

O USO DE CORPORA DE APRENDIZES NO ENSINO DE PRODUÇÃO ESCRITA EM LÍNGUA INGLESA

Adriane ORENHA
Diva C. de CAMARGO
Paula T. P. PAIVA

(Universidade Estadual Paulista/União das Faculdades dos Grandes Lagos)

RESUMO: O objetivo deste estudo é discutir aspectos relacionados a corpora de aprendizes e a características lingüísticas dos textos produzidos por aprendizes de língua inglesa com base em algumas evidências da influência da língua materna em sua produção textual. Serão propostos exercícios com o intuito de possibilitar uma reflexão, bem como enriquecer o ensino de produção textual.

PALAVRAS-CHAVE: corpora de aprendizes; produção de texto; colocações; lingüística de corpus.

ABSTRACT: This study aims at discussing aspects related to learner corpora and linguistic features of English learners texts based on some evidence of the influence of their mother tongue on their textual production. Exercises will be proposed in order to enhance reflection as well as enrich teaching of textual production.

KEYWORDS: learner corpora; textual production; collocations; corpus linguistics.

Introdução

A Lingüística de Corpus, por meio de sua abordagem teórico-metodológica, tem fornecido, nos últimos anos, subsídios importantes para as várias subáreas dos Estudos Lingüísticos, entre elas, a Lingüística Aplicada ao Ensino de Línguas e os Estudos da Tradução. O trabalho ora proposto¹ tem por finalidade

¹ Este artigo se baseia no minicurso *Corpora de aprendizes como ferramenta para o ensino de produção escrita em língua inglesa*, apresentado durante a

apresentar alguns pressupostos da Lingüística de Corpus voltados para o ensino de língua estrangeira, neste caso a língua inglesa, tomando como base um estudo sobre o uso desta língua por alunos da disciplina de Produção e Interpretação Textual em Língua Inglesa de um curso de Letras com habilitação em Tradução e Interpretação de uma universidade privada do interior paulista. Para tanto, apresentaremos a fundamentação teórica que permeia este trabalho e proporemos, em seguida, uma reflexão sobre como textos escritos e versões em língua inglesa poderiam ter seu conteúdo léxico e sintático mais adequado com a ajuda da Lingüística de Corpus e de análises com o auxílio do programa WordSmith Tools.

Dessa maneira, este trabalho teve como objetivo incentivar os alunos a empregar uma variedade maior de vocabulário e expressões, de maneira a contribuir para a produção de textos mais fluentes, assim como conscientizá-los de aspectos gramaticais, sintáticos e semânticos, favorecendo um aprendizado mais autônomo.

1. Fundamentação teórica

De acordo com Berber Sardinha (2004, p. 251), a forma como as aulas de língua estrangeira são ministradas poderiam ser revistas em razão das vantagens trazidas pela Lingüística de Corpus no que se refere à descrição lingüística.

Desse modo, o uso de corpora de línguas naturais favorece pesquisas em ensino e aprendizagem de línguas, ao possibilitar, conforme aponta Berber Sardinha (2004): 1) a descrição da linguagem nativa; 2) a descrição da linguagem do aprendiz; 3) a transposição de metodologias de pesquisa acadêmica para a sala de aula, e 4) o desenvolvimento de materiais de ensino, currículos e abordagens.

XVII Semana de Letras da Universidade Estadual Paulista – UNESP, Campus de São José do Rio Preto, realizada de 11 a 15 de setembro de 2006.

O segundo dado acima referido trata da descrição da linguagem, mas neste caso, a variedade em questão é a produção escrita ou falada de alunos de língua estrangeira, registrada em corpora de aprendizes (*learner corpora*). A descrição de corpora de aprendizes é uma atividade nova, que tem tomado impulso recentemente.

A idéia de que a utilização de corpora no ensino deve ser direta é defendida por John Sinclair (1978 *apud* Berber Sardinha, 2004, p. 256): “apresente somente exemplos reais. Esta talvez seja a primeira lição a ser aprendida do estudo de corpus. A linguagem não pode ser inventada; somente pode ser capturada”².

Os corpora de aprendizes são constituídos por textos de falantes não-nativos, os quais são utilizados para o estudo da linguagem que produzem. Geralmente possuem entre 20 e 200 mil palavras e costumam ser mais específicos em relação a corpora maiores, no que se refere aos seus tópicos e gêneros discursivos.

Entre os corpora de aprendizes de que temos conhecimento, podemos citar o ICLE (International Corpus of Learner English), compilado por Granger (1993), o CLC (Cambridge Learners' Corpus), construído por Nicholls (1999), e o Longman Learner's Corpus (Summers, 1999). No Brasil, podemos citar o Br-ICLE, subcorpus do ICLE, organizado pelo Prof. Antônio Berber Sardinha, da PUC/São Paulo, e o Corpus Multilíngue de Aprendizes, compilado pela Profa. Stella E. O. Tagnin, da USP.

Tribble (1990, 1991) foi um dos pioneiros no uso de pequenos corpora de aprendizes. O pesquisador investigou o uso de verbos relacionados à fala em um corpus de aprendizes de cerca de 54 mil palavras, a maioria extraída do *Longman Corpus of Learner English*. Notou um emprego excessivo e inapropriado do verbo *said* por parte dos aprendizes.

Kjellmer (1991), também por meio de pesquisa em corpus

² Present real examples only. This is perhaps the first lesson to be learned from corpus study. Language cannot be invented; it can only be captured.

de aprendizes, constatou que os alunos não nativos subusam pré-fabricados de tipos específicos, como os formulaicos *you know, sort of, I mean*.

Dessa forma, um corpus de aprendizes permite, em razão de possibilitar mapear a língua utilizada pelo falante não-nativo, suscitar também várias questões de pesquisa, tais como:

- a) Quais características lingüísticas da língua-alvo são empregadas com mais (sobretudo) ou menos (subuso) frequência em comparação com falantes nativos?
- b) Qual é a extensão da influência da língua nativa (transferência) na produção dos aprendizes?
- c) Em que áreas eles tendem a usar estratégias de evitação deixando de explorar a fundo o potencial da língua-alvo?
- d) Em que áreas eles tendem a demonstrar desempenho nativo ou não-nativo?
- e) Quais são as áreas nas quais os aprendizes de um dado país parecem necessitar de mais ajuda para desenvolver sua produção na língua-alvo? (Berber Sardinha 2004, p. 266)

Entre os itens acima apresentados, focaremos o item b e, de certo modo, o item e, uma vez que apontaremos aspectos sintáticos (proposta 2 – o emprego de *because* e *because of*) e léxico-semânticos (proposta 1 – colocações adverbiais), dos quais os aprendizes do Brasil parecem necessitar de práticas, a fim de melhorar sua produção textual na língua inglesa.

Outra questão a ser verificada em nosso corpus de aprendizes é a extensão da influência da língua portuguesa na produção de textos dos aprendizes (propostas 1 e 2). Pelo que pudemos averiguar, a língua materna está muito presente nos textos que formam o corpus em questão.

2. Duas propostas de ensino de produção escrita em língua inglesa com base em corpora de aprendizes.

Para esta pesquisa, compilamos um corpus de aprendizes de 21.309 palavras formados por textos em inglês produzidos por alunos do 3º ano diurno e noturno do Curso de Tradutor e Intérprete, turmas de 2005 e 2006. Este corpus se subdivide em dois subcorpora: 1) um subcorpus constituído de textos produzidos pela turma do diurno, somando um total de 10.897 itens; e 2) um subcorpus com as mesmas características produzidos por alunos do noturno, com um total de 10.412 itens.

Na proposta 2, destacamos que também foi elaborado um corpus composto por textos de base retirados do livro didático usado pelos alunos em sala de aula. Justifica-se a compilação de tal corpus em razão dos textos servirem como modelo aos alunos, visto que trazem a linguagem considerada adequada para a produção de seus textos.

2.1. Proposta 1

A partir dos dois corpora dos textos de alunos acima mencionados, foram extraídas duas listas de frequência de palavras com a ajuda do programa WordSmith Tools (Scott, 2004). Pudemos, então, observar quais eram as palavras mais utilizadas pelos alunos de língua inglesa das duas classes e fazer algumas comparações em relação à diversificação lexical das duas turmas. Observamos que as palavras utilizadas pelo grupo de aprendizes do diurno continham, ainda que em porcentagem pequena, mais palavras que poderiam ser consideradas de nível intermediário.

Em seguida, geramos uma lista de palavras-chave a partir do contraste da lista de palavras do corpus de aprendizes com a lista de frequência do corpus de referência *British National Corpus* (BNC Sampler, com 2.530.849 itens), corpus de língua composto por textos originalmente escritos em inglês, retirados de fontes variadas do cotidiano. Verificamos se as 200 primeiras palavras que constavam nessa lista também faziam parte da lista das 2.000 palavras mais utilizadas na língua inglesa (consideradas

pertencentes ao nível básico), a *General Service List*, criada por Michael West (1953). De 200 palavras-chave geradas, apenas 8% não pertenciam ao vocabulário da lista de 2.000 palavras.

Tendo em vista que os corpora de aprendizes podem ser explorados na reestruturação de orações, assim como no incentivo ao emprego de maior variedade vocabular, decidimos analisar alguns adjetivos empregados pelos alunos nesta lista de 200 primeiras palavras-chave, tais como: *beautiful, important, happy e interesting*. A seguir, geramos listas de concordâncias a fim de analisar o contexto no qual eram empregados, conforme mostra figura abaixo a partir do nóculo *important*:

Line	Text	Frequency
1	many singers, axe music and the most important is Carnival. Carnival is a big party in	343
2	the culture, to absorb the English; this is very important for my career, for I can become an	1.164
3	and many lakes, the joaquina beach is a very important stage of world cup of suif.	1.279
4	you will learn a lot about one of the biggest and important countries in the world. :: Enviado por	1.891
5	de Souza NATURE Environment is the most important thing in our lives, however, people	1.959
6	unacceptable! The environment is the most important thing in the whole world. Without it	2.320
7	thinks about them, and to have liberty is more important. All states are economically well and	3.863
8	in language. I know that this is very important to my improvement. In college we	4.361
9	without interrupt. That's way I think to much important we have the opportunity to travel to the	4.401
10	a variety of animals and plants which are very important for us human beings. But man is	4.606
11	are fantastici The soapoperas are very important for Brazil because we are considered	4.938
12	interesting man, or to talk about something important to her friend. Men pick up the remote	5.708
13	their families and so they forget to do a very important thing: relax! You must save some	7.348
14	a life of your own. That's why relaxing is so important! Dream I was in a big building, walking	7.407
15	It's important to relax when you get home after	2
16	Write a text about an important experience in your life An important	5
17	about an important experience in your life An important experience in my life was moving to	11
18	It's very important for a person nowadays, to learned a	2
19	a foreign language. In my opinion is very important you study this, even so a lot of school	46
20	this, even so a lot of school don't think this important, and people nowadays don't know	59
21	get a job in another country. But in here is very important too, because all the companies want	29
22	in whole world a second language is very important to get a job. Because of the	18

Figura 1: Lista de concordância do nóculo *important*

Conforme constatamos nas linhas de concordâncias geradas, não houve nenhum caso em que os aprendizes utilizaram colocações adverbiais com qualquer um dos adjetivos pesquisados (*beautiful, important, happy e interesting*). Essa constatação veio a corroborar nossa intuição, advinda de nossa experiência como professoras de língua inglesa e produção textual em língua inglesa, de que os aprendizes, de um modo geral, raramente fazem uso de colocações do tipo adverbial, seja ela formada por advérbio + adjetivo, quanto constituída por verbo + advérbio. Esse dado pode ser explicado em virtude de o falante do português fazer pouco emprego de colocações adverbiais em seus textos em língua portuguesa.

Para conscientizá-los do fator colocacional na língua, preparamos um exercício enfocando apenas a construção advérbio + adjetivo. Os alunos foram solicitados a produzir algumas frases com essa construção, a partir dos adjetivos *beautiful, important, happy e interesting*. Tais frases, todas com ocorrência de colocação adverbial teriam que ser traduzidas do português para o inglês. Esse tipo de exercício de versão se justifica em virtude de tratar-se de alunos de um curso de Tradutor e Intérprete. Dessa forma, o procedimento foi o seguinte: retiramos alguns exemplos de colocações adverbiais do corpus compilado da dissertação de mestrado de Orenha (2004), cuja base contivesse os adjetivos analisados, e traduzimos para o português. A partir dessas frases, os alunos teriam que realizar as versões para o inglês. Tal procedimento foi feito para assegurar o uso correto das colocações adverbiais em inglês. Veja abaixo uma pequena amostra do exercício:

1. Selvagem e, no entanto, *estonteantemente bonita*, era uma terra de mitos e lendas, de monges e monastérios, ...
Wild, yet STUNNINGLY BEAUTIFUL, it was a land of myths and legends, of monks and monasteries, ...

2. Acredito que nossa indústria tem um papel *altamente importante* na sociedade ...
I believe our industry has a HUGELY IMPORTANT role to play in society

3. Queria saber o que os fez *esplendidamente felizes*.
He wanted to know what made them RADIANTLY HAPPY.
4. A existência de extraterrestres é um assunto *imensamente interessante*, ...
5. The existence of extraterrestrials is an IMMENSELY INTERESTING topic, ...

Ao aplicarmos o exercício aos sujeitos da presente pesquisa, pudemos observar a dificuldade que esses aprendizes tiveram para produzir tais combinações. Para que pudessem concluir o exercício, tiveram que realizar pesquisas na *Internet* por meio do buscador *Google*, o que lhes permitiu perceber que as colocações adverbiais são agrupamentos lexicais fixos na língua e que, por essa razão, eles não teriam tanta liberdade de escolha como imaginavam. Além disso, tiveram a oportunidade de descobrir a alta frequência das colocações adverbiais formadas pelos adjetivos pesquisados: *stunningly beautiful* (1.060.000 ocorrências); *hugely important* (605.000 ocorrências); *radiantly happy* (15.600 ocorrências); *immensely interesting* (47.200 ocorrências).

Acreditamos, dessa maneira, que o exercício proposto tenha contribuído como forma de reflexão e conscientização para que aprendizes de uma língua estrangeira venham a elaborar textos mais fluentes e sintam-se mais confiantes em relação ao texto que estão produzindo.

2.2. Proposta 2

Nesta segunda proposta, escolhemos observar características lingüísticas apresentadas pelos alunos do 3º ano de Letras com Habilitação em Tradução do período noturno da mesma universidade. Como material, montamos dois corpora que consideramos relevantes e representativos para este estudo: a) um corpus com textos produzidos pelos alunos do 3º ano das turmas de 2005 e de 2006 (corpus A); b) um corpus composto

por textos de base para a produção de aprendizes retirados do livro *Passages – Student's Book 1* (corpus B).

Uma comparação com textos retirados de outras fontes poderá não ser apropriada ao apresentar variações no gênero textual inadequadas para tal análise. Além disso, o emprego de livros didáticos mostra-se pertinente por serem considerados por nós autênticos dentro do contexto de sala de aula.

Como forma de motivação para a compilação dos corpora, pedimos que os alunos elaborassem *blogs*, ou seja, diários pessoais que são disponibilizados na Web e que, hoje, já vêm sendo utilizados como ferramenta pedagógica em diferentes contextos educacionais. Desse modo, pedimos que os alunos enviassem uma primeira versão de seus textos à professora que, ao armazená-los, foi gradativamente montando um corpus de aprendizes. Uma cópia impressa foi requisitada para que a professora fizesse as sugestões e correções do texto. Em seguida, os alunos corrigiam seus textos em seus computadores e os enviavam a um aluno da classe, encarregado pela publicação on-line dos mesmos. Além disso, foi pedido que os alunos comentassem os textos dos colegas em inglês no próprio blog. Desta forma, também estariam praticando a escrita informal da língua inglesa de uma maneira que lhes é familiar.

Uma vez montados os dois corpora mencionados anteriormente, passamos à sua análise com a ajuda do programa WordSmith Tools. Este programa, conforme apresentado anteriormente, disponibiliza uma variedade de ferramentas eletrônicas que podem ser usadas por pesquisadores em suas investigações. Como metodologia, seguimos os seguintes passos: a) comparamos a variação lexical (razão forma/item) entre os corpora A e B, sendo que o número de formas se refere ao número de palavras sem repetição dentro do corpus e o número de itens se refere ao total de palavras que compõem o corpus; e b) analisamos aspectos pontuais de uso após *because* e *because of*, no corpus A.

Geramos os dados estatísticos dos corpora A e B com o programa WordSmith Tools e obtivemos as seguintes listas:

Corpus A
Corpus B

Corpus de Aprendizes (Corpus A) Corpus de Base (Corpus B)

Itens	10.412	Itens	2.796
Formas	1.970	Formas	901
Razão forma/item	18,92	Razão forma/item	32,22

Tabela 1: Dados estatísticos do corpus A e B

De acordo com a lista à direita, o número total de itens (palavras corridas) do corpus B, usado como referência, foi de 2.796 e o número total de formas (vocábulos) foi de 901. Ao compararmos estes dados com os da lista do corpus A, à esquerda, notamos que o número de itens do corpus A é quase quatro vezes maior, ou seja, de 10.412 itens. No entanto, ao observarmos o número de formas, notamos que o corpus A apresentou 1.970 formas, isto é, apenas dobrou o número de formas em relação ao corpus B. Essa disparidade na proporção de ocorrência de itens e formas traduz-se como variação lexical (razão forma/item). De acordo com os dados acima, a razão forma/item do corpus B foi de 32.22, um número maior que o do corpus A, que foi de 18,92. Com esta informação, podemos inferir que, apesar de o corpus A ser maior que o B, os textos do corpus B apresentam um vocabulário mais variado, ou seja, uma variação lexical mais rica; já o corpus A, composto pelos textos dos alunos, não variou tanto lexicalmente, ou seja, apresentou maior número de repetições de palavras.

Passamos então à análise de aspectos pontuais de uso após *because* e *because of*. O primeiro passo foi apresentar as linhas de concordância a partir do corpus A para toda a classe. Em seguida, lemos todas essas linhas e, em conjunto, verificamos onde havia a falta de objeto direto ou sujeito na oração após *because* ou ainda, a falta da preposição *of* em alguns casos. Vejamos algumas das linhas de concordância que os alunos sugeriram que fossem modificadas a partir do corpus de aprendizes:

01.The nights are dangerous *because can* occur several problems, accidents

02.Brazilians are more interested in studying, *because nowadays* is very difficult to find...

03.Something special to do I get so anxious and *because that* I have stomachache...

Nas duas primeiras linhas exemplificadas, há a falta de sujeito na segunda oração. Já na última linha, houve a omissão da preposição *of* após *because*. Interessante notar que, pelo fato de os alunos estarem fazendo este exercício em grupo, muitas vezes reconheciam seus erros ou outras estruturas que deveriam ser modificadas. Acreditamos que as aulas de produção escrita a partir de textos retirados de corpora de língua natural e de corpora de aprendizes contribuiu para uma conscientização do funcionamento do sistema da língua estrangeira.

Ao final do exercício, sugerimos aos alunos algumas estratégias para a produção escrita em língua inglesa como, por exemplo, partir de textos autênticos e de um levantamento de palavras com o auxílio de dicionários monolíngües para produzirem textos mais elaborados.

3. Considerações Finais

Notamos que os alunos empregam agrupamentos lexicais de natureza diferente da dos nativos, os quais seriam recorrentes na língua inglesa. Em decorrência, acabam, muitas vezes, produzindo textos que não soam naturais e convencionais para um falante nativo. Nesse caso, a observação da linguagem produzida pelo aluno a partir do uso de corpus pode incentivá-lo a empregar uma maior variedade vocabular e observar o uso real da linguagem.

Pudemos também verificar que a influência da língua materna é marcada nos textos produzidos por aprendizes brasileiros de língua inglesa. No caso específico do uso de *because*

e *because of*, nota-se que a elipse do sujeito da frase seguinte se dá de maneira freqüente. Também verificou-se uma menor variação lexical em seus textos em relação aos textos usados como modelo proposto pelo livro didático.

Acreditamos que as propostas analisadas neste estudo mostram como a Lingüística de Corpus pode auxiliar os aprendizes de uma língua estrangeira a observar, por meio da exploração de sua própria produção textual, desvios de padrões lingüísticos e, através de sua auto-correção, aprender de maneira mais consciente, produzindo textos mais adequados aos padrões da língua estrangeira tratada.

Referências

BERBER SARDINHA, A. P. *Lingüística de Corpus*. Barueri, SP: Editora Manole, 2004.

British National Corpus. Disponível em <http://www.hcu.ox.ac.uk/BNC>. Acesso em 2005.

CAMARGO, D. C.; ORENHA, A.; PAIVA, P. T. P. *Corpora de aprendizes como ferramenta para o ensino de produção escrita em língua inglesa*. Minicurso apresentado durante a XVII Semana de Letras da Universidade Estadual Paulista – UNESP, Câmpus de São José do Rio Preto, realizada de 11 a 15 de setembro de 2006.

ORENHA, A. *A compilação de um glossário bilíngüe de colocações, na área de jornalismo de negócios, baseado em corpus comparável*. Dissertação (Mestrado em Estudos Lingüísticos e Literários), São Paulo: FFLCH/USP, 2004.

RICHARDS, J. C. *Passages: an upper-level multi-skills course: student's book 1*. London: Cambridge University Press, 1998.

SCOTT, M. *WordSmith Tools: Version 4*. Oxford: Oxford University Press, 2004.

SINCLAIR, J. MCH; RENOUF, A. A lexical syllabus for language learning. In: CARTER, R.; MCMARTHY, M. (orgs.). *Vocabulary and language teaching*. Londres: Longman, p.140-60, 1988.

TORRES, N. *Gramática prática da língua inglesa: o inglês descomplicado*. São Paulo: Saraiva, 2002.

TRIBBLE, C. Concordancing and an EAP writing programme. *CALL Journal*, v.1, n.2, p.10-15, 1990.

_____. Some uses of electronic text in English for academic purposes. In: MILTON, J. & TONG, K.K. (Eds.). *Text Analysis in Computer Assisted Language Learning* Hong Kong: Hong Kong University of Science and Technology, p. 4-14, 1991.

WEST, M. *A general list of English words*. Londres: Longman, 1953.

BIBLIOTECA DO C. I. A.