CONTRIBUIÇÕES DA LINGUÍSTICA DE *CORPUS* PARA O PROCESSO DE ENSINO/APRENDIZAGEM DE IDIOMAS

Thereza Cristina de Souza Lima (UNINTER)
thereza.l@uninter.com

RESUMO

No que diz respeito à educação, são notórias as modificações pelas quais o mundo moderno vem passando a partir do desenvolvimento da tecnologia, mais especificamente em relação ao processo de ensino–aprendizagem de língua estrangeira. Um dos exemplos expressivos situa-se no campo das contribuições da Linguística do *Corpus* com vistas, principalmente, ao aprimoramento do aspecto lexical da língua alvo. Os enormes *corpora* disponibilizados na *internet* constituem uma inovação extremamente positiva, uma vez que os alunos têm a possibilidade de observar a língua em contexto real de uso e analisar as linhas de concordância em relação a vários aspectos, tais como, gramatical, semântico, pragmático entre outros. O objetivo desta comunicação, portanto, é demostrar a possibilidade de enriquecimento lexical do discente por meio da observação e descrição da língua alvo, apresentada em linhas de concordância e em sequência de colocados, não apenas em relação às regularidades, como também às irregularidades constitutivas da língua. Para tanto, baseamo-nos, principalmente, nos estudos de Sinclair (1993), Berber Sardinha (2004) e Tsui (2005).

Palavras-chave:
Enriquecimento lexical. Linguística de *Corpus*.
Regularidades e irregularidades da língua.

1. Introdução

No que diz respeito à educação, são notórias as modificações pelas quais o mundo moderno vem passando a partir do desenvolvimento da tecnologia, mais especificamente em relação ao processo de ensino aprendizagem de idiomas. Um dos exemplos expressivos situa-se no campo das contribuições da Linguística do *Corpus* com vistas, principalmente, ao aprimoramento do aspecto lexical da língua alvo. Os enormes corpora disponibilizados na internet constituem uma inovação extremamente positiva, uma vez que os alunos têm a possibilidade de observar a língua em contexto real de uso e analisar as linhas de concordância e em sequência de colocados em relação a vários aspectos, tais como, gramatical, semântico, pragmático entre outros. O objetivo desta comunicação, portanto, é demonstrar a possibilidade de enriquecimento lexical do discente por meio da observação e descrição da língua alvo, apresentada em linhas de concordância, não apenas em relação às regularidades, como também às irregularidades constitutivas da língua.

A Linguística de *Corpus* adota uma abordagem empírica, segundo a qual o conhecimento se origina na experiência e tem como elemento central a visão probabilística da linguagem. Na linguística, o empirismo significa dar primazia aos estudos provenientes da observação da linguagem, em geral reunidos sob a forma de um corpus. O empirismo coloca-se em oposição ao racionalismo, segundo o qual, em linhas gerais, o conhecimento provém de princípios estabelecidos a priori. Daí a oposição entre Chomsky (1965), expoente do racionalismo na linguística, e Halliday (1991), que segue a tradição empírica. Desse modo, a Linguística de *Corpus* coaduna-se, em certa medida, com a linguística de Halliday. Apesar de não se definir como um linguista de *corpus*, "parte de sua teoria se encaixa nos preceitos da linguística de *corpus* e serve como arcabouço teórico no qual ela pode se incluir" (BERBER SARDINHA, 2004, p. 34-5).

Em relação ao estatuto da Linguística de *Corpus*, de acordo com Berber Sardinha (2004), não se trata apenas de um conjunto de ferramentas nem de uma metodologia, mas sim de uma "nova abordagem de pesquisa e, na verdade, uma nova abordagem filosófica" (BERBER SARDINHA, 2004, p. 37).

A respeito da área de atuação da disciplina, Berber Sardinha (2004) entende que:

> A linguística de corpus ocupa-se da coleta e exploração de corpora, ou conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal dedica-se à exploração da linguagem através de evidências empíricas extraídas por meio de computador. (BERBER SARDINHA, 2004, p. 32)

Faz-se importante também mencionar a definição de corpus: o termo latino "corpus" significa "corpo, conjunto de documento sobre determinado assunto" (DICIONÁRIO LAROUSSE, 1999, p. 270). Segundo Berber Sardinha (2004, p. 3), estudos baseados em corpus existem desde a Antiguidade. Na Grécia Antiga, Alexandre o Grande definiu o Corpus Helenístico. Na Idade Média, produziam-se *corpora* (plural de *corpus*) de citações da Bíblia.

De acordo com Berber Sardinha (*Ibidem*), durante o século XX houve muitos educadores como Thorndike (1921) e linguistas como Fries (1952) que se dedicaram à descrição da linguagem por meio de corpora. A ênfase já era para o ensino de línguas. Atualmente, alinguística de corpus enfoca tanto a descrição da linguagem quanto a pedagogia.

A necessidade de corpus para o estudo da língua e da tradução parece, de maneira geral, partir da variação intrainterlinguística. Como enfatiza Marcuschi:

> A língua, sabidamente, não é um conjunto de rotinas e sim um contínuo muito diversificado e complexo de atividades sócio-interativas pelas quais os indivíduos em condições específicas produzem sentidos públicos partilháveis. Portanto, inerente a todas as línguas humanas, a variação é incontornável e torna condição necessária a utilização de corpora linguísticos por parte de quem se dedica ao estudo de atividades linguísticas situadas. (MARCUSCHI, 2001 *apud* CAMARGO, 2003, p. 77)

Há diferentes conceituações do termo "*corpus*". Devido à definição de Sanchez, a seguir, incorporar as características principais para a compilação de *corpus* eletrônico, Berber Sardinha (2004) considera-a uma das mais completas, portanto é a que se adotou para esta investigação:

> Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extenso em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (SANCHEZ, 1996, p. 8-9, apud BERBER SARDINHA, 2004, p. 18)

Em relação à contribuição da Linguística de *Corpus*, é possível observar a contribuição dessa abordagem em várias áreas do saber, tais como nos estudos lexicais e lexográficos entre os quais se destacam os estudos sobre colocações, a produção de dicionários, tal como o *Cambridge Dictionaryof American English* com mais de 40.000 palavras e verbetes baseados em corpus, as listas de palavras de uso mais frequente nas línguas etc.

Considerando-se a tradução como a quinta habilidade a ser abordada no processo de ensino e aprendizagem de um idioma (AYACHIA, 2018), no que tange à contribuição da Linguística de *Corpus* para os Estudos da tradução, uma das pesquisadoras de grande proeminência é a Drª Mona Baker, docente da Universidade de Manchester, cuja proposta inicial (1993) parte de duas principais correntes de pensamento, uma na área de investigação da tradução e outra na da linguística de corpus. A primeira baseia-se nas concepções de Toury (1978), para quem os estudos descritivos da tradução constituem o ramo da disciplina que deve fornecer uma metodologia coerente e procedimentos explícitos de pesquisa, de forma a permitir que os resultados de estudos descritivos indi-

viduais sejam expressos em termos de generalizações sobre o comportamento tradutório. A segunda vertente provém do linguista Sinclair (1991), o qual acredita que *corpora* computadorizados conseguem minimizar, em parte, as limitações do pesquisador e sua dependência da intuição. A partir dessas duas correntes de pensamento, Baker (1993) estabelece a tradução como objeto de pesquisa da disciplina, cujo objetivo principal passa a ser a identificação de traços do texto traduzido que levarão ao entendimento do que é, e de como funciona a tradução.Ainda com base em Baker (1993, p. 243), observamos que a pesquisadora menciona quatro categorias ou estratégias que "tipicamente ocorrem em textos traduzidos... e que não são resultado da interferência de sistemas linguísticos específicos". Mais especificamente, essas categorias são:

a) Normalização: A tendência do tradutor em exagerar as características da língua de chegada, adaptando a linguagem do texto original aos padrões típicos da linguagem do texto traduzido. A normalização pode ocorrer ao nível da microestrutura e afetar a macroestrutura do romance, como, segundo Scott (1998), aconteceu na obra espanhola *Don Quixote* em sua tradução para a língua holandesa. Baker (1996) afirma que há uma relação entre a normalização e o *status* da língua alvo, isto é, quanto mais alto o *status* da língua fonte, menor a tendência à normalização. Além disso, Baker (1996) observa que a normalização é mais evidente quando se relaciona a formas gramaticais, à pontuação e a padrões de combinação de palavras, ou seja, de colocações.

b) Explicitação: A tendência do tradutor de tornar a linguagem mais explícita, mais clara para o leitor do texto traduzido. A referida estratégia justificaria o fato de os textos traduzidos serem, em média, 10% mais longos do que os textos originais.

c) Simplificação: A tendência do tradutor de simplificar a linguagem usada na tradução, ou seja, tornar a leitura mais fácil (não necessariamente mais explícita) para o leitor. A simplificação envolve a análise do comprimento de sentença, ambiguidade, pontuação, densidade lexical e razão forma/ocorrência, ou seja, uma medida da variedade de vocabulário usada num texto ou *corpus*, possibilitando verificar se o texto traduzido apresenta um vocabulário mais ou menos variado do que o texto original na mesma língua. O uso de vocabulário menos variado é um traço dos textos direcionados para falantes não nativos de uma língua, para torná-los mais fáceis de processar.

d) Nivelamento: A referida estratégia diz respeito à tendência em encontrar um equilíbrio, em não exagerar características da linguagem do texto original, nem características da linguagem do texto traduzido. O nivelamento envolveria a tendência em trazer o texto traduzido para uma linguagem padrão sem privilegiar a língua de partida.

Essas categorias têm contribuído significativamente para a descrição tanto do processo tradutório quanto das opções do tradutor frente às dificuldades inerentes da tradução.

A utilização da Linguística de *Corpus* em sala de aula de língua estrangeira também tem trazido muitos benefícios para docentes e, sobretudo para discentes que se tornam pesquisadores interessados em observar a estrutura e o uso real da língua que estão aprendendo, uma vez que o uso dessa abordagem permite uma nova maneira de estudar e descrever a língua, bem como criar hipóteses e observar fenômenos raros ou ainda não constatados pelos meios de análises convencionais.

Além disso, a Linguística de Corpus permite que a língua seja analisada tanto em relação ao eixo paradigmático, por meio das frequências e lemas, quanto em relação ao eixo sintagmático, por meio das linhas de concordância. Essa interação entre os eixos saussurianos revolucionou a visão da linguagem modular, em que léxico e gramática eram vistos como pares indissociáveis. Nas palavras de Leech (2010):

> Nesse respeito, pode ser dito que o *corpus* tem revolucionado a nova perspectiva em estruturação linguística: em contraste com o paradigma de Chomsky (1965) por meio do qual a gramática e o léxico são dois componentes claramente distintos. Além disso, desafia-se a antiga tradição estabelecida nos estudos linguísticos, segundo a qual gramática e dicionários ofertam tipos distintos de informação sobre uma língua e são publicados com capas diferentes[1]. (LEECH, 2010, p. 12)

[1] In this respect, it can be said that the corpus revolution has introduced a new theoretical perspective on linguistic structuring: one in bold contrast to the mainstream paradigm of Chomsky (e.g. Chomsky 1965:84-88) whereby grammar and lexicon are two clearly distinct components. It also challenges a tradition long established in language study, whereby grammars and dictionaries provide distinct kinds of information about a language, and are published in separate covers (LEECH, 2010, pg. 12). (Todas as traduções foram efetuadas pela autora do artigo.)

Outra vantagem notável da utilização de corpus no ensino de segunda língua é referente à imparcialidade da intuição do autor. Para Gabrielatos (2003, p. 2), "a intuição do falante nativo nem sempre é confiável e a condição de falante nativo não nos garante, automaticamente, uma visão consciente, clara e abrangente da língua em todos seus contextos de uso". Por outro lado, é de valia ressaltar a autenticidade dos textos utilizados, conforme nos aponta Maciel:

> Nesse contexto, a Linguística de Corpus abre novos caminhos para que o professor e aluno percebam, a partir de realizações textuais autênticas, a complexidade do inter-relacionamento do léxico, da sintaxe e da semântica e possam fazer suas descobertas selecionando elementos lexicais e regras gramaticais de acordo com significado que desejam expressar na comunicação. Em tal integração, torna se possível desenvolver a conscientização linguística e a autonomia do aluno no uso da língua, tão valorizadas no processo pedagógico-didático da comunicação linguística (MACIEL, 2005, p. 129)

É de valia destacar a pesquisa de Tsui (2005), docente da Universidade de Hong Kong cuja investigação abordou a resolução de dúvidas de professores de inglês por meio do uso de *corpora* de Língua Inglesa. O quadro abaixo elenca os assuntos abordados e o número de questões concernentes a cada assunto:

Table 1. Categorization of teachers' questions

	Lexico-grammar	No. of Questions
1	Commonly confused expressions / words which are semantically close / meanings of words	225
2	Sentence structure / connectives	147
3	Countable and uncountable nouns	129
4	Prepositions	116
5	Agreement (singular and plural)	99
6	Adverbs	81
7	Adjectives	77
8	Tenses	70
9	Active / passive voice	70
10	Determiners	65
11	Collocation	42
12	Phrasal Verbs	42
13	Statements and Questions	38
14	Modals	35
15	Pronouns	29
16	Conditionals	29
	Total	1294

Como pode-se observar, as dúvidas abrangem os seguintes aspectos linguísticos: palavras cujos sentidos são semanticamente semelhantes, com 225 questões; uso de conectivos, com 147 questões; palavras contá-

veis e incontáveis, com 129 questões; uso de preposição, com 116 questões; dúvidas referentes ao singular e plural de palavras, com 99 questões; uso de advérbios adequados, com 81 questões; uso de adjetivos adequados com 77 questões; uso de tempos verbais e vozes dos verbos, com 70 questões cada; uso de determinantes, com 65 questões; uso de colocações e expressões frasais com 42 questões cada; composição da forma interrogativa, com 38 questões; uso de verbos modais com 35 questões; uso de pronomes e uso do condicional 29 questões cada.

A quantidade de dúvidas de professores de inglês que foram solucionadas nessa pesquisa alcançou o total de 1294. Trata-se, indubitavelmente, de um número bastante expressivo e que, sobretudo, demonstra que a utilização de corpora não se restringe apenas ao estudo de colocações lexicais; pelo contrário, as contribuições da Linguística de *Corpus* para o processo de ensino aprendizagem de idiomas são bem mais abrangentes.

Com o intuito de ilustrar nossa pesquisa, selecionou-se a palavra "informação/*information*" e a provável dúvida em relação à possibilidade de uso do artigo indefinido "*an*" posteriormente a essa palavra, ou seja, "aninformation" . A princípio imagina-se que a resposta seria que se trata de uma palavra que, na Língua Inglesa, não tem plural, e, consequentemente, não se usaria com o artigo indefinido anteposto; porém, o aluno ou mesmo o professor de nível básico ou intermediário deve considerar a estrutura da língua, pois, em pesquisa no *Corpus of Contemporary American Englishé* possível, à linha 8, encontrar: *... said Ken Moun, aninformation specialist with...* Certamente, o artigo indefinido "*an*" no exemplo em pauta refere-se ao substantivo contável *specialist*, e não à palavra "*information*". Diante disso, pode-se afirmar que a visualização do nódulo, ou seja, da palavra de busca no *corpus*, possibilita análises bem mais completas do que apenas uma simples explicação de forma oral.

Por outro lado, sabe-se que há críticas referentes à Linguística de *Corpus*, entra as quais se argumenta que a utilização de *corpus* possui pontos negativos sob a perspectiva pedagógica. A utilização e análise das linhas de concordância, por exemplo, são consideradas por alguns linguistas como um procedimento descontextualizado. Entretanto, boa parte das ferramentas que lidam com *corpus* permitem o prolongamento das linhas de concordância para sentenças maiores, oferecendo o cotexto dessas concordâncias a fim de contextualizá-las.

Outra crítica que vale destacar foi realizada pelo pesquisador chinês Wen-Shuenn Wu, em artigo intitulado [2]*O casamento entre a linguística baseada em corpus e instruções léxico gramaticais: o uso dos aconselhar, recomendar e sugerir em inglês como exemplos.*

Nesse estudo, o autor solicita que os alunos encontrem no corpus da língua inglesa as estruturas referentes ao uso dos três verbos supracitados e criem as regras de uso de cada um dos verbos. Durante a pesquisa, foi encontrada no corpus a seguinte frase: *Could you suggest someone to advise me how to do this?*

Embora, dependendo do contexto, a frase possa ser considerada correta e esteja presente no corpus da língua em questão, o verbo *suggest* foi usado diferentemente do modo como é usado usualmente, o que levou o pesquisador à conclusão de que o usuário de *corpus* deve ter muito conhecimento da língua em questão para poder perceber, entender e explicar as nuances presentes no *corpus*.

[2] The marriage between corpus-based linguistics and lexico-grammar instruction: Using advise, recommend, and suggest as an example.

2. Considerações finais

Por meio deste trabalho pode-se perceber o crescente uso e, consequentemente, a grande contribuição da Linguística de Corpus em várias áreas do saber, tais como o ensino de idiomas, os estudos da tradução, a pesquisa científica etc. É de valia destacar também outros aspectos positivos, tais como a autonomia que tende a se desenvolver no aluno, que passa a buscar por soluções referentes à língua estudada; a maior facilidade de memorização de itens lexicais por meio de pesquisa realizada pelo próprio aprendiz; o fato de que o pesquisador no corpus tende a melhorar a comunicação escrita, a leitura e a interpretação de textos devido à exposição à língua em uso real no corpus de pesquisa.

Por outro lado, conforme citado anteriormente, tal como acontece com novos paradigmas, há também questionamentos referentes ao uso e eficácia dessa abordagem.

De um modo geral, é visível a adesão a essa abordagem por mais e mais alunos, professores e pesquisadores que se beneficiam da Linguística de Corpus tanto em suas análises quantitativas, ou seja, com base na frequência das palavras existentes no corpus, quanto nas análises qualitativas, ou seja, com base nas análises pontuais realizadas, por exemplo, por meio das linhas de concordância.

REFERÊNCIAS BIBLIOGRÁFICAS

BERBER SARDINHA, T. *Linguística de Corpus*. São Paulo: Editora Manole Ltda. 2004.

CAMARGO, D. C. de. *Análise de um corpus paralelo de textos ficcionais brasileiros e dos respectivos textos traduzidos para o inglês*: uma investigação sobre o estilo do tradutor literário Gregory Rabassa. 01/nov./2002 a 28/mar./2003. 70 f. Pesquisa realizada para estágio pós-doutoral em Tradução e Linguística de Corpus, junto ao Programa de Estudos Pós-Graduados em Linguística Aplicada a Estudos da Linguagem – LAEL, PUC-SP, São Paulo, 2003.

BAKER, M. Corpus linguistics and translation studies – Implications and applications. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. *Text and technology*: in honour of John Sinclair. Amsterdam/Philadelphia: John Benjamins, 1993. p. 233-50

_____. *Corpora in translation studies*: an overview and some suggestions for future research. Target. 7:2, 1995. p. 223-43

_____. Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, H. (Ed.). *Terminology, LSP and translation studies in language engineering*, in honour of Juan C. Sager. Amsterdam/Philadelphia: John Benjamins, 1996. p. 175-86

_____. Linguística e Estudos Culturais: paradigmas complementares ou antagônicos nos Estudos da Tradução? In: MARTINS, M. A. P. (Org.). *Tradução e multidisciplinaridade.* Rio de Janeiro: Lucerna, 1999. p.15-34

_____. Towards a methodology for investigating the style of a literary translator. In: *Target*. 12:2, p.241-266, 2000.

_____. A corpus-based view of similarity and difference in translation. In: ARDUINI, S.; HODGSON, R. *Translating similarity and difference*. Manchester: St. Jerome, 2004.

GONZALEZ, Zeli Miranda Gutierrez. *Linguística de Corpus na análise do internetês*. Dissertação de Mestrado. Pontifícia Universidade Católica de São Paulo: 2007. Disponível em: http://www4.pucsp.br/pos/lael/lael-inf/teses/zeli_gonzales.pdf. Acesso em 11/09/2018.

Houda AYACHIA, Houda. The Revival of Translation as a Fifth Skill in the Foreign Language Classroom: A Review of Literature. https://pdfs.semanticscholar.org/93e2/ec6b373bc5b3b34767f111af72d605e6e4b7.pdf. Acessoem 05/08/2019.

LEECH, Geoffrey N. *Frequency, corpora and language learning*. In A Taste for Corpora: In honour of Sylviane Granger, Meunier, Fanny, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot (Eds), 2010.

LIMA, Thereza Cristina de Souza, OLIVEIRA, Vanderleia Stece; MÜLLER, Rodrigo. O Estágio supervisionadopara o profissional de secretariado executivo: uma investigação baseada em *corpus*. In: *Educação no século XXI*. http://poisson.com.br/bs/produto/educacao-no-seculo-xxi-volume-6/.

MACIEL, Anna Maria Becker. Novos Horizontes para o ensino do léxico. In: *PPG Letras UFRGS*: Revista Língua e Literatura: v. 6 e 7, n 10/11, 2004/2005. p. 123-30

MARTINS, Isabela. Documentação de estudos em Linguística Teórica e Aplicada. Delta: Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502007000200009. Acesso em 11/09/2018.

SCOTT, M. N. *Normalisation and Reader's Expectation*: A Study of LiteraryTranslation with Referenceto Lispector's A Hora da Estrela. Liverpool: 1998, 318f. Tese (Doutorado em Filosofia). Universidade de Liverpool. Liverpool.

SINCLAIR, J. *Corpus, Concordance, Collocation.* Hong Kong: Oxford University Press, 1991.

TOURY, G. The Nature and Role of Norms in Translation. 1978. In: VENUTI, L. *The Translation Studies Reader.* London: Routledge Press, 2000. p. 198-213

TSUI, Amy. Teacher's questions and corpus evidence. https://benjamins.com/catalog/ijcl.10.3.03tsu. Acesso em 05/08/2019.

WEN-SHUENN, Wu. The marriage between corpus-based linguistics and lexico-grammar instruction: Using advise, recommend, and suggest as an example. http://www.j-let.org/~wcf/proceedings/d-072.pdf. Acesso em 05/08/2019.