação de codificação em L2.

4. COLETA DE DADOS PRIMÁRIOS SOBRE O USO DO DICIONÁRIO NA CODIFICAÇÃO EM LÍNGUA ESTRANGEIRA

As revisões bibliográficas apresentadas nos capítulos 1, 2 e 3 forneceram elementos relevantes para a finalidade deste estudo. Contudo, esses elementos são dados secundários, resultados de pesquisas com propósitos diferentes daquele que se propõe aqui. Além disso, são poucas as pesquisas sobre o uso do dicionário em contexto brasileiro e nenhuma das que se tem conhecimento observou a escrita em L2 deixando livre o uso de recursos impressos e eletrônicos. Por essas razões, julgou-se essencial realizar uma pesquisa dessa natureza, focando especificamente as questões de interesse deste trabalho.

O objetivo da pesquisa sobre o uso do dicionário, aqui relatada, foi observar como os usuários codificam em L2, independentemente de se tratar de redação ou de tradução, focando os seguintes aspectos: que recursos utilizam, quando utilizam cada recurso, quais as dúvidas que motivam as consultas e, principalmente, quais as principais dificuldades não solucionadas pelos dicionários. Em suma, procurou-se, por meio de uma pesquisa exploratória, identificar oportunidades para desenvolver novas soluções que possam atender as necessidades lexicográficas de quem escreve em L2.

Essa fase do trabalho foi muito produtiva, principalmente na etapa de coleta dos dados, pois a oportunidade de observar pessoalmente a situação de escrita em L2 fornece uma visão mais real do processo, o que complementa a visão formada por meio da leitura de relatos de pesquisas.

4.1. Metodologia

As escolhas metodológicas foram feitas após a análise de diversas pesquisas sobre o uso do dicionário, o que tornou possível conhecer as vantagens e limitações dos métodos,

técnicas e instrumentos adotados. Como não se desejava verificar hipóteses, mas sim explorar a situação a fim de desvendar novos aspectos da codificação escrita em L2, a pesquisa qualitativa mostrou-se mais adequada, pois ela propicia ao pesquisador a oportunidade de investigar novos campos e aprofundar o conhecimento acerca do assunto pesquisado, motivo pelo qual é também chamada de pesquisa em profundidade.

A pesquisa qualitativa não tem o propósito de conhecer o todo pela parte, o universo pela amostra: não permite generalizações. Sua aplicação é adequada para a elucidação de questões complexas, quando muitas variáveis são desconhecidas. Gordon e Langmaid (1988) dizem que a pesquisa qualitativa é aquela que responde a perguntas como: *o quê?*, *por quê?* e *como?*, mas não consegue responder a pergunta *quantos?*

As pesquisas qualitativas foram muito criticadas no passado pelo fato de envolverem poucos sujeitos de pesquisa, o que, na visão daqueles que estavam acostumados com o paradigma das pesquisas quantitativas, comprometia a validade e a confiabilidade de seus resultados. No entanto, hoje esse tipo de pesquisa tem seu valor reconhecido em várias áreas científicas, pois por meio delas é possível ter acesso ao sujeito real de pesquisa, ao passo que nas pesquisas quantitativas aspectos da individualidade são apagados no processo de mensuração.

Dentre as técnicas mais utilizadas nas pesquisas de abordagem qualitativa sobre o uso do dicionário estão as técnicas introspectivas (v. WALLACE, 1998): o “pensar em voz alta”, a auto-reflexão e a auto-observação, mas há também estudos que utilizaram a observação e a entrevista em profundidade. Essas técnicas têm por objetivo descobrir a linha de raciocínio dos sujeitos de pesquisa durante a realização de uma determinada atividade. O que as difere são as questões relativas a: que dados são coletados, quando são coletados, como são registrados e quem os registra.

O interesse no presente caso era coletar dados simultaneamente à realização da atividade, posto que qualquer defasagem de tempo entre a ação e o relato da ação poderia implicar a perda de informações. Além disso, não se desejava atribuir aos sujeitos de pesquisa a tarefa de registrar os dados, pois isso os obrigaria a interromper a atividade, desviando a atenção da tarefa em si. Para atender a esses requisitos, optou-se por aplicar a técnica do “pensar em voz alta”, também conhecida como “protocolo verbal”. Por meio do protocolo verbal se registram em áudio os pensamentos verbalizados pelos sujeitos de pesquisa durante a atividade. Contudo, Welker aponta que: “Wiegand, com sua costumeira precisão, considera o termo [*think aloud*] errado, porque não existe o pensar alto, pois, na verdade, o que as pessoas fazem é falar, dizendo o que lhes passa pela cabeça” (WELKER, 2006, p. 30). Assim, os sujeitos desta pesquisa foram orientados a expressar em voz alta suas linhas de pensamento; suas dúvidas, suas estratégias de busca lexical, suas avaliações sobre os resultados das consultas e suas decisões tradutórias.

O primeiro pesquisador a utilizar o protocolo verbal na investigação do uso do dicionário foi Ard (1982). Welker (2006, p. 30) cita mais quinze pesquisas nessa área que utilizaram o mesmo recurso metodológico. No Brasil, Höfling (2006) utilizou o protocolo verbal para pesquisar o uso do dicionário na leitura em L2. Uma discussão detalhada sobre os problemas envolvidos na aplicação do protocolo verbal é apresentada por Uzawa (1996) na descrição de sua pesquisa. A técnica deve levar em conta que nem todos os sujeitos de pesquisa têm a mesma facilidade para expressar oralmente aquilo em que estão pensando. Por essa razão, Uzawa salienta a importância de entrevistas após a atividade para complementar e esclarecer os dados registrados.

Tendo por base essa recomendação, estabeleceu-se que, depois de concluída a tradução, o participante teria um tempo para comentar o que achou da atividade e dos recursos utilizados, opiniões que foram transcritas juntamente com os protocolos verbais.

Além disso, optou-se por adotar outro instrumento para complementar os registros do protocolo verbal. Esse instrumento é um formulário desenvolvido por Atkins e Varantola (1997) em que um observador registra por escrito determinadas ações e pensamentos verbalizados pelos usuários (Wiegand, *apud* WELKER, 2006, chamaria isso de protocolo escrito). A principal vantagem desse instrumento é dar um caráter mais sistemático à coleta de dados, a fim de facilitar a tabulação de dados.

O formulário sofreu algumas adaptações após a realização de um estudo-piloto e é composto de duas partes. A primeira é uma folha de rosto (Anexo 1), utilizada para registrar dados da atividade (data, horário, descrição da atividade) e dados do sujeito de pesquisa (nome, idade, curso, recursos que utilizados). A segunda parte (Anexo 2) é uma folha com capacidade para registro de dez consultas e que, portanto, foi reproduzida diversas vezes para cada atividade (utilizou-se cerca de 10 folhas por hora de atividade). Um exemplo dessa folha preenchida (aqui digitada, mas originalmente preenchida a mão) é apresentado no Anexo 3.

Enquanto na pesquisa de Atkins e Varantola (1997) o preenchimento foi realizado por um aluno-observador, aqui se optou por atribuir essa tarefa ao próprio pesquisador, eliminando intermediários no processo. A vantagem disso é a oportunidade de contato direto do pesquisador com a situação-alvo da pesquisa e a desvantagem é que isso limita a quantidade de sujeitos observados à disponibilidade de tempo do pesquisador.

A decisão de fazer um duplo registro da atividade, utilizando dois tipos de instrumento de pesquisa, foi essencial para a análise dos dados. O registro escrito mostrou-se mais completo que o registro oral, já que os sujeitos muitas vezes deixavam de verbalizar algumas etapas de consulta, confirmando o que Uzawa (1996) já havia observado. Por outro lado, o registro em áudio permitiu a análise das opiniões dos participantes, que ficaram excluídas do registro no formulário (o Anexo 4 apresenta um dos protocolos verbais transcrito na íntegra).

Assim, não houve uma superposição dos dois registros, mas uma complementação no fornecimento de dados para análise.

Propositalmente, não foi determinado se o exercício seria escrito em papel ou digitado no computador, nem quais obras de referência poderiam ser consultadas. Os sujeitos de pesquisa foram informados apenas de que o interesse da pesquisa era observá-los agindo como se estivessem em uma situação real. Pretendeu-se, assim, dar oportunidade para que suas preferências pessoais sobressaíssem.

As escolhas metodológicas mostraram-se satisfatórias, pois se preservou o máximo possível a naturalidade da atividade, já que o único esforço extra exigido dos sujeitos de pesquisa foi a verbalização de seus pensamentos. Além disso, o envolvimento do pesquisador com a situação de coleta de dados proporcionou uma experiência mais rica do que se o contato fosse apenas com os instrumentos nos quais ficaram registradas as atividades.

4.2. Estudo-piloto

Uma vez escolhido o método de pesquisa e os instrumentos de coleta de dados, mas antes da definição dos sujeitos de pesquisa e do tipo de atividade (redação, versão ou ambos), foi realizado um estudo-piloto com dois alunos do terceiro ano de um curso de Tradutor de francês. O primeiro realizou os dois tipos de tarefa típicos da codificação escrita em L2: a tradução de texto de L1 para L2 e a redação em L2. O segundo realizou apenas a tradução para L2. O objetivo desse estudo-piloto foi avaliar a adequação dos instrumentos de pesquisa e obter subsídios para decidir qual o melhor tipo de atividade para observar o uso do dicionário na codificação: a redação ou a tradução para L2.

A aluna que realizou a redação optou por escrever tudo e só depois consultar o dicionário. Segundo ela, dessa forma não se interrompia o processo criativo. As palavras sobre as quais tinha dúvida foram grifadas para posterior consulta e as palavras que não sabia

em L2 foram escritas em L1, também para posterior consulta. Ao todo, a aluna fez sete consultas no DM e uma no DB. Dessas consultas, três foram para verificar ortografia, uma para confirmar a existência da palavra empregada, três para buscar uma colocação e apenas uma, a do DB, para encontrar um equivalente. A aluna conseguiu utilizar diversas estratégias para expressar o que desejava, empregando apenas o vocabulário que já dominava, evitando, assim, consultar o dicionário.

Na tradução, ao contrário, a mesma aluna se viu obrigada a utilizar o dicionário, já que não era possível mudar o conteúdo a ser expresso para adequá-lo ao vocabulário já conhecido. A tradução constituiu-se de dois parágrafos de um texto jornalístico, sem vocabulário técnico. Durante uma hora de atividade de tradução foram feitas 20 consultas ao dicionário, o triplo da quantidade de consultas no mesmo tempo de atividade de redação.

Nessa experiência observou-se o que já havia sido apontado na pesquisa de Uzawa (1996) descrita no capítulo 1: como na tradução o sujeito de pesquisa não tem que gerar idéias e planificar um texto, ele pode concentrar-se mais nos problemas da língua do que na atividade de redação. A consequência desse fato foi um uso mais intensivo do dicionário na atividade de tradução que na atividade de redação.

O segundo participante do estudo-piloto foi convidado a realizar uma tradução, do mesmo texto jornalístico, para L2, a fim de ratificar a constatação de que a tradução para L2 obriga mais o usuário a usar o dicionário do que a redação. Em uma hora e quarenta e cinco minutos, 66 consultas foram realizadas. Tal resultado contribuiu para a escolha da tradução e não da redação em L2 como atividade mais adequada para verificar o uso do dicionário na codificação em L2. A diferença na quantidade de consultas feitas pelos dois alunos do estudo-piloto deve-se ao fato de que, embora o texto tenha sido o mesmo, o segundo dispunha de mais tempo e acabou traduzindo uma parcela maior do texto. O tempo não foi padronizado

nessa fase, pois não havia interesse em comparar os dois sujeitos, mas sim explorar a atividade individualmente.

Após o estudo-piloto, percebeu-se também que não havia necessidade de escolher um texto “difícil” para dar oportunidade de ocorrência de dúvidas lexicais que motivassem consultas a dicionários, pois as dificuldades dos sujeitos de pesquisa no estudo-piloto estavam relacionadas às palavras mais freqüentes da L2.

Concluída a fase de teste, passou-se à definição dos sujeitos de pesquisa e à escolha do texto a ser traduzido para L2.

4.3. Definição do texto para a atividade e dos sujeitos de pesquisa

O texto escolhido para tradução no estudo-piloto mostrou-se muito extenso e os sujeitos acabaram traduzindo apenas parte dele. Assim, a fim de que a pesquisa contemplasse a tradução de um texto completo, procurou-se outro texto de menor extensão, que pudesse ser traduzido em até duas horas de atividade. A escolha baseou-se na expectativa de verificar dificuldades relativas ao uso do léxico geral na codificação em L2, já que são os dicionários de língua (e não os de termos) que estão em foco neste trabalho. Foi escolhido um texto jornalístico, composto de dois parágrafos, que se encontra no Anexo 5.

A seguir, passou-se à definição dos sujeitos de pesquisa. Como já relatado anteriormente, as pesquisas sobre uso do dicionário apontam, de um lado, falhas e limitações dos dicionários e, de outro lado, a falta de habilidade de consulta por parte dos usuários. Como o objetivo desta pesquisa é buscar melhores soluções lexicográficas para a codificação em L2, considerou-se interessante tirar o foco da falta de habilidade dos usuários.

A fim de procurar atender essa premissa, foram convidados a participar da pesquisa todos os alunos do quarto ano de um curso de Tradutor em que as línguas ensinadas são o inglês, o francês, o espanhol e o italiano. Nesse curso as atividades de tradução para L2 são

objeto de disciplina específica. Escolhendo esse público, esperava-se encontrar sujeitos que já soubessem utilizar dicionários e, portanto, cujas dificuldades estariam mais associadas às deficiências dos dicionários do que à falta de habilidade para consultá-los.

Como apenas vinte por cento dos alunos, isto é, seis alunos, aceitaram o convite, procurou-se adequar a estrutura da pesquisa ao número e às características dos sujeitos participantes.

Tabela 1 - Sujeitos de Pesquisa

Identificação do Sujeito	Idade	Língua Estrangeira
Sujeito A	21	italiano
Sujeito B	22	espanhol
Sujeito C	21	inglês
Sujeito D	21	francês
Sujeito E	21	francês
Sujeito F	22	francês

As características da pesquisa, portanto, são:

- Número de sujeitos: 6;
- Línguas estrangeiras: italiano, inglês, espanhol e francês;
- Forma de coleta dos dados: protocolo verbal e protocolo escrito semi-estruturado preenchido pelo pesquisador;
- Ambiente: escolhido pelo sujeito de pesquisa (todos escolheram sala com computador e acesso à Internet);
- Pessoas presentes no ambiente de pesquisa: o sujeito de pesquisa que realizou a atividade e o pesquisador;

- Instrumentos de coleta de dados: gravador de áudio e formulário (modelo no Anexo 1);
- Variáveis de pesquisa:
 - dos sujeitos: idade, formação acadêmica, ano escolar (com os quais pretende-se garantir um nível semelhante de proficiência em L2);
 - L2;
 - aparato lexicográfico de L2 (dicionários bilíngües e monolíngües disponíveis para cada L2 no Brasil);
- dados coletados:
 - itens consultados;
 - quantidade de buscas;
 - quantidade de consultas por busca;
 - motivos de consulta;
 - fontes consultadas;
 - estratégias de consulta;

O tempo das atividades de tradução variou de 50 a 85 minutos. Inicialmente, havia-se pensado em comparar esse tempo. No entanto, problemas alheios à vontade dos participantes e do pesquisador acabaram influenciando a duração das atividades, como, por exemplo, o tempo de espera nas consultas na Web e, em alguns dos casos, problemas com o teclado do computador. Sendo assim, o dado ficou comprometido e considerou-se melhor desprezá-lo.

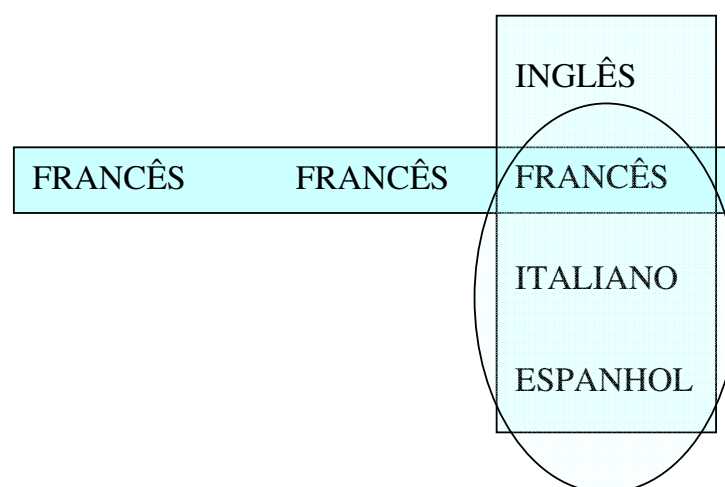
Foram visualizadas três perspectivas para análise dos dados. Assim, decidiu-se, em um eixo, analisar os três sujeitos que verteram o texto para o francês, comparando-se a quantidade de consultas e os itens lexicais consultados. Além disso, tradutores de uma mesma L2, teoricamente, têm à disposição os mesmos recursos para consulta e, por isso, podem ser comparadas também as estratégias de utilização desses recursos. Como constantes, nesse eixo,

estão a L1 (português do Brasil), a idade (entre 21 e 22 anos) e a formação acadêmica dos sujeitos de pesquisa, além da L2 (francês) e do texto a ser traduzido. Supostamente, esses alunos devem ter uma proficiência lingüística semelhante, já que estão prestes a obter o mesmo diploma. Nesse eixo busca-se verificar diferenças no uso do dicionário e identificar que variáveis podem explicar essas diferenças.

Em outro eixo comparam-se os sujeitos que realizaram a tradução para o francês com os sujeitos que realizaram a tradução do mesmo texto para o inglês, o italiano e o espanhol. Nesse eixo as constantes são a L1, a idade e a formação acadêmica extra-L2, bem como o texto a ser traduzido. As variáveis são: a L2 e os recursos lexicográficos disponíveis para cada uma das L2 envolvidas. Essa perspectiva permite que se levantem hipóteses a respeito da relação entre as estratégias de uso do dicionário e a qualidade do aparato lexicográfico para cada idioma.

Em outra perspectiva compara-se o conjunto de atividades em línguas neolatinas (francês, italiano e espanhol) com a atividade em inglês. A pergunta, no caso, é: o fato de o português (língua do texto de origem) ser uma dessas línguas (e, portanto, estar mais “próximo” delas do que do inglês) influencia o uso do dicionário?

Figura 4 - Perspectivas de análise dos dados



Finalmente, todas as consultas são analisadas e reunidas a fim de se produzir um repertório dos tipos de dúvidas que levam à consulta do dicionário.

4.4. Análise dos resultados

Os sujeitos de pesquisa, sem exceção, escolheram realizar a atividade em um computador e mostraram-se bem familiarizados com seus recursos. Alguns, porém, declararam não possuir seu próprio computador, o que não impediu sua inclusão digital, pois eles utilizam recursos alugados em *Lanhouses* ou compartilham recursos nos laboratórios das escolas. Assim, todas as atividades foram realizadas em salas onde havia um computador disponível com acesso à Web. Os sujeitos de pesquisa foram avisados que poderiam levar os dicionários que costumam utilizar, tanto impressos quanto em CDROM. É importante enfatizar também que, em função das escolhas metodológicas, as atividades não foram simultâneas, ou seja, cada sujeito de pesquisa foi acompanhado individualmente pelo pesquisador.

A facilidade de alterar o que foi escrito digitalmente parece ter desenvolvido nos sujeitos o hábito de fazer rascunho e adiar a solução de problemas, assinalando os pontos do texto a serem verificados no momento da finalização. Assim, enquanto o manuscrito exige reflexão e decisão no momento em que se escreve, pois toda alteração significa maior esforço físico, o texto digital facilita a escrita de várias versões até chegar à definitiva. No protocolo verbal do sujeito D, por exemplo, ocorre a seguinte fala:

é, eu tô pensando agora no *confiar*, como é que eu traduzo pro francês, porque tem o verbo *fier à quelqu'un*, que é o verbo *confiar*, só que eu preciso ver se aqui cabe nesse sentido o verbo *confiar os filhos a alguém* / então eu vou procurar no dicionário, primeiro no Robert & Clé / é, eu tô procurando, mas ainda não consegui achar pro *fier* esse sentido que o texto tá dando ainda, porque no dicionário tem geralmente ligado a *confiance*, é *acreditar* mesmo e não no sentido de *confiar*, mas por enquanto eu vou deixar meio por cima essa opção do *fier* pro verbo *confiar* /

O adiamento de questões problemáticas é apenas uma das inúmeras alterações de comportamento que a mídia digital pode estar provocando e foi observado em todos os sujeitos de pesquisa, corroborando o que foi apontado por Li (2006) e destacado no capítulo 1 desta tese.

Outra alteração de comportamento parece ser uma resistência à consulta ao dicionário impresso. Esse tipo de dicionário foi consultado, mas como última alternativa. Isso talvez se deva ao fato de os alunos estarem se acostumando à facilidade de consulta dos dicionários eletrônicos. Eles souberam, contudo, manusear adequadamente as obras impressas. Também na análise das falas gravadas, percebe-se que os alunos têm preferência por recursos acessíveis no computador, em detrimento dos dicionários impressos, mesmo nos casos em que reconhecem um impresso de qualidade superior a um dicionário *on-line*. Esses alunos parecem não ter investido muito na aquisição de dicionários, nem impressos nem eletrônicos e buscam sanar essa carência com obras disponíveis gratuitamente na Web.

Contudo, um aluno apontou uma vantagem dos dicionários impressos sobre os dicionários que imitam os impressos em mídia eletrônica: em verbetes de grande extensão, a tela é limitadora, não dá para ver o todo e, portanto, fica mais difícil correr os olhos em leitura dinâmica à procura do que interessa. Esse foi um problema enfrentado na maioria das consultas, pois grande parte das dúvidas ocorridas estava relacionada a itens lexicais altamente frequentes e que, por isso mesmo, apresentam verbetes mais extensos. Nesses casos, o recurso de busca dentro do verbete foi muito útil e elogiado pelos alunos como forma de poupar tempo e dar maior objetividade à consulta, mas não está disponível em todos os dicionários eletrônicos. Em mídia eletrônica, a tela teria que apresentar um volume menor de informações, para não ficar “poluída” e oferecer opções de aprofundamento que levassem à exibição de novas telas (recurso de *hyperlinks* ao invés de rolagem da mesma tela)

Essas percepções sugerem que o desenvolvimento de novos materiais lexicográficos de apoio à codificação escrita em L2 deva ser feito em ambiente digital, pois, cada vez mais, é nesse ambiente que as pessoas escrevem. Isso está de acordo com uma das conclusões do capítulo 2, quando se revisou as opiniões de teóricos da Lexicografia acerca de dicionários para codificação. Deve-se tentar, também, explorar ao máximo os recursos da nova mídia e não simplesmente “transportar” o formato de exibição impresso para o formato eletrônico.

Outro comportamento que pôde ser observado durante a pesquisa foi o de compor e testar. Os sujeitos compõem seqüências lexicais, com base em seus conhecimentos prévios e nas informações obtidas em dicionários e, em seguida, testam essas seqüências, digitando-as como palavra-chave no motor de busca Google. O resultado da busca no Google foi utilizado para verificar se seqüência existe, se é freqüente ou não no idioma e também em que tipo de contexto ocorre. É um uso da Web na qualidade de um *corpus* de língua. Essa prática mostra a oportunidade de desenvolvimento de ferramentas de apoio à escrita em L2 que extrapolam a visão estereotipada de dicionários.

Quanto à tarefa de tradução propriamente dita, a primeira impressão dos seis sujeitos de pesquisa foi de que se tratava de um texto simples que não apresentava nem idéias nem palavras difíceis. O sujeito C, que realizou a tradução para o inglês, disse sobre o texto a ser traduzido:

“até que o vocabulário dele é tranqüilo, se fosse um texto técnico aí seria o Google assim... direto, porque aí é mais difícil”.

Essa declaração sobre a necessidade de usar muito o Google na tradução de textos técnicos sugere que a carência de recursos terminológicos talvez seja ainda maior que a carência de dicionários de língua geral para a codificação em L2.

Grande parte dos itens lexicais consultados foi coincidente em todas as línguas:

- Babá e robô – à procura de palavras adequadas para compor “babá-robô”

- Deixar – no sentido de “deixar os filhos com alguém”
- Confiar – no sentido de “confiar os filhos aos cuidados de alguém”
- Ceder – no sentido de “ceder à vontade de alguém”
- Altura – no sentido de “metros de altura”
- Tomar – no sentido de “tomar conta de alguém”
- Função – no sentido de “ter por função” ou “ter como função”
- Apesar de
- Desenvolvido – no sentido de “máquina ou robô desenvolvido por alguém”
- Lançar – no sentido de “lançar mensagens”
- Advertência – no sentido de “mensagens de advertência”
- Câmera fotográfica

Os sujeitos de pesquisa com frequência conheciam equivalentes para os itens lexicais isolados, mas permaneciam com dúvidas em relação a seqüências lexicais, como se vê nos excertos a seguir:

“agora eu vou procurar o verbo *ceder* / eu imagino que seja *céder* mesmo, mas aqui está *ceder à vontade*, então eu vou ver se tem alguma coisa diferente no dicionário” (sujeito D)

“*com quem deixar seus filhos*, isso eu vou ter que procurar...” (sujeito C)

“agora, o *deixar aos cuidados*, *deixar* eu uso sempre o *laisser*, mas eu não sei se cabe nesse sentido” (sujeito E)

“então eu vou procurar esse *já tem acesso* / *acesso* eu já sei que é *accès*, deixa ver o que seria...” (sujeito F)

Essas seqüências confirmam o que foi apontado na revisão bibliográfica: a grande dificuldade de codificar em L2 está na combinação de itens lexicais.

Utilizando a mesma terminologia de Atkins e Varantola (1997), optou-se neste trabalho por chamar de “busca” o conjunto de consultas feitas para solucionar uma mesma

dúvida. Sendo assim, cada busca lexical pode envolver várias consultas. O número de buscas e de consultas realizadas por todos os sujeitos de pesquisa encontra-se no quadro a seguir.

Tabela 2 - Relação entre número de buscas e de consultas por usuário

Sujeito de Pesquisa	Língua	Número de Buscas	Número de Consultas	Média de consultas por busca
A	italiano	43	82	1,91
B	espanhol	36	53	1,47
C	inglês	23	29	1,26
D	francês	22	45	2,05
E	francês	23	43	1,87
F	francês	22	42	1,91
	TOTAIS	177	294	1,66

Observa-se uma grande variação no número de buscas e consultas entre as línguas, porém os três alunos de francês fizeram uma quantidade semelhante de buscas e consultas

Nota-se que o inglês apresentou a menor média de consultas por busca, indicando que o recurso utilizado prioritariamente foi suficiente para sanar a dúvida.

Os sujeitos de pesquisa utilizaram basicamente três tipos de recursos: DB, DM (de L2) e o motor de busca Google. Esse resultado difere dos resultados da pesquisa de Frankenberg-Garcia (2005), que apresentou maior diversidade de recursos utilizados. Salienta-se especialmente o fato de aquela pesquisa ter observado uso de *corpora* superior ao uso do Google, enquanto na presente pesquisa a ausência que se observa do uso de *corpora* parece ter sido compensada pelo maior uso do Google.

A tabela a seguir mostra como se distribuiu a quantidade de consultas entre esses três recursos. Essa distribuição indica que a decisão de consultar um ou outro tipo de recurso provavelmente está associada a fatores idiossincráticos, pois parece não haver nenhum elemento lógico que explique tamanha variação.

Tabela 3 - Consultas por participante/idioma, distribuídas por tipo de recurso utilizado.

Sujeito de Pesquisa	Língua da atividade	Consultas a DMs	Consultas a DBs	Consultas no Google	Total Consultas
A	italiano	29	34	19	82
B	espanhol	25	22	6	53
C	inglês	5	16	8	29
D	francês	19	8	18	45
E	francês	12	29	2	43
F	francês	11	14	17	42

Observa-se na Tabela 3 que, embora tenham apresentado um número de consultas semelhante, os três alunos que traduziram para o francês utilizaram estratégias diferentes: o sujeito D priorizou o DM e o Google; o sujeito E priorizou o DB e o sujeito F priorizou o Google, embora tenha utilizado bastante o DM e o DB também. O que explicaria essa falta de uniformidade de estratégias? Essa é uma pergunta difícil de responder apenas com os dados desta pesquisa, mas uma hipótese é a de que esses alunos não foram ensinados a utilizar determinadas estratégias e acabaram construindo cada um a sua.

Observou-se também que, quando o uso de um determinado recurso apresentava uma solução satisfatória, os alunos tendiam a priorizar sua utilização nas próximas consultas. De forma análoga, quando determinado recurso frustrava as expectativas por mais de uma vez, seu uso acabava sendo relegado pelos alunos nas próximas consultas. Esse fato parece mostrar que os alunos optaram por utilizar determinado recurso não *a priori*, por acreditarem ser a fonte mais adequada para resolver um tipo específico de problema, mas porque era o recurso que, durante a atividade, vinha conduzindo a um maior número de soluções satisfatórias. Isso demonstra a existência de um fenômeno em que os usuários, por meio de repetidas consultas, adquirem confiança (ou não) em um dicionário.

No inglês observa-se que o DB foi usado prioritariamente e com muito mais frequência que o DM e o Google. Como o sujeito C apresentou a menor média de consultas por busca, deduz-se a qualidade do DB utilizado na atividade de inglês seja superior à qualidade dos DBs utilizados nas atividades das demais línguas, pelo menos no que se refere à

codificação, pois dispensou consultas complementares. Uma fala que se repetiu no protocolo verbal da atividade em inglês foi: “só o bilíngüe já resolveu”, o que denuncia que o sujeito já estava condicionado a consultar o DB e depois o DM para confirmação e surpreendeu-se cada vez que essa rotina não precisou ser repetida. É importante ressaltar que esse sujeito esqueceu-se de levar um DB e acabou utilizando um DB que estava disponível na sala da atividade, mas que não conhecia: o *Longman Dicionário Escolar* (que será analisado no capítulo 5 por explicitar o apoio à codificação em sua proposta editorial).

Já a maior quantidade de buscas e de consultas dos sujeitos de pesquisa A e B, do italiano e do espanhol, respectivamente, pode ser explicada pelo fato de que tais línguas não são L2 de habilitação do curso de Tradutor em questão e, por isso, os alunos apresentam uma menor proficiência nelas do que nas línguas de habilitação, inglês ou francês. Esse mesmo fato pode também indicar que a proximidade entre o português, o espanhol e o italiano afeta os tradutores, levando-os a fazer muitas consultas para confirmar a existência de determinados itens e para descartar a suspeita de falsos cognatos.

A tabela 4 apresenta os tipos de dúvida que motivaram as consultas aos dicionários, classificados por ordem decrescente de número total de consultas.

Tabela 4 - Motivos de consulta aos dicionários, discriminando participantes e idiomas

Motivo	Inglês A	Italiano B	Espanhol C	Francês D	Francês E	Francês F	Total
procurar equivalente no DB	16	31	24	8	12	26	117
confirmar equivalente no DM	5	22	13	18	8	4	70
testar combinação no Google	8	19	1	15	17	2	62
conferir existência no DM	0	7	7	0	2	3	19
verificar ortografia	0	1	4	3	2	3	13
verificar conjugação	0	2	3	0	0	1	6
conferir plural/singular	0	0	1	0	1	2	4
conferir gênero	0	0	0	1	1	1	3
TOTAIS	29	82	53	45	43	42	294

O motivo de consulta mais freqüente foi a procura de um equivalente no DB, seguido do desejo de confirmar um equivalente no DM. Em alguns casos, o equivalente que se procurava confirmar foi obtido no DB, em outros casos os sujeitos de pesquisa possuíam um “candidato a equivalente” em seu vocabulário passivo, mas não se sentiam seguros para empregá-lo na escrita. A falta de segurança parece advir, sobretudo, do fato de que os itens lexicais são empregados em seqüências lexicais e os alunos desconhecem as seqüências recorrentes. Eles parecem possuir um “*corpus* mental” constituído de leituras e exposições a insumos em L2, visto que freqüentemente expressam uma espécie de “intuição” de que um ou outro uso não parece natural. Embora esse “*corpus* mental” não pareça ser suficiente para lhes dar confiança acerca do uso correto, ele tem o papel fundamental de lhes dar a desconfiança que deflagra as consultas. Alunos mais ingênuos possivelmente não teriam esse tipo de desconfiança e nem consultariam itens que presumem já conhecer. Isso explicaria o fato constatado por Atkins e Varantola (1997) e comentado no capítulo 3, de que alunos de nível avançado consultam mais os dicionários que os intermediários.

No eixo em que se comparam os três tradutores de francês, observam-se consultas para verificar ortografia, em especial o uso do acento agudo e o uso de consoantes duplas. Algumas consultas também foram feitas para confirmar o gênero dos substantivos, mostrando consciência sobre as desigualdades de gêneros entre português e francês. Para esses tipos de consulta, os sujeitos ora utilizavam o DB, ora o DM e, em ambos os casos, uma única consulta foi suficiente para esclarecer a dúvida. As variáveis que poderiam explicar as diferenças de comportamento entre esses sujeitos de pesquisa são: o conhecimento de recursos lexicográficos, o conhecimento de recursos computacionais, o conhecimento de mundo e preferências pessoais.

No inglês, observou-se a ausência de buscas para verificar acentos, ortografia, gênero, conjugação e suspeita de falso cognato, todas elas dúvidas comuns nas atividades relativas às línguas neolatinas.

4.5. Conclusões

A pesquisa sobre o uso de dicionários em situação de escrita em L2 permitiu verificar que todos os usuários optaram por escrever em ambiente eletrônico. Isso leva a crer que as pesquisas sobre o uso do dicionário que determinam o uso de um dicionário impresso em uma atividade de escrita em papel podem já não ser representativas do que ocorre atualmente na “era digital”. A expansão do uso dos computadores pessoais e o advento da Internet alteraram drasticamente as condições de produção de textos, incluídos aí os textos em L2. Embora eventualmente tenham utilizado algumas obras impressas, isso só foi feito na falta de recurso eletrônico ou quando o recurso eletrônico disponível não resolveu a dúvida.

Para alguns tipos de dúvidas, o DM mostrou-se perfeitamente adequado, por exemplo, verificar se um item conhecido tinha um determinado sentido. Não há razão para se pretender, portanto, o desenvolvimento de um DB que dispense totalmente o uso do DM na codificação.

Para dúvidas relativas a ortografia e concordância de gênero e número, poderiam ter sido habilitados os corretores ortográficos de L2 no editor de textos, o que evitaria muitas consultas a dicionários e muitos erros cometidos quando nem ao menos se cogita consultar o dicionário. Isso não foi feito pelos participantes, mas tanto os DBs quanto os DMs solucionaram adequadamente esses tipos de dúvida.

As dúvidas na situação de codificação parecem concentrar-se no emprego do vocabulário de alta frequência, composto em sua maioria de itens largamente polissêmicos. Esses itens provavelmente não apresentariam dificuldades em situação de decodificação, pois, apesar de não saberem empregá-los em determinados contextos, os sujeitos de pesquisa

demonstraram conhecê-los. A reflexão sobre esse fato confirma a conclusão a que se chegou no item 2.2 desta tese: a nomenclatura de um dicionário para apoio à produção em L2 deve privilegiar o vocabulário ativo em L1, pois é esse vocabulário que os usuários utilizam para se expressarem.

A análise da atividade mostrou que, na codificação, os alunos sentem necessidade de ver o item lexical equivalente em uso, para sentirem-se seguros da escolha. Por isso, os exemplos de uso são muito valorizados e, quando não se encontravam disponíveis no dicionário, foram procurados na Web por meio do Google. A possibilidade de consultar o emprego de um item lexical pode ser satisfeita por meio de exemplos no dicionário, consultas a um *corpus* de língua ou consultas com o motor de buscas na Web (tomada como se fosse um *corpus*). Nenhum aluno, porém, utilizou um *corpus* específico da L2 para a qual estava traduzindo, muito provavelmente por desconhecimento ou por falta de acessibilidade.

A maior dificuldade dos sujeitos de pesquisa foi quanto à busca de equivalentes para colocações, independentemente da L2 em questão. Foi nesse o tipo de dúvida que os sujeitos de pesquisa não se sentiram seguros de terem chegado a um resultado satisfatório. As consultas aos DBs, para essa finalidade, na maioria das vezes foram seguidas de uma consulta ao DM de L2 para confirmar o sentido e, não havendo exemplo de uso, a consulta ao DM foi seguida de consulta no Google para confirmar o uso. Os DMs não deram segurança aos usuários, pois muitas vezes apresentavam apenas a definição da colocação ou apenas um exemplo, insuficientes para confirmar se a colocação em L2 correspondia à colocação em foco na L1. O dicionário que melhor resolveu essas situações foi o DB pedagógico de língua inglesa, mas até mesmo este, em alguns casos, não apresentou a colocação procurada. Isso indica a grande relevância de levantar as colocações mais frequentes de cada palavra-entrada do DB L1-L2 e apresentar seus respectivos equivalentes em L2.

A realização da pesquisa de campo, portanto, mostrou-se frutífera para as reflexões produzidas neste trabalho. Em resumo, concluiu-se que:

- A mídia para o desenvolvimento de uma solução lexicográfica de apoio à codificação em L2 deve ser, preferencialmente, a eletrônica;
- A nomenclatura de um dicionário para apoio à codificação em L2 deve privilegiar os itens lexicais mais frequentes, aqueles que pertencem ao vocabulário ativo dos usuários em L1. Quanto mais rica a descrição sobre o uso desses itens, maior será a adequação desse tipo de dicionário ao seu propósito;
- Dificilmente uma solução lexicográfica seria suficiente para tornar totalmente dispensável o hábito apresentado pelos usuários de “testar” seqüências lexicais, por meio de uma busca na Web (tomada como *corpus*), a fim de verificar se essas seqüências são recorrentes (se existem no *corpus*), qual sua freqüência e em que contextos elas ocorrem;
- Os DBs, em maior ou menor grau, não conseguiram atender as necessidades de equivalentes para colocações;
- A maioria dos DBs carece de exemplos de uso que dêem ao usuário segurança sobre o emprego do item lexical equivalente e os DMs, muitas vezes, também não apresentam exemplos suficientes para isso.

Sendo assim, as evidências apontam que é possível obter um grande ganho de qualidade no aparato lexicográfico brasileiro de apoio à codificação em L2 se esforços forem concentrados no desenvolvimento de DBs português-L2 que apresentem, sobretudo, o vocabulário de maior freqüência do português do Brasil com suas respectivas colocações mais recorrentes e com ampla exemplificação de uso.

5. A ANÁLISE DE SOLUÇÕES

Como foi visto nos capítulos 2 e 3, muitos dicionários podem ser úteis nas atividades de codificação em L2, mesmo que não tenham sido elaborados especificamente para isso: DMs para aprendizes, DBs, dicionários de sinônimos, dicionários temáticos, dicionários de ilustrações, dentre outros.

Neste capítulo serão analisadas cinco obras lexicográficas que declaram expressamente a finalidade de dar apoio à codificação em L2, o que não se encontrou em outros dicionários. Serão salientados seus pontos fortes, que poderão ser aproveitados na proposta de uma nova solução, e seus pontos fracos, que deverão ser evitados. No item 4.1 será analisado o *Longman Language Activator* (LLA), uma obra monolíngüe, voltada exclusivamente para o uso produtivo da língua. No item 4.2 será analisado o *Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde* (DAFLES), uma obra multilíngüe que se propõe a apoiar em níveis diferentes a codificação e a decodificação da língua francesa por estrangeiros. Já no item 4.3 será analisado o protótipo do *Bilingual Dictionary Plus* (DB+), desenvolvido e testado por Laufer e Levitzky-Aviad (2006). O item 4.4 será dedicado ao *Dicionário de Aprendizagem Português Espanhol* (DAPE), da Universidade de Santa Catarina e, finalmente, no item 4.5, será analisada a seção português-inglês do *Longman Dicionário Escolar* (LDE).

5.1. Longman Language Activator (1993)

A primeira edição do LLA é de 1993, quando foi lançado exclusivamente sob a forma impressa (hoje já existe em CD-ROM). Sua confecção envolveu 15 lexicógrafos, sob a gerência de Michael Rundell, além da consultoria de vários lingüistas. No prefácio da

primeira edição encontram-se muitas informações sobre a confecção do dicionário, informações que foram suprimidas na segunda edição, de 2002.

Os editores pretendiam criar um dicionário que auxiliasse os estrangeiros a fazer a escolha lexical mais precisa para expressar a idéia pretendida e empregá-la adequadamente dentro da frase. Utilizaram um *corpus* de aprendizes para verificar quais eram as principais dificuldades dos estrangeiros em relação ao inglês. Foi assim que descobriram que os aprendizes apoiavam-se demasiadamente em itens lexicais do vocabulário básico, mais neutros e genéricos. Para melhorar a expressão em L2, no entanto, esses aprendizes precisavam conhecer formas mais precisas de dizer o que já vinham dizendo com o vocabulário básico. No entanto, não havia dicionários para orientar essas escolhas. Percebeu-se, então, a necessidade de apresentar uma série de paráfrases e de explicitar quais as sutis diferenças relacionadas ao emprego de cada uma delas.

No *corpus* de aprendizes, os editores também perceberam que os substantivos concretos, por exemplo, não apresentavam problemas e, por isso, resolveram não contemplá-los no dicionário:

Deve ser salientado, no entanto, que o *Activator* [LLA] não trata de palavras relativas a itens do ‘mundo real’, algumas das quais, é claro, também pertencem ao inglês fundamental. Acreditamos que substantivos concretos e palavras de conteúdo, em geral, apresentem menos problemas (e menos sérios) de uso correto pelos alunos. Assim, não se encontrarão aqui diferentes tipos de transporte, cães, maquinário ou construções. Isso fica a cargo do *Longman Lexicon*, que lida efetivamente com campos semânticos, incluindo itens do mundo real.⁶⁰ (DELLA SUMMERS, diretora de edição na página F8 do prefácio da 1ª edição do LLA)

O *corpus* de aprendizes serviu também para evidenciar quais as construções que os aprendizes de nível intermediário ou avançado já dominavam amplamente e que, portanto, não precisavam ser tratadas no LLA.

⁶⁰ It should be pointed out straight away, however, that the *Activator* does not address itself to words for ‘real world’ items, some of which, of course, also belong to the core of English. We believe that concrete nouns, and content words in general, present fewer, less serious problems of correct use for students, so you will not find different types of transport, dogs, machinery or buildings here. That is left to the *Longman Lexicon*, which deals effectively with semantic fields, including real world items.

5.1.1. A seleção da nomenclatura

Segundo está expresso no prefácio do LLA, o primeiro critério de seleção da nomenclatura foi estar incluído no *Longman Dictionary of Contemporary English*. Isso significa que os itens lexicais selecionados deviam ser de uso freqüente. Depois foram excluídos os substantivos concretos, as palavras de conteúdo e os itens que a análise do *corpus* de aprendizes demonstrou já serem de domínio do público-alvo (aprendizes de nível intermediário e avançado). A nomenclatura do LLA, portanto, inclui apenas itens lexicais freqüentes que apresentam dificuldade para os aprendizes e que não constituem substantivos concretos ou palavras de conteúdo.

5.1.2. A metodologia

O prefácio não explica como foi realizado o trabalho, apenas esclarece que foi utilizada uma estratégia *bottom-up* (ascendente), ou seja, foi criada uma rede semântica, ligando itens lexicais mais específicos a itens mais genéricos, chegando a 1052 palavras-chave pertencentes ao vocabulário de 2000 palavras que, supostamente, os aprendizes de nível intermediário ou avançado já dominam. Assim, consultando uma dessas palavras-chave, o aprendiz tem acesso a um verbete que especifica vários sentidos e, para cada sentido, apresenta várias paráfrases apropriadas para expressar determinada idéia. É possível também consultar diretamente itens mais específicos, pegando um “atalho” nesse percurso.

A editora informa que a equipe de lexicógrafos explorou alguns aspectos levantados pelas pesquisas em semântica lexical e psicolinguística, mas o trabalho, em si, foi mais orientado pela própria prática. O agrupamento de conceitos e subconceitos levou cerca de cinco anos. Os conceitos a que se chegou constituem as 1052 palavras-chave e sob eles se

reúnem uma série de itens lexicais e expressões semanticamente relacionados. O nome desses conceitos são palavras ou expressões familiares ao aprendiz, o que garante sua acessibilidade.

Os itens lexicais e expressões selecionados foram consultados em vários *corpora* de nativos, de língua escrita e de língua falada, tanto da variedade britânica quanto da americana. Esses itens foram analisados com a ajuda de um “sistema de análise semântica apoiado em computador, com o qual pudemos formalmente fazer distinções entre os significados de palavras semelhantes”⁶¹ (DELLA SUMMERS, diretora de edição, na página F8 do prefácio da 1ª edição do LLA).

Segundo a mesma fonte, “o processo de análise começou pela atividade de ligar palavras e expressões a seus conceitos apropriados”⁶² (DELLA SUMMERS, diretora de edição, na página F8 do prefácio da 1ª edição do LLA).

Todo o trabalho do LLA foi baseado em *corpora*:

- *corpus* de aprendizes para identificar as dificuldades dos aprendizes que deveriam inspirar as soluções;
- *corpora* de nativos para analisar as nuances de emprego e significado dos itens lexicais;
- *corpus* de língua falada para identificar a variedade oral da língua (a qual foi marcada no dicionário introduzindo-se a definição com um “*you say...*”);
- *corpus* de gêneros mais formais comparado com *corpus* de gêneros informais para identificar que itens lexicais deveriam ser marcados como “formal” ou “informal” por não ocorrerem com a mesma frequência nos dois *corpora*.

⁶¹ computer-aided semantic analysis system with which we can formally distinguish between the meanings of similar words.

⁶² The whole analysis process began with the assignment of words and phrases to their appropriate concepts.

5.1.3. A microestrutura

Os exemplos do LLA foram baseados em *corpus*, ou seja, “foram todos fortemente influenciados por evidências de *corpus*, mas nem todos eles foram tirados diretamente de *corpora*⁶³” (DELLA SUMMERS, diretora de edição, na página F10 do prefácio da 1ª edição do LLA). Vários exemplos são fornecidos para um mesmo sentido, sempre começando com um exemplo mais claro e mais pedagógico e aumentando a complexidade dos demais gradativamente.

As colocações são indicadas em negrito e a estrutura sintática, entre parênteses. As formas mais freqüentes de expressar uma determinada idéia são apresentadas antes das menos freqüentes, não importando tratar-se de palavra simples ou expressão. Aliás, embora não seja uma regra, os editores perceberam que, normalmente, no inglês as palavras mais longas são utilizadas na escrita formal e as mais curtas na escrita informal. Percebeu-se também que os itens lexicais multi-palavras, por sua vez, são mais utilizados na variedade oral do Inglês.

Poucas marcas gramaticais são utilizadas: é informada a classe gramatical de cada item ou expressão; se os substantivos são contáveis ou incontáveis; se os verbos são transitivos ou intransitivos; se o verbo não admite a forma progressiva; se um adjetivo é usado apenas antes de substantivos e se um adjetivo não pode ser utilizado antes de um substantivo.

As notas sobre restrições de uso são explicitamente apresentadas e as definições são elaboradas utilizando-se um vocabulário de definição, ou seja, um vocabulário básico de domínio dos aprendizes.

Para o objetivo a que se propôs, o LLA foi muito bem sucedido. Conhecer a proposta do dicionário é fundamental para que se faça sua análise crítica. Caso não se tivesse acesso ao prefácio da primeira edição do LLA, que explica o porquê da supressão dos substantivos

⁶³ The examples in this dictionary are all strongly influenced by corpus evidence, but not all of them are taken directly from the corpora.

concretos do dicionário, provavelmente esta tese apontaria esse fato como uma falha, como fez Bogaards (1996). O autor testou três DMs para aprendizes em uma simulação de atividade de codificação em L2 e o melhor resultado que conseguiu foi resolver 50% de suas dificuldades. Depois fez o mesmo teste com o LLA e o resultado foi ainda pior. Para o autor, “o valor desse dicionário parece, aliás, estar mais no lado do vocabulário abstrato que no lado do vocabulário concreto⁶⁴” (BOGAARDS, 1996, p. 302).

Um *thesaurus*, como a *Wordnet*⁶⁵ do português do Brasil, tem muito em comum com a proposta do LLA de agrupar o léxico pelo conceito, porém normalmente é elaborado para nativos e, por isso, não explica qual a diferença entre itens próximos nem mostra como empregá-los.

O LLA preencheu uma lacuna no conjunto de dicionários oferecidos pela Longman e foi exclusivamente projetado para as necessidades de codificação em L2, o que não significa, no entanto, que seu uso exclusivo possa resolver todas as necessidades de codificação. É freqüente, no prefácio, os autores dizerem que para esse ou aquele tipo de dúvida um ou outro dicionário seria mais apropriado.

Embora o LLA tenha uma nomenclatura relativamente restrita, possui a mais completa descrição lexical para fins de codificação se comparado aos demais dicionários. Contudo, explorá-lo demanda tempo, mas esse tempo não é gasto em vão, pois a forma como os itens estão relacionados pode facilitar a aquisição, já que “imita” a forma de organização do léxico mental.

O LLA declarou-se “o primeiro dicionário de produção do mundo” e com certeza inspira todos aqueles que buscam soluções lexicográficas para apoio à codificação. É o caso, por exemplo, de Binon e Verlinde (2000a), que citam o LLA como única obra genuinamente voltada às necessidades de codificação naquela época.

⁶⁴ The value of this dictionary seems indeed to be more on the side of abstract than of concrete vocabulary.

⁶⁵ <http://www.nilc.icmc.usp.br/nilc/projects/wordnetbr.htm>

No entanto, o fato de não ser bilíngüe faz com que o LLA tenha duas grandes limitações: ignora as diferenças contrastivas entre a L1 do usuário e a L2 em questão, e impossibilita o acesso via L1.

Para superar essas limitações, Binon, Selva e Verlinde (2002) desenvolveram dicionários de acesso multilíngüe à língua francesa, cuja proposta é apoiar simultaneamente a codificação e a decodificação, como é o caso da obra analisada a seguir.

5.2. Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde (2000)⁶⁶

O DAFLES é um dicionário eletrônico disponível gratuitamente para consultas na Web, a partir de seis idiomas: francês, inglês, alemão, espanhol, italiano e holandês, sendo que as instruções e a metalinguagem encontram-se em francês. A estrutura do DAFLES está pronta, mas como o sítio ainda estava em construção, muitos recursos encontravam-se inoperantes quando se procedeu à sua consulta. Por não ser o português uma das línguas de acesso, foram realizadas algumas consultas aleatórias em inglês. Após seis tentativas com palavras freqüentes (*forbidden, point out, make, do, try, leg*), obteve-se o primeiro resultado (as demais consultas retornavam, em francês, a mensagem “palavra tal não está no dicionário”). Isso mostra que o acesso via L2 ainda está em fase de implementação, porém o conteúdo em francês já está acessível para consultas a partir de palavras-chave em francês.

O resultado da consulta, a partir de uma palavra-chave em L1, é um ou mais equivalentes em francês com *hiperlink*. Por exemplo, consultando a palavra-chave *leg*, a partir da janela de consulta de língua inglesa, obtém-se: *leg, jambe* e *leg, étape*. Assim, clicando no equivalente, tem-se acesso a outra tela onde é apresentada uma lista dos sentidos. Cada sentido abre a opção de consultar as colocações, classificadas por verbo/tipo/modo/outro,

⁶⁶ <http://www.kuleuven.be/dafles/acces.php?id=null>

como segue (os sublinhados são *hiperlinks*), na Figura 5, resultado de um item lexical consultado diretamente em francês:

Figura 5 - Resultado de consulta ao DAFLES (1)

<p>retraite Sens 1a: période après la vie professionnelle => + verbe, type/façon de, autre Sens 1b: somme d'argent => type/façon de, autre Sens 2a: fait qu'une armée quitte un endroit => + verbe Sens 2b: résultat du retrait d'une armée Toutes les collocations de retraite</p>

No DAFLES, é possível também consultar as colocações formadas com o item lexical consultado. Uma barra inferior oferece acesso a:

- definições (curtas ou longas, à escolha do usuário);
- sinônimos e antônimos;
- derivados e actantes;
- colocações e funções lexicais;
- ajuda.

A exibição do conteúdo em francês com metalinguagem em francês torna esse dicionário um DM, exceto pela possibilidade de acessá-lo a partir de palavras-chave em outros idiomas. Há também um *link* (ainda não ativado) com a etiqueta “Trad.” que, presume-se, seja para a tradução dos exemplos.

Para um dicionário pedagógico, o DAFLES apresenta muitas abreviações, o poderia dificultar a interação do usuário com o material, principalmente porque são abreviações de palavras em francês. A forma de apresentação dos resultados talvez exija algum treinamento do usuário, já que pressupõe conhecimento do que sejam sinônimos, antônimos, derivados, colocações, actantes e funções lexicais.

A seguir encontra-se um quadro padrão do resultado da consulta a derivados e actantes para o item lexical *retirer*, que parece ser um tanto complexa para aprendizes:

Figura 6 - Resultado de consulta ao DAFLES (2)

Mots utilisés avec **retirer** :

		action	N1:sujet	verbe	N2	prép N3	Ntemps	Nrésultat	Nmoyen
1a	faire disparaître quelque chose	<u>le retrait</u>	qqn	<u>retirer</u>	qqch.				
1b	priver une personne de quelque chose	<u>le retrait</u>	qqn	<u>retirer</u>	qqch.	qqn			
2	quitter une fonction		qqn => <u>un retraité</u>	<u>se retirer</u>			<u>l'âge de la retraite</u>	<u>la retraite</u>	
3a	s'en aller vers un autre endroit	<u>la retraite</u>	qqn	<u>se retirer</u>		qqpart		<u>une retraite</u>	
3b	retrouver un niveau moins élevé		<u>un gaz</u> <u>un liquide</u>	<u>retirer</u>		qqpart			
3c	éloigner d'un endroit		qqn	<u>retirer</u>	qqch.	qqpart			
4	prendre une somme d'un compte	<u>le retrait</u>	qqn	<u>retirer</u>	<u>une guerre</u> <u>sans merci</u>	<u>un compte</u>		un débit	<u>une carte bancaire</u>
5	décider de ne plus participer	<u>le retrait</u>	qqn <u>un concurrent</u> <u>une équipe</u>	<u>se retirer</u>		<u>une compétition</u> <u>une épreuve</u>		<u>un abandon</u>	

O DAFLES apresenta uma excelente análise do léxico francês, principalmente no que concerne a colocações e sentidos dos itens lexicais. Percebe-se que há muitas informações a serem exibidas e agora o desafio é como fazer isso de forma clara e didática para que o material desenvolvido seja efetivamente aproveitado pelos aprendizes estrangeiros. Produzir conteúdo e exibir conteúdo são, como mostra esse dicionário, etapas diferentes de um processo. A forma de exibição de um dicionário eletrônico é diferente daquela de um dicionário impresso e talvez exija, dos desenvolvedores, competência didática especializada em ambiente digital.

5.3. Bilingual Dictionary Plus (2006)

Um protótipo do dicionário proposto por Laufer (1995) foi testado por Laufer e Levitzky-Aviad (2006) e, segundo avaliação das autoras, apresentou resultados superiores aos demais tipos de dicionários em atividade de produção em L2. Esse dicionário tem os quatro elementos sugeridos por Laufer (1995) para um DB de apoio à codificação em L2:

1. Entrada em L1 e equivalentes de tradução em L2, sendo fornecidos equivalentes para todos os sentidos da palavra-entrada;
2. Informações fonológicas, semânticas, gramaticais, definição e exemplo de uso de cada equivalente fornecido, tudo em L2. As autoras dizem que essas informações são idênticas às apresentadas nos dicionários bilingualizados, (como o *Password* inglês-português);
3. Palavras semanticamente relacionadas aos equivalentes fornecidos em L2 (apresentadas com suas respectivas definições e traduções para L1);
4. Sentidos adicionais dos equivalentes fornecidos em L2.

Na verdade, o DB+ foca mais a forma em L1 do que o significado em L1. Os sentidos do item lexical de L1 são indicados pelos diferentes equivalentes de tradução em L2. São apresentados itens lexicais semanticamente relacionados a cada um dos sentidos apresentados em L2, bem como outros sentidos dos equivalentes de tradução em L2. Isso torna as informações (3) e (4) um pouco confusas se o usuário não prestar atenção. A avaliação dos resultados da pesquisa, aliás, revelou que as informações (3) e (4) foram pouco utilizadas pelos usuários, o que surpreendeu as autoras.

O exemplo de verbete apresentado por Laufer e Levitzky-Aviad (2006) é para hebraico-inglês, contudo, para ilustrar melhor essa proposta, decidiu-se adaptar o exemplo para português-inglês:

Figura 7 - Modelo de verbete do DB+

<p>REVELAR</p> <p>1. develop (...) Have you sent the films to be developed yet?</p> <p>2. reveal (...) All their secrets have been revealed.</p> <p>RELATED WORDS (to 2.):</p> <p>disclose to reveal or make known <i>She refused to disclose his identity.</i></p> <p>expose to discover and make known <i>It was a newspaper that exposed his spying activities.</i></p> <p>unfold to (cause to) be revealed or become known <i>She gradually unfolded her plan to them.</i></p> <p>OTHER MEANINGS OF</p> <p>expose to leave unprotected from (eg weather, danger, observation) Paintings should not be exposed to direct sunlight. Don't expose children to danger. (expor)</p> <p>☞ by releasing the camera shutter, to allow the light to fall on (a photographic film) (expor)</p> <p>unfold:</p> <p>☞ to open and spread out (a map, etc.): He sat down and unfolded his newspaper (abrir, desdobrar).</p>
--

De forma similar ao DAFLES, no DB+ a L1 é utilizada apenas como um acesso à parte monolíngüe do dicionário. O restante das informações é todo composto em L2.

5.4. Dicionário de Aprendizagem Português-Espanhol (2007)

A primeira iniciativa de um dicionário para apoio à codificação em L2 no Brasil foi feita por Philippe Humblé em seqüência a sua tese de doutorado sobre o assunto, em 1996. O *Dicionário de aprendizagem português-espanhol* (DAPE), da Universidade Federal de Santa Catarina, foi concebido sob o formato eletrônico e reúne uma série de exemplos bilíngües. Segundo Humblé (2002), os exemplos foram retirados, em sua maioria, do *corpus* Varsul⁶⁷. Por serem exemplos de língua falada, sofreram algumas adaptações para a linguagem escrita,

⁶⁷ O Projeto VARSUL (Variação Lingüística Urbana na Região Sul) foi constituído oficialmente em 1990 e visa à instalação de um Banco de Dados lingüísticos a partir da documentação do português falado nas áreas urbanas lingüisticamente representativas dos estados do Paraná, Santa Catarina e Rio Grande do Sul. (<http://www.cce.ufsc.br/~varsul/objetivo.htm>)

sem prejuízo do sentido e, só então, foram traduzidos para o espanhol. As consultas podem ser feitas por palavras-chave tanto em espanhol quanto em português.

Nesse dicionário eletrônico, ao digitar uma palavra-chave para busca, o usuário obtém uma relação de exemplos que a contêm. Um mesmo exemplo pode servir para ilustrar várias palavras-chave, ou seja, um exemplo não é exclusivo de uma entrada, como ocorre nos dicionários impressos. Como o dicionário é constituído exclusivamente por exemplos bilíngües, a simetria permite que tanto o público nativo da L1 quanto o público nativo da L2 possam utilizá-lo na mesma medida. Esse fato está expresso no próprio dicionário:

Este dicionário serve para ajudar alunos lusofalantes a produzirem espanhol e ajudar alunos que falam espanhol a produzirem português. Ele consiste em mais de 25.000 frases autênticas de português brasileiro, eventualmente modificadas, traduzidas para o espanhol. Ocasionalmente foi feito o contrário: frases em espanhol foram traduzidas para o português.⁶⁸

Embora represente um grande avanço, o DAPE não contempla informações explícitas sobre propriedades do léxico, cuja importância é defendida por vários autores citados no capítulo 2. Há de se reconhecer, contudo, que muitas dessas propriedades podem ser inferidas a partir dos exemplos.

O conteúdo do dicionário é, na verdade, não um conjunto de verbetes, mas um *corpus* paralelo português-espanhol de frases isoladas. Nada garante, contudo, que um *corpus* de 25.000 frases contenha ocorrências de todos os itens lexicais mais freqüentes e muito menos que contemple a maioria dos sentidos desses itens lexicais. Uma consulta aleatória mostrou, por exemplo, que itens lexicais como *realçar*, *ênfaticar*, *listar*, *brindar*, *estimar*, *internar* não apresentam nenhuma ocorrência. Por outro lado, ao se consultar *jeito*, obteve-se 98 exemplos,

A utilidade do DAPE parece ser grande, já que tem um grande volume de acessos, segundo informação verbal do próprio Humblé (o dicionário gera estatísticas de acesso para

⁶⁸ <http://www.dicionario-online.cce.ufsc.br/descricao.php>

seus administradores). A simplicidade da consulta, a forma clara com que os resultados são apresentados e o fato de estar gratuitamente disponível na Web são os pontos fortes da obra.

Entretanto, como um mesmo exemplo pode ser exibido a partir de vários argumentos de consulta, não é possível estabelecer um filtro para apresentar apenas um exemplo para cada sentido dos itens lexicais. O resultado disso é que uma consulta pode apresentar diversos exemplos em que a palavra-chave tem sentidos iguais ou muito próximos. No caso de “jeito”, por exemplo, o usuário talvez tivesse que ler os 98 exemplos até encontrar uma equivalência para o sentido pretendido. Ao que parece, também não há controle de que a grande maioria dos sentidos de cada item lexical seja exemplificada.

Pelo que se observou, como dicionário, faltou para o DAPE definir uma nomenclatura. A partir disso teria sido possível listar os sentidos de cada item da nomenclatura e então coletar exemplos para ilustrá-los, fazendo um caminho inverso ao que foi feito. Nada impede, contudo, que isso seja feito durante o aperfeiçoamento da obra. Outra característica desejável seria agrupar exemplos semanticamente relacionados, pois a opção de incluir semelhantes, na tela de consulta, contempla apenas a forma (consultando-se *suar*, obtêm-se também exemplos contendo *usuário*; consultando *pintar*, exemplos com *carpintaria* etc.).

O formato de exemplos bilíngües, contudo, é excelente para dar segurança ao usuário sobre o sentido dos itens lexicais empregados. Projetos semelhantes poderiam, portanto, seguir um percurso diferente para superar as pequenas limitações já citadas:

- selecionar uma nomenclatura que reflita o léxico ativo da língua (critérios para isso são um grande desafio);
- enumerar os sentidos de cada item da nomenclatura;
- coletar exemplos de uso que ilustrem cada sentido dos itens da nomenclatura em português;
- elaborar traduções de cada exemplo para a L2.

Além disso, sempre que determinado sentido em português pudesse ser expresso por mais de uma forma em L2, as paráfrases deveriam ser fornecidas. Nos casos em que um mesmo sentido tivesse que ser codificado de diferentes formas em L2, respeitando algum tipo de restrição, poderia ser acrescentada uma nota de uso explicando isso.

Seria importante, sobretudo, que projetos semelhantes aproveitassem o visual proposto pelo DAPE, não “poluindo” a tela de consulta e de exibição dos resultados.

5.5. Longman Dicionário Escolar (2002)

O LDE é um dicionário inglês-português e português-inglês dirigido especificamente a estudantes brasileiros. Ele está disponível sob forma impressa e em CD-ROM. Seu prefácio é apresentado nos dois idiomas e termina com a frase: “O velho dicionário bilíngüe está morto. Que viva o novo!”.

A proposta do dicionário observa muito do que vem sendo pregado dentro da Lexicografia Pedagógica: uso reduzido de abreviaturas, abolição do uso do til “~” para substituir a repetição da palavra-entrada, o uso de cores (azul, vermelho e preto), tipos (negrito, itálico) e fontes para separar e destacar as informações da microestrutura etc.

As críticas que motivaram esse “novo dicionário bilíngüe” estão expressas no prefácio e, dentre elas, uma que diz respeito às necessidades de codificação:

A principal crítica, porém, é que tais dicionários não são confiáveis especialmente para os alunos que, desejando expressar-se em inglês utilizando palavras novas, ou que conheçam apenas parcialmente, consultam-nos em busca de palavras em inglês equivalentes às de seu idioma. (JEREMY HARMER, no prefácio do LDE, 2002)

Atender as necessidades de codificação, portanto, é um dos principais objetivos do LDE:

...o que torna esse dicionário significativamente diferente de outros DBs é que é tão útil para os alunos que queiram encontrar o significado de uma palavra em inglês quanto para aqueles que queiram encontrar o equivalente em inglês de uma palavra em português. (JEREMY HARMER, no prefácio do LDE, 2002)

Analisando-se a seção português-inglês da obra, constata-se que o objetivo foi contemplado adequadamente, porém de forma parcial. Alguns itens lexicais são cuidadosamente descritos e exemplificados, agregando notas de uso de grande valor para quem necessita codificar. Mas outros itens lexicais, que também requerem esse tratamento, não apresentam grandes avanços em relação aos “velhos” DBs. Dentro da obra, portanto, coexistem verbetes muito bem elaborados, verbetes parcialmente melhorados e verbetes do tipo *entrada* → *equivalente*. Faltou, pois, aplicar exaustivamente a fórmula que encontraram para verbetes realmente úteis na codificação.

Da mesma forma que o LLA, o LDE utilizou *corpus* de aprendizes para descobrir e tratar as dificuldades do público-alvo (no caso, um *corpus* de textos em inglês produzidos por estudantes brasileiros). Para formar seu *corpus* de aprendizes, a Longman tem uma política de premiar, com dicionários, professores que enviam as redações em inglês de seus alunos⁶⁹.

Segundo o prefácio do dicionário, os exemplos que ilustram os verbetes na seção português-inglês foram extraídos de um *corpus* de português do Brasil, de língua falada e escrita, provavelmente formado pela própria Longman, pois essa é sua política internacional. Esse diferencial é marcante quando se compara os exemplos do LDE com exemplos fornecidos em outros DBs de linha pedagógica editados no Brasil.

Todas as explicações do LDE são feitas em português do Brasil e com um vocabulário adequado aos adolescentes (público-alvo desse dicionário). Essas características, segundo o editor, têm o propósito de fazer com que o estudante brasileiro sinta que o dicionário “fala” com ele.

Os exemplos e as colocações no LDE são apresentados em português e inglês.

Entender v 1 (compreender) to understand: *Não entendo alemão.* I don't understand German. / *Você entendeu?* Did you understand? / **entender mal** to misunderstand 2 **entender de algo** to know about sth 3 **dar a entender que** to make it clear that : *Ele deu a entender que estava a fim dela* He made it clear that he was interested in her.

⁶⁹ <http://www.pearsonlongman.com/dictionaries/corpus/learners.html>

entender-se v **entender-se (com alguém)** (a) (dar-se bem) to get along (with sb): *Nunca me entendi bem com ele.* I've never gotten along well with him. (b) (conciliar-se) to make up (with sb): *Ela acabou se entendendo com o namorado.* In the end she made up with her boyfriend.

As informações gramaticais são dadas de forma clara e simples, como, por exemplo:

jamais adv never : Ele jamais mentiria. *He would never lie* Precedido de outra palavra negativa, a tradução correta é **ever**: Ninguém jamais o viu. *Nobody has ever seen him.* (LDE, 2002)

O LDE, portanto, é um avanço em relação aos DBs tradicionais, principalmente no que se refere a necessidades de codificação. A simplicidade da metalinguagem, em português; a notas tratando de problemas típicos dos brasileiros em relação à L2; a exploração dos sentidos e colocações do item lexical em L1 e não do equivalente em L2; colocações e exemplos em L1 e L2, são todas características desejáveis em uma obra de apoio à codificação em L2.

5.6. Conclusões

A análise desses dicionários, somada às demais evidências de pesquisas, mostrou, conforme já constatado, que é ingenuamente pretensioso ter como objetivo projetar um dicionário que seja suficiente para todas as necessidades de codificação em L2.

No entanto, como a maior parte das consultas a dicionários durante a codificação em L2 é motivada pela busca de equivalentes, parece ser possível projetar obras que atendam essa necessidade mais adequadamente do que foi feito até agora, aproveitando todos os pontos fortes das obras analisadas e superando seus pontos fracos.

O melhor a ser aproveitado do LLA é sua exaustiva descrição dos itens da nomenclatura, apresentando-os de forma semanticamente relacionada. A rede semântica, porém, é estabelecida do ponto de vista do léxico da língua inglesa e nem sempre isso contempla as necessidades dos estrangeiros. Caso o usuário desejasse codificar em inglês “entender-se com alguém”, não encontraria no LLA possíveis equivalentes para o verbo

entender dentre os sentidos de *understand*. Da mesma forma, o verbo *apontar*, de “apontaram os responsáveis pelo acidente”, não possui nenhum equivalente dentre os sentidos de *show* e *point*.

A possibilidade de acesso, via L1, ao conteúdo em L2, conforme as propostas do DAFLES e do DB+ também não parece ser totalmente satisfatória, posto que a multiplicidade de sentidos e colocações da L1 não é contemplada e nem sempre a definição e os exemplos em L2 são suficientes para dar segurança ao usuário no momento da escolha. A solução apresentada pelo DAPE e pelo DPE mostra-se mais adequada, pois o ponto de vista do falante da L1 é valorizado, o que é fundamental, já que é desse ponto de vista que as idéias serão codificadas em L2. Além disso, exemplos baseados em *corpora* do português do Brasil, como os do DAPE e do DPE, parecem ser essenciais para dar ao usuário segurança sobre o sentido a ser escolhido, mesmo que não sejam todos exemplos completos (alguns casos, no DPE, são glosas).

O que ocorre no LLA ocorre também no DAFLES e no DB+: o usuário tem acesso à polissemia do item lexical de L2. Já no DAPE e no DPE, o usuário tem acesso à polissemia do item lexical de L1. Desses dois dicionários, o DAPE é o único que dá acesso às diversas formas de codificar uma mesma idéia em L2, mesmo que isso não se dê de forma sistemática e exija certa capacidade de inferência do usuário. O DPE, por sua vez, apresenta apenas uma possibilidade de tradução para cada sentido ilustrado, mas tem a vantagem de possuir uma nomenclatura criteriosamente selecionada.

O uso de *corpus* de aprendizes, utilizado pelo LLA e pelo DPE parece ser um recurso muito bom para orientar a produção de notas de uso.

Quanto à mídia das obras, parece evidente a superioridade da mídia eletrônica, visto que o conteúdo desejável para codificação é muito extenso e consultas que permitam especificar o tipo de informação procurado poderia economizar o tempo do usuário.

Finalmente, as informações para codificação fazem uso de muitas referências cruzadas, o que é facilmente implementado com os *hiperlinks* na mídia eletrônica, sem exigir esforço de consulta do usuário.

Concluindo o percurso desta pesquisa, no próximo capítulo será apresentada uma proposta para desenvolvimento de soluções lexicográficas para codificação em L2.

6. FERRAMENTAS DE SUPORTE PARA CODIFICAÇÃO EM LÍNGUAS ESTRANGEIRAS

A idéia inicial deste trabalho era propor os fundamentos para o desenvolvimento de uma única obra lexicográfica que atendesse todas as necessidades dos usuários em situação de codificação em L2 e que substituísse, com vantagens, as obras atualmente utilizadas. Essa expectativa, no entanto, não se confirmou, pois se percebeu que diferentes dificuldades pediam soluções de diferentes naturezas e algumas das dúvidas freqüentes já eram adequadamente atendidas por determinadas obras existentes. Uma obra única seria demasiadamente complexa e representaria a perda da oportunidade de focar esforços no desenvolvimento de recursos para problemas que ainda não contam com soluções satisfatórias. Assim, de uma idéia de inovação, com substituição do que já existia, partiu-se para a idéia de complementação dos recursos existentes, com reaproveitamento de informações já levantadas e recursos já desenvolvidos.

O percurso de pesquisa, portanto, desconstruiu parcialmente a premissa do trabalho, ou seja, de que os dicionários não atendem adequadamente as necessidades de codificação.

Na verdade, há dicionários que atendem algumas das necessidades de quem codifica em L2. No que diz respeito a informações sobre o léxico e sobre o comportamento do léxico em L2, muitas línguas já possuem DMs feitos para estrangeiros. Contudo, no que se refere à necessidade de busca de equivalentes a partir do léxico de L1, os DBs ainda não são suficientes. Essa afirmação pode se aplicar aos DBs brasileiros e, acredita-se, também aos DBs editados em outros países, pelo que se conhece do mercado editorial por meio das revistas dedicadas à Lexicografia.

Ao mesmo tempo, percebeu-se que DBs e DMs não são intersubstituíveis no suporte à codificação, mas sim complementares, juntamente com outros recursos (*corpora* e corretores automáticos, por exemplo).

Além disso, constatou-se que, dentre os problemas de quem codifica em L2, estão tanto a falta de informações quanto a falta de acesso fácil às informações. As soluções lexicográficas úteis para a codificação em L2 estão disponíveis em lugares diferentes e suportes diferentes (papel, CD-ROM, rede), exigindo muito esforço do usuário para serem consultadas de forma combinada durante uma mesma atividade. Essa situação agrava-se à medida que aumenta a adesão das pessoas à escrita em mídia eletrônica, o que parece ser acompanhado de uma resistência à utilização de dicionários impressos durante esse tipo de atividade.

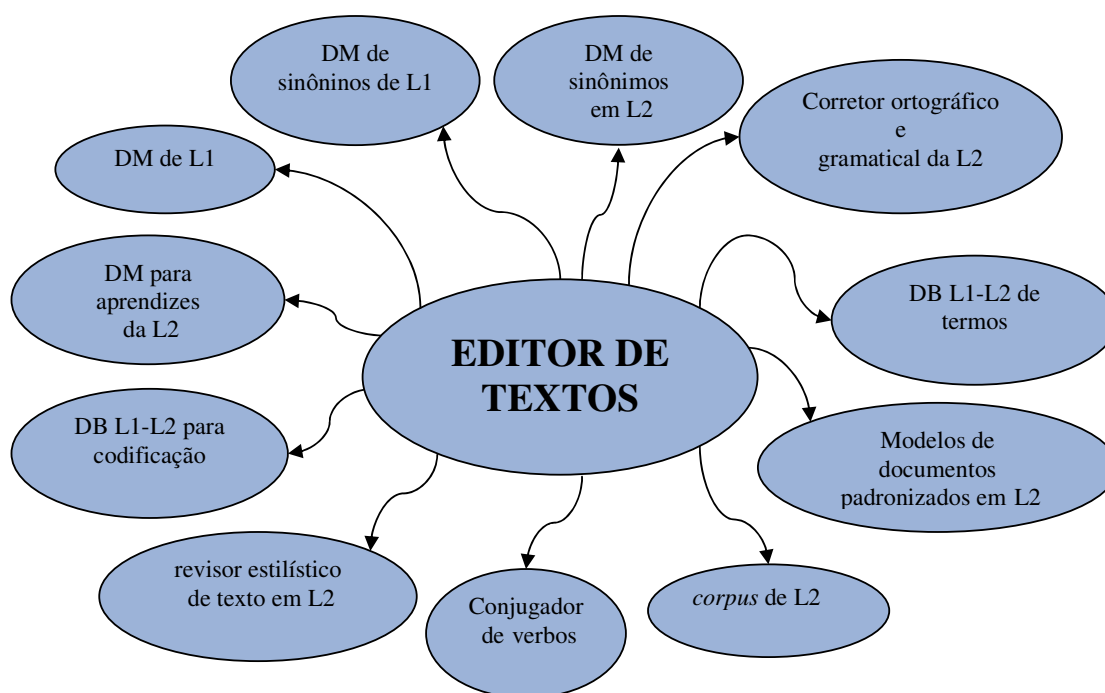
Visto dessa forma, percebeu-se que o problema poderia ser abordado por uma estratégia de constituição de um conjunto de módulos destinados a atender necessidades específicas de quem codifica em L2, dentre os quais dicionários bilíngües L1-L2.

Tais módulos poderiam ser, além disso, reunidos em um mesmo ambiente computacional, facilitando o acesso dos usuários a todos eles durante uma atividade de codificação escrita em L2. Dessa forma, o usuário poderia habilitar os recursos que julgasse convenientes. Isso eliminaria a necessidade de abrir várias sessões simultâneas de tarefas no computador, assim como a necessidade de consultas a obras impressas.

Alguns módulos reaproveitariam soluções que já atendem algumas das necessidades de escrita em L2. Essas soluções, existentes para alguns idiomas, poderiam ser tomadas como modelo pelas línguas que ainda não possuem soluções similares. Para isso, seria preciso definir a forma e o conteúdo dessas soluções, a fim de orientar novos desenvolvimentos dependentes de língua.

Em resumo, a idéia seria reunir, em um mesmo ambiente virtual, um conjunto de recursos para escrita em L2, como:

Figura 8 - Recursos de um ambiente para escrita em L2



Esses módulos são dependentes de língua (talvez com parcial exceção do editor de textos) e, por isso, aquelas que já apresentam avanços no Processamento de Línguas Naturais (PLN) exigiriam um esforço relativamente menor para serem implementadas em um ambiente computacional dedicado à codificação escrita em L2.

O ideal seria desenvolver um ambiente para cada público de L1. Assim, partindo-se do português do Brasil, por exemplo, apresentar-se-iam recursos para escrita em vários idiomas que atendessem a demanda nacional de comunicação com o exterior.

O esforço exigido para definição e desenvolvimento de um ambiente dessa natureza é multidisciplinar e multinacional. Por isso, limita-se aqui a exemplificar em linhas gerais o que se pretende sugerir com um ambiente desse tipo. O par de línguas tomado para exemplificação é o português-inglês. O português por ser a língua materna do público

brasileiro, que se prioriza atender com o resultado deste estudo; o inglês por ser a língua que apresenta maior desenvolvimento de recursos de apoio à escrita em L2, uma vez que é considerada uma língua franca.

Como já mencionado, essa abordagem do problema envolveria tanto desenvolvimento quanto reaproveitamento de recursos. Alguns deles são destinados à consulta do usuário que precisa fazer uma escolha lexical: os dicionários, os *corpora*, os conjugadores e os modelos de documentos por gênero textual. Esses recursos, chamados aqui de *recursos pré-escrita*, serão discutidos no item 6.1.

Por outro lado, há recursos para escrita em L2 que já são automatizados, como o corretor ortográfico/gramatical e o revisor estilístico. São recursos que verificam aspectos de adequação do texto já escrito e, por isso, serão chamados aqui de *recursos pós-escrita* e discutidos separadamente no item 6.2.

6.1. Recursos pré-escrita

Uma das evidências apontadas pelas pesquisas sobre o uso do dicionário na codificação em L2 é a de que um mesmo usuário pode optar por diferentes estratégias para resolver um mesmo tipo de problema. Essas estratégias, contudo, envolvem basicamente três tipos de recursos: consulta a DMs, a DBs e a *corpora*. Sendo assim, o ideal é que os três tipos estejam disponíveis em um ambiente computacional para codificação escrita em L2. Muitas vezes o usuário quer consultar diretamente um item lexical de L2 e, para isso, pode utilizar o DM. Outras vezes deseja ter uma ampla exemplificação do uso de um item em L2 e, para isso, consulta um *corpus*. Na maior parte das vezes, porém, conforme levantado nos capítulos 3 e 4, o usuário necessita saber como determinado sentido de um item lexical de L1 pode ser codificado em L2 e, para isso, necessita de um DB L1-L2 adequado.

Os recursos pré-escrita para busca de equivalentes deveriam contemplar tanto as consultas relativas a itens lexicais da língua geral quanto a termos. A busca de termos em L2 é uma das grandes dificuldades dos tradutores. Dicionários de termos técnico-científicos nem sempre se encontram disponíveis, o que obriga esses profissionais a utilizar estratégias inusitadas na tentativa de encontrar os equivalentes adequados, muitas vezes sem sucesso. Contudo, há muitos levantamentos de termos já realizados e que se encontram distribuídos, com acesso livre ou restrito.

Na União Européia, um banco de termos multilíngüe foi construído para reunir terminologias de diversas áreas, com equivalentes em vários idiomas: o IATE⁷⁰ (*Inter Active Terminology for Europe*). O acesso a um banco dessa natureza, que contemplasse também o português do Brasil, seria um bom recurso a ser incorporado num ambiente computacional de apoio à codificação escrita em L2. Embora se reconheça essa necessidade, a definição de um módulo de busca de equivalentes terminológicos não será tratada neste trabalho, pois se entende que isso requer o conhecimento de especialistas em terminologia.

Quanto aos *corpora*, tanto de língua geral quando de áreas de especialidade, são recursos que vêm sendo criados para diversas finalidades em várias línguas. Tornar disponível o acesso a *corpora* já consagrados seria um grande diferencial para o ambiente computadorizado aqui discutido, pois nem sempre os usuários conseguem fazê-lo, ou por falta de conhecimento ou por não terem permissão para isso (o acesso não é livre).

Em se optando, porém, por desenvolver *corpora* específicos para o ambiente, especialistas teriam que ser consultados para desenhar a estrutura e a composição dos mesmos, já que isso impactaria diretamente na qualidade e na confiabilidade das consultas neles baseadas.

⁷⁰ disponível no endereço: <http://iate.europa.eu/>.

Além dos DMs, DBs e *corpora*, os recursos pré-escrita poderiam contemplar também dicionários de sinônimos em L2, modelos de documentos e conjugadores de verbos para as línguas a que se aplicam.

Outro recurso já desenvolvido no Brasil e relevante para esse contexto é o *Amadeus*⁷¹, do Núcleo Interinstitucional de Lingüística Computacional, cujo propósito é apoiar estrangeiros na atividade de escrita no gênero acadêmico em inglês. Esse aplicativo atende a necessidade de aquisição de gêneros textuais, já apontada no capítulo 1 (v. TARDY, 2006).

Dentro do contexto dos recursos pré-escrita, DBs de L1-L2 que forneçam opções de equivalentes e elucidem seu emprego ainda não contam com modelos a serem seguidos. Por isso, é oportuno desenvolver um modelo que possa servir de base para diversos pares de línguas, como, por exemplo, português-inglês, português-francês, português-espanhol etc.

Reiterando, defende-se aqui a idéia de que esses DBs sejam desenvolvidos em mídia eletrônica e constituam módulos integrados a um ambiente computacional dedicado à escrita em L2. Uma desvantagem comumente associada a essa mídia, do ponto de vista financeiro, é a dificuldade de inibir cópias não-autorizadas, o que talvez possa desestimular investimentos privados.

A mídia eletrônica, no entanto, apresenta possibilidades de projeto que seriam inimagináveis na mídia impressa, além de uma série de facilidades para o usuário. As possibilidades de relacionar as palavras-chave (referências cruzadas) também são mais sofisticadas do que nas obras impressas, podendo ampliar os resultados das consultas sem exigir mais do armazenamento de informações.

Considerando o tempo de maturação de um projeto lexicográfico de grande porte, torna-se imperativo desenvolverem-se as novas soluções nessa mídia. Dessa forma, poderiam ser entregues versões incompletas do dicionário, tornando novas versões disponíveis à medida que avança seu desenvolvimento, o que seria inviável na mídia impressa. Isso seria uma

⁷¹ www.nilc.icmc.usp.br/nilc/projects/amadeus.htm

vantagem, pois é muito grande o esforço demandado para a elaboração de uma obra lexicográfica bilíngüe com o nível de detalhamento requerido para o apoio à codificação e, assim, os usuários poderiam se beneficiar dos primeiros resultados num menor prazo. Evitar-se-ia, assim, ter de esperar o final da obra para divulgá-la, mesmo porque a idéia de “final da obra” talvez seja uma utopia, já que um dicionário sempre pode ser melhorado e ampliado.

A fim de explicar a solução concebida para o desenvolvimento de DBs L1-L2 de apoio à codificação escrita em L2, estabeleceu-se uma série de definições, separadas em dois blocos. No primeiro (6.1.1), contemplam-se as características desses DBs; no segundo (6.1.2), contempla-se a forma de desenvolvê-los.

6.1.1. Características do dicionário bilíngüe para apoio à codificação

O objetivo do DB L1-L2 para codificação aqui definido é atender o usuário que deseje buscar um equivalente em L2 que comunique determinado sentido de um item lexical de L1. Não tem o objetivo de substituir as funções dos demais módulos já citados para um ambiente computacional para escrita em L2, porém tem a pretensão de evitar a necessidade de consultas subseqüentes. Isso significa que, uma vez que o usuário opte pela busca de um equivalente, deverá obter uma resposta confiável consultando apenas esse dicionário.

No dicionário aqui idealizado, a busca por equivalentes será iniciada com a digitação de uma palavra-chave em L1. Como nos dicionários eletrônicos mais modernos, uma vez digitada a palavra-chave, sem acentuação, o aplicativo deverá retornar possibilidades com e sem acento, se houver. Se a busca não apresentar nenhum resultado, a hipótese de erro de digitação deverá ser considerada e possibilidades de palavras próximas à digitada serão oferecidas.

Além disso, se a palavra-chave tiver mais de um sentido possível, a primeira etapa da consulta será a escolha do sentido do item lexical em L1 para o qual se procura equivalente

em L2. Nessa etapa, será apresentada uma primeira tela de resultados, contendo elementos indicadores de cada sentido possível, para que o usuário proceda à desambigüização, assinalando uma das opções de sentidos.

Se a palavra-chave pertencer a mais de uma categoria gramatical, poderão ser apresentadas as possíveis opções para assinalar. Se a palavra-chave for tanto um termo quanto um item do léxico geral, será apresentada a possibilidade de assinalar *termo* e redirecionar a consulta para o módulo de busca de equivalentes terminológicos (essa opção dependeria do desenvolvimento do módulo terminológico e de se estabelecer um relacionamento entre ele e o do DB L1-L2).

Em um DB tradicional, a desambigüização às vezes não é feita ou então é feita apenas em parte, simultaneamente com a apresentação de equivalentes. Em um DM para aprendizes, a desambigüização costuma ser feita em relação à polissemia do item lexical de L2, o que não corresponde à polissemia do item lexical na L1. Portanto, o ideal, para a atividade de codificação, seria desambigüizar os sentidos em L1, a fim de focar o conteúdo que é de interesse imediato na busca de equivalentes, e é nesse aspecto que esta proposta inova. Isso aumentaria a objetividade da busca e pouparia o tempo do usuário. Por exemplo, se o usuário precisasse dizer em inglês *tomar conta das crianças* e decidisse consultar o verbo *tomar*⁷², seria apresentado às informações da Tela 1.

No mesmo exemplo, se o usuário utilizasse como palavra-chave *tomar conta de*, iria direto para esse subconjunto de sentidos do verbo *tomar* (sob fundo cinza claro no exemplo da Tela 1), na qual poderia assinalar, então, o sentido correspondente a sua necessidade (*tomar conta de alguém*). Se o usuário digitasse uma colocação como chave de consulta, como *tomar banho*, o percurso seria abreviado, dando acesso direto aos exemplos bilíngües. Não havendo

⁷² O exemplo foi elaborado apenas para fins de ilustração, sem a utilização de técnicas específicas para desambigüização de sentidos.

resultado para a colocação digitada, seria apresentado o menu normal das acepções para que ele encontrasse uma colocação próxima à que estava procurando.

Tela 1 - Desambiguiização de sentidos

<p>TOMAR</p> <p>ADOTAR TIPO DE COMPORTAMENTO <u>tomar cuidado</u> <u>tomar precaução</u></p> <p>ALTERAR ESTADO MENTAL OU EMOCIONAL <u>tomar conhecimento de</u> <u>tomar consciência de</u> <u>tomar ciência de</u> <u>tomar coragem</u> <u>tomar juízo</u> <u>tomar vergonha</u></p> <p>DECIDIR <u>tomar decisão</u></p> <p>EXPOR-SE A <u>tomar sol</u> <u>tomar banho de sol</u> <u>tomar friagem</u></p> <p>BANHAR-SE <u>tomar banho</u> <u>tomar uma ducha</u></p> <p>INGERIR <u>tomar água, leite, café, refrigerante, suco, álcool (líquidos)</u> <u>tomar sopa</u> <u>tomar sorvete</u> <u>tomar café da manhã</u> <u>tomar remédio, injeção, anestesia, vacina, comprimido, medicamento, dose de</u> <u>tomar ar, tomar fôlego</u></p> <p>EMBARCAR EM MEIO DE TRANSPORTE <u>tomar o ônibus, tomar o avião, tomar o trem, tomar um táxi</u></p> <p>PRATICAR AÇÃO <u>tomar atitude</u> <u>tomar iniciativa</u> <u>tomar medida</u> <u>tomar providência</u> <u>tomar a frente</u> <u>tomar a dianteira</u></p> <p>RECEBER <u>tomar uma surra</u> <u>tomar uma sova</u> <u>tomar uma bronca</u> <u>tomar uma advertência</u></p> <p>ROUBAR <u>tomar a bola</u> <u>tomar o namorado de alguém</u></p> <p>SER DOMINADO POR <u>ser tomado por (ex: terras foram tomadas pelas plantações de cana)</u> <u>ser tomado pela raiva, pelo amor, pelo medo</u></p> <p>TOMAR COMO <u>tomar como exemplo</u> <u>tomar como base</u> <u>tomar como objetivo</u> <u>tomar como referência</u> <u>tomar como ofensa</u></p> <p>TOMAR CONTA cuidar <ul style="list-style-type: none"> ▪ <u>tomar conta de alguém,</u> ▪ <u>tomar conta do dinheiro,</u> ▪ <u>tomar conta da casa</u> ▪ <u>tomar conta do negócio</u> <u>espalhar-se por um espaço</u> <ul style="list-style-type: none"> ▪ <u>tomar conta de um lugar (ex: A Internet tomou conta do planeta)</u> dominar <ul style="list-style-type: none"> ▪ <u>um sentimento toma conta de alguém ou de um conjunto de pessoas (ex: a raiva tomou conta dele)</u> ▪ <u>alguém toma conta de um lugar, impondo sua vontade : (ex: ela chegou e tomou conta da casa)</u> <p>OUTROS <u>tomar aulas</u> <u>tomar posse de</u> <u>tomar partido</u> <u>tomar posição</u> <u>tomar empréstimo</u> <u>tomar alguma coisa emprestada</u> <u>tomar rumo</u> <u>tomar o rumo de (lugar)</u> <u>tomar o depoimento de (alguém)</u></p> </p>
--

Cada sentido poderá ser especificado por *links* sucessivos. Por exemplo, cada vez que o cursor ou o mouse repousar sobre um dos sentidos apresentados em maiúsculas na Tela 1, serão apresentados exemplos para aquele sentido, podendo ser especificados ainda em maior nível de detalhe, como é o caso da seqüência TOMAR CONTA → cuidar → tomar conta de alguém. Uma vez que se chega ao último nível de especificação do sentido, obtém-se um *link* para um conjunto de exemplos bilíngües (Tela 2).

Nessa próxima etapa, o usuário terá acesso à face bilíngüe do dicionário, constituída por exemplos nas duas línguas, podendo, para um mesmo exemplo em L1, ser exibida mais de uma versão em L2 (paráfrases). Na mesma tela de exibição desses exemplos será disponibilizada uma janela com notas de uso, que poderá, sob opção do usuário, ser aberta sob a forma de um pop-up, ou seja, sobreposta aos exemplos. Suponha-se que o usuário assinale o sentido *tomar conta de alguém* na Tela 1. Isso lhe dará acesso a:

Tela 2 - Exemplos bilíngües e link para notas de uso

TOMAR CONTA DE ALGUÉM (clique sobre um dos exemplos para mais exemplos em inglês)

Esta noite é sua vez de **tomar conta** das crianças.

It's your turn **to look after** the kids tonight

Você poderia **tomar conta** das crianças para mim esta tarde?

Could you **take care of** the children for me this afternoon?

Tereza teve que sair do emprego para **tomar conta** da mãe doente.

Tereza had to leave her job **to care for** her sick mother.

A filha **tomou conta** dela enquanto ela estava no hospital.

Her daughter **nursed** her while she was in the hospital.

Deixamos nossos maridos **tomar conta** das crianças enquanto saímos.

We left our husbands **to mind** the children while we went out

Você poderia **tomar conta** da minha filha enquanto vou ao banheiro?

Could you please **keep an eye on** my daughter while I go to the toilet?

Ela **toma conta** de crianças todas sexta à noite pra ganhar um dinheiro extra.

She **babysits** every Friday night to earn some extra money.

Ela **tomou conta** das crianças para mim ontem à noite.

She **sat** for me last night.

NOTAS DE USO

Os exemplos bilíngües ilustrarão diversas possibilidades de emprego de equivalentes para o sentido escolhido em L1. Poderão ser baseados em *corpora*, mas selecionados e adaptados com propósitos pedagógicos.

Tela 3 – Notas de uso (abertura em *pop-up*)

NOTAS DE USO

look after e **take care of** têm o mesmo sentido de *tomar conta de alguém*, podendo referir-se a cuidar de crianças, de idosos, incapacitados e animais de estimação. Porém, *look after* é mais freqüente no inglês britânico e *take care of* no inglês americano.

care for tem o mesmo sentido de *look after* e *take care of*, porém é mais formal.

nurse tem o sentido de *tomar conta de alguém* muito doente ou ferido

mind tem o sentido de *tomar conta de alguém* por um curto período de tempo, embora possa ser habitual. Seu uso é menos freqüente que *look after* e *take care of*

keep an eye of tem o sentido de *tomar conta de alguém* de forma eventual e por um curto período de tempo.

babysit e **sit** têm o mesmo sentido de *tomar conta de crianças* e implicam o recebimento de dinheiro pelo trabalho; *sit*, porém, é usado mais no inglês americano.

Essas notas serão redigidas em português e terão por objetivo prevenir erros comuns aos brasileiros que aprendem a L2 em questão, bem como fornecer informações sobre registro que orientem a escolha do usuário na possibilidade de duas ou mais formas semelhantes de comunicar o mesmo conteúdo.

Não serão apresentados exemplos semelhantes na Tela 2, mas cada um deles terá um *link* para exemplos semelhantes em L2, apresentados na Tela 4. Aí, sim, poderão ser coletados exemplos de *corpora* que tenham o mesmo sentido (não se trata de uma consulta igual à consulta a um *corpus*, pois os exemplos não serão selecionados por palavra-chave, mas por sentido-chave). A perspectiva seria, portanto, semasiológica na fase de desambigüização em L1 e onomasiológica na fase de exibição dos resultados bilíngües e em L2. Essas perspectivas de exibição de conteúdo, contudo, não precisam ser mantidas na armazenagem das informações. A grande vantagem da mídia eletrônica, aliás, é desvincular forma de armazenagem de forma de exibição de informações.

Supondo-se, a título de exemplificação, que o usuário clique sobre o exemplo *It's your turn to look after the kids tonight*, indicado por um fundo cinza na Tela 2, terá acesso a outros exemplos em L2, conforme mostrado na Tela 4:

Tela 4 - Extensão de exemplos em L2

look after no sentido de *tomar conta de alguém*

She's sixteen, and feels that she is old enough to **look after** herself.

Your dog will get sick if you don't **look after** him properly.

It is the people's obligation in China to **look after** their parents.

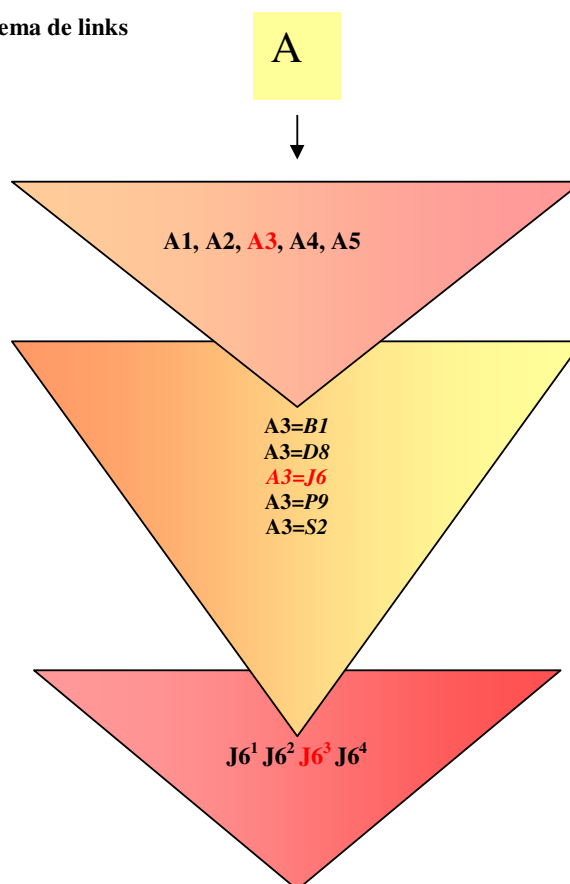
Lots of young people help their parents to **look after** younger brothers.

He asked me to **look after** his daughter while he was away.

Com base nesses exemplos, o usuário poderá inferir uma série de informações a respeito do uso dos equivalentes apresentados, como a regência nominal e verbal (preposições que acompanham nomes e verbos), tipos de sujeitos e de objetos passíveis de ocorrer com o verbo nesse sentido, combinações freqüentes com outros itens lexicais etc.

Em suma, a busca de equivalentes a partir da L1 consistiria de um processo de afunilamento, proporcionado pelas escolhas de *links*. O afunilamento em uma dimensão abriria espaço para o desdobramento na dimensão subsequente. Assim, por exemplo, ao especificar a acepção desejada do item lexical em L1, abrir-se-ia a possibilidade de explorar em maior detalhe os itens lexicais em L2 que poderiam equivaler àquela acepção. Da mesma forma, ao escolher um dos equivalentes, abrir-se-ia a possibilidade de apresentar vários exemplos em que o equivalente escolhido codifica sentido semelhante à acepção especificada em L1. Em uma representação esquemática desse caminho, tem-se:

Figura 9 - Esquema de links



Sendo:

- **A**: item lexical em L1 utilizado como argumento da busca;
- **A1, A2, A3, A4, A5**: sentidos do item lexical **A**, para desambigüização (tendo sido assinalado o sentido **A3**, grifado de vermelho)
- **B1, D8, J6, P9, S2**: sentidos dos itens lexicais **B, D, J, P** e **S**, da L2, que expressam sentido similar ao sentido **A3** da L1 (tendo sido assinalado o sentido **J6**)
- **J6¹, J6², J6³, J6⁴**: exemplos de uso em que se realiza **J6** na L2

Enfatiza-se aqui o que foi observado por ocasião da análise do dicionário DAFLES: a exibição de conteúdos em mídia digital requer conhecimento dos recursos dessa mídia, caso contrário a visualização em tela poderia ficar igual ou pior que a visualização em papel. No exemplo criado, vê-se que a quantidade de material para desambigüização de sentidos do

verbo *tomar* extrapolaria o limite de uma tela. Seria desejável, no caso, que os desenvolvedores de uma solução desse tipo evitassem a necessidade de rolar a tela e implementassem uma estrutura de exibição em camadas ou em três dimensões, como é feito, por exemplo, no dicionário eletrônico *Visual Thesaurus*⁷³. O que se descreve em telas neste trabalho, portanto, são propostas de conteúdo e não propriamente de exibição desse conteúdo.

6.1.2. Desenvolvimento dos dicionários bilíngües para codificação escrita em línguas estrangeiras

Para desenvolver um produto com as características descritas no item anterior, é fundamental que os lexicógrafos responsáveis pelo projeto trabalhem de forma colaborativa com profissionais da computação. A definição da estrutura da base de dados, por exemplo, deve anteceder a coleta de material lexicográfico, pois, do contrário, as obras eletrônicas poderiam seguir estruturas semelhantes às das obras impressas, o que é um desperdício do potencial oferecido pela mídia eletrônica.

A seqüência de passos do trabalho de desenvolvimento lexicográfico da obra seria a seguinte:

- 1) seleção da nomenclatura em L1;
- 2) levantamento de todos os sentidos (ou pelo menos de todos os que se conseguir identificar) e das colocações mais freqüentes de cada item lexical da nomenclatura;
- 3) seleção e adaptação de exemplos de *corpora* de L1 que ilustrem, pedagogicamente, cada um dos sentidos e colocações levantados para os itens da nomenclatura;
- 4) tradução dos exemplos para L2;

⁷³ www.visualthesaurus.com (dicionário inovador na forma de exibição, mas os dados são os mesmos da WordNet)

- 5) seleção de exemplos em L2 que tenham sentido semelhante a cada um dos sentidos dos exemplos bilíngües e que ilustrem diferentes aspectos do emprego dos equivalentes;
- 6) elaboração de notas de uso para cada sentido ou colocação, baseadas em evidências de *corpus* de aprendizes.

Os três primeiros passos desse trabalho são a fase monolíngüe do desenvolvimento e, por isso, são comuns ao desenvolvimento de soluções lexicográficas bilíngües para diversas L2. Sendo assim, o resultado dessa fase poderá ser reaproveitado para todas as L2 para as quais se deseje fazer DBs da L1.

A seleção da nomenclatura para constituir uma obra desse tipo deve ter como foco o vocabulário ativo em L1, com suas respectivas acepções e colocações ilustradas por exemplos pedagógicos. Nesse vocabulário devem estar incluídos os itens lexicais mais freqüentes da língua em uso (lexias simples e complexas), incluindo registros formal e informal, pois, com certeza, são os mais consultados em situação de codificação, devido a seu alto grau de polissemia. O “peso” de um item lexical de alta freqüência na percepção da qualidade de um dicionário para apoio à codificação em L2 é relativamente maior que o de um item lexical de baixa freqüência. Isso é diferente do que ocorre ao se consultar um dicionário para decodificação, quando a probabilidade de busca de itens lexicais menos freqüentes é maior e uma nomenclatura mais abrangente é decisiva para a percepção de qualidade.

Definir critérios para a seleção dessa nomenclatura e aplicá-los na geração do vocabulário ativo é uma tarefa que exige muita reflexão e muitas decisões. O trabalho teria que ser baseado em *corpus* e a primeira decisão seria quanto a utilizar um *corpus* já existente ou constituir um novo, especialmente para essa finalidade. Em qualquer das duas possibilidades, como o objetivo da obra é o apoio à escrita, o *corpus* para seleção da nomenclatura deve ser constituído de gêneros escritos e de textos atuais. Ressalta-se,

inclusive, a relevância de incluir no *corpus* gêneros típicos da Web, como *blogs*, *e-mails* e *chats*, que costumam reproduzir um registro menos formal da língua e constituem gêneros nos quais comumente muitos usuários desejam se comunicar em L2.

Uma vez definida a nomenclatura, inicia-se a desambigüização dos sentidos dos itens lexicais que a constituem. Esse trabalho deve ser baseado em *corpus* de L1, podendo ser aproveitado o mesmo *corpus* utilizado para seleção da nomenclatura. O resultado provavelmente será um volume muito grande de material lexicográfico, pois, como diz Shcherba, “toda palavra merece uma monografia” (1940, p. 328). Embora a mídia eletrônica não apresente problemas quanto à capacidade de armazenagem, a tela do computador apresenta restrições quanto à capacidade de exibição de conteúdo. Isso, como já foi dito, representa um desafio para os especialistas em leiaute de exibição digital.

A desambigüização desses itens lexicais é uma fase complexa e, por isso, sugere-se que, no caso brasileiro, seja estudada a possibilidade de reaproveitar as informações já levantadas no projeto *Wordnet*⁷⁴ do português do Brasil (v. DIAS-DA-SILVA & DE MORAES, 2003).

Outro projeto que faz a descrição semântica do léxico e que poderia inspirar essa fase é a *Framenet*⁷⁵ (v. ATKINS, RUNDELL & SATO, 2003 e BAKER, FILLMORE & LOWE, 1998), que difere da *Wordnet* por ser totalmente baseada em *corpus*, porém até o momento não houve nenhuma iniciativa brasileira nesse sentido.

Simultaneamente à desambigüização dos sentidos, ou em seqüência, serão levantadas, por meio de análise de *corpus*, as colocações mais freqüentes formadas pelos itens lexicais da nomenclatura. O ideal é estabelecer um índice de freqüência a partir do qual as colocações devam ser consideradas.

⁷⁴ Situação mais atualizada do projeto pode ser obtida na página: pln.br.tripod.com/new_page_4.htm

⁷⁵ <http://framenet.icsi.berkeley.edu/>

Durante o processo de análise de *corpus* de L1 para levantar acepções e colocações, serão escolhidos os exemplos de uso a serem traduzidos para constituir os exemplos bilíngües.

Após essa seleção é que se inicia a fase bilíngüe do dicionário, constituída da tradução dos exemplos (podendo haver mais de uma opção de tradução para cada exemplo de L1). Para essa tradução, podem ser consultados *corpus* de L2, *corpus* paralelo de L1-L2 e DMs de L2. Pode ser necessária, também, uma assessoria de lexicógrafos e/ou tradutores nativos de L2 que conheçam a L1 do público-alvo do dicionário. Se um *corpus* paralelo apresentar exemplos adequados, parte do trabalho de tradução dos exemplos poderá ser poupado.

A seguir, uma pesquisa em *corpus* de L2 fará a seleção de exemplos em L2 que apresentem os mesmos equivalentes propostos nos exemplos bilíngües, empregados no mesmo sentido, porém em contextos diferentes. Esses exemplos serão oferecidos a partir do *hiperlink* de cada exemplo bilíngüe. Resultados da pesquisa de Yon e Hirvella (2004) abordados no capítulo 1, apontam a complementaridade entre a consulta a *corpus* e a consulta a dicionário, o que sugere ser interessante combinar os benefícios dos dicionários aos benefícios de um *corpus* em uma mesma ferramenta eletrônica. No caso dos exemplos em L2, se houvesse (ou quando houver) um *corpus* semanticamente anotado, ao invés de se extraírem exemplos de *corpus* para exibir no dicionário, poder-se-ia exibir o resultado de uma consulta ao *corpus*, o que faria com que o usuário tivesse acesso direto ao *corpus* e eliminaria os vieses que toda seleção de material lingüístico feita pelo homem corre o risco de conter.

A idéia é começar a trabalhar com os itens mais freqüentes, tornando-os disponíveis gradativamente. Em se tratando de mídia eletrônica, a atualização de versões é menos complexa e tem um custo infinitamente menor do que na mídia impressa. Por isso, não seria necessário “fechar” uma nomenclatura antes de iniciar o desenvolvimento da obra, mas apenas definir um ponto de partida.

Uma vez definidos, por exemplo, os primeiros 1000 itens lexicais a serem trabalhados para compor a obra, esses itens terão que ser profundamente analisados, a fim de que a grande maioria das acepções e colocações seja levantada.

6.2. Recursos pós-escrita

Alguns aspectos da revisão de textos escritos podem ser racionalizados e automatizados. A primeira forma de revisão automática a ser implementada foi a revisão ortográfica, hoje já desenvolvida para diversos idiomas. Existem também, atualmente, revisores gramaticais e estilísticos, porém o número de idiomas que os possuem é relativamente menor, pois são recursos de desenvolvimento mais complexo.

Os verificadores ou corretores ortográficos e gramaticais são aplicativos que monitoram a escrita, corrigem alguns erros e apontam outros prováveis, dando ao usuário a oportunidade de corrigi-los. Problemas de ortografia, de concordância e de regência podem, dessa forma, ser resolvidos sem a necessidade de uma consulta. Basta que o usuário digite uma tentativa e espere que o aplicativo aponte uma eventual impropriedade. Pode-se dizer que os recursos pós-escrita são sucedâneos das consultas a dicionários, a *corpora* e a gramáticas, pois substituem a necessidade de alguns tipos de consulta, automatizando o que antes era feito manualmente. Os aplicativos que constituem os recursos pós-escrita costumam ser agregados aos recursos disponíveis para edição de textos.

Entretanto, no que diz respeito aos corretores ortográficos, o apoio à escrita em L1 é priorizado, sendo previstos erros típicos de nativos da língua. Em um ambiente dirigido à escrita em L2, seria oportuno tornar igualmente disponíveis recursos de correção que previssem erros típicos de cada grupo de L1 na escrita em L2. Os erros de transferência negativa, por exemplo, poderiam ser levantados e tratados adequadamente. Nesse caso, se um brasileiro escrevesse em inglês *recognizement*, transferindo a regra de composição *reconhecer*

+ *mento* = *reconhecimento* para *recognize* + *ment* = *recognizement*, poderia ser apresentado à opção correta *recognition*, já que *recognizement* não existe. Como esse tipo de erro é dependente da L1 do aprendiz, essa proposta representaria uma adequação do aplicativo às necessidades de um público específico. Tais erros poderiam ser levantados submetendo-se um *corpus* de aprendizes a um revisor ortográfico. Analisando-se a lista de palavras inexistentes gerada, seriam, então, identificadas quais delas obedecem a algum critério de transferência de L1 para L2. Tais erros, por serem previsíveis pela lógica, poderiam ter seu reconhecimento agregado aos corretores ortográficos destinados a apoiar a escrita em L2, gerando a apresentação de alternativas de correção. Esse tipo de especialização de corretores ortográficos de L2 para um determinado público falante de L1 seria um diferencial significativo no caso de línguas muito próximas. Na pesquisa relatada no capítulo 4, percebeu-se que grande parte das consultas a dicionários feitas pelos sujeitos de pesquisa que codificaram em outras línguas neolatinas foram motivadas pela necessidade de confirmar se existem em L2 determinados itens lexicais, intuídos por eles a partir de simetrias contrastivas com a L1.

Já o revisor estilístico de texto em L2 é um tipo de aplicativo menos comum. O único exemplo de que se tem conhecimento nessa categoria é o *StyleWriter*⁷⁶, do inglês, compatível com o editor de textos *Word*, da Microsoft. Para utilizá-lo, seleciona-se um texto, assinala-se um gênero de texto de uma lista de opções e submete-se a consulta. O revisor analisa a distância entre características do texto escrito e características de um *corpus* de textos do gênero escolhido: o léxico empregado, o tamanho das frases, a quantidade de passivas etc. Como resultado da análise, o aplicativo atribui um conceito para o texto escrito, sugere algumas alterações e aponta questões críticas a serem revistas pelo usuário.

6.2.1. Considerações adicionais

⁷⁶ <http://www.stylewriter-usa.com/>

Além da possibilidade de solução aqui apresentada para atender as necessidades de codificação escrita em L2, há de se considerar que existem outras possibilidades. A mais desejável delas (pois pouparia esforço do usuário), seria o desenvolvimento de um tradutor automático eficiente, o que, ao que tudo indica, ainda é uma utopia. Tradutores automáticos, porém, apesar de suas limitações, parecem já ser utilizados como um recurso a mais à disposição de quem lida com outras línguas.

Dentre as possibilidades mais factíveis, no entanto, visualizam-se dois caminhos: o primeiro seria pesquisar em *corpus*, levantar evidências do comportamento lexical e sistematizar as informações em um dicionário. Seria criada, para isso, uma série de marcas de uso. Esse caminho não foi escolhido para a proposta porque, apesar de as evidências apontarem a necessidade de informações sobre o uso, apontam também o fato de que, quanto mais complexas as informações do dicionário, menor a probabilidade de serem usadas. Por isso, concorda-se com a afirmação de Philippe Humblé sobre a suficiência dos exemplos e, ao invés de se optar por dizer ao usuário como as palavras se comportam, optou-se por mostrar isso a ele. Assim, o usuário poderá ver com seus próprios olhos e tirar suas conclusões, o que é, aliás, um dos avanços da Lexicografia Pedagógica relatados por Worsch (1999) e Katzaros (2004), como visto no capítulo 2. Para complementar os exemplos, optou-se por notas de uso que chamem a atenção para problemas mais comuns no uso da palavra-entrada (o que não se confunde com marcas de uso, que seriam uma descrição sistematizada).

O segundo caminho alternativo que se visualiza para desenvolvimento de uma solução de apoio à codificação em L2 seria etiquetar um *corpus* de maneira a facilitar sua consulta pelos usuários. Uma vez que avançam os esforços no sentido de se implementar a *web semântica*, o próprio motor de busca poderia dar a opção de restringir a pesquisa a um subconjunto de documentos. Poderiam ser oferecidas, por exemplo, as opções de argumentos de pesquisa que restringissem a pesquisa por área de conhecimento ou ao conjunto de

documentos que apresentam versões bilíngües ou multilíngües, da mesma forma como hoje já é possível restringir a pesquisa pelo argumento “idioma”. Essa solução substituiria em grande parte as necessidades de consulta a dicionários e não seria, portanto, uma solução lexicográfica. A solução lexicográfica aqui proposta, no entanto, ao apresentar exemplos bilíngües adequados para cada sentido e colocação da palavra-entrada, abreviaria o esforço exigido do usuário em consultas diretas a um *corpus*.

Ressalta-se ainda que a construção de um ambiente computacional para codificação escrita em L2, conforme sugerido no início deste capítulo, poderia ser iniciativa de uma editora, de um centro de pesquisa, de um desenvolvedor *softwares* (principalmente de editores de textos), de um motor de buscas ou, até, ser um esforço conjunto de todos eles.

A propósito, cita-se a discussão acerca do futuro dos dicionários, foco de quatro artigos publicados entre 2001 e 2004 pela *KDictionaries News*. O autor de dois deles é Charles Levine, da *Random House Dictionaries*, e os autores dos outros dois são Joseph Sposito, consultor da *Merriam Webster's dictionaries* e Julian Parish, gerente de negócios da central de dicionários da Microsoft.

No primeiro artigo, Levine (2001) previa um novo fôlego para a indústria dos dicionários impressos de inglês, garantido pelos dicionários destinados a falantes de inglês *língua estrangeira* ou inglês *segunda língua*.

Sposito (2002), contudo, contestou as previsões de Levine. Segundo ele, não importa quanto tempo os dicionários impressos têm de sobrevida, o fato é que eles estão condenados a desaparecer e a serem substituídos pelos dicionários eletrônicos. O autor previu que as editoras de dicionários iriam, aos poucos, sucumbindo: primeiro as pequenas (deixando sua fatia de mercado para as grandes), mas no final todas elas sofreriam falta de investimentos e acabariam em dificuldades. Sposito argumentou ainda que a maior ameaça para as editoras de dicionários é a Microsoft, pois a empresa vê os dicionários como um assunto estratégico

dentro de seu negócio e tem a política de agregar dicionários a seus aplicativos. Em resposta, Levine (2003) reconheceu o acerto das previsões de Sposito, à luz dos acontecimentos no mercado editorial de dicionários naquele momento.

Parish (2004), por sua vez, respondeu à seqüência de artigos, dizendo que, realmente, a Microsoft vê a área de dicionários como um assunto estratégico, mas que isso não é uma ameaça às editoras e sim uma oportunidade. Segundo ele, não faz parte da política da Microsoft levantar informações sobre o léxico, mas sim adquiri-las de terceiros.

Se tais opiniões estiverem corretas, hoje as editoras de grandes dicionários deverão estar investindo menos no negócio e procurando meios de estender seus níveis de venda por mais alguns anos. Talvez isso explique a atitude que surpreende Binon e Verlinde:

Em um colóquio ocorrido em Lyon em setembro de 2007, Pierre Corbin do SILEX de Lille denunciou a relativa inércia dos editores franceses que muito freqüentemente se contentam em dar uma maquiada, fazer um “lifting”, mas **não investem muito na renovação lexicográfica** (...) Quando os editores e dicionaristas franceses vão enfim assumir os numerosos desafios que lhes são lançados? ⁷⁷(BINON & VERLINDE, 2008, p. 9, grifo meu).

Ao mesmo tempo, uma observação dos autores aponta o que, talvez, seja um dos últimos redutos da pesquisa lexicográfica, principalmente de outras línguas que não o inglês: “Temos a impressão que é sobretudo nas instituições universitárias que tais inovações são concebidas e implementadas.”⁷⁸ (BINON & VERLINDE, 2008, p. 9).

O que esse cenário sugere é que os dicionários podem sobreviver às mudanças provocadas pelas novas tecnologias, mas o endereço e o perfil de os quem faz está mudando. As editoras estão sofrendo concorrência das empresas que desenvolvem *softwares* (em especial a Microsoft) e os lexicógrafos tradicionais estão sendo substituídos por lingüistas

⁷⁷ Lors d'un colloque qui s'est tenu à Lyon en septembre 2007 P. Corbin du SILEX de Lille a dénoncé cette relative inertie des éditeurs français qui se contentent trop souvent d'un toilettage, d'un lifting mais n'investissent pas assez dans un renouvellement de la dictionnaire, (...). Quand les éditeurs et les dictionnaristes français vont-ils enfin relever les nombreux défis qui leur sont lancés?

⁷⁸ Nous avons l'impression que c'est surtout dans des institutions universitaires que ces innovations sont conçues et mises au point.

com habilidades em ciências da computação ou por informatas com habilidades lingüísticas, se não por equipes multidisciplinares.

Schriver (2003), aliás, realizou uma revisão acerca dos trabalhos críticos sobre os dicionários eletrônicos e mostrou como os “sonhos” dos lexicógrafos vão, aos poucos, sendo realizados. Sonhos sobre aumentar o espaço de armazenamento e a velocidade de processamento da consulta; agregar elementos multimídia; acoplar um *corpus*; integrar o dicionário a outros *softwares*, desenvolver novas formas de acesso e tornar o dicionário mais fácil de usar e passível de ser adaptado às necessidades e preferências de cada usuário.

Espera-se que as propostas aqui apresentadas se enquadrem nessa categoria de “sonhos”.

CONSIDERAÇÕES FINAIS

A necessidade de comunicação em L2 intensificou-se nas últimas décadas em função do estreitamento das relações entre nações. Isso criou uma demanda pelo desenvolvimento de competência oral e escrita em L2. No que diz respeito à escrita, uma série de fatores contribuem para o processo, dentre eles o aparato lexicográfico disponível. Conforme evidenciam as pesquisas discutidas neste trabalho, há dificuldades, entretanto, que não podem ser sanadas pelo uso do dicionário e mesmo aquelas que o podem, dependem do aperfeiçoamento dos dicionários e do ensino de seu uso.

Enquanto os dicionários para decodificação evoluíram razoavelmente nas últimas décadas, os DBs para suporte à codificação são praticamente inexistentes. No caso brasileiro, em especial, há carência de dicionários L1-L2 partindo do ponto de vista do português do Brasil.

Muitos teóricos já discutiram a questão de dicionários para apoio à codificação. Grande parte dessa discussão foi a respeito de qual dicionário seria mais adequado, o DM ou o DB. Se, por um lado, demonstrou-se que o dicionário L1-L2 é necessário em várias situações de escrita em L2, em outras situações, os DMs para estrangeiros poderiam ser mais úteis.

Acredita-se que, por detrás do discurso que defende o uso exclusivo do DM para aprendizes, ao invés do DB, nas atividades de codificação em L2, estejam interesses profissionais e empresariais. DMs de inglês para aprendizes estrangeiros representam uma grande parcela do mercado lexicográfico mundial. Muito investimento é feito para aperfeiçoá-los continuamente. Enquanto os DMs atendem todas as nacionalidades (representando um mercado global), DBs dividiriam esse mercado.

A iniciativa da Kernerman Dictionaries de bilingualizar DMs para aprendizes foi um sucesso de mercado. Porém, a iniciativa só produziu obras L2-L1 e não L1-L2, como seria o ideal para as necessidades de codificação. Além disso, enquanto o empreendimento da Kernerman exigiu acordos com tradutores e editoras de cada país onde foi lançado, dicionários L1-L2 exigiriam muito mais esforço de profissionais nativos da L1. Isso representaria uma divisão de trabalho intelectual e, acima de tudo, uma divisão de lucros.

Investigando a situação de codificação por meio de uma atividade de tradução L1-L2, percebeu-se que muitas das dúvidas são adequadamente atendidas pelos recursos já disponíveis para consulta. As dificuldades concentram-se na busca de equivalentes em L2, a partir de palavras-chave em L1, acompanhados de informações sobre seu emprego. Concluiu-se também que não é necessário optar entre DM e DB, mas sim oferecer os dois, para que o usuário possa optar por um deles, dependendo da estratégia que decida adotar em cada busca lexical.

Teóricos e pesquisadores do uso parecem concordar que uma gama de informações sobre o uso do léxico é necessária para a situação de codificação em L2. Contudo, também se obtiveram várias indicações de que, quanto mais simples e direto um recurso, maior a probabilidade de ser consultado.

Os usuários, em sua maioria, não conseguem expressar como gostariam que fosse um dicionário de apoio à codificação. Embora não se tenham dados suficientes para generalizar uma afirmação, observou-se empiricamente que o DB pedagógico parece atender melhor as necessidades dos aprendizes do que os DBs tradicionais. Cabe ao lexicógrafo romper o paradigma do que tradicionalmente se convencionou chamar de dicionário e propor soluções que facilitem a vida dos usuários que precisam dos conhecimentos sobre o léxico para expressar suas idéias adequadamente na escrita em L2.

Embora haja vários caminhos de busca lexical que possam levar a uma resposta satisfatória, o melhor para o usuário é aquele que requer menor esforço e demanda menor tempo, de forma a não desviar a atenção da tarefa principal, que é escrever o texto em L2.

Diante dessas evidências, sugere-se focar esforços no desenvolvimento de dicionários L1-L2 para atender a finalidade específica de busca de equivalentes para uso, o que preencheria uma grande lacuna no aparato lexicográfico necessário para os brasileiros codificarem em L2.

Como o aspecto do uso deve ser destacado, sugerem-se apenas dois componentes para o dicionário: exemplos e notas de uso. As consultas a esse dicionário teriam, em português, a desambigüização dos sentidos, filtrando a apresentação de resultados (exemplos) em L1 e L2. As notas de uso, em português do Brasil, seriam baseadas em evidências colhidas na análise de *corpus* de aprendizes brasileiros.

Contudo, além da necessidade de desenvolvimento de dicionários L1-L2, percebeu-se também outra oportunidade de melhorar o atendimento aos redatores e tradutores que codificam em L2. Uma vez que a mídia eletrônica vem substituindo a escrita manual, sugere-se reunir, em um mesmo ambiente, os recursos mais relevantes para a escrita em L2: um editor de textos e, para cada L2 considerada: um corretor ortográfico/gramatical, um revisor estilístico, acesso a *corpora*, DBs L1-L2 de língua geral e de termos, DM de L1, DM de L2 para aprendizes estrangeiros, dicionário de sinônimos, modelos de documentos etc. Esses recursos são dependentes de idioma e, por isso, seria preciso definir quais as línguas que deveriam ser priorizadas no desenvolvimento desse tipo de ambiente. Logicamente, o estágio do PLN de cada idioma afeta diretamente esse tipo de decisão.

Uma vez que se consiga desenvolver obras fáceis e rápidas de consultar, os usuários poderão diminuir sua resistência ao uso do dicionário durante a redação, atividade em que, como foi constatado, são utilizadas várias estratégias para evitar o uso do dicionário. Embora

na atividade de tradução L1-L2 o dicionário já seja intensamente utilizado, as melhorias lexicográficas poderão diminuir o tempo exigido e aumentar a qualidade dos resultados das consultas.

Acredita-se que são diversos os encaminhamentos a serem feitos a partir dos resultados desta tese. A definição de um vocabulário ativo da língua portuguesa, por exemplo, seria um dos mais importantes, já que serviria de base para a construção de dicionários de codificação do português para diversas L2.

Outro pré-requisito para a implementação das idéias aqui sugeridas, seria que os *corpora* de aprendizes disponíveis no Brasil ganhassem robustez e passassem a inspirar trabalhos de análise de dificuldades de aprendizes que, por sua vez, forneceriam subsídios para as notas de uso a serem incluídas nos dicionários L1-L2.

Não menos importante é a etiquetagem semântica do português do Brasil, fundamental para a desambigüização das palavras-chave a partir das quais os usuários consultam os dicionários português-L2. Esse processo já está em curso, posto que também é pré-requisito para vários aplicativos baseados no processamento computacional do português.

A construção de *corpora* paralelos português-L2 de língua escrita geral e sincrônica seria outro desenvolvimento significativo para a implementação desta proposta, pois esses *corpora* poderiam ser utilizados tanto para a seleção de exemplos bilíngües como servir de base para estudos contrastivos. Os resultados desses estudos, porém, precisariam ser coordenados a fim de fornecer insumo para o apoio à escrita em L2, não só para os dicionários, mas também para as ferramentas pós-escrita discutidas no capítulo 6.2.

Espera-se, acima de tudo, que o conteúdo apresentado consiga sensibilizar os leitores para a necessidade de construir um aparato de apoio aos brasileiros que codificam em L2.

Em havendo um projeto que articule a construção de um aparato conforme aqui descrito, muitos dos pré-requisitos citados terão seu desenvolvimento impulsionado. Em não

havendo um projeto, acredita-se que os pré-requisitos acabarão sendo desenvolvidos para outras finalidades e, no futuro, talvez esse aparato envolva mais reaproveitamento do que desenvolvimento de aplicativos, *corpora* e material lexicográfico, o que, espera-se, torne sua implementação mais factível.

REFERÊNCIAS BIBLIOGRÁFICAS

ABASI, A. R. ; AKBARI, N. ; GRAVES, B. Discourse appropriation, construction of identities, and the complex issue of plagiarism: ESL students writing in graduate school. **Journal of Second Language Writing**, v.15, n.2, p. 102-117. Elsevier Science Direct, 2006.

ALIANDRO, H. (org.) **Dicionário português-inglês The Portuguese-English dictionary**. Pocket Book, 1960.

ARD, J. The use of bilingual dictionaries by ESL students while writing. **ITL, Review of Applied Linguistics**, v. 58, p. 1-27, 1982.

ATKINS, B. T. S.; VARANTOLA, K.; Monitoring Dictionary Use. **International Journal of Lexicography**, v. 10, n. 1, p. 1-45. Oxford University Press, 1997.

ATKINS, B. T. S.; RUNDELL, M.; SATO, H. The contribution of Framenet to practical Lexicography. **International Journal of Lexicography**, v. 16, n. 3, p. 333-357. Oxford University Press, 2003.

BAKER, C. F.; FILLMORE, C. J.; LOWE, J. F. The Berkeley FrameNet Project. **Proceedings of the Computational Linguistics 1998 Conference**, University of Montréal, p. 86-90. Association for Computational Linguistics (ed.). Disponível em: <http://framenet.icsi.berkeley.edu/papers/acl98.pdf>. Acesso em: 13 set. 2006.

BAKER, M. Corpus-based translations studies: The challenges that lie ahead. In: **Terminology, LSP and Translation Studies ind language engineering in honour of Juan C. Sager**. Amsterdam & Philadelphia, 1996, p. 175-186.

BAKER, M. Corpora in Translation Studies: An overview and Some Suggestions for Future Research. **Target** v. 7, n. 2, p. 223-243. Amsterdam: John Benjamins, 1995.

BAKER, M. Corpus Linguistics and Translation Studies: Implications and Applications. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (eds.) **Text and Technology. In honour of John Sinclair**. Philadelphia/Amsterdam: John Benjamins, 1993, p. 233-250.

BÉJOINT, H. **Modern Lexicography: an introduction**. Oxford: Oxford University Press, 2000.

BERBER SARDINHA, T. **O corpus de aprendiz BR-ICLE**. São Paulo: Puc-Lael, 2001. Disponível em: www.lael.pucsp.br/~tony/2001bricle. Acesso em: 31 jul. 2005.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.

BINON, J. ; VERLINDE, S. La contribution de la Lexicographie Pédagogique du français langue étrangère ou seconde (FLES) à la dictionnaire du français langue maternelle (FLM). 2008 (no prelo).

BINON, J.; VERLINDE, S. A contribuição da lexicografia pedagógica à aprendizagem e ao ensino de uma língua estrangeira ou segunda. In: LEFFA, V. J. (org.) **As palavras e sua companhia – o léxico na aprendizagem**. Pelotas: EDUCAT, 2000, p. 95-118.

BINON J., SELVA T., VERLINDE S.: Tendances et innovations récentes en lexicographie pédagogique: la contribution des dictionnaires d'apprentissage DAFA et DAFLES. **I Symposium Internacional de Lexicografía** (Universitat Pompeu Fabra, Barcelona, 16-18 mai 2002). Disponível em: <<http://www.kuleuven.ac.be/grelep>>. Acesso em: 25 nov. 2004.

BINON J., VERLINDE, S., SELVA T. Lexicographie pédagogique et enseignement/apprentissage du vocabulaire en français langue étrangère ou seconde (FLES): Un mariage parfait. **Cahiers de Lexicologie**, v. 78, n. 2, p. 41-63, 2001.

BLATTNER, N. Demystifying Writing Assessment: Empowering Teachers and Students. **Assessing Writing**, v. 6, n. 2, p. 229-237. Elsevier, Science Direct, 1999.

BOGAARDS, P. Dictionaries for Learners of English. **International Journal of Lexicography**, v. 8, n. 4, p. 277-310. Oxford University Press, 1996.

BOGAARDS, P.; HANNAY, M. Towards a new type of bilingual dictionary. In: EURALEX, Lorient, 2004. **Proceedings of Euralex 2004**. Lorient, 2004, p. 463-474.

BUENO DE ABREU *et al.*, **Longman dicionário escolar inglês-português português-inglês para estudantes brasileiros**. Longman, 2002.

CHAN, A. Y. W. TAYLOR, A. Evaluating learners' dictionaries: what the reviews say. **International Journal of Lexicography** v. 14, n. 3, p. 163-180. Oxford University Press, 2001.

CHRISTIANSON, K. Dictionary use by EFL writers: what really happens? **Journal of Second language Writing**, v. 6, n. 1, p. 23-43. Elsevier Science Direct, 1997.

CLAHSEN, H.; FELSEN, C. How native-like is non-native language processing? **Trends in Cognitive Sciences**, v. 10, p. 564-570, 2006

COWIE, A. P. **English dictionaries for foreign learners: a history**. Oxford University Press, 1999, p. 1-50.

DE COCK, S. ; GRANGER, S. Computer Learner Corpora and Monolingual Learners' Dictionaries: the Perfect Match. In TEUBERT W.; MAHLBERG M. (eds.) **The Corpus Approach to Lexicography**. Special issue of *Lexicographica* 20, 2005, p. 1-15.

DIAS-DA-SILVA, B. ; MORAES, H. R. A construção de um *thesaurus* eletrônico para o português do Brasil. **Alfa**, n. 47 , v. 2, p. 101-115. São Paulo, 2003.

DOLEZAL F.T., MCCREARY D.R. **Pedagogical Lexicography today**: a critical bibliography on learners' dictionaries with special emphasis on language learners and dictionary users. *Lexicographica Series Maior* 96. Tübingen: Niemeyer, 1999.

DURAN, M. S. **Dicionários bilíngües pedagógicos: análise, reflexões e propostas**. 2004. Dissertação (Mestrado em Estudos Lingüísticos) Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto, Universidade Estadual Paulista.

EAST, M. The impact of bilingual dictionaries on lexical sophistication and lexical accuracy in tests of L2 writing proficiency: quantitative analysis. **Assessing Writing**, v. 11, n. 3, p. 179-197. Elsevier, Science Direct, 2006.

ENGBER, C. The relationship of lexical proficiency to the quality of ESL compositions. **Journal of Second language Writing**, v. 4, n. 2. Elsevier Science Direct, 1995.

EZER, H. ; SIVAN, T. "Good" academic writing in Hebrew: The perceptions of pre-service teachers and their instructors. **Assessing Writing**, v. 10, n. 2, p. 117-133, 2005.

FLOWERDEW, J. Problems in Writing for Scholarly Publication in English: The Case of Hong Kong. **Journal of Second Language Writing**, n. 8, v. 3, p. 243-264, 1999.

FORTKAMP, M. B. M.; PREBIANCA, G. V. V.; PIRES, A. P. O. Lexical access in speech production: a brief literature review. **Matraga**, ano 13, n. 19, p. 15-32. Rio de Janeiro: Caetés, 2006.

FRANKENBERG-GARCIA, A. A peek into what today's language learners as researchers actually do. **International Journal of Lexicography**, v. 18, n. 3, p. 335--355. Oxford University Press, 2005.

GOMES, D. **O uso do dicionário bilíngüe na produção escrita em alemão como língua estrangeira**. Dissertação (Mestrado em Lingüística Aplicada). Universidade de Brasília, Brasília, 2006.

GRANGER, S. Computer learner corpus research: current status and future prospect. In: Connor U. and T. A. Upton (eds.) **Applied Corpus Linguistics: A multidimensional Perspective**. Amsterdam & Atlanta: Rodopi, 2004, p. 123-145.

GRANGER, S. From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. In: Aijmer K.; Altenberg B. and Johansson M. (eds.) **Languages in Contrast. Text-based cross-linguistic studies**. Lund Studies in English. Lund: Lund University Press, 1996, p. 37-51.

HAAN, P. J. M. ; ESCH, C. J. M. The development of writing in English and Spanish as foreign languages. **Assessing Writing**, v. 10, n. 2, 100-116, 2005.

HARTMANN, R. What is “dictionary research”? Review article of Wiegand, H. E. Wörterbuchforschung. Untersuchungen zur wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband. Berlin: Walter de Gruyter. 1998. **International Journal of Lexicography**, v. 12, n. 2 p. 155-160. Oxford University Press, 1999.

HARVEY, K.; YUILL, D. A study of the use of a monolingual pedagogical dictionary by learners of English engaged in writing. **Applied Linguistics** n.18-3 p. 253-278, 1997.

HÖFLING, C. **Traçando um perfil de usuários de dicionários – estudantes de Letras com Habilitação em Língua Inglesa: um novo olhar sobre dicionários para aprendizes e a formação de um usuário autônomo**. Tese (Doutorado em Linguística e Língua Portuguesa). Universidade Estadual Paulista “Júlio de Mesquita Filho”, Araraquara, 2006.

HUMBLÉ, P. O dicionário bilíngüe português-espanhol da Universidade Federal de Santa Catarina. In: BORBA, H. M.; CARDOSO, T. M. G. **Desenvolvimento de um Dicionário Eletrônico de Apoio à Produção de Textos em Língua Estrangeira**. Projeto de conclusão de curso de Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 2003, p. 24-34.

HUMBLÉ, P. Produção versus compreensão no dicionário bilíngüe português-espanhol da Universidade Federal de Santa Catarina. In: Maria José Damiani Costa; Meta Elizabete Zipser; Marta Elizabete Zanatta; Angelita Mendes. (Orgs.). **Línguas: ensino e ações**. Florianópolis, 2002, v. 1, p. 181-192.

HUMBLÉ, P. **A new model for a foreign language learners' dictionary**. Tese (Doutorado em Letras). Universidade Federal de Santa Catarina, 1997.

JACOBSEN J. ; PEDERSEN V. H. Examples in the Bilingual Dictionary. F.J. Hausmann *et al.* (eds.) **An International Encyclopedia of Lexicography**, vol. 3. Berlin: Walter de Gruyter, 1991, p. 2782-2789.

JOHNSON, K.; JOHNSON, H. **Encyclopedic Dictionary of Applied Linguistics**. London: Blackwell, 1998.

KATZAROS, V. The different functions of illustrative examples in learners' bilingual dictionaries. In: EURALEX, Lorient, 2004. **Proceedings of Euralex 2004**. Lorient, 2004 p. 487-494.

KOMURO, Y.; YAMADA, S. Dictionary use for production among Japanese college students of English. **Kernerman Dictionary News**, n. 8, Tel-Aviv, jul. 2000. Disponível em : kictionaries.com/newsletter/kdn8-3.html. Acesso em: 29 set. 2006.

LAUFER, B. A case for a semi-bilingual dictionary for productive purposes. **Kernerman Dictionary News**, n. 3, Tel-Aviv, jul. 1995. Disponível em: <http://www.kictionaries.com/newsletter/kdn3-2.htm>. Acesso em: 15 ago. 2002.

LAUFER, B. The effect of dictionary definitions and examples on the use and comprehension of new L2 words. **Cahiers de Lexicologie**. n. 63, p. 131-142, Paris, 1993.

LAUFER, B. *Corpus-based versus lexicographer examples in comprehension and production of new words*. In: EURALEX, Stuttgart, 1992. **Proceedings of Euralex 1992**. Stuttgart, 1992 p. 71-76.

LAUFER, B.; HADAR, L. Assessing the effectiveness of monolingual, bilingual and "bilingualised" dictionaries in the comprehension and production of new words. **Modern Language Journal**. n. 81, p. 189-196, Tel-Aviv, 1997.

LAUFER, B.; LEVITZKY-AVIAD, T. Examining the effectiveness of 'bilingual dictionary plus – a dictionary for production in a foreign language. **International Journal of Lexicography**, v. 19, n. 2, p. 135-155, 2006.

LEVELT, W. J. M. ; ROELOFS, A. ; MEYER, A.S. A theory of lexical access in speech production. **Behavioral and Brain Sciences**, n. 22, p. 1-75. Cambridge University Press, 1999.

LEVINE, C. M. The coming boom in English Lexicography – reconsidered. **Kernerman Dictionary News**, n. 11, jul. 2003. Disponível em: <http://kictionaries.com.kdn11-1.html>. Acesso em: setembro 2003.

LEVINE, C. M. The coming boom in English Lexicography: some thoughts about the World Wide Web. **Kernerman Dictionary News**, n. 9, jul. 2001. Disponível em: <http://kictionaries.com.kdn9-1.html>. Acesso em: setembro 2003.

LI, J. The mediation of technology in ESL writing and its implications for writing assessment. **Assessing Writing**, n. 11, p. 5-21. Elsevier, Science Direct, 2006.

LIU, Y. **The cognitive process of translation in L2 writing**. Tese (Doutorado) Purdue University, Indiana, Estados Unidos, 2004.

LONGMAN. **Longman Language Activator**. Essex: Longman, 1993.

MCCARTHY, M. **Vocabulary**. Oxford: Oxford University Press, 1990.

MCCREARY D. R.; DOLEZAL, F.T. A study of dictionary use by ESL students in na American university. **International Journal of Lexicopgraphy**, vol. 12, n. 2. Oxford University Press, 1999.

NESI, H. The effect of language background and culture on productive dictionary use. In: EURALEX, Amsterdam, 1994. **Proceedings of Euralex 1994**. Amsterdam: Vrije Universiteit, 1994

NESI, H. Do dictionaries help students write? In: Bloor, T. ; Norrish, J. (eds.) *Written Language: papers from the Annual Meeting of the British Association for Applied Linguistics*. London: CILT, 1987, p. 85-97.

NUNES, J. H. **Discurso e instrumentos lingüísticos no Brasil: dos relatos de viajantes aos primeiros dicionários**. Campinas. IEL-UNICAMP, 1996.

OLIVEIRA, H. F. Ensino do léxico: o problema de adequação vocabular. **Matraga**, v. 19, p. 49-68, Rio de Janeiro, 2006.

PARISH, J. Microsoft and Dictionary Publishers: Defining Partnerships. **Kernerman Dictionary News**, n. 12, jul. 2004. Disponível em: <http://kictionaries.com.kdn12-1.html>. Acesso em: agosto 2007.

PECORARI D. Good and original: Plagiarism and patchwriting in academic second-language writing. **Journal of Second Language Writing**, V. 12, Issue 4, p. 317-345. Elsevier Science Direct, 2003.

RAJAGOPALAN, K. O ensino de línguas na contemporaneidade. Conferência proferida em 24/11/2006, no **XI Simpósio Nacional de Letras e Lingüística (XI SILEL) e I Simpósio Internacional de Letras e Lingüística**.

RUNDELL, M. Recent trends in publishing monolingual learners' dictionaries. In: HARTMANN, R. R. K. (ed): **Thematic Network Projects, Sub-project 9 – Dictionaries - Dictionaries in Language Learning, Final Report Year Three**, 1999a, p. 83-98. Disponível em: <<http://www.fu-berlin.de/elc/tnp1/SP9dossier.doc> tnp 1> Acesso em: 04 jul. 2003.

RUNDELL, M. Dictionary use in production. **International Journal of Lexicography**, vol. 12, n. 1, 1999b. Oxford University Press.

RUNDELL, M. GRANGER, S. From corpora to confidence. **English Teaching Professional**. Issue 50, May 2007.

SASAKI, M. Toward an empirical model of EFL writing processes: and exploratory study. **Journal of Second Language Writing**, vol. 9 n.3, p. 259-291, 2000.

SCHMITZ, J. R. A problemática dos dicionários bilíngües. In: OLIVEIRA, A. M. P. P.; ISQUERDO, A. N. (org.). **As ciências do léxico**. Campo Grande: Editora UFMS, 2001, p. 161-170.

SCHRYVER, G. M. Lexicographers' dreams in the electronic-dictionary age. **International Journal of Lexicography**, v. 16, n. 2, p. 143-199. Oxford University Press, 2003.

SHCHERBA, L. V. Towards a General Theory of Lexicography (traduzido por D. FARINA a partir do original russo de 1940). **International Journal of Lexicography** n. 8, v. 4, p. 314-350. Oxford University Press, 1995.

SIEPMANN, D. Collocation, colligation and encoding dictionaries. Part II: lexicographical aspects. **International Journal of Lexicography**, v. 19, n. 1, p. 1-39, 2006.

SIEPMANN, D. Collocation, colligation and encoding dictionaries. Part I: lexicological aspects. **International Journal of Lexicography**, v. 18, n. 4, p. 409-443, 2005.

SILVA, T.; BRICE, C. Research in Teaching Writing. **Annual Review of Applied Linguistics**, n. 24, p. 70-106. Cambridge University Press, 2004.

SPOSITO, J. J. Dictionaries, another Netscape? **Kernerman Dictionary News**, n. 10, jul. 2002. Disponível em: <http://kictionaries.com.kdn10-1.html>. Acesso em: setembro 2003.

TAGNIN, S. E. O. A multilingual learner corpus in Brazil. CORPUS LINGUISTICS 2003 CONFERENCE. **Learner Corpus Workshop**. Lancaster University. Disponível em: <http://www.fllch.usp.br/dlm/comet/artigos/A%20multilingual%20learner%20corpus%20in%20Brazil.pdf>. Acesso em: 16/11/2006.

TARDY, C. M. Researching first and second language genre learning: a comparative review and a look ahead. **Journal of Second Language Writing**, V. 15, p. 79-101. Elsevier Science Direct, 2006.

TONO Y. Learner corpus: design, development and applications. In Archer *et al.* (eds.) **Proceedings of the Corpus Linguistics 2003 Conference** (CL 2003). Technical Papers 16.

Lancaster University: University Centre for Computer Corpus Research on Language, 800-809, 2003.

TOURY, G. **Descriptive Translation Studies and beyond**. Amsterdam/Philadelphia: John Benjamins, 1995 (revisão de 1978), p. 198-211.

UZAWA, K. Second language learners' processes of L1 writing, L2 writing and translation from L1 to L2. **Journal of Second Language Writing**, vol.5, n. 3, p. 271-294, 1996

VARANTOLA, Krista. Linguistic corpora (databases) and the compilation of dictionaries. In **A Practical Guide to Lexicography**, Sterkenburg, Piet van (ed.), 2003, p. 228-239.

WALLACE, M. **Action Research for Language Teachers**. Cambridge: Cambridge University Press, 1998.

WANG, W. ; WEN. Q. L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. **Journal of Second Language Writing**, v. 11, n. 3, 225-246, 2002.

WELKER, H. A. **O uso de dicionários**. Brasília, Thesaurus: 2006.

WINKLER, B. Students working with an English learners' dictionary on CD-ROM. **Papers from the ITMELT 2001 Conference**. Disponível em: <http://elc.polyu.edu.hk/Conference/papers2001/winkler.htm>. Acesso em: 13 fev. 2008.

WONG, A. T. Y. Writers' mental representations of the intended audience and of the rhetorical purpose for writing and the strategies that they employed when they composed. **System**, n. 33, p. 29-47. 2005.

WOODALL, B. R. Language switching: using the first language while writing in a second language. **Journal of Second Language Writing**, n. 11, p. 7-28, 2002. Pergamon.

WORSCH, W. Recent trends in publishing bilingual learners' dictionaries. In: HARTMANN, R. R. K. (org.): **Thematic Network Projects Sub-project 9 – Dictionaries - Dictionaries in Language Learning, Final Report Year Three**, 1999, p. 99-107. Disponível em: <www.fu-berlin.de/elc/tnp1/SP9dossier.doc tnp 1>. Acesso em: 04 jul. 2003.

XATARA, C. M. Os dicionários bilíngües e o problema da tradução. In: OLIVEIRA, A M. P. P.; ISQUERDO, A. N. (org.). **As ciências do léxico**. Campo Grande: Editora UFMS, 2001, p. 181-188.

YOON, H. ; HIRVELA, A. Hirvela. ESL student attitudes toward corpus use in L2 writing. **Journal of Second Language Writing**, v. 13: p. 257-283. 2004.

ANEXOS

Anexo 1: Modelo de folha de rosto do formulário de pesquisa.

Anexo 2: Modelo do formulário para registro das consultas.

Anexo 3: Exemplo de formulário preenchido.

Anexo 4: Exemplo de protocolo verbal.

Anexo 5: texto para a atividade de tradução para língua estrangeira.

Anexo 6: Exemplo de tradução produzida na atividade.

ANEXO 1

MODELO DE FOLHA DE ROSTO DO FORMULÁRIO DE PESQUISA

ATIVIDADE DE CODIFICAÇÃO ESCRITA EM LÍNGUA ESTRANGEIRA

DESCRIÇÃO DA ATIVIDADE:

SUJEITO DE PESQUISA:

- NOME :
- IDADE :
- ESCOLARIDADE:

HORÁRIO DE INÍCIO:

HORÁRIO DE TÉRMINO:

OBRAS DE REFERÊNCIA UTILIZADAS:

- A: _____
- B: _____
- C: _____
- D: _____
- E: _____
- F: _____
- G: _____

ANEXO 2

MODELO DO FORMULÁRIO PARA REGISTRO DAS CONSULTAS

NÚMERO DO SUJEITO:

BUSCA*

<input type="checkbox"/>	1-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	2-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	3-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	4-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	5-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	6-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	7-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	8-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	9-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____
<input type="checkbox"/>	10-	CONSULTA EM _____	ITEM _____	MOTIVO _____	RESULTADO _____

* Campo preenchido com números sequenciais após a aplicação do protocolo, analisando-se quais consultas pertencem a uma mesma busca.

ANEXO 3

**EXEMPLO DE FORMULÁRIO
PREENCHIDO**

Sujeito de pesquisa: D

Nº BUSCA	Nº CONSULTA	OBRA CONSULTADA	ITEM CONSULTADO	MOTIVO	RESULTADO
1	1	ROBERT (DM)	JAPONAIS	Conferir se tem dois “n”	Tem um “n” só
2	2	ROBERT (DM)	NOUNOU	Ver qual o termo mais adequado, <i>nounou</i> ou <i>nourrice</i>	Ficou em dúvida
	3	MICHAELIS (DB)	BABÁ	Ver equivalente mais adequado para <i>babá</i>	Encontrou <i>nounou</i> , <i>babysitter</i> e <i>nurse</i> ; Optou por <i>nounou</i>
3	4	ROBERT (DM)	FIER	Ver se tem o sentido de <i>confiar</i> <i>alguém a alguém</i>	Não achou o sentido
4	5	MICHAELIS (DB)	CEDER	Procurar equivalentes para <i>ceder</i>	<i>Céder</i> , <i>livrer</i> , <i>se soumettre</i> , <i>succomber</i>
	6	ROBERT (DM)	LIVRER	Ver em que contexto <i>livrer</i> tem o mesmo sentido de <i>céder</i>	Sentido de se entregar. Não achou conveniente. Optou por <i>céder</i> .
5	7	MICHAELIS (DB)	ALTURA	Procurar equivalente para altura na colocação <i>metros de altura</i>	Encontrou <i>hauteur</i> e <i>élévation</i> , mas <i>nenhum exemplo com metros de altura</i> .
	8	ROBERT (DM)	HAUTEUR	Procurar exemplo no sentido de <i>metros de altura</i>	Não encontrou, mas optou por <i>hauteur</i> assim mesmo.
	9	ROBERT (DM)	HAUTEUR	Ver se o “h” é aspirado	Encontrou um asterisco, mas não soube interpretar a informação.
	10	GOOGLE	D’HAUTEUR	Testar se existe a forma para verificar se <i>hauteur</i> admite elisão	Encontrou 23.300 casos e codificou <i>mètres d’hauteur</i> .

ANEXO 4

EXEMPLO DE PROTOCOLO VERBAL

Sujeito de pesquisa: D

Idade: 21 anos

Convenções para a transcrição:

Utilizou-se “/” para marcar as pausas e “...” para marcar as hesitações.

Os esclarecimentos da pesquisadora, quando necessários, foram inseridos entre “[]”, bem como as perguntas, ao final, na breve entrevista.

Bom, *japonais*, então tem que colocar *des japonais* / agora eu vou procurar, eu vou ver se *japonais* tem dois *n* ou um *n* só / um *n* só / agora eu vou procurar *babá* porque eu tô em dúvida de colocar *nourrice* ou só *nounou*, que é a forma mais informal de falar *babá* / só que esse dicionário Michaelis *on-line* você tem que colocar toda a acentuação porque eu coloquei *baba* e ele deu *saliva*, né, então... / ele dá como... as acepções... vem... *babysitter*, em inglês, *nourrice* e *nounou* / então eu vou colocar *nounou*, até porque eu acho que fica legal pra formar essa palavra *babá-robô* / é uma palavra curta, ao invés de colocar *nourrice* / agora a palavra *filhos*, a mesma coisa de *japoneses*, eu vou colocar o *des*: *des enfants* / eu tô com dificuldade com essa estrutura da frase: *que não têm com quem* / tô pensando no *personne* e tô pensando no pronome relativo *avec qui* / é, eu tô... eu vou optar pela decisão da negativa, é... *japonais qui n'ont personne pour laisser leurs enfants* então eu optei pela negativa *ne... personne* / é... eu tô pensando agora no *confiar*, como é que eu traduzo para o francês, porque tem o verbo *fier à quelqu'un*, que é o verbo *confiar*, só que eu preciso ver se aqui cabe nesse sentido o verbo *confiar os filhos a alguém* / então eu vou procurar no dicionário, primeiro no Robert & Clé / é... eu tô procurando, mas eu não consegui achar pro *fier* esse sentido que o texto tá dando ainda, porque no dicionário tem geralmente ligado a *confiance*, é... *acreditar* mesmo e não no sentido de *confiar*, mas por enquanto eu vou deixar meio por cima essa opção do *fier* pro verbo *confiar* / agora eu vou procurar o verbo *ceder* / eu imagino que seja *céder* mesmo, mas aqui está *ceder à vontade*, então eu vou ver se tem alguma coisa diferente no dicionário / é, nas acepções que eu encontrei no dicionário *on-line* [DB Michalis português-francês *on-line*] tem o *céder*, tem o *se sumettre*, o *succomber*, tem o *conceder*, *donner*, *livrer* / é, daí eu vou optar pelo *céder*, que eu acho que cabe no sentido, mas eu fiquei em dúvida nesse *livrer*, que eu acho que tem alguma coisa, é uma acepção muito diferente pra ter aparecido, então eu vou ver se tem alguma coisa [no Robert & Clé impresso] / é, eu tô em dúvida agora, porque o *livrer*, ele tem o sentido de *entregar* e no texto eu tô vendo que é assim, *que jamais se cansa ou cede à vontade das crianças*, então esse *céder* eu fiquei em dúvida porque ele tem o sentido de *se cansa ou se entrega à vontade das crianças* / só que eu acho que *se entregar* é mais forte que *ceder*, então vou optar pelo *céder* mesmo, que é o verbo *céder* e não o *livrer* / porque *se entregar* dá o sentido que você vai se envolver com aquilo, que você vai ceder mas vai se envolve, tipo você se entrega a uma paixão e você se envolve / agora, eu acho que *ceder* não necessariamente, você pode deixar tal pessoa fazer aquilo porque você cedeu, mas não tá envolvido naquilo / agora eu vou procurar a palavra *altura* em francês, porque eu esqueci / tem *hauteur* e *élévation*, mas *élévation*, construção, essas coisas, é usado para coisas físicas, então eu vou optar por *hauteur* [essa diferença não estava no dicionário] / então deixa eu verificar no Clé / o dicionário dá como *élévation du niveau des eaux* elevação do nível das águas, elevação de um terreno, e *hauteur*, tô lendo tudo / é, vou optar por *hauteur* mesmo porque tá se referindo à altura do robô / bom, agora eu vou ter que decidir, porque é 1,4 metros de altura, então esse *de*, que é da preposição, no francês, quando começa por vogal a gente tem que juntar, só que aqui a palavra *hauteur* começa com *h*, então eu vou ter que ver se é aspirada ou não pra poder fazer o apóstrofo lá / aqui tá um asteriscozinho, um asterisco em *hauteur*, então eu tenho que ver se quando tem asterisco é

aspirado ou não, então eu vou jogar uma palavra que eu sei que tem a *élision* no Google e depois vou procurar essa palavra no dicionário e vou ver se ela tem o asterisco ou não / bom, vou mudar a preferência pro francês / vou jogar direto a *altura* com apóstrofo pra ver se existe / sim, quer dizer que quando tem o asterisco no dicionário quer dizer que tem a elisão / bom, aqui é... agora eu tenho *que tem como função* / eu sei que *que tem como função* em francês é *qui a comme fonction*, só que essa *função*, quer dizer que o robô tem o papel de fazer alguma coisa, então em francês a gente fala também *qui a le rôle de quelque chose*, então eu vou ver o que fica melhor: *qui a fonction* ou *qui a le rôle* / eu vou procurar primeiro no dicionário Robert Clé, na palavra *fonction* / então, aqui eu acabei de achar o seguinte, que era a minha dúvida: *il fait fonction de directeur, il joue le rôle de directeur*, que era a dúvida que eu tava, que é desempenhar um papel e fazer uma determinada função / por questão de estilo eu vou optar por *jouer le rôle* e eu não jogaria no Google pra ver questão de ocorrência, porque pra mim é a mesma coisa, *avoir la fonction de*, como o dicionário tá dando e *jouer le rôle*, só que eu sempre ouço falar mais desempenhar papel é *jouer le rôle*, até mesmo no teatro, quando a gente estuda teatro francês, então eu vou optar, opção tradutória, *jouer le rôle* / agora *tomar conta de criança* eu tô em dúvida de três coisas... é... do... de... três verbos, é *garder les enfants*, é *prendre les enfants* ou *soigner les enfants* / então vou procurar tudo no dicionário / eu vou procurar primeiro o *soigner*, que é o que eu tenho mais certeza / bom, *soigner* tem esse sentido mesmo aqui no dicionário e agora eu vou pro *garder* / *garder les enfants*... só que olhando esse... *garder les enfants*, já vem do lado *elle s'occupe d'enfants* e tem uma setinha falando *babysitting*, que tem a ver com babá, que esse seria o papel do robô, então agora gerou mais uma dúvida minha, porque além do *soigner* e do *garder*... é, eu tô em dúvida e eu já vou pro *prendre* que é *prendre soin de*, pode ser *une personne* ou inanimado / é, agora eu tô em dúvida dos três / bom, é... eu acho que eu vou optar pela questão de estilo, só que eu não vou tomar ainda nenhuma decisão, eu vou deixar pro final, vou pro outro parágrafo / ah, eu não procurei o *prendre* porque o no *soigner* eu já achei o *prendre soin de* e a opção *s'occupe de* eu não gostei, então eu não vou procurar / agora eu achei esse *deve* aqui no texto / eu não sei se se pode falar esse tipo de dúvida / também eu pensei um pouco pra ver a quem ele remetia, se era os pais ou se era uma outra coisa, porque tá longe... / é, isso que tem aqui, eu não fiquei em dúvida, mas sempre vem à cabeça as duas formas, o *cela* e o *ça*, só que o *ça* a gente não usa porque é muito informal, mais da linguagem oral / é, o verbo *falar* aqui, eu sei que é *parler*, só que sempre que a gente fala *falar*, vem na cabeça o *bavarder*, conversar, né... / bom agora que eu li *apesar do seu vocabulário*, eu fiquei pensando o que é que eu ia colocar no *apesar de*, que no francês tem várias formas, né, o... posso colocar o *bien que*, *malgré*, então eu vou procurar no dicionário *on-line* / é, eu achei *malgré* porque *bien que*, que é o que eu tava em dúvida, é mais *ainda que*, então eu vou optar pelo *malgré* mesmo / e também agora me veio na cabeça agora o uso do *malgré*, se com ele que tipo de conjugação verbal eu teria que usar, porque tem algumas conjunções que eu tenho, eu sou obrigado a jogar o subjuntivo, mas o *malgré* eu uso o indicativo mesmo / bom aqui tá a *máquina desenvolvida* e a primeira palavra que eu pensei foi o *developpée*, só que esse desenvolvida é no sentido de construída e criada, então não cabe o *developpée*, então eu tô em dúvida se eu vou ficar com o *créée* ou com o *construite* / esse *junto* eu vou procurar, porque eu tô em dúvida se é o *ensemble* / procurar no *on line* / *Michaelis on-line* / tô em dúvida aqui / tem *join*, *uni*, *lié*, nem *uni* nem *lié* eu vou pegar, porque não tem sentido de *junto com* / eu tô em dúvida entre o *join* e o *ensemble*, então eu vou procurar no Robert & Clé / eu achei a expressão *join à* / eu vou colocar aqui do lado e vou ver o *ensemble* / também tem esse sentido, só que eu tô em dúvida agora se é *créée ensemble à* com a preposição ou se é *ensemble* sozinho / então eu vou jogar no Google pra ver a ocorrência que tem / vou jogar com acento pra ver se dá diferença / deu diferença / então eu já vou optar, porque como eu ouço falar mais o *ensemble* que o *join*, eu vou usar o *ensemble à* / é, aqui tem um *projektor em um de seus olhos*, então eu tô pensando no

que é que eu vou colocar nesse *em um de*, que seria dentro, né, e eu fiquei em dúvida da preposição, se eu coloco a preposição *dans* ou o *en*, daí eu vou procurar no dicionário, no Robert / eu acho que vou começar pelo *en*, porque *dans* não tem esse sentido necessariamente, pode ser *na superfície* / é, aqui o sentido de *en* tá dando o sentido de *dentro* mesmo / então eu vou jogar no Google pra ver se existe *en ses yeux*, essa expressão / não, o *en ses yeux* não tem, só que eu vou colocar *en les yeux* / também não tem / eu vou colocar *dans ses yeux* / só que tem um problema com o *dans ses yeux*, porque aqui as ocorrências que tem, por exemplo, *olhe nos meus olhos* é... não sei se fica bem claro a idéia *dentro dos olhos*, porque olhar você olha pra superfície do olho, você não olha pra dentro do olho / então eu tô tentando pensar em alguma coisa... é, eu vou tentar jogar um problema dentro do olho, uma catarata, alguma coisa, eu vou colocar a palavra *problema* / é, não deu com o *dans*, deu uma coisa aqui, *problème aux yeux*, eu fiquei em dúvida e vou ver se cabe o *aux* aqui pra *dentro*, né / aqui colocou *mostrar fotos que tirou*, só que eu acho que não precisa colocar *que tirou*, *tiradas* mesmo, que a máquina tirou, né, só que eu não vejo necessidade em de deixar esse *tirou*, é pra mostrar as fotos tiradas, porque se eu colocar *que tirou* ia ter que colocar *a pris* só que eu... fica até melhor *pour montrer les photos prises*, fica mais enxugado o texto / agora eu vou procurar *caméra* é masculino ou feminino / procurei *le câmera* só que [incompreensível] é, *caméra* é feminino / bom, agora eu vou dar uma revisada no texto, eu já terminei / bom, eu ainda to com dúvida com esse negócio do texto, então eu vou ver como que o título, como é de jornal, uma notícia, uma coisa mais rápida, vou ver como seria uma notícia mais ou menos assim em frances, se tem o artigo ou se pode ficar sem o artigo / inventei qualquer coisa, só pra... / não deu certo / então eu vou procurar algum jornal, o *Le Monde* mesmo / é, aqui eu achei um que tem o artigo: *Des photographies* / só que eu achei um que não tem: *Legislatif...* / eu vou optar por deixar mesmo / e, ao invés de *les enfants*, eu não consigo traduzir isso como uma coisa indefinida / eu vou colocar o *les enfants* / aqui é *leur* [incompreensível] *robot que ne se fatigue jamais ou...* no francês eu não sei se é muito comum como a gente, a gente não coloca, subentende que é o mesmo sujeito que é carregado junto com a frase anterior, mas acho que no francês não é assim, acho que é melhor repetir *ou qui ne cede pas à la volonté des enfants* / pronto, eu acho que tá certo.

Entrevista

[O que você gostaria de comentar sobre os recursos que você utilizou para a atividade?]

Eu acho que falta sempre um dicionário bom do português para o francês porque até mesmo assim... quando a gente vai fazer redações, o Michaelis, ele é bom, mas é muito limitado / eu não aconselho mais o Michaelis / que nem, quando a gente entra, os professores falam “compra o Michaelis” / eu sei que é para quebrar galho, mas esse quebra-galho, ele na verdade quebra o que a pessoa teve que aprender, porque, por exemplo, eu até hoje eu uso o meu Michaelis porque eu tenho e eu sei que ele vai quebrar um galho / então eu não vou comprar um dicionário de duzentos reais / então eu acho que desde o começo tem que incentivar a não usar o Michaelis / eu acho que tem que ser o contrário, sabe, porque daí a pessoa já acostuma com outro e ela vai sempre usar aquele outro, não vai usar o Michaelis, que é pequenininho, tá sempre ali, até ajuda... acento... coisas... só que o problema é que, quando tem expressão, esse tipo de coisas, às vezes a gente acha, mas é muito fraco, assim, sabe? e a gente acaba pondo aquilo porque não tem mais nada pra por / eu acho que o melhor que tem é o CD on-line do Petit Robert / acho até... seria...

[o que você faz em uma situação em que não conhece uma palavra? Como vai usar um monolíngüe?]

É, então no caso, se não existisse a possibilidade de eu ter um dicionário de português para francês, supondo que não existe o Michaelis, eu ia tentar por aproximação de alguma coisa do

tipo, mas mesmo assim eu acharia difícil pra expressão alguma coisa mais com o Michaelis eu ia salvar, né, porque se eu não sei o termo rádio, eu sei que no Michaelis vai ter rádio, já teria um termo científico, uma ou outra coisa teria que colocar, mas eu acho que nem os monolíngües eu... / tem um outro que eu usei da editora Porto, dicionário português-francês, eu achei muito bom esse tipo de ferramenta, mas eu acho impossível fazer sem Internet, tradução e versão não tem como [Internet, você diz assim...] Google, Google, não tem como, é praticamente impossível fazer sem Google e Google... qualquer dúvida que você tem, você cria um exemplo e joga, entende? que ele existe com outra coisa / mas de qualquer jeito precisa ver se fica legal ou não, principalmente quando é tradução, que o francês às vezes a gente erra, porque como nós não somos nativos, nós não temos noção de como falar aquilo, né? / é, então fica mais difícil ainda, por isso é que versão é considerada ainda mais difícil, porque é como um francês falaria, né? / agora, português, se vê que aquilo fica legal, você vai lá, joga no Google e geralmente consegue achar, né? mas o que eu uso e acho que é suficiente é Google, dicionário on-line do Petit Robert e é... um bom dicionário/ é um bom dicionário português-francês para fazer versão que falta/ porque até agora... porque às vezes até a gente tenta procurar aquele, não é o Azevedo esse que eu tô falando, que tem na biblioteca/ o Azevedo que é o cinza, é o grosso, mas aquilo já saiu de moda, aquilo você acha em Portugal só, pra comprar, aqui você não acha/ eu acho ele bom / até agora esse outro que eu tô falando é um da Porto Editora mesmo, é um laranja, custa na faixa de cento e noventa reais mais ou menos, acredito que seja melhor que o Michaelis, né? / mas acho que por enquanto seria essas ferramentas mesmo/ pouca gente compra por causa do preço/ às vezes o pessoal... não sei, nem sabe às vezes porque acho que é não muito antigo / português de Portugal, mas é o que tem, o que a gente pode ter acesso / e eu acho que, por exemplo, uma coisa que foi criada, achei muito interessante, foi aquele site da Xatara de provérbios, expressões, porque às vezes você se bate com expressões que até... você acha, só que nem sempre / você precisa do contexto pra saber se é informal ou um pouquinho mais formal / e aquele site que ela criou ajuda muito a conseguir isso, principalmente provérbio, que fica difícil, porque tem provérbio que é muito brasileiro, né? que nem, quando eu trabalhei no consulado, eu me deparei com um que o Chacrinha falava “quem não se comunica se estrumbica” e isso não tem lá fora, então pelo site dela eu consegui achar uma outra coisa que fazia parte do domínio científico, mas que dava para passar mais ou menos a mesma idéia, e é bom isso, né? e falta também, pelo menos eu não tenho acesso, eu sei que inglês tem, português, tem, inglês não sei se é muito ou pouco, mas eu sei que tem, é um *corpus* de francês [tem sim] é bom saber, não é divulgado, né? [então você acha que é fundamental ter o Google para usar a Internet como *corpus*?] Mas aí nem sempre... sempre com um olho atrás, porque Google também não é tudo / às vezes ele passa algumas coisas que a gente não... não cabe né? tem que olhar o contexto.

ANEXO 5

TEXTO PARA A ATIVIDADE DE TRADUÇÃO PARA LÍNGUA ESTRANGEIRA

Notícia do Jornal **Bom Dia**, da cidade de São José do Rio Preto, edição de 26/03/2008, p. 32.

Japoneses já têm acesso a babá-robô para filhos

Os pais japoneses que não têm com quem deixar seus filhos poderão confiá-los agora a uma simpática babá-robô que jamais se cansa ou cede à vontade das crianças. A Aeon Co. anunciou ontem que começou a vender um robô de 1,4 metro de altura que tem como função tomar conta de crianças.

Se os pais quiserem deixar seus filhos com esta máquina, que tem o tamanho de uma criança, devem programá-la com códigos especiais que determinarão o que o robô deve fazer. Isso também permite que a babá identifique as crianças por seus nomes e idades e fale com eles, apesar de seu vocabulário ainda limitado no momento. A máquina, desenvolvida junto com o principal criador de robôs Tmusk, também pode utilizar um projetor em um de seus olhos para lançar mensagens de advertência ou mostrar fotos que tirou com a câmera instalada no outro olho.

ANEXO 6

TRADUÇÕES PRODUZIDAS NA
ATIVIDADE

Sujeito de pesquisa: A

Giapponesi genitori hanno già accesso a una babysitter robot per i figli

I giapponesi genitori che non hanno nessuno per lasciare i figli, potranno affidargli a una simpatica robot babysitter che mai si stanca o cede alla volontà dei bambini. La Aeon Co. ha annunciato ieri che ha iniziato la vendita di una babysitter robot 1,4 metri di altezza la cui funzione è badare i bambini.

Se i genitori vogliono lasciare i figli con questo robot, che ha la taglia di un bambino, devono programmarlo con alcuni codici speciali che determineranno la attività che il robot deve fare. Questo anche permette che la babysitter identifichi i bambini per i suoi nomi e età, e che parli con loro, nonostante il suo vocabolario limitato in questo momento. La macchina, sviluppata insieme al principale creatore di robot, la Tmusk, anche può utilizzare un proiettore in uno dei occhi per inviare messaggi di avvertenza o per mostrare le fotografie che ha fatto con la camera fotografica.

Sujeito de pesquisa: B

Japoneses ya tienen acceso a la niñera-robot para sus hijos

Los padres japoneses que no tienen con quien dejar sus hijos podrán confiarlos ahora a una simpática niñera-robot que jamás se cansa o cede a la voluntad de los chicos. Aeon Co. anunció ayer que empezó a vender un robot de 1,4 metros de altura que tiene la función de cuidar de los niños.

Si los padres quieren dejar sus hijos con esta máquina, que tiene el tamaño de un niño, ellos deben programarla con códigos especiales que determinarán qué el robot debe hacer. Eso también permite que la niñera identifique los niños por sus nombres y edades y hable con ellas, aunque su vocabulario sea limitado en el momento. La máquina, desarrollada junto al principal creador de robots Tmusk, también puede utilizar un proyector en uno de sus ojos para lanzar mensajes de advertencia o mostrar fotos que hizo con la cámara instalada en el otro ojo.

Sujeito de pesquisa: C

Japanese people has access to a robot babysitter for their sons

Japanese parents who do not have someone to let their sons with will entrust them to a simpatic robot babysitter, which never get tired or does what children wants. Aeon Co. announced yesterday that it started to send a 1,4 meters hight robot which function is to take care of children.

If parents want to let their children with this machine, which has the size of a child, they should program it with special codes that will determine what the robot will have to do. This resource also allows the babysitter to identify the children by their names and ages, and to talk with them, despite its vocabulary be limited at the moment. The robot, developed together with Tmusk, can also use a projector in one of its eyes to launch warning messages, or to show photos which it has taken with the camera installed in its other eye.

Sujeito de pesquisa: D

Des japonais ont déjà accès à une nounou-robot pour les enfants

Les pères japonais qui n'ont personne pour laisser leurs enfants pourront, maintenant, leur fier à une nounou-robot qui ne se fatigue jamais ou qui ne cède pas à la volonté des enfants. L'Aeon Cie. a annoncé hier qu'elle a commencé à vendre un robot de 1,4 mètres d'hauteur, qui joue le rôle de faire des gardes d'enfant.

Si les parents veulent laisser leurs enfants à la garde de cette machine, ayant la taille d'un enfant, ils doivent la programmer avec des codes spéciaux qui vont déterminer ce que le robot devra faire. Cela permet aussi à la nounou d'identifier les enfants pour leurs noms et leurs âges, et de leur parler, malgré son vocabulaire encore limité en ce moment. La machine, créée ensemble au principal créateur de robots, Tmusk, peut utiliser aussi un projecteur dans l'un de ses yeux pour jeter des messages d'avertance ou pour montrer des photos prises avec la caméra installée dans l'autre oeil.

Sujeito de pesquisa: E

Des Japonaises ont déjà d'accès à nounou robot pour ses enfants

Des parents japonaises que n'ont pas avec qui laisser ses enfants pourront maintenant faire confiance à une sympathique nounou robot qui jamais se fatigue ou cède à l'envie des enfants. L'Aeon Co. a annoncé hier qu'elle a commencé à vendre un robot de 1,4 mètre de hauteur qui a comme fonction s'occuper des enfants.

Si les parents veulent laisser ses enfants avec cet appareil, qui est dans la taille d'un enfant, ils doivent la programmer avec des codes spéciaux qui détermineront ce que le robot doit faire. Ceci permet aussi que la nounou identifie les enfants pour ses noms et l'âge et qu'elle parle avec eux, malgré que son vocabulaire est encore limité dans le moment. L'appareil, développé avec le principal créateur de robots Tmusk, peut aussi utiliser un projecteur dans un de ses yeux pour lancer messages d'avertissement ou montrer des photos qu'il a pris avec la caméra installée dans l'autre œil.

Sujeito de pesquisa: F

Des japonais ont déjà accès à une baby-sitter robot pour les enfants

Des parents japonais qui n'ont personne pour laisser ses enfants pourront les confiés maintenant à une sympa baby-sitter robot qui ne se fatigue jamais ou cède à la volonté des petits. L'entreprise Aeon Co a annoncé hier qu'elle a commencé à vendre un robot de 1,4 mètres de haut qui a la fonction de surveiller des enfants.

Si les parents voulaient laisser ses fils avec cette machine, qui a la taille d'un enfant, ils doivent programmer des codes spéciaux de façon à déterminer ce que le robot devra faire. Cela permet aussi que la baby-sitter les identifie par ses prénoms et âges et parle avec elles, en dépit de son vocabulaire encore limité à ce moment. La machine, développée en collaboration avec le principal créateur de robots Tmusk, peut utiliser aussi un projecteur dans un de ses yeux pour lancer des messages d'avertissement ou montrer des photos qu'elle a pris avec un appareil installé dans l'autre œil.

Autorizo a reprodução xerográfica para fins de pesquisa.

São José do Rio Preto, 16 de dezembro de 2008

MAGALI SANCHES DURAN