

Produção de atividades didáticas *corpus-based* em língua inglesa com foco em produção escrita

Mayra Aparecida dos Santos

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP),
São José do Rio Preto, São Paulo, Brasil
mayrafavaro@gmail.com
<http://orcid.org/0000-0001-8858-5100>

DOI: <http://dx.doi.org/10.21165/el.v47i2.2015>

Resumo

Discutimos neste trabalho características da escrita acadêmica em língua inglesa (LI) de alunos universitários brasileiros. Elaboramos um *corpus* para análise com 82 textos redigidos por alunos brasileiros universitários com proficiência no nível B1 de LI e, usando como base o Quadro Europeu Comum de Referência (QECR) e a Gramática Longman com base em *corpus* de Biber et al. (1999), realizamos uma observação quanto ao uso dos quantificadores *some*, *many* e *any*. Para coleta e observação dos dados, foi utilizada a ferramenta computacional *AntConc*®. Nossa dissertação considera os trabalhos sobre Escrita Acadêmica e Linguística de corpus de Berber Sardinha (2004, 2010), Dayrell (2011), Dew (2010), Romer (2004) e Viana (2011). O segundo estágio da pesquisa é a realização de atividades didáticas com base em *corpus*.

Palavras-chave: escrita acadêmica; língua inglesa; Linguística de corpus.

Production of *corpus-based* didactic activities in English focusing written production

Abstract

In this work, we discuss the characteristics of the academic writing in English language (EL) of Brazilian university students. We built a corpus for analysis with 82 texts written by Brazilian university students with proficiency at level B1 of EL and, based on the Common European Framework of Reference (CEFR) and the Longman Grammar based on corpus of Biber et al. (1999), we investigated about the use of the quantifiers *some*, *many* and *any*. For data collection and analysis, the *AntConc*® computational tool was used. The design of our dissertation considers the works on Academic Writing and Corpus Linguistics of Berber Sardinha (2004, 2010), Dayrell (2011), Dew (2010), Romer (2004) and Viana (2011). The second stage of the research is the execution of didactic activities based on corpus.

Keywords: academic writing; english language; Corpus Linguistics.

Introdução

Na presente pesquisa, realizamos um estudo de caso com textos de alunos universitários. Formamos um *corpus* eletrônico com 82 textos argumentativos, escritos em língua inglesa (LI) por estudantes categorizados no nível B1 de LI, segundo o Quadro Europeu Comum de Referências para Línguas (QEQR). O foco deste estudo é analisar o uso dos quantificadores *some*, *many* e *any* nas redações e se houve inadequação ou não.

O objetivo deste trabalho é contribuir para a prática do ensino de Inglês com Fins Acadêmicos (IFA), uma vez que o Brasil é considerado um dos países com o pior índice de proficiência em língua inglesa no mundo. Segundo pesquisa realizada pelo

British Council (2014, p. 7), apenas 5% da população brasileira com mais de 16 anos “afirma possuir algum conhecimento do idioma inglês”.

Ressaltamos que este trabalho contribui com um projeto maior que já está em andamento intitulado CorIFA – *Corpus* de Inglês com Fins Acadêmicos, que foi desenvolvido e é gerenciado pela Profa. Dra. Deise Prina Dutra, da Universidade Estadual de Minas Gerais – UFMG.

1. Fundamentação teórica

Apresentaremos aqui a abordagem teórico-metodológica da Linguística de *Corpus* (LC) e traremos conceitos referentes à escrita acadêmica e textos argumentativos, que configuram nosso material coletado. Outro tópico abordado será o Quadro Europeu Comum de Referência para Línguas (CEFR) e os descritores de competência relativos ao nível B1 de proficiência escrita, no caso deste trabalho, os quantificadores *some*, *many* e *any*.

1.1. A abordagem teórico-metodológica da Linguística de *Corpus*

Realizamos neste trabalho uma observação da língua escrita por alunos universitários brasileiros. Para tanto, foi necessária a compilação de um *corpus*, uma vez que nos norteamos pela LC e, portanto, faz-se necessário trazer o conceito de *corpus*, para que o trabalho seja melhor compreendido pelo leitor.

Pela definição de Sanchez (1995, p. 89 *apud* BERBER SARDINHA, 2000, p. 104), podemos considerar o *corpus* como sendo:

[...] um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos 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 poder propiciar resultados vários e úteis para a descrição e análise.

Dessa forma, entendemos que, para uma análise justa dos dados, é preciso preocupar-se com a configuração do *corpus* e o pesquisador deve montá-lo de acordo com os critérios que sua pesquisa exige.

Com relação à compilação dos *corpora*, vivemos hoje um momento em que a tecnologia nos favorece nesse sentido. Anterior ao advento da tecnologia e dos computadores, a compilação dos *corpora* era feita de forma manual. Atualmente, com a ajuda das ferramentas tecnológicas, o trabalho de criação de um *corpus* é realizado muito mais rapidamente. Também contamos hoje com *softwares* de análise, como o *WordSmith Tools* e o *AntConc*®.

O avanço nos estudos de LC categorizou diferentes tipos de *corpus*. Citamos aqui a definição de *corpus* de aprendiz, que será o alvo de nossa pesquisa, pelas palavras de Delegá-Lúcio (2006, p. 21):

Um corpus de aprendiz é aquele formado por textos naturais falados ou escritos por estudantes de uma língua estrangeira, que atenda a critérios que possibilitem seu estudo e que tenha sido coletado de modo que possa ser lido por computador. Os textos de um corpus de aprendiz são considerados naturais porque são produzidos por estudantes sem controle do que escrevem, ou seja, nenhum ponto gramatical (por exemplo) específico deve ser elicitado. Há, no entanto, critérios para a coleta desse corpus de acordo com o tipo de texto que se quer representar.

Esse cenário de estudos, onde a tecnologia se alia à LC, nos leva ao caminho que faremos para desenvolver nosso trabalho de observação de *corpus* de aprendizes, visando uma pesquisa de caráter quanti-qualitativa no que se refere às inadequações cometidas em textos escritos pelos estudantes brasileiros.

A abordagem teórico-metodológica da Linguística de *Corpus* visa à coleta do material linguístico, podendo ser texto escrito ou oral, com o objetivo de que observemos, por exemplo, a escolha de léxico ou de agrupamentos lexicais em uso. Acreditamos na importância desta ferramenta de análise e usaremos uma célebre citação de Firth para exemplificar a razão: “a totalidade de nosso comportamento linguístico é melhor compreendida se vista como uma rede de pessoas, coisas e eventos conectadas, apresentando estruturas e sistemas assim como percebemos em nossa existência” (FIRTH, 1968 *apud* ANDERMAN; ROGERS, 2007, p. 9).

A citação acima ilustra como a observação do léxico em diferentes contextos pode nos fazer compreender melhor a dinâmica da língua e essa é uma das razões pela qual as pesquisas envolvendo a Linguística de *Corpus* estão em ascensão e cada vez mais presentes em estudos pedagógicos. Também esta será usada como base teórico-metodológica para a análise e elaboração das atividades pedagógicas de nosso projeto.

1.2. Escrita Argumentativa

Sabemos que a habilidade de escrita é de grande importância no meio acadêmico e que nesta etapa surge a necessidade de desenvolver textos mais elaborados, uma vez que a exigência é maior. Alguns dos tipos de textos que podem ser requeridos de um aluno são redações de comparação e contraste, definição, opinião, divisão lógica, histórias narrativas e de processo (como fazer algo) (DEW, 2010).

Os textos de opinião, que aqui chamamos de argumentativos, têm como característica fornecer uma opinião *formal* e sustentar sua opinião¹ (DEW, 2010). Ao elaborar um texto argumentativo, o autor prevê que irá convencer seu leitor acerca de seu ponto de vista, e, portanto, deve fazê-lo por meio de um texto coeso, valendo-se de uma organização textual que pode se dividir, segundo Dew (2010) em: unidade, coerência, sinais de transição, variação no início das orações e estruturas paralelas.

Uma vez que nosso objetivo é trabalhar com a prática de produção escrita e também contribuir com o *corpus* CorIFA, dentre os tipos de texto que um aluno precisa elaborar durante o período em que está na universidade, optamos por observar a escrita de um texto argumentativo.

¹ Texto original: “[...] to provide an opinion formally and to support their opinion.”.

1.3. Quadro Europeu Comum de Referência para Línguas (QECR)

Usaremos como base para avaliar o nível de proficiência dos participantes os parâmetros estabelecidos internacionalmente pelo Conselho Europeu. Esses parâmetros foram definidos como Quadro Europeu Comum de Referência (QECR) ou *Common European Framework of Reference* (CEFR) e o quadro é dividido em seis níveis de conhecimento: A1 e A2, B1 e B2, C1 e C2.

Resultado de um estudo de 20 anos realizado pelo Conselho Europeu, o QECR tem como finalidade colaborar com a prática de ensino e aprendizagem de línguas, avaliação e nivelamento de estudantes e produção de atividades e materiais didáticos.

O documento descreve de forma pormenorizada as competências comunicativas inerentes ao usuário da língua, de acordo com os níveis e as habilidades. As habilidades de comunicação são divididas em produção e compreensão oral e escrita.

Como forma de mensurar o avanço do conhecimento dos aprendizes de uma língua, o QECR divide os níveis de proficiência das habilidades conforme apontado abaixo:

Quadro 1: Níveis de proficiência do QECR

BÁSICO	A1	Iniciante
	A2	Básico
INDEPENDENTE	B1	Intermediário
	B2	Usuário independente
PROFICIENTE	C1	Proficiência operativa eficaz
	C2	Domínio pleno

Fonte: Elaborado pela autora

1.3.1. Nível B1 – Proficiência Escrita

Como forma de interagir com o outro, é preciso que o utilizador da língua desenvolva as habilidades de compreensão escrita e oral, e que também seja capaz de produzir de forma oral e escrita. Ocorre que o aprendiz nem sempre desenvolve de maneira linear e equivalente todas as habilidades, inclusive não sendo estritamente necessário dependendo de seu objetivo de uso do idioma.

Sendo a língua meio de interação social, o QECR divide as atividades linguísticas em quatro grandes áreas, chamadas de *domínios*, que são os domínios público, privado, profissional e educacional.

Uma vez conhecidas as habilidades de compreensão e produção e quiçá os domínios de língua pertinentes ao desenvolvimento do aprendiz, delimitamos nosso alvo de pesquisa na observação da habilidade de produção escrita, no que tange ao domínio educacional.

Para que possamos realizar uma análise efetiva dos textos coletados e uma observação do *corpus* de referência para a elaboração das atividades didáticas, no que se refere à prática de produção escrita de um texto argumentativo, consideramos que o aluno proficiente no nível B1, segundo o QECR,

É capaz de escrever redações simples e curtas sobre tópicos de seu interesse. É capaz de, com alguma confiança, resumir, relatar e dar sua opinião sobre informações baseadas em fatos acumuladas que sejam de sua rotina familiar e de assuntos não pertinentes à sua rotina dentro da sua área de interesse.

É capaz de escrever, de forma muito breve, relatos em um formato convencional padrão que transmitam informações rotineiras com base em fatos e estabelecer razões para ações. (COMMON EUROPEAN FRAMEWORK OF REFERENCE, 1995, p. 62, tradução nossa)².

Conforme descrito na citação acima, podemos considerar que um aprendiz da língua inglesa, que esteja categorizado no nível B1 de proficiência na habilidade de produção escrita, seja capaz de escrever um texto argumentativo de assuntos que lhe sejam familiares ou não, mas que estejam em sua área de interesse.

2. Metodologia

Nessa seção, apresentaremos nosso *corpus* de análise, o *Corpus* de Inglês com Fins Acadêmico – CorIFA (2.1) e nosso *corpus* de referência *Michigan Corpus of Upper-Level Student Papers* – MICUSP (2.2) que é o *corpus* de escrita acadêmica que escolhemos para a elaboração de nossas atividades didáticas, uma vez que nosso objetivo é realizar uma comparação entre a escrita de dois públicos não-nativos em língua inglesa. Por fim, na seção de Procedimentos Metodológicos adotados neste estudo (2.3), discorreremos sobre a ferramenta computacional *AntConc* e fazemos uma breve apresentação de suas funções.

2.1. *Corpus* de análise: recorte do CorIFA

Nosso *corpus* de análise é formado por 82 textos argumentativos, escritos por alunos brasileiros em nível B1 de proficiência em língua inglesa. São alunos universitários, regularmente matriculados no 1º e 3º ano do curso de graduação Licenciatura em Letras – Português / Inglês e também alunos do 3º ano de Bacharelado em Letras com Habilitação de Tradutor, da UNESP – S. J. Rio Preto.

Os 82 textos geraram um *corpus* pequeno (BERBER SARDINHA, 2003), com um total de 13.244 formas (*word tokens*) e 1.752 itens (*word types*) e, para a análise e extração dos dados, utilizaremos a ferramenta computacional *AntConc*®.

Os textos produzidos por alunos do 3º ano de Licenciatura em Letras Português / Inglês e 3º ano de Bacharelado em Letras com habilitação de Tradutor, foram coletados através do *corpus* eletrônico de inglês para fins acadêmicos – CorIFA, que vem sendo desenvolvido sob orientação da profª. Dra. Deise Prina Dutra da Universidade Federal de Minas Gerais – UFMG.

Por se tratar de um projeto de coleta de dados, antes de realizar as atividades integradas, os alunos precisam preencher seus dados na plataforma e dar o aceite no

² Texto original: “Can write short, simple essays on topics of interest. Can summarise, report and give his/her opinion about accumulated factual information on familiar routine and nonroutine matters within his/her field with some confidence. Can write very brief reports to a standard conventionalised format, which pass on routine factual information and state reasons for actions”.

Termo de Consentimento Livre e Esclarecido (TCLE), para que seus dados possam ser utilizados nas pesquisas realizadas com o *corpus*.

Uma outra parcela do conteúdo para a elaboração do *corpus* desta pesquisa foi coletada em aulas de língua inglesa oferecidas pela autora da dissertação em seu estágio de docência realizado na disciplina de Língua Inglesa I do 1º ano de Licenciatura em Letras Português / Inglês, sob supervisão da profa. Dra. Cláudia Ceneviva Nigro. Este material, apesar de ter sido coleta *in loco*, também é parte do CorIFA, pois foi enviado para integrar o *corpus* eletrônico.

2.2. *Corpus* de referência MICUSP

Criado na Universidade de Michigan, o MICUSP é um *corpus* que tem sido utilizado para a pesquisa e ensino de Inglês para Fins Acadêmicos (em inglês *English for Academic Purposes – EAP*). É possível encontrar no MICUSP artigos variados, como os argumentativos, que serão nosso foco, escrita criativa, relatórios, entre outros. Esses textos foram escritos por estudantes de 1ª a 4ª série da graduação de quatro grandes áreas, sendo elas Humanas e Artes, Ciências Sociais, Ciências Biológicas e da Saúde e Ciências Físicas. Abaixo apresentamos a figura da tela do MICUSP:

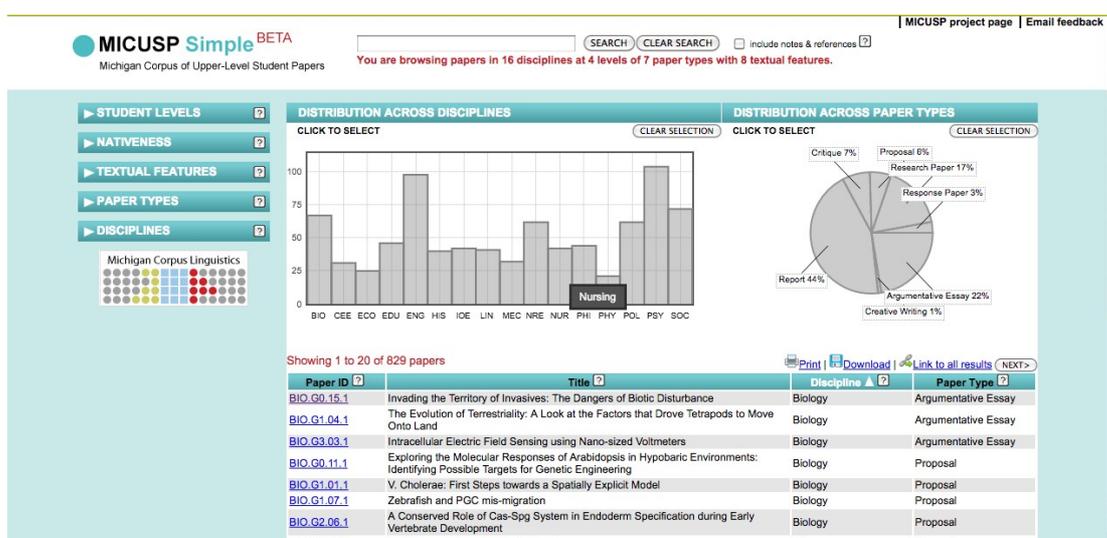


Figura 1: Tela com o *Corpus* do MICUSP

Conforme vemos na tela acima, o MICUSP apresenta os textos divididos nas seguintes categorias: nível dos alunos que contribuíram para o *corpus*, se este é falante nativo ou não-nativo de língua inglesa, tipo do texto, tipo do artigo e áreas de estudo. Os textos que compõem o MICUSP são argumentativos, escrita criativa, críticas/avaliações, propostas, relatórios, artigos de pesquisa e artigos de resposta. Os textos que fazem parte do MICUSP são amostras em que os estudantes receberam nota A em sua escrita. Essa seleção nos forneceu um total de 31 textos e gerou um *corpus* eletrônico contendo 71.980 itens e 7.862 formas.

Sabemos que essa pode ser considerada uma comparação injusta, se considerarmos que os alunos que escreveram os textos selecionados para fazer parte do MICUSP são alunos com notas altas e, possivelmente, com um nível de proficiência maior do que nossos alunos. Entretanto, ressaltamos que iremos observar o uso de

competências linguísticas pertinentes aos usuários da língua categorizados em nível de proficiência B1, que é o caso dos alunos que escreveram os textos que usamos para formação de nosso *corpus* de análise. Também o fato de o *corpus* de referência conter uma menor quantidade de textos, porém maiores em número de palavras, visto que são textos de conclusão de curso, não irá influenciar em nossa análise.

2.3. Procedimentos Metodológicos utilizados neste estudo

Escolhemos para a observação e extração de dados a ferramenta computacional *AntConc*® 3.4.1m (versão Macintosh OS X) 2014, desenvolvida pelo Professor Doutor do Departamento de Ciências e Engenharia da Universidade de Waseda, no Japão, Laurence Anthony.

A seguir, apresentamos a imagem da tela inicial do programa *AntConc*®:

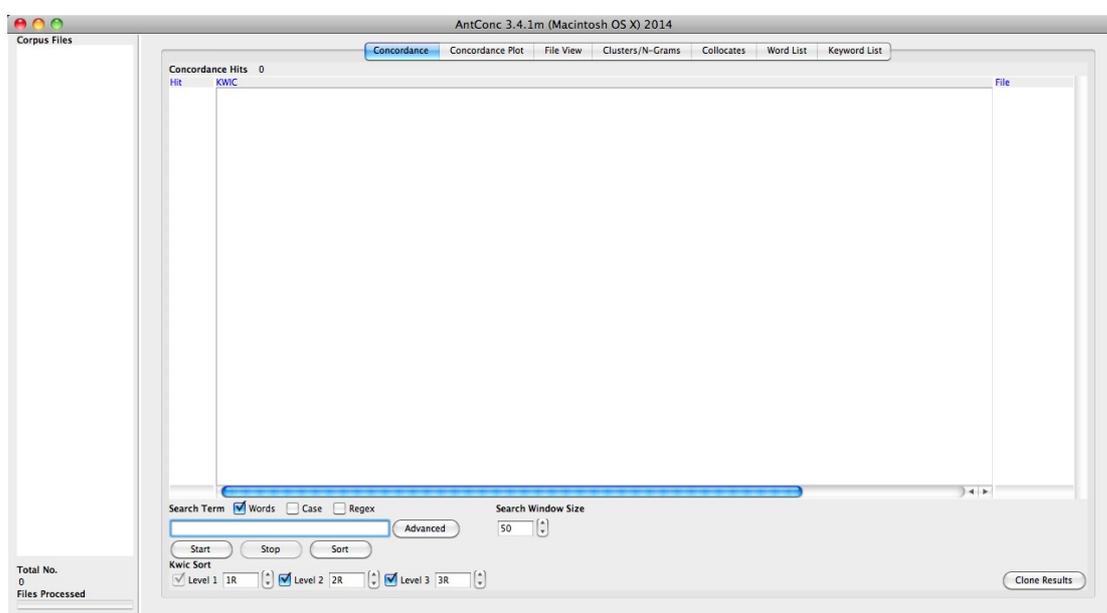


Figura 2: Tela inicial do programa *AntConc*

Como podemos observar na figura acima, o *AntConc* nos possibilita a análise do *corpus* por meio das seguintes telas: *Concordance*, que nos apresenta linhas de concordância com o termo utilizado na busca; *Concordance Plot*, tela que apresenta, em um gráfico linear, o termo pesquisado distribuído conforme a disponibilização deste no *corpus*. Quando clicamos em uma palavra na tela *Concordance*, o *AntConc* nos direciona para a tela *File view*, que nos apresenta a palavra dentro do trecho onde foi originalmente escrito. Contamos ainda com a aba *Clusters/N-grams*, que nos oferece a possibilidade de pesquisar uma composição de dois ou mais termos juntos. *Collocates* nos permite gerar uma lista de palavras próximas àquela que estamos pesquisando e, clicando em uma das palavras da lista, somos redirecionados novamente à aba *Concordance*, para que possamos observar a linha de concordância. *Word List* gera uma lista das palavras que existem no *corpus* apresentada em ordem de frequência e, por último, *Keyword List* apresenta uma lista de palavras mais ou menos frequentes no *corpus* de análise em comparação a um *corpus* de referência.

Para realizar a análise, adicionamos os dois *corpora* no programa *AntConc* e selecionamos a tela *Concordance Plot*. Colocamos a palavra a ser observada no buscador e assim observamos em uma imagem dois gráficos lineares, semelhantes a um código de barras, um posicionado abaixo do outro e, na frente, os números de ocorrência da palavra buscada. Podemos observar a tela do *Concordance Plot* contendo essas informações abaixo:

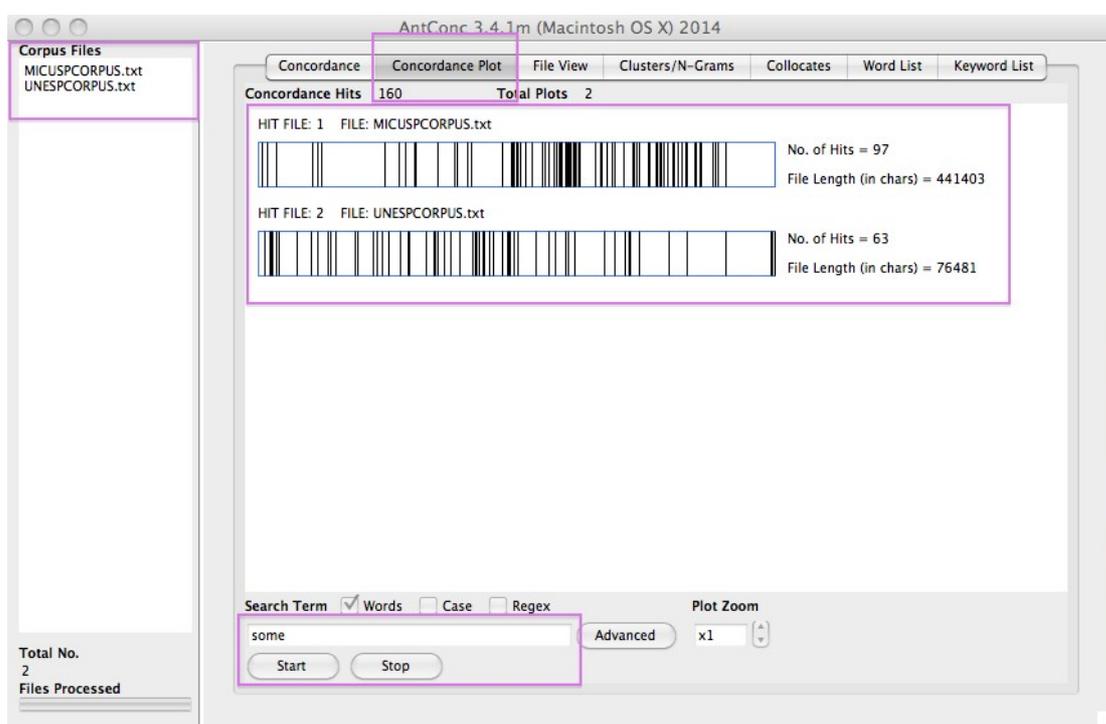


Figura 3: Tela do *Concordance Plot*

Em seguida, limpamos o programa com o objetivo de excluir os dois *corpora* e adicionamos novamente somente o *corpus* CorIFA, para realizar as análises dos dados na tela *Concordance*, ilustrada na Figura 2.

3. Análises preliminares

Uma vez que nosso objetivo final é elaborar atividades didáticas direcionadas a alunos de nível B1, focamos em algumas competências linguísticas pertinentes à habilidade de produção escrita referentes a este nível, segundo o QEER.

Dentre estas competências, optamos por trabalhar com os quantificadores *some*, *many* e *any*, visto que são considerados uma das classes mais importantes de determinantes e constam no arcabouço de conhecimento, acerca de elementos gramaticais, necessário para a proficiência de nível B1.

Para averiguar a porcentagem de ocorrência dos quantificadores pesquisados no *corpus*, faremos o cálculo usando a regra de 3:

$$\frac{\text{n}^\circ \text{ itens}}{\text{forma}} \cdot 100 \rightarrow \text{n}^\circ \text{ itens } x = \text{n}^\circ \text{ forma} \cdot 100 \rightarrow \text{n}^\circ$$

$$x = \frac{\text{n}^\circ \text{ forma} \cdot 100}{\text{n}^\circ \text{ itens}} \rightarrow x = \%$$

Realizado o cálculo para identificarmos a porcentagem do uso da palavra em cada *corpus*, passamos para um novo cálculo de comparação entre os *corpora*, para sabermos onde a palavra foi mais usada. Para alcançarmos esse resultado, faremos a divisão entre os resultados obtidos por meio do cálculo anterior e então teremos o número de vezes em que a palavra foi mais ou menos utilizada no *corpus* CorIFA. Segue exemplo da fórmula:

$$\frac{\text{resultado CorIFA}}{\text{resultado MICUSP}} = \text{número de vezes em que a palavra foi mais ou menos utilizada}$$

3.1. Análise do quantificador *some*

No que tange às palavras funcionais, fizemos um recorte dentro da categoria dos determinantes, uma vez que esta também é considerada pelo QEER uma competência linguística inerente aos alunos B1. Delimitamo-nos quanto ao uso de determinantes quantificadores, que notadamente causam certa confusão nos aprendizes de língua inglesa.

Observamos na figura abaixo uma comparação entre o uso do determinante quantificador *some* em ambos os *corpora*:

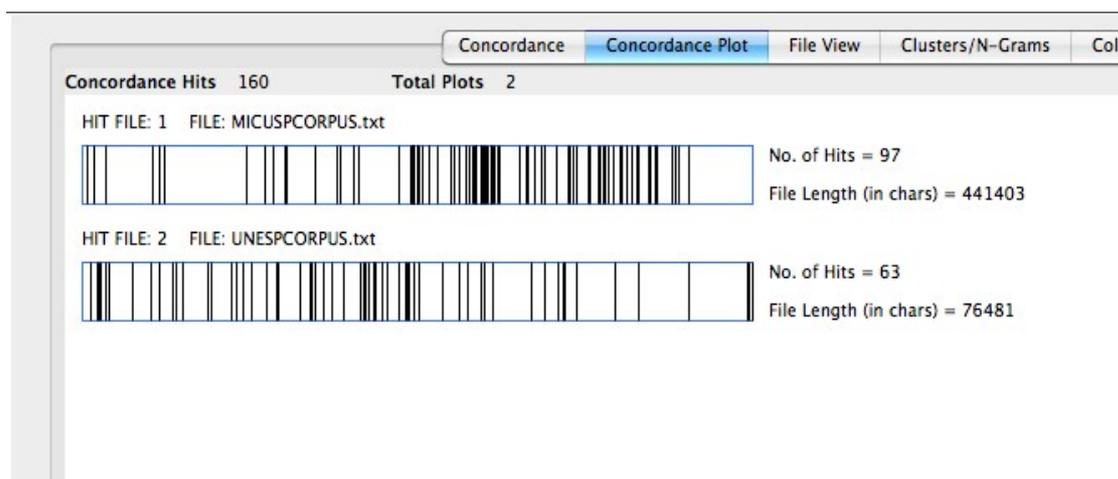


Figura 4: Tela do *AntConc* com comparação dos *corpora*

A palavra *some* no *corpus* CorIFA aparece 63 vezes, ao passo que no *corpus* MICUSP temos 97 ocorrências. Realizando um cálculo no qual dividimos o número de ocorrência da palavra pelo número de itens, obtivemos os seguintes resultados:

$$\text{Corpus CorIFA: } 63 / 13.244 = 0.48\%$$

$$\text{Corpus MICUSP: } 97 / 71980 = 0.14\%$$

Concluimos que no *corpus* CorIFA, *some* tem uma ocorrência aproximadamente 3.4 vezes maior.

Nosso segundo passo foi observar todas as linhas de concordância com *some* para identificar de que forma os alunos estavam usando o quantificador e propor discussões a respeito desses dados. Observamos a seguinte oração “*some specialist says*”, conforme ilustrado na figura abaixo:

anything, even if we do not really need *some* things, these ads keep pushing products (come capitalization and development of a country. *Some* specialist says that advertising is changing : the death or life of the classical genre. *Some* experts affirm that China, the most populous cheaper and easier than ever before. However, *some* people still believe that this old genre is

Figura 5: Captura de tela do *AntConc*

Segundo o dicionário *on-line* Cambridge, *some* é um determinante que se refere a um número não específico de substantivos e, quando antecede um substantivo contável, indica a obrigatoriedade do uso deste em seu plural. Ainda segundo o mesmo dicionário, *specialist* é um substantivo contável, portanto, estando antecedido por *some*, deveria estar no plural. Em uma consulta a outro dicionário *on-line* chamado WordReference³, encontramos *specialist* também como sendo um substantivo que pode ser usado como adjetivo⁴. Como nosso objetivo é fazer uma análise da língua em uso e contextualizada, estendemos o trecho em observação e notamos que o autor usou o verbo *to say* em seguida. O verbo foi conjugado concordando com a palavra *specialist* indicando que seu objetivo era usar a palavra *specialist* como substantivo e, portanto, usando o determinante *some* de forma inadequada. Estando o substantivo no singular, o autor deveria ter optado pelo artigo definido *a*.

3.2. Análise do quantificador *many*

Quanto ao uso do quantificador *many*, apresentaremos a tela abaixo:

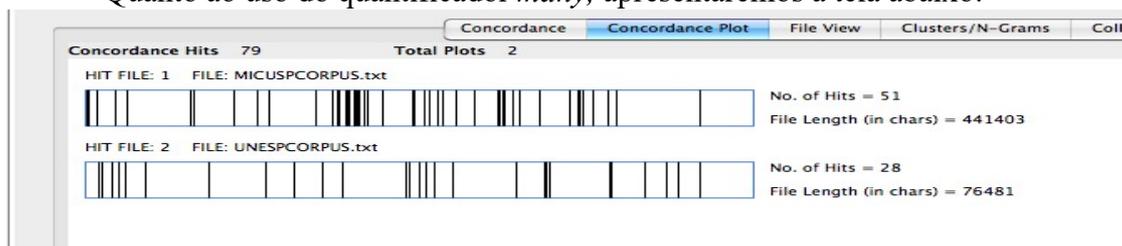


Figura 7: Tela do *AntConc* com comparação dos *corpora*

O quantificador *many* aparece no CorIFA em 28 linhas de concordância, ao passo que no MICUSP, o encontramos 51 vezes. Realizando os mesmos cálculos com o objetivo de comparar a diferença de ocorrência entre os *corpora*, teremos:

Corpus CorIFA: $28 / 13.244 = 0.21\%$

Corpus MICUSP: $51 / 71980 = 0.07\%$

Não muito diferente da palavra anterior, os dados indicam que *many* teve uma ocorrência 3 vezes maior no *corpus* CorIFA.

³ Disponível em www.wordreference.com.br. Acesso em: 9 jun. 2017.

⁴ Texto original: *noun as adjective*.

Encontramos a seguinte oração escrita por um aluno brasileiro:

good messages in ways that they can reach many people. For example, informative advertisement and the options in choosing not half as many options. Examples of successful advertisement ads made for children, and also, there are many advertisings known by it's mysoginy (as beer damages are being widely discussed nowadays, and many consumers are starting to question the values , I believe that As is just one of many ways to make us think about our lives.

Figura 8: Tela do *AntConc*

Conforme podemos visualizar na figura acima, o aluno escreveu: “[...] *there are many advertisings* [...]”. Em uma busca pela palavra *advertising* no dicionário *on-line* de Cambridge, encontramos a informação de que se trata de uma palavra incontável, portanto não pode ser antecedida pelo quantificador *many*.

3.3. Análise do quantificador *any*

Quando fizemos a busca pela palavra *any*, entretanto, notamos que a palavra foi utilizada 2.75 vezes menos pelos alunos brasileiros, conforme apontam os cálculos a seguir:

Corpus CorIFA: $5 / 13.244 = 0.04\%$

Corpus MICUSP: $78 / 71980 = 0.11\%$

Abaixo temos a tela *Concordance Plot* ilustrando os dados extraídos:

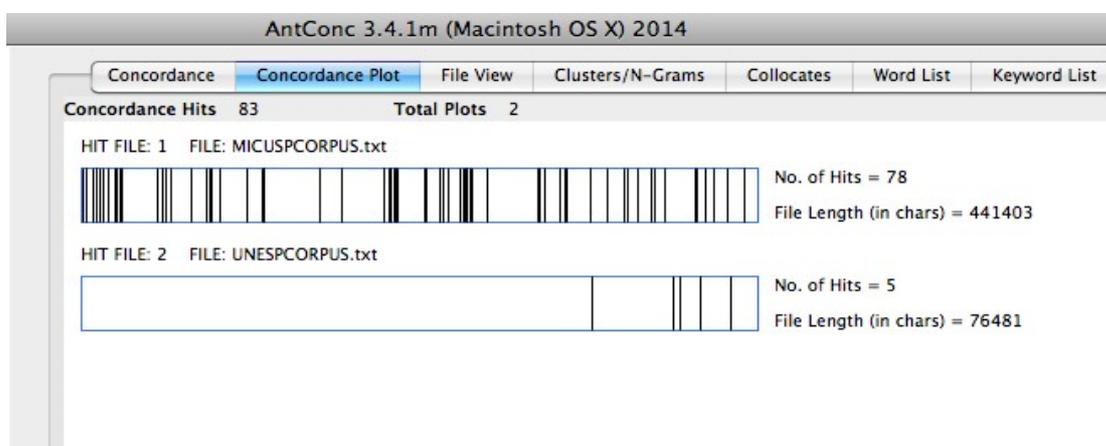


Figura 9: Tela do *AntConc* com comparação dos *corpora*

Trazemos a figura com a captura de tela do *AntConc* na qual podemos observar todos os usos da palavra *any* em nosso *corpus*:

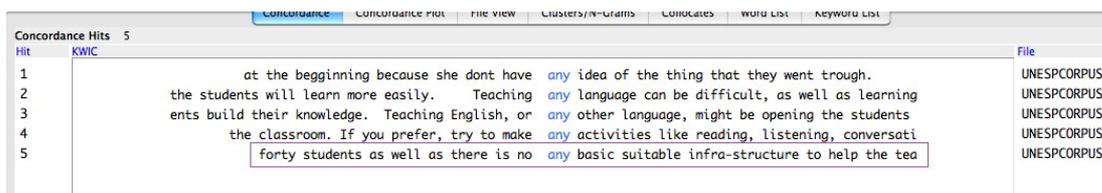


Figura 10: Captura de tela do *AntConc*

Das 5 orações, apenas a última emprega a palavra de forma inadequada, conforme apontado abaixo: “[...] *there is no any basic suitable infra-structure* [...]”.

Observando a oração destacada, notamos que o aluno pretende falar sobre a ausência completa de “infra-estrutura básica”, uma vez que ele lança mão do uso de *no*. Entretanto, segundo a LGSWE (BIBER et al., 1999), *no* tem significado equivalente a *not any*, implicando em repetição errônea e desnecessária. Neste caso, o ideal seria excluir o uso de *any* na oração, ou substituir *no* por *not*.

Considerações finais

Com base na observação do *corpus* composto por textos acadêmicos de estudantes brasileiros aprendizes de língua inglesa em nível B1, segundo o QCRE, fizemos uma análise quanto ao número de ocorrências dos determinantes quantificadores *some*, *many* e *any*.

Em uma análise mais detalhada, observamos que os alunos usaram os quantificadores *some* e *many* antecedendo substantivos incontáveis, destacando a dificuldade que existe entre os aprendizes de língua inglesa em usar adequadamente esses elementos gramaticais. Dentre os exemplos destacados neste artigo, entendemos que os alunos conseguem se comunicar, mas seria necessário o uso adequado da linguagem acadêmica de escrita para que obtivessem sucesso em um exame de proficiência classificatório para mobilidade internacional. Com relação ao quantificador *any*, apesar de pouco usado, quase não foram encontrados desvios de uso, exceto por uma situação de repetição.

Esse é um estudo com análises realizadas até o momento e que está em andamento. Iremos, juntamente com as leituras que realizamos a respeito de Linguística de *Corpus*, Ensino e Aprendizagem de Língua Estrangeira e Escrita Acadêmica, desenvolver atividades didáticas com base em *corpus* e foco em produção escrita.

Parte do objetivo deste artigo é contribuir com os textos coletados para o *corpus* do site do CorIFA, ampliando o material de apoio para outros professores e alunos de programas de pós-graduação que possuem projetos de estudo semelhantes ao nosso.

REFERÊNCIAS

- ANDERMAN, G.; ROGERS, M. *Incorporating Corpora The Linguist and the Translator*. [S.l], Multilingual Matters, 2007.
- ANTHONY, L. *Antconc* (Version 3.4.4) [Computer Software]. Tokyo. Japan: Waseda University, 2011.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Manole, 2004.
- _____. Linguística de corpus: histórico e problemática. *D.E.L.T.A.*, São Paulo, EDUC, v. 16, n. 2, p. 323-367, 2000.
- BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press, 1998.
- BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman Grammar of Spoken and Written English*. London: Pearson, 1999.
- British Council*. Disponível em: <<https://www.britishcouncil.org.br/>>. Acesso em: 8 ago. 2016.

- BRITISH COUNCIL. *Demandas de Aprendizagem de Inglês no Brasil*, [S.l.], 2014. Disponível em: <https://www.britishcouncil.org.br/sites/default/files/demandas_de_aprendizagempesquisacompleta.pdf>. Acesso em: 16 jun. 2017.
- _____. Corpus based approaches to issues in applied linguistics. *Applied Linguistics*, Oxford University Press, v. 15, n. 2, p. 169-223, 1994.
- CAMARGO, D. C. de; ROCHA, C. F.; PAIVA, P. T. P. (Orgs.). *Pesquisas em estudos da tradução e corpora eletrônicos no Brasil*. São Paulo: Editora UNESP, 2012.
- CorIsf*. Disponível em: <<https://sites.google.com/site/corpusisf/>>. Acesso em: 4 ago. 2016.
- DAYRELL, C. Corpora no ensino de inglês acadêmico: padrões léxico-gramaticais em abstracts de pós-graduandos brasileiros. In: GILQUIN, G.; GRANGER, S.; PAQUOT, M. Learner corpora: the missing link in EAP pedagogy. *Journal of English for Academic Purposes*, v. 6, n. 4, p. 319-355, 2007.
- DELEGÁ-LUCIO, D. *A relexicalização de adjetivos nas redações de alunos de inglês – um estudo baseado em corpus de aprendiz*. 2006. 105 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2006
- DEW, S. E. *Practical Academic Essay Writing Skills*. 1st ed. [S.l.] HBICambodia, 2010.
- DUTRA, D. P.; GOMIDE, A. R. Compilation of a University Learner Corpus. *Brazilian English Language Teaching Journal*, Porto Alegre, v. 6, p. s21-s33, dez. 2015.
- GUEDES, A. S. The Academic Writing of University English Learners: A Corpus Based Study. *Revista do SELL*, UFTM, [S.l.], v. 5, n. 1, 2016.
- FIRTH, J. R. Linguistics and translation. In: PALMER, F. R. (Ed.). *Selected Papers of J.F. Firth, 1952-59*. London and Harlow: Longmans, Green and Co, 1968. p. 84-95.
- FLOWERDEW, L. The Exploitation of small learner corpora in EAP materials design. In: GHADESSY, M.; HENRY, A.; ROSEBERRY, R. L. (Eds.). *Small Corpus Studies and ELT [Studies in Corpus Linguistics 5]*. Amsterdam: John Benjamins, 2001. p. 363-379.
- GRANGER, S. A bird's eye view of learner corpus research. In: GRANGER, S.; HUNG, J.; PETCHTYSON, S. (Eds.). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam/Philadelphia: John Benjamins Publishing, 2002. p. 3-33.
- GRANGER, S.; TRIBBLE, C. Learner corpus data in the foreign language classroom: form focused instruction and data driven learning. In: GRANGER, S. (Ed.). *Learner English on computer*. London: Longman, 1998. p. 199-209.
- MICUSP*. Disponível em: <<http://micusp.elicorpora.info/>>. Acesso em: 29 ago. 2016.
- PAIVA, P. T. P. Instrumentos e atividades realizadas para a coleta de um *corpus* de aprendizes em língua inglesa para integrar o Br-ICLE (Brazilian Portuguese Sub-corpus of ICLE). *Estudos Linguísticos*, São Paulo, v. 1, p. 312-322, 2013
- THORNDIKE, E. L. *Teacher's wordbook*. Nova York: Columbia Teachers College, 1921.

WordSmith Tools. Versão 5. Oxford: Oxford University Press, 2007.

STTUBBS, M. Grammar, text, and ideology: computer assisted methods in the linguistics of representation. *Applied Linguistics*, v. 15, n. 2, p. 201-223, 1994.

Recebido em: 24/09/2017

Aprovado em: 15/02/2018