LEXICOGRAFIA E LINGÜÍSTICA DE CORPUS

Magali Sanches DURAN[1] (UNESP São José do Rio Preto)

RESUMO: A lexicografia foi uma das primeiras áreas da lingüística a fazer uso de corpora eletrônicos. Porém, as possibilidades de exploração de um corpus são inúmeras e outras áreas da lingüística têm descoberto evidências que podem ser aproveitadas pela lexicografia. Essas evidências podem fornecer tanto matéria para integrar os dicionários quanto subsídio para decisões de projeto de dicionários. Este artigo procura identificar interesses comuns entre a lexicografia e as abordagens baseadas em corpus eletrônico.

ABSTRACT: Lexicography was the first linguistic field to benefit from computerized corpus. There is a large range of possibilities to explore a corpus and other linguistics fields have revealed profitable evidences to lexicography. Such evidences provide dictionary content and subsides decisions about dictionary projects. This article is an effort to identify common interests between lexicography and computerized corpus approaches.

1. Introdução

A atividade lexicográfica tradicional sempre foi baseada em corpus. Antes da automação, a "varredura" de grandes corpora era feita manualmente e envolvia, quase sempre, grandes equipes de colaboradores.

Os corpora eletrônicos foram inicialmente produzidos visando, principalmente, aos interesses lexicográficos: descobrir a maior lista de palavras de uma língua com suas respectivas freqüências. Até hoje grandes empreendimentos em torno da construção de corpora de línguas são mantidos por editoras de dicionários, conforme observa Rizo-Rodriguez (2004). Com base na freqüência, definiram-se critérios lógicos para selecionar a nomenclatura e determinar seu tamanho, estabelecendo as noções de léxico básico e fundamental, vocabulário ativo e passivo etc. Descobriu-se a "*core language*", ou vocabulário básico, que passou a ser usada para as definições (palavras mais freqüentes, mais comuns, para explicar os itens mais raros).

Além desses usos, os corpora também passaram a ser utilizados pelos lexicógrafos para a seleção dos exemplos que integram os verbetes dos dicionários. Existem críticas a esse método de seleção, pois há frases de corpus que não são adequadas para integrar dicionários, ou seja, o fato de constar do corpus não garante que possam ser consideradas típicas da língua ou adequadas para qualquer público-alvo. Laufer (1992) constatou, aliás, que exemplos produzidos por lexicógrafos parecem ser mais eficientes para aprendizes do que exemplos retirados de corpus. O ideal, segundo Rundell (2003), é fazer exemplos baseados em corpus, o que permite ao lexicógrafo promover pequenas adaptações a fim de torná-los adequados para a finalidade de mostrar o uso típico.

As abordagens baseadas em corpus, porém, desenvolveram-se e ultrapassaram em muito esses usos pioneiros de corpus pela lexicografia. As ferramentas de exploração de corpus apresentam capacidade de processar com precisão um volume de dados cada vez maior, permitindo a identificação de padrões de uso das línguas. Assim, conhecimentos epilingüísticos dos falantes nativos passaram a ser explicitados pelas pesquisas. Quando divulgados, principalmente em dicionários, esses padrões de uso são fundamentais para auxiliar aprendizes estrangeiros a produzir um discurso mais próximo daquele produzido pelos falantes nativos.

Pesquisas baseadas em corpus empreendidas por lingüistas não ligados aos interesses lexicográficos revelam informações inéditas sobre o léxico e os lexicógrafos devem estar atentos à possibilidade aplicar esses resultados no aperfeiçoamento e na inovação dos dicionários.

[1] magali.duran@uol.com.br

2. Abordagens baseadas em corpora e seus benefícios para a lexicografia

Enxergo basicamente dois tipos de benefícios que os resultados de estudos baseados em corpus podem fornecer à lexicografia: aqueles que subsidiam decisões de projetos lexicográficos, que chamarei de benefícios indiretos e aqueles que constituem matéria passível de integrar dicionários e que chamarei aqui de benefícios diretos.

Os benefícios indiretos impactam as decisões acerca do tipo de dicionário a ser produzido: a quem se destina, o que deve conter, como deve ser organizado etc. É o caso das informações sobre as necessidades dos aprendizes de língua estrangeira inferidas pela análise de suas produções. Como afetam a lexicografia no momento anterior à confecção dos dicionários propriamente dita, esses benefícios serão discutidos no item 2.1. Já os benefícios diretos, representados pelas informações a respeito do comportamento do léxico, suas propriedades combinatórias, suas diversas realizações de sentido, suas ocorrências nos diversos registros da língua serão discutidos no item 2.2.

2.1 Benefícios indiretos: corpus de aprendizes

Corpus de aprendizes é uma abordagem muito utilizada em pesquisas relacionadas ao ensino e à aquisição de línguas estrangeiras. Em pesquisas relacionadas às línguas maternas, porém, seu uso é mais raro, principalmente no caso da aquisição que, pelo fato de ocorrer na tenra infância e na oralidade, dificulta a coleta de dados.

A idéia de conhecer o aprendiz de uma língua estrangeira 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 foca as relações bilíngües e 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.

Assim, os estudos da tradução ganharam maior fundamento quando passaram a analisar, além do eixo bilíngüe, da tradução em relação a seu respectivo original, o eixo monolíngüe, ou seja, a tradução em relação aos textos produzidos originalmente na mesma língua. Com isso, as traduções deixaram de ser vistas apenas como espectros de seus textos originais e ganharam status de textos que compartilham características típicas.

Com os corpora comparáveis criou-se também a possibilidade de testar diversas hipóteses para descobrir se as características identificadas poderiam realmente ser atribuídas ao texto traduzido ou se estavam associadas à ocorrência de outras variáveis.

Dois corpora podem ser comparáveis se compartilham uma parcela considerável de características. A idéia de Baker (1995) foi discutida e executada por Laviosa (1997), que descreve as dificuldades do processo e as decisões adotadas para contorná-las, relato que contém questões relevantes para qualquer interessado em construir um corpus.

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.

Depois de mais de uma década, a proposta de Granger já orientou muitas pesquisas e foi ganhando maior sofisticação[2]. 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. 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 língua estrangeira. O ideal seria aplicar um teste padrão para medir a proficiência dos sujeitos, mas a exeqüibilidade disto em larga escala é pequena até o momento (TONO, 2003).

O primeiro projeto de corpus de aprendizes de âmbito internacional é o ICLE (International Corpus of Learners English), constituído de vários subcorpora de redações de aprendizes de inglês de diferentes línguas maternas. 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 verificar características típicas de textos produzidos em inglês por estrangeiros em geral (independentemente de suas respectivas línguas maternas) e características típicas de textos produzidos em inglês por estrangeiros falantes de uma determinada língua materna. 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 conhecer o usuário dos dicionários. Até pouco tempo, os dicionários continham o que o lexicógrafo considerava importante, mas agora se começa a pesquisar o que é importante para o cliente. A partir das dificuldades identificadas pela análise das redações de aprendizes de língua estrangeira, por exemplo, pode-se cogitar que tipo de necessidade de ação ou de informação elas representam e, assim, desenvolver 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 sobreuso de vocabulário de alta freqüência, 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 língua materna.

Segundo Tono (2003), o desenho 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 língua materna na interlíngua;
- a produção de aprendizes em ambiente de língua estrangeira (quando a língua aprendida é chamada de segunda língua) com a produção de aprendizes em ambiente de língua materna (quando a língua aprendida é chamada de língua estrangeira), 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

[2] 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.

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 língua estrangeira são devidas à falta de obras lexicográficas adequadas. Existem dificuldades que só podem ser superadas com o ensino, como por exemplo, aquelas relacionadas ao uso de tempos verbais. Há outras que nem o ensino consegue superar, como é o caso de tendências que não constituem propriamente "erros". Seria o caso, por exemplo, das tendências no texto do aprendiz semelhantes às apontadas pelos estudos da tradução (BAKER, 1996), como a simplificação, a explicitação, a normalização e a estabilização.

Na verdade, as pesquisas baseadas em corpus de aprendizes focam o produto e é sobre os resultados da análise desse produto que o lexicógrafo infere as necessidades dos aprendizes e avalia quais poderiam ser providas pelos dicionários. Acredito que essas pesquisas com foco no produto poderiam ser complementadas por pesquisas com foco no processo para levantar as causas mais prováveis das dificuldades verificadas (cf. HARVEY & YUILL, 1997; CRHISTIANSON, 1997; MCCREARY & DOLEZAL, 1999; NESI, 2002; LAUFER, 2006).

Para o lexicógrafo, é importante saber, por exemplo, se:

- O aprendiz consultou o dicionário, mas não obteve resultado satisfatório, pois:
  - faltava informação no dicionário ou
  - faltava habilidade (ao aprendiz) em lidar com o dicionário;
- O aprendiz não consultou o dicionário, pois:
  - a palavra não suscitou dúvidas;
  - achou que poderia inferir com alguma segurança;
- As dificuldades estão relacionadas a motivo que não envolve o uso de dicionário.

Dependendo das causas mais recorrentes, diferentes medidas poderiam ser tomadas para melhorar a situação, como, por exemplo:

- Melhorar o ensino de habilidades de consulta a dicionários;
- Melhorar a forma de organização dos dicionários;
- Agregar novas informações aos dicionários.

Para a lexicografia bilíngüe, o corpus de aprendizes poderia ser utilizado também como abordagem para testar a eficiência de diferentes tipos de dicionários ou formas de exibição de conteúdo (cf. LAUFER, 1997 e 2006).

Embora a aplicação lexicográfica dos resultados de corpus de aprendizes seja principalmente no que diz respeito a decisões orientadoras de projeto, sabe-se de pelo menos um caso em que os "erros" dos aprendizes formam reunidos em forma de dicionário: o *Longman Dictionary of Common Errors*, disponibilizado no CDROM do *Longman Interactive English Dictionary* (RIZO-RODRÍGUEZ, 2004).

2.2 Benefícios diretos: padrões de uso das línguas

Berber Sardinha (2004) destaca que os padrões detectados pela lingüística de corpus deveriam constar dos dicionários bilíngües, 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) e "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" (p. 250).

Segundo o autor, os padrões revelados a partir da análise de corpus podem ser classificados em três tipos:

- 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).

A descrição desses fenômenos requer um árduo trabalho dos lingüistas. Vejamos o exemplo da prosódia semântica, apontado pelos estudos de Michael Hoey, Michael Stubbs e John Sinclair, dentre outros, no inglês e de Berber Sardinha (2000) no português. Em síntese, a prosódia semântica é a expectativa de

seqüência que orienta a escolha lexical. Essa seqüência não admite apenas uma palavra, mas um conjunto de palavras que compartilham determinada característica. É preciso identificar que tipo de característica é essa a fim de definir a prosódia semântica de uma palavra. O fenômeno ainda não foi extensivamente estudado devido à dificuldade de se estabelecer um método sistemático para fazer isso.

Atkins, Rundell & Sato (2003) demonstram, no entanto, que as ferramentas desenvolvidas pelo projeto *Framenet* (baseado na semântica de *frames* de Fillmore), para exploração de um corpus organizado em forma de base de dados textual, aceleram várias das etapas do trabalho dos lexicógrafos, inclusive a identificação de prosódias semânticas. Afirmam, no entanto, que "há muito a ser feito ainda em termos de transformar *insights* sobre prosódia semântica em informação de dicionário que seja útil e aplicável[3]" (p. 342).

Os exemplos de prosódia semântica mais citados são relacionados a verbos, mas nos perguntamos: todas as categorias gramaticais possuem prosódia semântica? Seria interessante informar sempre a prosódia semântica em um dicionário bilíngüe ou apenas quando ela fosse diferente entre as línguas? No sempre citado caso do verbo *cometer*, que tem prosódia semântica negativa em português, qualquer palavra que tenha "sentido" negativo poderia preencher a lacuna? Poderíamos dizer, por exemplo, *cometer uma desgraça*? Parece-nos que não. Então, talvez, a informação da prosódia semântica no dicionário bilíngüe pudesse, inclusive, induzir a um erro de seleção. Não seria, então, mais interessante exibir um resumo das colocados mais freqüentes junto à palavra-chave ao invés de fazer generalizações?

Ainda não temos respostas para essas perguntas, mas é indiscutível a relevância do que vem sendo observado no comportamento combinatório do léxico.

Atkins, Rundell & Sato (2003) mostram também os benefícios da *Framenet* para a desambigüização e para a detecção do padrão lógico de seleção de equivalentes no caso de anisomorfismo. Os autores citam os verbos *to say* e *to tell*, do inglês, que constituem grande dificuldade para os aprendizes de outras línguas européias, pois possuem um único verbo equivalente nas respectivas línguas maternas (francês *dire*; espanhol *decir*, italiano *dire*, português *dizer*, alemão *sagen* etc). Pelo estudo baseado em corpus, verificou-se que *told* sempre especifica um receptor da mensagem, enquanto que *say* poucas vezes o faz. A seleção de *told* ocorre quando se quer dar destaque à pessoa que está recebendo a informação, enquanto a seleção de *say* ocorre quando se quer focar a atenção na informação que está sendo transmitida.

Os autores fornecem ainda um exemplo contrário, em que o francês tem dois equivalentes para uma palavra do inglês: o verbo *cook* deve ser traduzido por *cuire* se o sentido for de "aplicar calor a" e por *préparer* se o sentido for de "criação culinária".

Ambas conclusões foram produzidas pela análise da descrição de valências fornecida pela *Framenet*. Existem outros relatos sobre os benefícios lexicográficos de bases textuais baseadas na semântica de *frames*, dentre os quais Baker, Fillmore & Lowe (1998), Fontenelle (2000) e Boas (2005).

A organização em torno de *frames*, além de vir mostrando resultados satisfatórios para auxiliar na descrição do léxico, como no caso da *Framenet*, é adequada para dicionários bilíngües de suporte à produção em língua estrangeira, pois o aprendiz sabe o sentido do que quer dizer e procura a forma de dizê-lo. Essa parece ser a lógica na organização daquele que se proclama o "primeiro dicionário para produção do mundo", o *Longman Language Activator* (LONGMAN, 1996) que é, no entanto, monolíngüe.

Com o avanço das pesquisas, podemos observar que os corpora vêm ganhando uma rica etiquetagem externa (atributos de cada texto como um todo) e interna (atributos das palavras ou de segmentos do discurso dentro dos textos). As possibilidades de pesquisa e de extração de informações se multiplicam com um corpus etiquetado, pois ele já agrega o resultado de um trabalho de análise. Cada corpus é etiquetado de acordo com sua finalidade. Seria interessante que um corpus para fins de exploração lexicográfica tivesse seus textos etiquetados com o estilo e o gênero, por exemplo. Já as partes do texto receberiam etiquetas com: **marcas de uso**, ou seja, indicações de que determinado uso da unidade lexical é literário, figurado ou coloquial (marcas diafásicas); regionalismo (marcas diatópicas); familiar, popular, gíria ou linguagem de baixo calão (marcas diastráticas), bem como etiquetas com a **classificação morfossintática**, o **grau de fixidez** das colocações, os **campos semânticos** a que pertence, a **prosódia semântica**, a **valência** etc. Corpora etiquetados podem vir a constituir bases de conhecimento lexical, em que as palavras não são mais apenas palavras, mas dados lexicais relacionados com outros dados. Outra tendência é que as ferramentas para exploração de corpora etiquetados suplantem os já tradicionais concordanciadores em possibilidades de consulta e extração de resultados.

[3] *There is still much to be done in terms of transmutting insights about semantic prosody into useful and usable dictionary text.*

Tudo isso nos leva a acreditar que os dicionários do futuro serão construídos sobre essas sofisticadas bases de conhecimento lexical. Com isso, teríamos a possibilidade de criar diversas consultas pré-formatadas para que o usuário de dicionário pudesse acessar o conhecimento armazenado. Assim, como o usuário não precisaria "ver", de uma só vez, tudo o que está armazenado sobre um item lexical, seriam superados os problemas que restringem o volume das informações selecionadas para constituir um dicionário.

Consultas diferentes poderiam, então, ser disponibilizadas, como por exemplo:

- O usuário testaria uma construção (um sintagma) e o dicionário "responderia" se ela ocorre ou não em seu corpus (como hoje se faz na Internet para confirmar codificações em língua estrangeira).
- O usuário solicitaria a freqüência de cada posicionamento de determinado advérbio (inicial, central, final) ou adjetivo (antes ou depois do substantivo) e suas respectivas paráfrases, a fim de verificar diferenças de sentido.
- O usuário solicitaria as companhias mais freqüentes de uma determinada palavra, por categoria (por exemplo, verbos mais freqüentes com o substantivo "medo": sentir, dar, provocar; substantivos mais freqüentes com o verbo "praticar": *esporte*; (um) *ato* (de), etc.).

3. Considerações finais

Vimos que os resultados das pesquisas baseadas em corpus são em grande parte aproveitáveis para a lexicografia. Alguns estão prontos para constituir informações dicionarísticas, outros são idéias a respeito de propriedades do léxico que ainda têm que ser exploradas em larga escala e outros são revelações sobre as dificuldades enfrentadas pelo público-alvo dos dicionários em sua atividade de produção lingüística e que podem subsidiar decisões de projeto.

Esses resultados podem inspirar no lexicógrafo novas possibilidades de temas e de métodos de pesquisa, inovações dicionarísticas e até uma mudança de paradigma. Pensemos o seguinte: hoje o corpus é fornecedor de insumo que, selecionado, analisado e organizado é armazenado e exibido nos dicionários. Mas com o processo de etiquetagem o próprio corpus já está incorporando o produto das análises do léxico, o que poderia dispensar, no futuro, a necessidade de seleção e armazenamento de informações em outro lugar. O dicionário, por sua vez, em sua forma eletrônica, rompeu as amarras que restringiam o trabalho lexicográfico: diferentemente do que acontece na mídia impressa, o que está armazenado não é necessariamente o que tem que ser exibido. Isso dispensa a necessidade de selecionar muito para fins de economia de espaço. No futuro, portanto, o corpus etiquetado poderia substituir a função de armazenador de dados lexicais do dicionário. Assim, o dicionário ganharia maior flexibilidade para disponibilizar novas formas de busca e exibição das informações solicitadas por seus usuários.

Portanto, a relação entre corpus e dicionário, que poderia nos parecer até de concorrência, quando o ensino de línguas estrangeiras e de tradução passou a sugerir a utilização de corpora como recurso de consultas sobre o uso das línguas (FRANKENBERG-GARCIA, 2000; TAGNIN, 2002; VANTAROLA, 2002), parece-me mais de sinergia. A crítica mais freqüente aos dicionários é a omissão de informações que poderiam ser encontradas no corpus. Ao selecionar material no corpus para exibição no dicionário, o lexicógrafo está sempre condenado a perder alguma coisa.

Mas sem a rígida limitação de espaço para armazenamento dos dicionários impressos, os dicionários eletrônicos e os dicionários *on-line* podem minimizar as perdas de conteúdo inerentes à seleção. Assim, pensando as funções SELECIONAR, ARMAZENAR E EXIBIR do trabalho lexicográfico, poderíamos ter a supressão da primeira, a segunda seria desempenhada pelo próprio corpus e a terceira pelo dicionário.

Parece que a lexicografia começa a caminhar nesse sentido: há dicionários monolíngües, como o *Collins Cobuild on CD-ROM* de 2001 e o *Longman Dictionary of Contemporary English* CD-ROM de 2003, que incorporaram um corpus em seus "pacotes" de recursos (RIZO-RODRIGUEZ, 2004). Já no Brasil, temos o Dicionário de Produção Português-Espanhol da Universidade de Santa Catarina, constituído por um corpus paralelo e cujas consultas permitem visualizar uma

Nesse cenário, dentre as atividades do lexicógrafo estariam: construir e etiquetar um corpus para fins lexicográficos; definir menus e fluxos de consultas; desenhar telas de consulta e de exibição de resultados, além de empreender pesquisas sobre as necessidades dos usuários de dicionários.

4. Referências Bibliográficas

ATKINS, 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, 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.

BAKER, M. Corpora in Translation Studies: An overview and Some Suggestions for Future Research. **Target** 7:2, p. 223-243. Amsterdam: John Benjamins, 1995.

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, 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.).

BERBER SARDINHA, T. Computador, corpus e concordância no ensino da léxico-gramática de língua estrangeira. In: LEFFA, V. J. (org.) **As palavras e sua companhia – o léxico na aprendizagem**. Pelotas: EDUCAT, 2000, p. 71-94.

BERBER SARDINHA, T. **Lingüística de Corpus**. Barueri: Manole, 2004.

BERBER SARDINHA, T. O corpus de aprendiz BR-ICLE. São Paulo: Puc-Lael, 2001. Disponível em: www.lael.pucsp.br/~tony/2001bricle. Acesso em: 31/07/2005.

BOAS, H. C. Semantic frames as interlingual representations for multilingual lexical databases. **International Journal of Lexicography**, V. 18, n. 4. Oxford University Press, 2005.

CHRISTIANSON, K. Dictionary use by EFL writers: what really happens? **Journal of second language writing**, 6 (1), 23-43. Elsevier, 1997.

FONTENELLE, T. A bilingual lexical database for frame semantics. **International Journal of Lexicography**, v. 13, n. 4. Oxford University Press, 2000.

FRANKENBERG-GARCIA, A. Using a translation corpus to teach English to native speakers of Portuguese. **Journal of Anglo-American Studies** vol. 3: pp 65-78. Associação Portuguesa de Estudos Anglo-Americanos (APEAA) 2000. Disponível em http://www.linguateca.pt/COMPARA/Ana /AnaPublications.html.

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: Na 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 Sudies in English. Lund: Lund University Press, 1996, p.37-51.

HARVEY, K. YUILL, D. A study of the use of a monolingual pedagogical dictionary by learners of English engaged in writing. **Applied Linguistics**, v. 18, n. 3. Oxford University Press, 1997.

HUMBLÉ, P. O dicionário bilíngue 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. Florianópolis, 2003, p. 24-34.

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.; KIMMEL, M. Bilingualised dictionaries: How learners really use them. **System**, n. 25, 3, p. 361-369, 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, v. 3, p.135-155. Oxford University Press, 2006.

LAVIOSA, S. How Comparable can 'Comparable Corpora' be? **Target** 9:2, p. 289-319. Amsterdam: John Benjamins, 1997.

LONGMAN. **Longman Language Activator**. Essex: Longman, 1996.

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.; HAILL, R. A study of dictionary use by international students at a British university. **International Journal of Lexicography**, v. 15, n. 4, p. 277-305. Oxford University Press, 2002.

RIZO-RODRÍGUEZ, A. Current lexicographical tools in EFL: monolingual resources for the advanced learner. **Language Teaching**, 37, 29-46. United Kingdom: CUP, 2004.

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**, 1999, p. 83-98. Disponível em: <http://www.fu-berlin.de/elc/tnp1/SP9dossier.doc tnp 1> Acesso em: 04 jul. 2003.

TAGNIN, S. E. O. A multilingual learner corpus in Brazil. Lancaster: **Corpus Linguistics - Learner Corpus Workshop, 2003**. Disponível em: http://www.fflch.usp.br/dlm/comet/. Acesso em: 14, nov. 2006.

TAGNIN, S. E. O. Os corpora: instrumentos de auto-ajuda para o tradutor. In: Tagnin, S. E. O. (Org.). **Cadernos de Tradução**: Corpora e Tradução. Florianópolis: NUT, 2002, v. 1, n. 9, p. 191-218.

TONO, Y. Learner corpora: design, development and applications. In: Archer et al. (eds.) **Proceedings of the Corpus Linguistics 2003 Conference. Technical Papers 16**. Lancaster University: University Centre for Computer Corpus Research on Language, 2003, p. 800-809.

TOURY, G. **Descriptive Translation Studies and beyond**. Amsterdam/Philadelphia: John Benjamins, 1995 (revisão de 1978), p. 198-211.

VARANTOLA, K. Disposable corpora as intelligent tools in translation. In: TAGNIN, S. E. O. (Org.). **Cadernos de Tradução**: Corpora e Tradução. Florianópolis: NUT, 2002, v. 1, n. 9, p. 171-189.