

Adriane Orenha
adrianeorenya@gmail.com

A compilação de corpora comparáveis na área de negócios e sua relevância para a tradução e terminologia

The compilation of comparable corpora in the business field and its relevance to translation and terminology

RESUMO – Este artigo tem como objetivo ressaltar a importância da Linguística de Corpus para a área da Tradução, em particular a compilação de corpora comparáveis especializados, assim como a prática tradutória. Para tanto, pretendemos relatar as etapas de construção e de organização de um corpus comparável bilíngue na área de Negócios, bem como discutir suas aplicações, no intuito de igualmente realçar sua relevância para dois de nossos públicos-alvo: tradutores e pesquisadores da área de Terminologia/ Terminografia.

Palavras-chave: linguística de corpus, corpora comparáveis especializados, tradução, negócios, terminologia, terminografia.

ABSTRACT – This article aims at highlighting the importance of Corpus Linguistics particularly to the compiling of specialized comparable corpora to the field of Translation as well as to the practice of translation itself. Hence, we report the stages of the compilation and the organization of bilingual comparable corpora in the Business field and its applications, with the purpose of also highlighting its relevance to two of our target audiences: translators and also researchers in the Terminology/Terminography field.

Key words: corpus linguistics, specialized comparable corpora, translation, business, terminology, terminography.

Introdução

Neste artigo, pretendemos salientar a importância da interdisciplinaridade entre a Tradução e a Linguística de Corpus, confluência cada vez mais promissora para pesquisadores envolvidos nas referidas áreas.

O corpus linguístico vem sendo amplamente utilizado para fins investigativos da linguagem e, desse modo, tem sido considerado uma ferramenta de apoio para a Tradução, assim como um recurso imprescindível para o tradutor, levando em consideração que a busca por novas ferramentas e tecnologias é um fator primordial para um bom profissional da área em questão. Bowker (2000) defende que o tradutor que não utilizar corpora eletrônicos em suas traduções, poderá ser considerado negligente. Segundo a autora, embora corpora e ferramentas de processamento de corpus possam não substituir o tradutor, o tradutor que utilizar tais recursos poderá substituir aqueles que não os utilizam.

Krishnamurthy (1996) comparou o advento dos enormes corpora à invenção do telescópio, por fornecerem o instrumental necessário para o estudo do léxico. Atualmente, com o auxílio dos corpora eletrônicos e de programas computacionais para análise linguística, podemos descobrir, em virtude do fácil acesso a contextos diversos e da rapidez com que são realizadas as análises

linguísticas, regularidades e, até mesmo, irregularidades no uso da língua, o que antes era impossível, na prática. Conforme Ortiz (Ortiz *et al.*, 1998), os corpora podem, inclusive, ajudar-nos a ver as irregularidades de maneira mais uniforme, com uma perspectiva mais ampla e com índices de frequência relativa mais confiáveis.

Nesse sentido, descrevemos, neste artigo, a construção de corpora comparáveis bilíngue na área de jornalismo de Negócios e seu potencial para a Tradução. Os corpora em questão foram construídos com artigos extraídos dos jornais *Gazeta Mercantil* (<http://www.gazetamercantil.com.br>) e *Financial Time* (www.financialtime.com), e das revistas *Exame* (<http://www.exame.com.br>) e *Businessweek* (<http://www.businessweek.com>), por serem específicos da área de Negócios, conjuntamente com os jornais *Estadão* (<http://www.estadao.com.br>) e *The New York Times* (<http://www.nytimes.com>), e as revistas *Veja* (<http://www.veja.com.br>) e *Time* (<http://www.time.com>), selecionados conforme resultado apontado por uma pesquisa feita com alunos, executivos de empresas em que lecionávamos – Avanade, Cargill S.A., Cooperpower Systems; AT&T–, entre outras. Tal pesquisa tinha como propósito verificar os jornais e revistas que os referidos alunos mais liam em português e inglês (Orenha, 2004b). A título de ilustração, apresentamos as Figuras 1 e 2, com o desenho dos referidos corpora comparáveis.

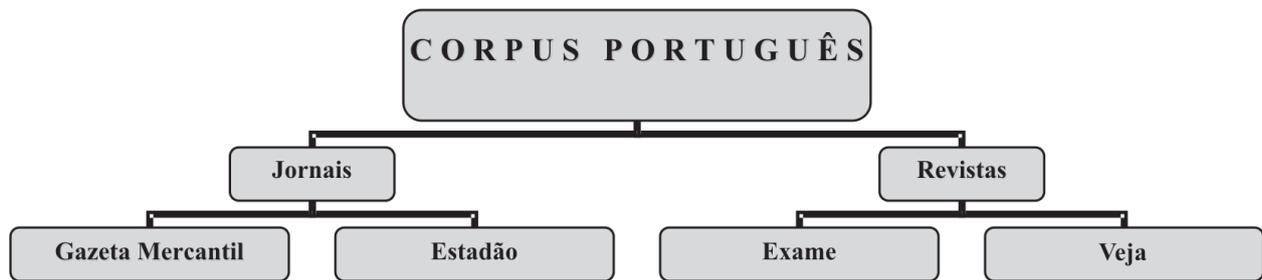


Figura 1: Composição do corpus comparável do português.

Figure 1: Composition of the Portuguese comparable corpus.

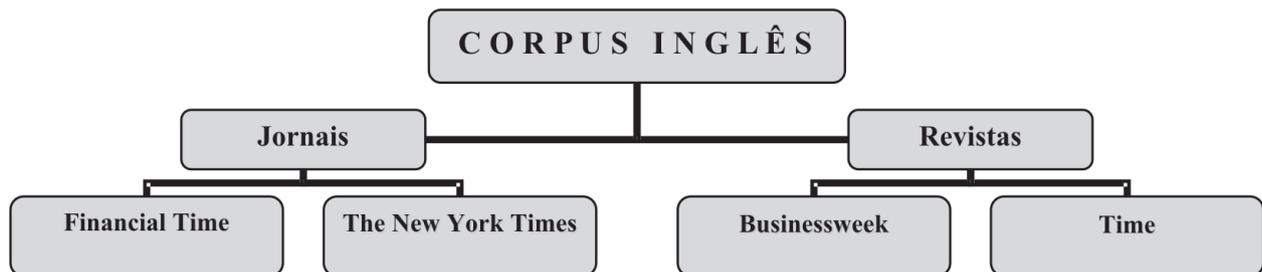


Figura 2: Composição do corpus comparável do inglês.

Figure 2: Composition of the English comparable corpus.

No que se refere ao número total de palavras/ itens (*tokens*), o corpus do inglês possui aproximadamente 553.000, e o corpus do português apresenta cerca de 568.000, perfazendo um total de 1.106.000 palavras, resultado este que permite classificá-lo como sendo um corpus de porte médio-grande. O critério utilizado para a escolha das seções dos jornais e das revistas foi o fato de abordarem assuntos relacionados à grande área de Negócios.

Quanto à disposição interna do corpus, trata-se de corpora comparáveis bilíngues, conforme mencionamos anteriormente. Empregamos o termo “comparáveis”, uma vez que contém textos de diferentes línguas (inglês e português, no caso desta pesquisa), os quais foram selecionados, em razão de possuírem algumas características em comum que, no âmbito deste artigo, estão todas relacionadas à área de Negócios. De acordo com Ulrych (1997), corpora comparáveis bilíngues são formados por textos originais escritos de maneira independente, em línguas diferentes, selecionados com base em fatores de similaridade de conteúdo, gênero, função comunicativa, tamanho etc. Segundo Bowker e Pearson (2002), a única característica que distingue, em um corpus comparável, um conjunto de textos de outro, é a língua na qual os textos estão escritos. Os corpora em questão são formados por textos escritos em inglês e em português. As característi-

cas semelhantes em um corpus comparável incluem, de modo geral, o assunto ou tópico. Também podem incluir características como: tipo de texto, período em que os textos foram escritos, grau de tecnicidade, etc. Nesta investigação, ambos os corpora de pesquisa trazem textos de linguagem jornalística, na área específica de Negócios, coletados no mesmo período (2002-2003) e, na medida do possível, são equivalentes no que concerne ao grau de tecnicidade dos textos.

Objetivos da construção dos corpora comparáveis

Bowker (1998) realizou uma pesquisa com um grupo de alunos aspirantes a tradutores, os quais traduziram o mesmo texto de uma dada linguagem de especialidade. Porém, um grupo utilizou recursos mais convencionais como, por exemplo, dicionários monolíngues e bilíngues, manuais e folhetos; outro grupo, além do dicionário bilíngue, contou com o auxílio de um corpus especializado monolíngue eletrônico e da ferramenta *WordSmith Tools*, de Scott (2004). Com base nos resultados, a pesquisadora notou que o grupo que utilizou o corpus especializado como recurso para auxiliar nas traduções apresentou traduções melhor elaboradas, além de uma melhora em relação à habilidade linguística e à consciência da língua.

Por isso, acreditamos que esta pesquisa pode atestar a necessidade de se incorporar o uso de corpora especializados na Tradução.

Traduzir e, principalmente, verter não é uma tarefa simples. No entanto, quando se trata de traduzir ou verter um determinado texto em uma dada linguagem de especialidade, essa tarefa se torna ainda mais difícil, especialmente se considerarmos a falta de domínio do tradutor em um contexto especializado, bem como os poucos recursos disponíveis para realizar tais traduções ou versões (como exemplo desses recursos, podemos citar dicionários de linguagem de especialidade eficazes, além de dicionários de combinatória igualmente especializados).

Foi neste contexto que entrou em cena a compilação dos referidos corpora comparáveis bilíngues. Um de seus objetivos é disponibilizar para tradutores, principalmente para tradutores pouco experientes, um material de pesquisa que retrate a linguagem em uso. Destacamos que essa disponibilidade será feita apenas para fins de pesquisa acadêmica, devido a questões relacionadas a direitos autorais.

Corpora comparáveis podem, também, ser utilizados com o propósito de buscar equivalentes na tradução de textos especializados como, por exemplo, na área em questão – a de Negócios. Além de localizar os equivalentes por meio das linhas de concordâncias geradas, o tradutor aprendiz ou o profissional da área de tradução tem a vantagem de observar, em ambas as línguas, os mais diversos contextos em que os termos buscados estão inseridos e, desse modo, decidir pela opção de tradução mais apropriada ao contexto pesquisado.

A necessidade da compilação e o potencial de corpora comparáveis especializados para a Tradução e Terminologia/Terminografia

Por uma questão de delimitação, não entraremos em detalhes acerca das vantagens teóricas do uso e da exploração de corpora paralelos em estudos de Tradução, no que se refere, por exemplo, às características dos textos traduzidos (simplificação, explicitação, normalização e nívelamento), propostos na literatura e discutidas por Baker (1995, 1996). Entretanto, deteremo-nos nas questões práticas, no sentido de como os corpora eletrônicos, mais especificamente, os comparáveis, podem auxiliar no processo de tradução.

Primeiramente, destacamos o caráter prático dos corpora, uma vez que o tradutor passa a ter um rápido acesso aos dados linguísticos necessários para uma tradução. Ao gerarmos, por exemplo, uma lista de palavras ou uma lista de palavras-chave, tanto na língua-alvo quanto na língua fonte, podemos encontrar termos equivalentes na linguagem de especialidade em questão. Essas listas podem ser geradas com o auxílio do programa computacional *WordSmith Tools* (2004), o qual pode ser adquirido pelo acesso ao *site* <http://www.lexically.net/wordsmith>.

Um corpus comparável bilíngue igualmente pode ajudar a identificar e levantar, de maneira muito mais rápida, colocações e unidades multipalavras (Orenha, 2003, 2004a, 2004b; Tagnin, 2000), já que tanto dicionários de língua geral quanto de linguagem especialidade fracassam nesse propósito, segundo aponta pesquisa realizada por Orenha (2002). A título de ilustração, citamos um levantamento de colocações que fizemos a partir do termo *business*, em nosso corpus comparável, subcorpus de inglês, a partir do qual conseguimos arrolar 249 colocações. Entre as 249 colocações, os tipos mais frequentes de colocações levantadas foram as nominais (96 colocações), seguidas pelas adjetivas (90) e pelas verbais (57); sendo 49 formadas com o termo/base *business*, com função de objeto, e 8 colocações com o termo/base *business*, com função de sujeito. Foram levantadas também 6 colocações formadas por um quantificador combinado com o termo/base *business*. Nota-se que, apenas com um termo (*business*), foi-nos possível superar o número de colocações inseridas em qualquer dicionário de língua geral ou de linguagem de especialidade atualmente disponíveis. Dessa maneira, evidencia-se que corpora comparáveis são fontes de pesquisa riquíssimas, em razão de fornecem subsídios necessários para a compilação de obras léxico-terminográficas, uma vez que tais colocações foram inseridas em um glossário compilado por Orenha (2004a). As Tabelas 1 e 2 mostram alguns exemplos de colocações extraídas do corpus comparável do inglês.

Além disso, por meio de uma análise de listas de concordâncias é possível, a partir de um corpus comparável, fornecer ao tradutor contexto necessário até mesmo para buscar definições e/ou explicações de termos desconhecidos na área desejada. Ademais, o tradutor poderá investigar o estilo de linguagem do texto inserido nesta linguagem de especialidade. A fim de ilustrar a questão de análise de contexto para uma possível definição de um termo, veja Figura 3, com linhas de concordância da colocação *to win a business* (= “atrair negócios”), geradas a partir da ferramenta *Concord*, do programa *WordSmith Tools*.

Outrossim, podemos levantar estruturas morfosintáticas de colocações, ou seja, uma vez definida a base como o núcleo da colocação, é possível delimitar as unidades de significado funcionalmente completas – “*functionally complete units of meaning*” (Tognini-Bonelli, 2001) –, determinando quais os outros elementos que a acompanham, de maneira que formem um padrão regular e constante. Vejamos, resumidamente, algumas das estruturas de colocações verbais encontradas em nosso corpus comparável para os casos em que o substantivo é objeto de verbo em inglês, tendo em mente as quatro formas possíveis de colocações verbais: verbo _{colocado} + substantivo_{base}; substantivo_{base} + verbo _{colocado}; verbo _{colocado} + Preposição + substantivo_{base}; verbo _{colocado} + adjetivo_{base}.

Tabela 1: Colocações nominais mais frequentes no corpus comparável do inglês.**Table 1:** Most frequent nominal collocations in the English comparable corpus.

COLOCAÇÕES NOMINAIS	FREQUÊNCIA NO CEI
<i>business school</i> (faculdade de administração de empresas)	13
<i>business world</i> (mundo dos negócios)	12
<i>business model</i> (modelo empresarial)	11
<i>business travel</i> (viagem de negócios)	11
<i>business investment</i> (investimento comercial)	11
<i>business fare</i> (tarifa comercial)	8
<i>business leader</i> (líder empresarial)	8
<i>business plan</i> (plano de negócios)	8
<i>business practice</i> (prática comercial)	8

Tabela 2: Colocações verbais mais frequentes no corpus comparável do inglês.**Table 2:** Most frequent verbal collocations in the English comparable corpus.

COLOCAÇÕES VERBAIS	FREQUÊNCIA NO CEI
<i>to do business</i> (fazer negócios)	22
<i>to start a business</i> (iniciar um negócio)	7
<i>to run a business</i> (dirigir um negócio)	6
<i>to win a business</i> (atrair negócios)	6
<i>to operate a business</i> (operar um negócio)	5
<i>to grow a business</i> (ampliar, expandir um negócio)	4

Concord	
File Edit View Compute Settings Windows Help	
N	Concordance
1	People to Watch in International Business A software winner; Banesto's new chair; an
2	Buffalo, N.Y., for its chicken wings, but in the airline business it's famous for ferocious winter storms that
3	he recalls, "I swore I was going to get out of the wine business." But he still had the stuff in his veins. After
4	(as it does to the chains), but Sullivan wins additional business through service. His sales force helps people
5	in Pembroke Pines, Fla., should have a going out of business sign in its window. A family-owned shop that
6	(as it does to the chains), but Sullivan wins additional business through service. His sales force helps people
7	in Pembroke Pines, Fla., should have a going out of business sign in its window. A family-owned shop that
8	executives in a bribelike manner to win banking business, devising complex financing to help Enron
9	it has yet to win general acceptance from major business application makers. "It's moving the ball
10	investors by issuing biased research to win banking business. Spitzer has also accused the firm of bribing
11	to tarnish both UBS Warburg's ability to win new business in Malaysia and Tenaga's standing in
12	needed help. "I wasn't programmed to be in the wine business," she said. "Italian girls get married and have
13	watchdog. "He's using his position on the board to win business." Mr Perle said he was only advising Global
14	investors. It has since picked up 7% of the lucrative business. So who's winning? Right now, it's cheaper to
15	Of course, battered rivals are also chasing services business and winning contracts, such as a \$40 million
16	didn't issue rosy research to win investment banking business. Concerns over conflicts are part of what drove

Figura 3: Concordância da colocação *win a business*.**Figure 3:** Concordance of the collocation *win a business*.

V + N = *to establish a company*
 V + N_(sing) = *to do business*
 V + N_(U) = *to waste money*
 V + N_(pl) = *to issue shares; to own shares*
 V + PA + N = *to take over a company*
 V + PA + N_(pl) = *to dole out shares, to prop up shares*
 V + N_(pl) + PA = *to push shares down*
 V + N_(U) + PA = *to draw money in*
 V + N_(U) + Prep = *to invest money in sth*
 V + PA + Prep + N_(pl) = *to hold on to shares*

Conforme pode ser notado, as vantagens acima apresentadas, entre outras não mencionadas no escopo deste artigo, levam a um caminho: os corpora eletrônicos possibilitam um aumento da velocidade de pesquisa e do desempenho por parte do tradutor. Além disso, as traduções e versões baseadas em corpora provavelmente serão um reflexo da linguagem em uso, objetivo primeiro da Linguística de Corpus.

Segundo sustenta Zanettin (1998) e Pearson (1996), corpora eletrônicos constituídos por textos especializados são fontes ricas em terminologia e em conceitos. De acordo com Zanettin (1998), corpora comparáveis podem ser utilizados em sala de aula para confirmar hipóteses de tradução, além de oferecer possíveis soluções para problemas de tradução encontrados em determinados textos especializados.

Considerações finais

Por meio da exposição da necessidade de se aplicar os recursos oferecidos pela Linguística de Corpus na área de Tradução em uma dada área de especialidade, bem como pelas várias sugestões de uso de corpora comparáveis apresentadas neste artigo, esperamos ter permitido ao leitor vislumbrar a relevância do uso e da exploração de corpora para a Tradução, para que, futuramente, estes profissionais venham fazer uso de tais recursos em sua própria prática tradutória produzindo, assim, traduções de melhor qualidade e, conseqüentemente, tornando-se tradutores mais fluentes.

No que concerne às contribuições relativas às áreas de Terminologia (na pesquisa terminológica) e Terminografia (no fazer terminográfico), acreditamos que o uso e a exploração de corpora comparáveis especializados possibilitam um trabalho mais rápido e resultados mais confiáveis. Além disso, corpora comparáveis especializados podem tornar-se fontes de pesquisa valiosas e extremamente úteis para profissionais da área em questão.

Referências

BAKER, M. 1995. Corpora in Translation Studies: an overview and some suggestions for future research. *Target*, 7(2):223-243.

- BAKER, M. 1996. Corpus-based translation studies: the challenges that lie ahead. In: H. SOMERS (ed.), *Terminology, LSP and Translation Studies in Language Engineering in Honor of Juan C. Sager*. Amsterdam/Philadelphia, John Benjamins Publishing Co., p. 177-186.
- BOWKER, L. 1998. Using specialized monolingual native-language corpora as a translation resource: a pilot study. *Meta*, XLIII(4):631-651.
- BOWKER, L. 2000. Towards a methodology for exploiting specialized target language corpora as translation resources. *International Journal of Corpus Linguistics*, 5:17-52.
- BOWKER, L.; PEARSON, J. 2002. *Working with specialized language: a practical guide to using corpora*. London, Routledge, 242 p.
- KRISHNAMURTHY, R. 1996. Keeping good company: collocation, corpus and dictionaries. *Cicle de Conferències 95-96 Lèxic, corpus i diccionaris*, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, p. 31-56.
- ORENHA, A. 2002. *Uma análise comparativa das colocações em dois dicionários de língua geral, um dicionário de linguagem especializada na área de negócios, dois dicionários de colocações, e três corpora eletrônicos*. São Paulo, SP. Monografia. FFLCH/USP, 28 p.
- ORENHA, A. 2003. A fraseologia e a elaboração de um glossário bilíngüe de colocações na área de negócios. *Estudos Linguísticos*, XXXI. Disponível em: <http://www.gel.org.br/estudoslinguisticos/volumes/32/htm/comunica/cc002.htm>. Acesso em 25/11/2004.
- ORENHA, A. 2004a. *A compilação de um glossário bilíngüe de colocações, na área de jornalismo de Negócios, baseado em corpus comparável*. São Paulo, SP. Dissertação de Mestrado. Universidade de São Paulo, USP, 236 p.
- ORENHA, A. 2004b. O uso de corpora e o ensino de ESP. *Estudos Linguísticos*, XXXII:1036-1041.
- ORTIZ, A.M.; HERNANDEZ, C.P.; BENÍTEZ, P.F. 1998. Lexicografia Computacional y lexicografia de corpus. *Revista Española de Lingüística Aplicada (RESLA)*, Volumen monográfico:175-214.
- PEARSON, J. 1996. Teaching Terminology using Electronic Resources. In: S. BOTLEY; J. GLASS; T. MCENERY; A. WILSON (eds.), *Proceedings of Teaching and Language Corpora 1996*. UCREL Technical Papers, 9. Lancaster, UCREL, p. 203-216.
- SCOTT, M. 2004. *WordSmith Tools: version 4.0*. Oxford, Oxford University Press.
- TAGNIN, S.E.O. 2000. Collecting data for a bilingual dictionary of verbal collocations: from scraps of paper to corpora research. In: B. LEWANDOWSKA-TOMASZCZYK; P.J. MELIA (eds.), *PALC '99: Practical Applications in Language Corpora. Papers from the International Conference at the University of Lodz*. Frankfurt am Main: Peter Lang GmbH, p. 399-407.
- TOGNINI-BONELLI, E. 2001. *Corpus Linguistics at Work. Studies in Corpus Linguistics*, vol. 6. Amsterdam, John Benjamins Publishing Company, 223 p.
- ULRYCH, M. 1997. The impact of multilingual parallel concordancing on translation. In: B. LEWANDOWSKA-TOMASZCZYK; P.J. MELIA (eds.), *PALC '97: Practical Applications in Language Corpora. Papers from the International Conference at the University of Lodz*. Lodz University Press, p. 421- 436.
- ZANETTIN, F. 1998. *Bilingual comparable corpora and the training of translators*. *Meta*, 43(4):616-630.
- WORDSMITH TOOLS. 2004. Oxford University Press. Disponível em <http://www.lexically.net/wordsmith/index.html>. Acesso em: 12/12/2008.

Submissão: 31/05/2009
 Aceite: 27/10/2009

Adriane Orenha

UNESP – IBILCE/Campus de São José do Rio Preto
 Rua Cristóvão Colombo, 2265, Jardim Nazareth
 15054-000, São José do Rio Preto, SP, Brasil