

CORPORA LINGÜÍSTICOS COMO FERRAMENTAS DE APOIO À TRADUÇÃO DE COLOCAÇÕES

MIRIAM ANGEL GOLDSCHMIDT¹; ROBERTA REGO RODRIGUES²

¹Universidade Federal de Pelotas – miriam.golds01@gmail.com

²Universidade Federal de Pelotas – betareseau@gmail.com

1. INTRODUÇÃO

O termo colocação, introduzido por J.R. Firth em 1957, designa combinações entre palavras que desenvolveram uma relação semântica baseada em sua coocorrência frequente (BUSSMANN, 1998). Essa combinação de palavras, segundo BAKER (2006), é própria de cada idioma, de modo que nem sempre as colocações de uma língua equivalem às exatas colocações de outra. Assim, no processo de tradução, enfrenta-se o problema de encontrar os termos mais usuais da língua de chegada e não mesclá-los com os da língua de partida, pois, como observado por TAGNIN (2002), é a correta utilização desses termos que oferece naturalidade a uma língua. São exemplos de colocações em língua portuguesa combinações tais como: *disparar um alarme, fiel escudeiro, mentira deslavada, chorar copiosamente*¹.

O tema das colocações é amplamente tratado em bibliografia de língua inglesa, tendo-se conhecimento de inúmeros trabalhos e obras lexicográficas nesse idioma². No entanto, os estudos dedicados a esta área em língua espanhola e em língua portuguesa ainda são incipientes. Dessa forma, ao trabalhar com esse par de línguas, o tradutor normalmente se vê com dificuldades para decidir qual a melhor seleção lexicográfica devido à escassez de materiais de consulta.

Este estudo propõe-se a trabalhar o tema das colocações a partir da problemática da tradução, em contrapartida à abordagem habitualmente empregada que visa ao ensino e à aprendizagem de línguas. Acredita-se que este trabalho possa contribuir para os estudos tradutórios oferecendo mais uma metodologia e ferramenta de pesquisa capazes de auxiliar o tradutor a encontrar soluções e facilitar o processo da tradução de colocações.

Para alcançar tal objetivo, optou-se pela metodologia proposta pela Linguística de Corpus, utilizando-se, como objeto de pesquisa, o corpus linguístico legível por computador, por meio do qual se podem verificar diferentes fenômenos linguísticos, dos mais comuns aos mais infrequentes (MCENERY, 2014). A Linguística de Corpus é um dos campos que mais influenciam a pesquisa linguística contemporânea, uma vez que coloca “à disposição do analista quantidades de dados antes inacessíveis” (BERBER SARDINHA, 2006, p. 06).

Vários autores, como TAGNIN (2002, 2007) e VARANTOLA (2002), sugerem a utilização de corpora linguísticos como ferramentas de apoio à tradução, pois, como mencionado por VARANTOLA (2002, p. 180), “o corpus destaca convenções terminológicas, oferece ao escritor vocabulário periférico

¹ Exemplos de ALLEGRO et al. (2010).

² São exemplos o *Oxford Collocations Dictionary*, da Oxford University Press; o *Longman Collocations and Thesaurus*, da Editora Pearson; e o *Happy Couples: dicionário de colocações lexicais adjetivas (inglês-português/ português-inglês)*, da editora Disal.

adicional e sugere formas idiomáticas de expressar ideias³". Com esse propósito, os critérios de compilação de corpus se dariam de acordo com os objetivos da tradução. O tradutor seria capaz de compilar corpora linguísticos e utilizá-los conforme sua necessidade, visto que são inúmeras as possibilidades de pesquisa passíveis de realização a partir deles. Esta seria uma maneira de suprir a escassez de recursos lexicográficos em certas áreas, certos pares de línguas e certas estruturas linguísticas, como as colocações.

O objetivo específico deste trabalho, então, é analisar a possibilidade do uso de corpora comparáveis bilíngues para a validação de traduções de colocações lexicais que tenham um substantivo como base entre o par de línguas espanhol-português. Para tanto, será feita a identificação das possíveis colocações desse tipo através de pesquisa com corpus em língua espanhola e a posterior validação dos termos traduzidos em um corpus em língua portuguesa, utilizando a ferramenta gratuita de manipulação de corpus *AntConc*. Tenciona-se, assim, verificar a utilidade do uso de corpora linguísticos como ferramentas de apoio à tradução.

Acredita-se que, se o tradutor fizer uso de corpora comparáveis bilíngues para a validação de traduções de colocações lexicais entre o par de línguas espanhol-português, ele poderá beneficiar-se, no que diz respeito à qualidade das traduções, e será uma maneira de suprir, em parte, a falta de material lexicográfico no par de línguas mencionado.

2. METODOLOGIA

BERBER SARDINHA (1999, s/p.) conceitua as colocações como a “coocorrência significativa de itens lexicais verificada computacionalmente em um corpus eletrônico”. Antes do advento do método de pesquisa baseado em corpus, era necessário observar a recorrência desses itens através da leitura, até que se percebesse que se tratava de unidades fixas. Hoje em dia, por meio da manipulação de grandes quantidades de texto, essas unidades convencionais são facilmente identificáveis (TAGNIN, 2005). Assim, a metodologia da Linguística de Corpus apresenta-se como ideal para a realização deste trabalho.

MCENERY e HARDIE (2012) discorrem sobre a possibilidade da utilização da internet como um corpus, por meio de mecanismos de busca, como o *Google*, ou de ferramentas especialmente desenvolvidas para essa finalidade, como o *WebCorp*⁴, que apresenta os resultados da busca em linhas de concordância. Essa abordagem, no entanto, apresenta problemas inerentes à própria natureza da internet, como a possibilidade de qualquer usuário expor seus textos *online*, sem zelo gramatical ou ortográfico, e a presença de materiais resultantes da tradução automática de textos escritos originalmente em língua estrangeira, oferecendo interferência lexical e gramatical aos resultados obtidos, entre outros.

Apesar do supracitado, segundo os mesmos autores, pesquisas utilizando a internet como corpus são vantajosas em vista do volume de textos disponíveis para pesquisa e da possibilidade de identificação de léxico e de combinações de baixa frequência, que dificilmente seriam encontrados até mesmo em corpora de grande extensão. Textos retirados da internet, cujas fontes tenham sido

³ “The corpus highlights the terminological conventions, it gives writers additional peripheral vocabulary and suggests idiomatic ways of expressing ideas” (tradução minha).

⁴ Disponível em: <<http://www.webcorp.org.uk/live/>>.

cuidadosamente selecionadas, são adequados para a compilação de corpora para diferentes propósitos.

Dessa maneira, optou-se pela captação de textos disponíveis *online*, retirados dos sites de oito renomados jornais brasileiros e estrangeiros, para a construção de um corpus comparável bilíngue, constituído, como definido por FRANKENBERG-GARCIA (2006), por dois subcorpora de textos produzidos originalmente nas línguas A e B (não traduzidos) e de mesmo gênero textual e tipo. Esse tipo de corpus é considerado pela autora como funcional para pesquisa terminológica no dia a dia do tradutor.

A técnica utilizada para a compilação do corpus foi a de *download manual*, referida por MCENERY (2014). Foram selecionados os textos redirecionados pelas manchetes da versão online dos jornais e copiados para arquivos do bloco de notas (formato *.txt*). Para o subcorpus de língua espanhola, utilizaram-se textos dos jornais *El País*, do Uruguai; *Clarín*, da Argentina, *ABC Color*, do Paraguai; e *El Universal*, da Venezuela; para o de língua portuguesa, dos jornais brasileiros *O Estado de São Paulo*, *Zero Hora*, *Folha de São Paulo* e *Estado de Minas*. O subcorpus em língua espanhola servirá para a extração das possíveis colocações mais frequentes nessa língua, enquanto o subcorpus em língua portuguesa será utilizado para a validação da tradução dessas colocações.

Para a identificação das colocações presentes nos textos deste corpus, será feita uma análise quantitativa e qualitativa através do programa de manipulação de corpus *AntConc* (versão 3.3.4w). Os candidatos à colocação serão selecionados através de sua frequência de coaparição.

A verificação das colocações se dará, em primeiro lugar, com a utilização da função *Word list*, com a qual se obterá a lista das palavras por ordem de frequência de aparição no texto. Em seguida, será efetuada a busca dos substantivos presentes no texto, por ordem de aparição. Serão localizados os substantivos, pois, segundo HAUSMANN (1985, *apud* TAGNIN, 2002) as colocações são formadas por não mais que duas palavras, sendo uma delas a base, a palavra de maior carga semântica e a outra o *colocado*, palavra que acompanha a base. Como, neste caso, os tipos de colocações escolhidos têm como base um substantivo, é interessante que a pesquisa parte desses termos.

Segundo TAGNIN (2005), a ferramenta *KWIC* (*Key Word in Context* – palavra-chave em contexto) é a que melhor permite a observação de estruturas recorrentes, pois apresenta os resultados na forma de concordância, exibindo a palavra em que se tem interesse no centro da tela, dentro de seu contexto natural de ocorrência. Os padrões de recorrência podem ser verificados a partir da observação das palavras que ocorrem à direita ou à esquerda da palavra em questão. Dessa forma, por meio do uso da função *Concordance*, como é denominada a ferramenta *KWIC* no programa utilizado, serão verificados os substantivos em contextos, e poderão ser observadas e quantificadas as duplas de palavras que aparecem juntas com maior frequência.

A partir da lista de colocações obtida pela pesquisa, será realizada a tradução desses sintagmas por seus correspondentes usuais em língua portuguesa (por meio de dicionários, corpus de língua portuguesa e demais pesquisas necessárias). Finalmente, de posse das traduções das estruturas, será efetuada sua validação no corpus de textos em português, a qual poderá ser feita pela colocação completa ou partindo-se de sua base.

3. RESULTADOS E DISCUSSÃO

A pesquisa encontra-se em andamento. Por conseguinte, ainda não apresenta resultados.

4. CONCLUSÕES

Este trabalho está em andamento. Portanto, ainda não apresenta conclusões.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ALLEGRO, Alzira L. V.; BREZOLIN, Adauri; MOBAID, Rosalind. **Happy couples dicionário de colocações lexicais adjetivas:** português-inglês, inglês-português. Barueri, SP: DISAL, 2010.

BAKER, Mona. **In other words:** a coursebook on translation. London: Routledge, 2006.

BERBER SARDINHA, Tony. **Estudo baseado em corpus da padronização lexical no português brasileiro:** colocações e perfis semântico. 1999. Disponível em: <<http://www2.lael.pucsp.br/~tony/temp/publications/>>. Acesso em: 25 out. 2013.

_____. **Pesquisa em Lingüística de Corpus com WordSmith Tools.** 2006. Disponível em: <http://sis.posugf.com.br/AreaProfessor/Materiais/Arquivos_1/13879.pdf>. Acesso em: 15 dez. 13.

BUSSMANN, Hadumod. **Routledge dictionary of language and linguistics.** Routledge, 1998. (formato e-book).

FRANKENBERG-GARCIA, Ana. **Corpora e Tradução.** Material didático utilizado na primeira Escola de verão da Linguateca, Módulo 3, 2006. Disponível em: <<http://www.linguateca.pt/escolaverao2006/Corpora/EDV2006CorporaeTraducao.pdf>>. Acesso em: 20 nov. 2013.

MCENERY, Tony. **Corpus linguistics:** method, analysis, interpretation. Curso online. Lancaster: Future Learn, 2014. Disponível em: <<https://www.futurelearn.com/courses/corpus-linguistics/>>. Acesso em: 3 abr. 2014.

MCENERY, Tony; HARDIE, Andrew. **Corpus linguistics:** method, theory and practice. Cambridge University Press, 2012. (formato e-book).

TAGNIN, Stella. Os Corpora: instrumentos de auto-ajuda para o tradutor. **Cadernos de Tradução**, Florianópolis, v. 1, n. 9, 191-219, 2002/1.

_____. **O jeito que a gente diz:** expressões convencionais e idiomáticas. São Paulo: Disal, 2005.

_____. A Identificação de equivalentes tradutórios em corpora comparáveis. In: **Anais do I Congresso Internacional da ABRAPUI.** Belo Horizonte, 2007. Disponível em: <http://www.fflch.usp.br/dlm/comet/Novo/Stella_Abrapui%202007_artigo.pdf>. Acesso em: 25 out. 2013.

VARANTOLA, Krista. Disposable corpora as intelligent tools in translation. **Cadernos de Tradução**, Florianópolis, v. 1, n. 9, 171-189, 2002/1.