

O desenvolvimento de atividades de compreensão escrita: um *corpus* de temática política

Luana Aparecida Nazzi Laranja¹
Paula Tavares Pinto²

Resumo

Baseando-se nos estudos da Linguística de Corpus (LC) o objetivo do presente trabalho é desenvolver atividades de compreensão escrita voltadas para o vestibular baseadas no uso de agrupamentos lexicais. Para isso, foi compilado um *corpus* de textos jornalísticos autênticos advindos dos sites CNN® e BBC® com o tema *politics* (política) e este foi comparado a um *corpus* de inglês geral com o intuito de extrair as palavras-chave positivas, ou seja, do *corpus* de estudo. A partir de então, foram levantados os *clusters* - agrupamentos lexicais - dessas palavras para posterior criação das atividades de compreensão escrita na língua inglesa, de modo que os exercícios desenvolvidos tiveram como base as combinações lexicais mais frequentes do tema dos textos compilados. Como base para o desenvolvimento dos exercícios foi utilizado o vestibular da Unesp, contendo questões em português e a língua-alvo. Em função disso, foi esperado responder à seguinte pergunta: Como o uso de um *corpus* criado a partir de textos jornalísticos autênticos com o tema *política* pode auxiliar no levantamento de agrupamentos lexicais para o vestibular? A aplicação das atividades criadas foi feita com alunos do terceiro ano do ensino médio em uma escola privada situada em uma cidade do interior de São Paulo. E, pode-se afirmar que as atividades tiveram um impacto positivo em relação ao desenvolvimento da leitura dos estudantes.

Palavras-chave: Compreensão escrita. Linguística de corpus. Política.

Abstract

Based on the studies of Corpus Linguistics (LC), the objective of the present work is to develop written comprehension activities aimed at entrance exams based on the use of clusters. To this end, a corpus of authentic journalistic texts from the CNN® and BBC® sites was compiled with the theme politics and compared to a corpus of general English in order to extract the positive keywords, that is, of the corpus of study. From then on, the clusters of these words were raised for later creation of the comprehension activities written in the English language, so that the exercises developed were based on the most frequent lexical combinations of the theme of the compiled texts. The basis for the development of the exercises was the Unesp entrance exam, including questions in Portuguese and the target language. As a result, it was expected to answer the following question: How can the use of a corpus created from authentic journalistic texts with the political theme help in the raising of lexical groupings for the entrance exam? The application of the created activities was made with third year high school students in a private school located in a city in the interior of São Paulo. And, it can be said that the activities had a positive impact on the development of student reading.

Keywords: Written comprehension. Corpus Linguistics. Politics.

¹ Mestranda em Estudos Linguísticos - Universidade Estadual Paulista - UNESP

² Doutora em Estudos Linguísticos - Universidade Estadual Paulista - UNESP

1 Introdução

Com o passar do tempo, tem sido possível perceber que a língua inglesa (LI) tem se expandido em todo o mundo, sendo um dos idiomas mais exigidos no âmbito acadêmico e/ou profissional (JENKINS, 2006, 2009; CRYSTAL, 2003). Assim, cursos de idiomas têm se voltado para os interesses de cada aluno, procurando alcançar os seus objetivos, sejam eles voltados para turismo, negócios, motivos pessoais ou simplesmente passar em determinado exame, como é o caso em que a presente pesquisa mantém o foco – o vestibular – prova que abre as portas para o ensino superior no Brasil e que têm exigido a LI.

Ainda que a abordagem LinFE (Línguas para Fins Específicos) (HARVEY, A.; HORSELLA) possa abranger as quatro habilidades linguísticas, produção oral e escrita e compreensão oral e escrita, as escolas mantêm foco nessa última habilidade, visto que, como mostra o PCN

[...] o uso de uma língua estrangeira parece estar, em geral, mais vinculado à leitura de literatura técnica ou de lazer. Note-se também que os únicos exames formais em Língua Estrangeira (vestibular e admissão a cursos de pós-graduação) requerem o domínio da habilidade de leitura. Portanto, a leitura atende, por um lado, às necessidades da educação formal, e, por outro, é a habilidade que o aluno pode usar em seu contexto social imediato. (MEC/ SEF, 1998, p. 20).

Nesse contexto, sendo parte integrante da abordagem LinFE, o Inglês para Fins Específicos (IFE), mantém o foco no objetivo do aluno envolvendo situações de usos reais da LI. É assim, dentre outras formas, que a Linguística de *Corpus* (LC) vem ganhando espaço nas pesquisas linguísticas, especialmente no que diz respeito ao contexto de ensino aprendizagem de línguas estrangeiras (LE). Berber Sardinha (2011), discorrendo acerca da LC no ensino, explica que o professor/linguista de *corpus* não pode ficar dependente somente de uma lousa e oratória em sala de aula, outros métodos podem ser adotados para que o aluno tenha um aprendizado mais efetivo.

Portanto, levando em consideração a relevância de um curso nos moldes LinFE e também a compreensão escrita como habilidade que permite ao aluno o acesso à uma graduação, o objetivo do presente trabalho é desenvolver atividades de compreensão escrita voltadas para o vestibular baseadas no uso da LC.

Para isso, foi compilado um *corpus* de textos jornalísticos autênticos advindos dos sites CNN® e BBC® com o tema *politics* (política) e este foi comparado a um *corpus* de inglês geral para que fossem extraídas as palavras-chave positivas, ou seja, do *corpus* de

estudo. A partir de então, foram levantados os agrupamentos lexicais dessas palavras para posterior criação das atividades de compreensão escrita, de modo que os exercícios tiveram como base as combinações lexicais mais frequentes do tema dos textos compilados. Em função disso, foi esperado responder à seguinte pergunta: 01. Como o uso de um *corpus* criado a partir de textos jornalísticos autênticos com o tema *política* pode auxiliar no levantamento de agrupamentos lexicais para o vestibular? A aplicação das atividades desenvolvidas foi feita com alunos do terceiro ano do ensino médio em uma escola privada situada em uma cidade do interior de São Paulo.

2 Fundamentação Teórica

A relevância da abordagem LinFE se baseia nas necessidades e objetivos do aprendiz em relação à língua-alvo. Nesse contexto, as aulas são voltadas especificamente para o que aluno espera aprender em determinado idioma e quais habilidades deseja desenvolver, bem como o professor toma consciência em levar essa necessidade do aprendiz em consideração no momento de elaboração do material.

Segundo o estudioso Jordan (1997, p. 143), “Quando os alunos leem, é para um propósito³” e é esse fator que impulsiona a leitura dos estudantes, seja esse propósito o prazer pela leitura em si ou, como tem sido destacado na presente pesquisa, a intenção de prestar o vestibular. Ainda de acordo com Jordan (1997, p. 51), quando os estudantes são submetidos à autoavaliações e/ou questionários acerca de habilidades em LI, há a tendência de respostas afirmando que a compreensão escrita é a habilidade que eles têm menos dificuldade em comparação às outras três habilidades linguísticas – compreensão oral, produção oral e produção escrita. Ademais, o autor explica que essa realidade não significa que os alunos saibam ler perfeitamente, ou seja, que não há nenhum tipo de problema quando a habilidade em questão é a compreensão escrita. Assim sendo, como o professor pode ajudar seus alunos utilizando a abordagem IFE em conjunto com a LC, a melhorarem o seu desenvolvimento em compreensão escrita?

Em consonância com o que foi dito anteriormente, Hutchinson e Waters (1987) afirmam que se tanto professor quanto aluno sabem o motivo pelo qual estão estudando a

³ Tradução nossa para: “*When students read, it is for a purpose*” (JORDAN, 1997, p. 143).

língua, essa consciência terá uma influência positiva na seleção do material que será utilizado em sala de aula (SA), bem como as habilidades que terão ênfase durante o aprendizado. Todavia, um dos fatores considerados desfavoráveis nessa abordagem é o tempo que o professor leva para organizar materiais que sejam específicos para cada aluno, tendo de se organizar muito previamente.

Nesse contexto, a Linguística de Corpus (LC) passa a ser uma grande aliada, auxiliando professores e pesquisadores a desenvolverem inúmeros tipos de estudos, sendo o foco de um deles a elaboração de materiais, de forma que a LC

ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica -se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (BERBER SARDINHA, 2004, p. 3)

De acordo com Berber Sardinha *et al* (2012), a LC pode ser empregada em qualquer tipo de vertente do ensino de idiomas, incluindo o Ensino de Línguas para Fins Específicos, já que a LC não conta especificamente com nenhuma abordagem ou método de ensino, podendo ser utilizado em materiais/aulas de diferentes perspectivas teóricas. Römer (2011) discorre acerca da contribuição da LC na área do ensino dividindo as suas aplicações entre diretas e indiretas. Essa última diz respeito aos recursos que não desenvolvidos exatamente para o ensino, mas que podem contribuir nesse contexto, como o dicionário, por exemplo. Já as aplicações diretas têm, como o próprio nome sugere, ligação direta com o ensino, como materiais didáticos. A presente pesquisa está inserida no primeiro tipo.

Segundo Viana (2011), além de ser possível obter informações acerca do funcionamento de uma língua por meio de contextos reais de uso, o que é chamado de dado autêntico da língua (O'KEEFFE; McCARTHY; CARTER, 2007), as pesquisas baseadas em *corpus* mostram como a língua é padronizada, ou seja, a relação existente entre uso e contexto. Cada contexto da língua traz probabilidades e padrões diferentes (VIANA, 2011), o que pode ser utilizado de modo construtivo em materiais didáticos para alunos que objetivam aprender uma determinada língua.

Ainda no que tange ao contexto de ensino-aprendizagem de LE e o uso autêntico da língua, Dutra e Silero (2010) concordam que a LC tem muito a acrescentar aos estudos da Linguística Aplicada, já que, a utilização de *corpus* em sala de aula mostra o uso real da língua,

que, para professor e aluno é uma ferramenta de grande utilidade, especialmente em contextos específicos, visto que o aprendiz tem a chance de lidar com textos/materiais que são selecionados criteriosamente para atender suas necessidades e de acordo com a sua área de especialidade.

Portanto, em função das vantagens advindas da união da abordagem LinFE e da teoria e metodologia da LC, espera-se que o material selecionado para a presente pesquisa seja relevante para os alunos do terceiro ano que pretendem prestar o vestibular, aprendendo técnicas de leitura em conjunto com um material autêntico, como o que é utilizado nesses tipos de exame. Dessa maneira, tanto professor quanto aluno têm a possibilidade de aproveitar melhor o tempo em sala de aula, visto que o conteúdo do material será importante para alcançar seus objetivos durante a prova. Nesse contexto, o aluno busca pelos seus interesses e o professor tem um papel colaborativo, na tentativa de solidificar o conhecimento do aluno de modo que o mesmo veja sua evolução.

Para tanto, o desenvolvimento das atividades dessa pesquisa apresenta aparato teórico em Inglês Instrumental dos estudiosos Celani (1984), Hutchinson e Waters (1987), Dudley-Evans e St. Johns, (1998), e em Linguística de Corpus com McCarthy; Carter (2007); Berber Sardinha (2009), O’Keeffe; Dutra e Silero (2010) e Viana e Tagnin (2011).

3 Palavra-chave: qual a sua relevância?

A presente pesquisa buscou extrair 14 agrupamentos lexicais de palavras-chave por meio da comparação entre o *corpus Politics* e o *corpus* de referência de inglês geral *EnglishTenTen*[®]. De acordo com Berber Sardinha, as palavras-chave são “aquelas cuja frequências são diferentes, de modo estaticamente significativo, das frequências encontradas em um *corpus* de referência” (BERBER SARDINHA, 2009, p. 193). O autor explica que a palavra-chave nem sempre é uma palavra individual, como por exemplo *pact* – exemplo extraído do *corpus* de estudo – mas também, o que chamamos de *clusters*, ou seja, agrupamentos lexicais, como *trade pact*, que tende a focalizar ainda mais o contexto do tema em si e que serão abordados em breve.

As palavras-chave relevantes para a presente pesquisa são as denominadas positivas, isto é, aquelas cuja frequência é mais alta no *corpus* de estudo do que no *corpus* de

referência⁴, já que este representa a língua de maneira generalizada. As palavras-chave negativas, que não são o foco do trabalho, possuem uma frequência muito mais alta no *corpus* de referência em relação àquele compilado pelo pesquisador.

Berber Sardinha (2009) elucida que as palavras-chave não são absolutas, ou seja, não são chave independentes do *corpus* de referência. Assim, o *corpus* escolhido pelo pesquisador para ser referência da língua tem forte influência na busca pelas palavras-chave. Em outras palavras, ao utilizar *corpora* de referência distintos, as palavras-chave serão outras. Como já mencionado, o *corpus* de referência desta pesquisa é o *EnglishTenTen*[®].

Segundo Scott (2009), as palavras-chave podem ser classificadas entre temáticas, gramaticais e identificativas. As primeiras dizem respeito ao tema, ao assunto do *corpus* de estudo, que são, como afirma Berber Sardinha (2009) inicialmente identificadas. O segundo tipo, apesar de não estar entre as primeiras da lista, já que ocorrem tanto no *corpus* de estudo quanto no *corpus* de referência, são úteis na análise do estilo do *corpus* em questão. Em relação às identificativas, essas são os nomes próprios que aparecem no *corpus* e têm caráter exclusivo. A presente pesquisa faz uso do primeiro tipo de palavras-chave, as temáticas, visto que o objetivo é desenvolver atividades de compreensão escrita que abordem um único tema, política, e foquem em seu léxico.

Scott (1997) afirma que as palavras-chave oferecem uma maneira muito útil de caracterizar um determinado texto, tendo grande utilidade para diferentes áreas da Linguística, assim como no ensino de línguas, em que os professores podem utilizar as palavras-chave para desenvolver exercícios (como em aulas de abordagem LinFE) em que o léxico pode ser identificado a partir de um texto, por exemplo.

Nesse contexto, o termo “chavicidade” se torna essencial para a pesquisa. Segundo Viana (2011, p. 64) a chavicidade (*keyness*) diz respeito à importância que uma palavra-chave positiva tem no *corpus*. Assim, quanto maior for o valor da chavicidade, a relevância de determinada palavra também aumenta estatisticamente. Esse valor será de muita relevância na metodologia do presente trabalho no que tange à escolha do texto que será abordado, como será esclarecido mais adiante.

⁴ Quando um corpus de amostra afirma ser um repositório razoavelmente confiável de todos os recursos de um idioma, pode ser chamado de corpus de referência. Hoje em dia terá que ser bem grande - 100 milhões de palavras é o tamanho típico - e conterá quantidades substanciais de os principais tipos de linguagem encontrados em uma sociedade (TOGNINI-BONELLI, 2010, p. 20).

4 O uso de agrupamentos lexicais e a sua importância no desenvolvimento de atividades de compreensão escrita

Voltando-se agora para os agrupamentos lexicais, base para o desenvolvimento das atividades, pode-se afirmar que esses são “sequências de ocorrência regular de duas ou mais palavras que parecem possuir significados ou funções unitárias”⁵ (O’KEEFFE; McCARTHY; CARTER; 2007, p. 46) ou ainda, “são palavras que se seguem mais frequentemente do que se espera acontecer por acaso, ajudando a moldar significados de texto e a contribuir para o nosso senso de distinção em um registro”⁶ (HYLAND, 2008, p. 42).

Como é possível verificar, os agrupamentos lexicais são como blocos de palavras que, juntos, correspondem à apenas uma ideia. Entretanto, por que seria importante abordar agrupamentos de palavras ao invés de palavras soltas?

De acordo com O’Keeffe, McCarthy e Carter (2007) em várias situações os agrupamentos são tão frequentes quanto as palavras consideradas “soltas”, isso quando não aparecem mais frequentemente considerando o léxico geral de uma língua. Segundo Sinclair (1991, p. 55), isso é positivo, pois os “agrupamentos lexicais estão prontos para o uso em qualquer momento e não precisam ser reunidos toda vez que são utilizados”⁷.

É desse modo que, quando se tratando do contexto de ensino-aprendizagem de línguas, o aluno passa a “enxergar” a língua por uma perspectiva diferente, ou seja, não tentando traduzir ou entender palavra por palavra, mas sim entender sentenças como um todo. Esse novo ponto de vista auxilia os alunos, inclusive, a aceitar melhor a língua como ela é, pois, muitas vezes, os aprendizes de LI ficam se perguntando “mas por que é assim?”, “por que não pode falar desse jeito”? A partir da utilização de agrupamentos lexicais fica mais “simples” explicar aos alunos que frequentemente esses agrupamentos de palavras coocorrem e que possuem, na maior parte das vezes, um único sentido.

⁵ Tradução nossa para “regularly occurring strings of two or more words which seem to possess unitary meanings or functions” (O’KEEFFE; McCARTHY; CARTER; 2007, p. 46).

⁶ Tradução nossa para “Essentially, these are words which follow each other more frequently than expected by chance, helping to shape text meanings and contribute to our sense of distinctiveness in a register” (HYLAND, 2008, p. 42).

⁷ Tradução nossa para “Chunks are ready for use at any moment and do not need re-assembling every time they are used” (SINCLAIR, 1991, p. 55).

Visto as principais características dos agrupamentos lexicais, faz-se necessário explicar a sua importância dentro do contexto de ensino-aprendizagem de LI, já que esse é o foco de seu uso no presente trabalho.

Como asseguram O’Keeffe McCarthy e Carter (2007), o ensino de LE tinha como base as famosas listas de palavras que, muitas vezes, advinham de textos construídos a partir da intuição daqueles que desenvolviam os materiais pedagógicos, portanto a confiança atribuída a esses materiais era menor, visto que sempre havia a dúvida de que algo poderia estar divergindo do que os falantes nativos realmente utilizavam. A partir do avanço de pesquisas nesse contexto e com o auxílio da LC, textos autênticos passaram a ser utilizados na construção dos materiais pedagógicos, bem como a língua falada passou a ser reportada e ensinada em sala de aula.

Sinclair (1991) elucida a relevância do uso de agrupamentos afirmando que o acesso aos padrões da língua permite que o processamento linguístico seja mais rápido e esse fator sustenta uma produção e compreensão mais fluente da língua, já que não é necessário atentar-se a palavras individuais do texto. Além do mais, como já foi mencionado, dessa maneira os alunos tendem a compreender a LI de um modo melhor e mais eficiente.

Corroborando os pensamentos de Sinclair (1991), Berber Sardinha afirma que

Normalmente, os aprendizes de língua estrangeira tendem a ver a linguagem com um conjunto de palavras individuais sustentadas por regras gramaticais; desse modo, a compreensão e a produção são, em muitos casos, com base na junção de palavras individuais. (BERBER SARDINHA, 2009, p. 13).

Segundo o autor, utilizar *corpora* no ensino de LE pode auxiliar a compreensão dos alunos em relação à língua-alvo, já que esses têm um contato maior com unidades já ‘pré-fabricadas’ e não individuais, sendo que essas podem dificultar o entendimento dos textos, sejam eles orais ou escritos, em LE. Assim sendo, acredita-se que, quando um aluno passa a apreender o conteúdo da LI por meio de agrupamentos lexicais, há uma tendência mais eficaz de aprendizagem, pois o estudante não tenta “fragmentar” a língua, mas, ao contrário, procura entendê-la por meio de uma ideia mais geral. Desse modo, quando o aluno ouve ou lê um agrupamento como *trade deal*, já é possível que esse aprendiz consiga inferir o contexto desse agrupamento e entender que se tratam de acordos comerciais, além do mais, com esse

processo é menos provável que o estudante utilize agrupamentos inexistentes e/ou menos frequentes, como *commercial deal*, por exemplo.

Tendo vista a importância desses agrupamentos no contexto de ensino-aprendizagem de LI, especialmente quando diz respeito à compreensão escrita na língua-alvo, será explicado a seguir como foram utilizados os agrupamentos no desenvolvimento das atividades bem como a descrição do *corpus* utilizado e os procedimentos metodológicos referentes a esta pesquisa.

5 Metodologia

5.1 O *corpus* – *politics*

O *corpus* denominado *Politics* devido o tema das atividades (política), foi compilado por meio do *software AntConc*® (ANTHONY, 2018) e possui, no total, 55.861 *types* e 64,811 *tokens*. Foram compilados três textos semanalmente durante o período de cinco meses (agosto a dezembro de 2018) totalizando 66 textos retirados dos sites BBC® e CNN®. É importante ressaltar que todos os textos eram atuais, ou seja, escritos na mesma semana da compilação quando não, da semana anterior.

Como é possível verificar no quadro abaixo, baseada nos critérios de Berber Sardinha (2000) para a construção de um *corpus*, temos o conteúdo e o propósito do *corpus* da presente pesquisa de um modo resumido:

MODO	Escrito – o <i>corpus</i> será composto de textos eletrônicos
TEMPO	Contemporâneo
SELEÇÃO	Composto por textos escritos, sendo eles jornalísticos e autênticos, retirados dos sites BBC® e CNN®
CONTEÚDO	O conteúdo do <i>corpus</i> é de caráter especializado, todos textos são do tema <i>política</i> .
AUTORIA	Os autores são falantes nativos ou não-nativos.
FINALIDADE	O <i>corpus</i> foi utilizado para o desenvolvimento de atividades de compreensão escrita.

Quadro 1. Critérios para a compilação do *corpus* de estudo
Fonte: Berber Sardinha (2000, p. 340)

Como pode-se notar, o *corpus* é composto por textos eletrônicos retirados dos sites BBC® e CNN®, configurando textos do tipo jornalístico e autêntico, visto que são retirados do contexto real de uso. O tema dos textos refere-se ao conteúdo de notícias políticas circuladas nos sites mencionados, configurando o *corpus* de estudo como especializado, já que aborda um único tema e são extraídos agrupamentos que fazem menção a esse tema em específico, não sendo, assim, de caráter geral.

Faz-se relevante mencionar que, apesar do conteúdo específico, tem-se a pretensão de trabalhar com os alunos conteúdos de notícias atuais, ou seja, estudar a LI voltada para a compreensão escrita, já que é a habilidade exigida nos exames de vestibular, e o tema escolhido foi política por ser um assunto muito em voga no mundo todo. A partir de então verificou-se a necessidade de compilar um *corpus* atual com as notícias acerca desse assunto, a fim de abordar com os alunos a língua autêntica ao mesmo tempo que eles se familiarizam com vocabulário mais utilizado nessas notícias.

Dessa maneira, os instrumentos da pesquisa foram organizados como consta no Quadro 2.

Objetivos	Instrumentos
Coleta de dados	- Compilação do <i>corpus</i> - Extração das palavras-chave e agrupamentos lexicais
Aplicação das atividades	- A aplicação das atividades se deu em uma escola particular de ensino regular situada em uma cidade do interior de São Paulo. As aulas foram ministradas durante o primeiro semestre de 2019. - Público – alunos do segundo e terceiro anos do ensino médio.
Análise de dados	- Uso do software <i>Sketch Engine</i> ® para análise do <i>corpus</i> . - Avaliações diagnósticas como forma de obter um <i>feedback</i> sobre a aprendizagem dos alunos.
Apresentação dos resultados	- Apresentação dos dados obtidos pela análise do <i>corpus</i> . - <i>Feedback</i> dos alunos por meio de questionário.

Quadro 2. Organização da pesquisa
Fonte: Dados da presente pesquisa

5.2 Critérios obrigatórios na preparação de atividades didáticas baseadas em *corpora*

As atividades propostas pela presente pesquisa se baseiam nos critérios obrigatórios para atividades com base em *corpora* (DELFINO, 2016) como exhibe o quadro abaixo⁸:

- O1 – O exercício faz uso de *corpus*;
- O2 – O exercício precisa ter enunciados claros;
- O3 – O exercício tem como foco principal o padrão léxico-gramatical;
- O4 – O exercício é ético;
- O5 – O exercício é replicável;
- O6 – O exercício é motivador;
- O7 – O exercício não simplifica a língua usada nos textos/concordâncias/lista de palavras, etc.;
- O8 – O exercício deve apresentar nível de dificuldade adequado;
- O9 – O exercício contém conteúdo relevante para o aluno e para a construção do conhecimento em inglês;
- O10 – O professor é facilitador e não distribuidor de conhecimento;
- O11 – O aluno é descobridor, pesquisador e não recipiente de conhecimento.

Quadro 3. Lista dos critérios obrigatórios para a preparação de atividades

Fonte: Delfino (2016, p. 56)

Primeiramente, além de fazer uso de um *corpus*, os enunciados dos exercícios contidos na atividade são claros, às vezes na LI, por vezes na língua portuguesa – como é abordado no vestibular da Unesp – instituição cujo vestibular foi base para o desenvolvimento dos exercícios. Ademais, a atividade busca focar no léxico fornecido pelos textos compilados, especialmente em relação aos agrupamentos lexicais extraídos do *corpus* de pesquisa. Os exercícios além de éticos, são replicáveis, ou seja, é possível que qualquer professor (a) utilize a atividade proposta nesta pesquisa em sala de aula, desde que com a devida menção/citação das pesquisadoras e criadoras da atividade.

Os exercícios são motivadores na medida em que se trata de um assunto relevante para o contexto dos alunos que estão prestes a prestar o vestibular, visto que política (principalmente atual) pode ser um dos possíveis temas a serem abordados em algum exercício de inglês nesse tipo de exame.

Os exercícios desenvolvidos nesta pesquisa também não simplificam a língua inglesa, ou seja, os textos retirados do *corpus* para atividade de compreensão escrita e/ou os

⁸ Para mais detalhes e informações, vide Delfino (2016) e Silva, Pinto e Dias (2018).

agrupamentos lexicais para a criação dos exercícios lexicais não mudaram em nada. Em outras palavras, foram utilizados exatamente dados retirados do *corpus* sem nenhum tipo de simplificação.

Em relação ao nível de dificuldade das atividades, pode-se dizer que estava adequado aos alunos, já que, esse fato pôde ser constatado por meio dos questionários entregues aos alunos, como será visto mais adiante. Além disso, em consonância com o critério número 9, os exercícios contêm conteúdo relevante para os alunos que irão prestar o vestibular, como explicado anteriormente.

Pode-se afirmar também que, o (a) professor (a) passa a ser facilitador neste tipo de atividades – elaborada com *corpora* – visto que o aluno é descobridor do seu próprio conhecimento, ou seja, os exercícios levam os alunos a inferirem os padrões que, em outros tipos de abordagens, seriam apenas transmitidos aos estudantes pelo (a) professor (a).

Tendo em vista que os exercícios propostos nesta pesquisa estão de acordo com os critérios obrigatórios sugeridos por Delfino (2016), a próxima seção abordará o passo-a-passo do desenvolvimento da atividade.

5.3 Elaboração das atividades

Como mencionado, para a elaboração das atividades foram utilizadas as ferramentas palavras-chave (*keywords*) e agrupamentos lexicais (*clusters*). Como já foram explicitadas as suas utilidades e relevância, objetiva-se nessa subseção decorrer a respeito do passo-a-passo da criação dos exercícios bem como eles serão empregados.

Assim sendo, a primeira etapa para a construção do material foi compilar o *corpus* do assunto abordado – política. Como mencionado, foram compilados 66 textos dos sites BBC® e CNN® salvos em formato txt. (texto sem formatação) no *software Sketch Engine*® para gerar a lista de palavras-chave. Como explicado na subseção “Palavras-chave”, há a necessidade de um *corpus* de referência para que as mesmas sejam extraídas, assim, na utilização da maioria dos concordanciadores, é necessário que esse *corpus* de referência também seja compilado, para que a comparação entre esse e o *corpus* de estudo seja realizada. Entretanto, utilizando o *Sketch Engine*®, essa comparação é feita automaticamente pelo *software* que já possui um *corpus* de referência em sua base, sendo ele o *English TenTen*®.

Dessa forma, foram geradas duas listas de palavras-chave – uma referente a unigramas (apenas uma unidade lexical) e outra com os agrupamentos lexicais mais frequentes. A segunda lista foi a base para a criação das atividades, como pode-se verificar na imagem abaixo, com um exemplar de 15 palavras:

Word	Focus corpus	Reference corpus
1 military parade	10	14
2 administration official	11	47
3 trade war	10	24
4 special counsel	10	26
5 trade deal	10	51
6 no-deal brexit	7	0
7 senior administration	8	43
8 9th circuit	7	35
9 withdrawal agreement	6	0
10 senior administration official	6	14
11 crown prince	6	34
12 transition period	8	135
13 campaign chairman	5	5
14 defense official	5	9
15 new leadership	7	130

Figura 1. Agrupamentos lexicais retirados do *corpus Politics*
Fonte: Dados do *corpus* de estudo *Politics*

Na observação dessa lista de blocos de palavras, a pesquisadora escolheu 10 que apresentavam chavicidade relevante, ou seja, que tinham mais probabilidade de serem chaves no contexto em que estavam empregadas. Assim, para a criação do primeiro exercício, o texto escolhido para a compreensão escrita dos alunos foi o que possuía a palavra-chave que, na lista, tinha maior chavicidade, a palavra *military parade*. Portanto, o texto escolhido para a atividade foi *The Trump military parade was always a really bad idea*.

Para iniciar o exercício, incitando os alunos a falarem o que eles sabiam acerca do assunto, foram feitos dois exercícios de *warm-up*. O primeiro se trata de um exercício de inferência por meio das palavras-chaves encontradas no texto abordado em sala de aula. Para esse exercício em específico foi necessário compilar apenas o texto *The Trump military parade*

was always a really bad idea e extrair as palavras-chave desse. O intuito era que os estudantes pudessem identificar as palavras cognatas e inferissem o conteúdo e o assunto abordado no corpo do texto.

O segundo exercício de *warm up* foi desenvolvido em português, pois buscou-se apenas abrir uma discussão a respeito de paradas militares. Faz-se importante mencionar que essa discussão não foi realizada na língua-alvo, pois, tratando-se de um contexto em que alunos têm diferentes níveis de inglês, seria provável que muitos alunos não participassem da discussão, por, por exemplo, não possuir vocabulário suficiente para tal tarefa. Além disso, o foco do presente trabalho foi abordar a habilidade de compreensão escrita e não de produção oral. Na imagem abaixo pode-se ver o exercício:

POLITICS



1. *As palavras abaixo são consideradas chave no texto que você irá ler a seguir. Você consegue imaginar do que se trata o texto?*

military parade
bad local government
democratic-controlled government
French military roll
democratic city

2. *Das palavras-chave, alguma lhe é familiar ou um cognato?*

1. **Warm-up**

a. O que são as paradas militares e por que elas acontecem?

b. Esse tipo de evento acontece(u) em seu país?



Figura 2. Exercício de *warm up*
Fonte: Exercício da própria pesquisa

Posteriormente, antes de iniciar a leitura, foi feita a seguinte pergunta: “O que você espera ler no texto levando em consideração o título?”, a fim de verificar se os alunos tinham entendido o título do texto, além de incitá-los a utilizar a técnica da inferência novamente, além de ser uma técnica importante quando se refere à compreensão escrita.

Para a elaboração dos exercícios foi utilizado como base os vestibulares da UNESP – Universidade Estadual Paulista – dos últimos três anos, 2016, 2017 e 2018. Assim sendo, as atividades propostas nesse estudo tiveram questões de múltipla escolha em português e em inglês, como é exigido nos exames da universidade citada.

Os exercícios até então desenvolvidos voltaram-se especificamente para a compreensão escrita do texto a partir do agrupamento lexical *military parade*, bem como acontece na segunda atividade, como podemos ver na imagem abaixo. Todavia, neste exercício, houve uma “ampliação” lexical, ou seja, não foi apenas abordado *military parade*, mas sim os verbos que ocorriam mais frequentemente com esse agrupamento, fazendo o uso de linhas de concordâncias retiradas do *corpus* de estudo *Politics*.

Dessa forma, o exercício pede para que os alunos leiam as frases (linhas de concordância) e, analisando o padrão/contexto em que ocorrem, preencher as lacunas com as expressões apresentadas na caixa:

To watch – was held – was staged – attended the – organized

A military parade was staged on the former airfield, involving 270 armored vehicles and 800 troops.

Kim Jong Un applauds as *he watches a military parade* in Pyongyang on April 15, 2012.

More than 20,000 people *attended the military parade* organized with the occasion of the National Day.

A military parade was then held in a well-attended.

More than 100,000 people *attend* notably *the main parade* with over 3,500 performers.

Another military parade _____ across downtown Seoul later in the afternoon.

In 1987, an estimated 100,000 people came _____ the parade.

An enormous parade _____ downtown the day of the game.

Pro-administration officials _____ parades and demonstrations in favor of the tax.

More than 20,000 people _____ military parade organized with the occasion of the National Day.

Figura 3. Exercício com o agrupamento *military parade*
 Fonte: Exercício da própria pesquisa

As questões nessa fase das atividades passam a ser todas em inglês, especialmente para que o aluno pratique, na maior parte do tempo, a língua-alvo. Isso se deve ao fato de que quanto mais contato os alunos tenham com o léxico, mais facilidade terão para lembrar e utilizar as palavras em algum outro momento.

Foi dada muita atenção em relação ao contexto a que o aluno é exposto nessa atividade, visto que é analisando as linhas de concordância que os alunos têm mais chance de associar aos blocos de palavras. Em outras palavras, quanto maior o texto à volta do agrupamento lexical, acredita-se que maior será a probabilidade de o aluno conseguir identificar o contexto e acertar a expressão a ser inserida na lacuna. É relevante mencionar que todos os excertos utilizados nessa atividade foram retirados dos textos compilados no *corpus*. Para isso, basta clicar em cima do agrupamento lexical na lista de palavras-chave (na interface do *AntConc*) para visualizar o contexto em que está inserido.

O exercício a seguir tem o intuito de despertar no aluno a curiosidade pelos agrupamentos lexicais, ou seja, qual palavra “combina” com qual? Desse modo, foi pedido aos alunos para que eles ligassem as palavras dadas à unidade *trade*.

5. Which words below are related to “trade”? Complete the chart.

pact – record – future – peace – police – war – deal – parties

<i>Trade</i>

Figura 4. Atividade de associação com *trade*.
Fonte: Exercício da própria pesquisa

É interessante mencionar que, nesta atividade, apesar de muitos alunos conhecerem a palavra *trade*, eles não sabiam quais unidades combinavam a ela, mesmo que essas unidades estivessem relacionadas ao mesmo tema.

No último exercício, os alunos utilizavam os agrupamentos da atividade anterior para completar as lacunas das frases apresentadas. Vale lembrar que, novamente, os excertos utilizados nas atividades foram extraídos do próprio *corpus* de estudo.

- a. German Chancellor Angela Merkel said "substantial progress" was needed on the UK's withdrawal agreement by the next European Council meeting in October, with "still a large piece of work" on the separate issue of _____ relations with the UK.
- b. Moreover, with the end of the Obama-negotiated Trans-Pacific Partnership, there are really no other reliable _____ involving the US across Asia. Still, as Trump told a news conference the morning after his NAFTA victory, Japan and India have indicated an interest in negotiating tariff agreements.
- c. During this time, the two sides hope to negotiate a permanent _____. The UK and the EU want to avoid a hard Northern Ireland border whatever happens, so they agreed to a "backstop" - described as an insurance policy by Mrs May.
- d. Finally, of course, there is the mother of all trade disputes, already developing into a _____, between the world's two leading trading nations, China and the US. Neither side here seems to be anywhere near an early round of productive negotiations.

Figura 5. Exercício de lacuna com os agrupamentos de *trade*.

Fonte: Exercício da própria pesquisa

Assim sendo, além da atenção voltada para os agrupamentos lexicais, outras técnicas de leitura como inferência, palavras familiares, repetição de palavras-chave entre outras, também podem ser úteis e foram utilizadas durante a aplicação dos exercícios aqui propostos, porém não serão descritas aqui por questão de espaço.

6 Análise do Questionário

Como obtenção de *feedback* por parte dos alunos, foi entregue a eles um questionário com três questões: 1. O que você achou mais interessante na aula de hoje? 2. A dificuldade das atividades sendo 1= *muito fácil* e 5=*muito difícil*, e 3. As atividades de vocabulário te ajudaram a compreender o texto? com as opções de resposta: sim, em partes e não. No total foram 29 respondentes, sendo eles alunos do segundo e terceiro anos do ensino médio.

Em relação à primeira questão, 14 se referiram ao tema/assunto abordado no texto, 10 alunos mencionaram a abordagem e/o aprendizado de novas palavras, enquanto 4 alunos aludiram ao entendimento/técnicas que auxiliaram na compreensão textual, como pode-se verificar no gráfico abaixo:

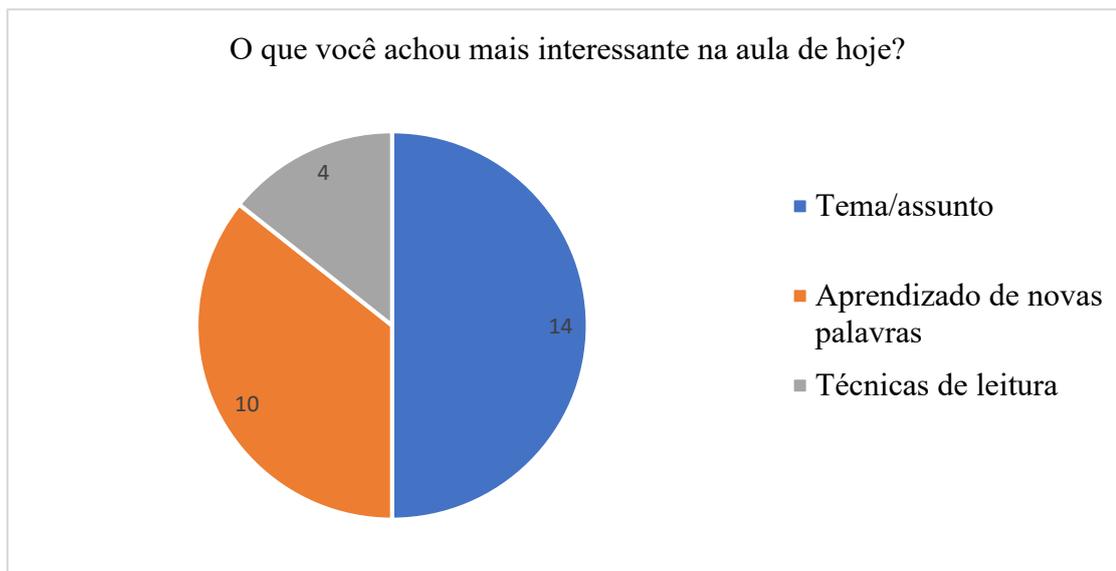


Gráfico 1. Questão 1
Fonte: Dados da presente pesquisa

No que diz respeito à segunda questão, 4 alunos acharam a atividade muito fácil, escolhendo o número 1 do questionário, 18 estudantes escolheram o nível 2, ou seja, fácil, enquanto 5 optaram pelo nivelamento 3, intermediário, e, por fim, 2 respondentes acharam a atividade difícil, optando pelo 4, como pode-se visualizar no gráfico abaixo:

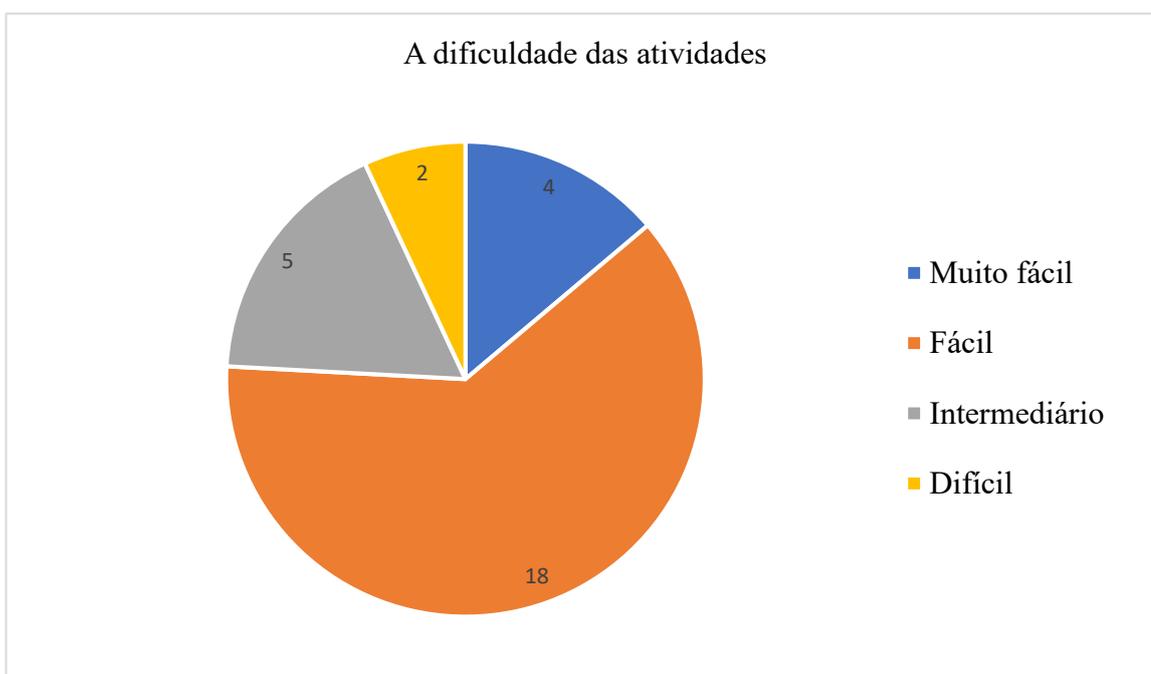


Gráfico 2. Questão 2
Fonte: Dados da presente pesquisa

Esse resultado foi considerado positivo, visto que as atividades não poderiam estar nem muito fáceis tampouco muito difíceis para a maioria dos alunos, tendo em vista o diferente nível de inglês de cada um.

No que diz respeito à terceira e última questão, 24 alunos afirmaram terem sido auxiliados pelos exercícios de vocabulário no momento de leitura e compreensão textual e 5 acreditam que os exercícios ajudaram em partes na compreensão escrita e nenhum aluno respondeu negativamente a essa pergunta. Abaixo temos o gráfico que representa essa análise.

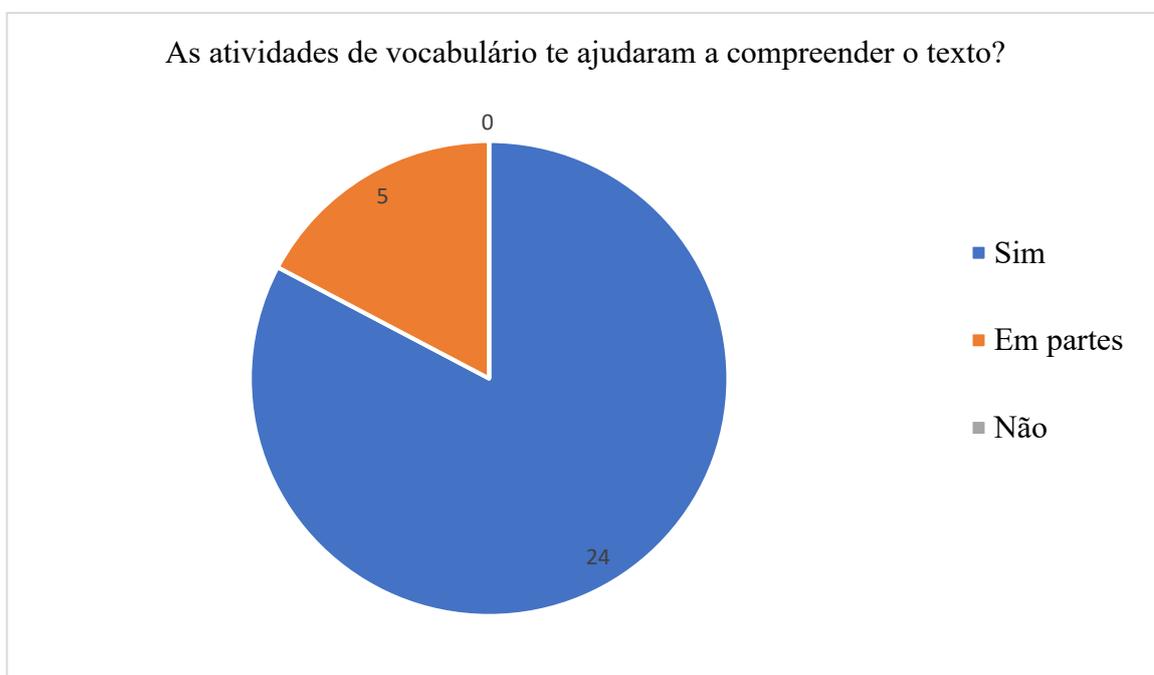


Gráfico 3. Questão 3
Fonte: Dados da presente pesquisa

Novamente, acredita-se que o resultado alcançado até então foi muito positivo, pois os alunos se sentiram auxiliados pelas atividades pré-texto, ou seja, antes de lerem o texto e responderem às questões de compreensão escrita, eles tiveram a oportunidade de estudarem sobre o tema desse texto e, além disso, o léxico encontrado nele.

7 Conclusão

Analisando os dados de um modo geral, pode-se afirmar que o *feedback* dado pelos alunos foi positivo e construtivo, já que a maior parte dos respondentes gostaram do tema abordado e a maneira como foi guiado pelas atividades pré-texto. Grande parte dos estudantes não acharam as atividades muito difíceis nem muito fáceis, o que é um fator muito interessante para esse contexto, todavia é possível que, por meio desse retorno dos alunos, as próximas atividades tenham um nível um pouco mais complexo com o intuito de se aproximar mais do nível intermediário. Por fim, um dos pontos mais relevantes do questionário foi o *feedback* dado pelos alunos em relação ao auxílio das atividades na compreensão textual. Pode-se afirmar que 82,75% dos alunos acreditam que os exercícios de vocabulário (anteriores aos de compreensão textual) ajudaram no entendimento do texto.

Desse modo, respondendo à pergunta de pesquisa “Como o uso de um *corpus* criado a partir de textos jornalísticos autênticos com o tema *política* pode auxiliar no levantamento de agrupamentos lexicais para o vestibular?”, pode-se afirmar que o *corpus* precisa estar em consonância com o que o aluno precisa saber naquele determinado momento, ou seja, qual sua necessidade em estar estudando inglês? Assim, o tema/assunto do *corpus* precisa ser relevante ao contexto desse aluno, bem como o nível de dificuldade das atividades, caso contrário, o ensino-aprendizagem fica falho, ou seja, com uma lacuna, já que tanto professor (a) quanto aluno (a) acabam “perdendo” tempo em leituras que não interessam ao aluno tampouco atende às suas necessidades.

8 Encaminhamentos futuros

Foi esperado, por meio dessa pesquisa, que os exercícios lexicais tivessem relevância no contexto dos aprendizes, possibilitando-os amadurecer no que diz respeito à perspectiva da LI como uma língua composta por vários agrupamentos e não somente por palavras soltas, o que, muitas vezes, pode prejudicar o aluno no entendimento de um texto.

Como encaminhamento futuro desta pesquisa, objetiva-se aumentar o *corpus* de estudo e atualizá-lo para a elaboração de mais atividades de compreensão escrita, bem como compilar *corpora* sobre diferentes temas de textos jornalísticos publicados on-line, para que,

dessa forma, os alunos entrem cada vez mais em contato com a língua inglesa e suas diversas temáticas, bem como tenham conhecimento de mais agrupamentos lexicais na língua-alvo, especialmente em textos autênticos.

Agradecimentos

Agradecemos à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro recebido durante a realização da presente pesquisa.

Referências

ALLWRIGHT, D.; BAILEY, K. M. *Focus on the Language Learner*. Cambridge: Cambridge University Press, 1991.

BERBER SARDINHA, T. *Linguística de corpus: histórico e problemática*. Revista D.E.L.T.A., São Paulo, v. 16, n. 2, p. 323-367, 2000.

_____. *Linguística de Corpus*. Barueri, SP: Editora Manole, 2004.

_____. *Pesquisa em linguística de Corpus com WordSmith Tools*. Campinas, SP: Mercado de Letras, 2009.

_____. Como usar a Linguística de Corpus no ensino de língua estrangeira. Por uma Linguística de Corpus educacional brasileira. In: VIANA, V e TAGNIN, S. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011.

_____. *Tecnologias e Mídias no ensino de inglês: o corpus nas "receitas"*. Cotia, SP: Macmillan do Brasil, 2012.

BRASIL. Secretaria de Educação Fundamental. *Parâmetros curriculares nacionais: terceiro e quarto ciclos do ensino fundamental: língua estrangeira / Secretaria de Educação Fundamental*. Brasília: MEC/SEF, 1998.

CELANI, M. A. A. *Considerações sobre a disciplina "Prática de Ensino de Inglês" nos cursos de licenciatura*. In: Celani (orgs) *Ensino de Línguas*. Cadernos PUC, 17. São Paulo, SP: EDUC, 1984.

CRYSTAL, D. *English as a Global Language (Second edition)*. Cambridge University Press, 2003.

DELFINO, M. C. N. *Uso de música para o ensino de Inglês em um ambiente baseado em corpus*. Orientador: Tony Berber Sardinha. 2016. 159 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2016.

Orientador: Mario Ferreira Junior. 2009. 82 f. Dissertação (Mestrado em Fisiopatologia Experimental) – Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2009.

DIAS, E. *Avaliação: meio para intervenção pedagógica na formação do professor de língua estrangeira e (re)construção da sala de aula*. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2011.

DUDLEY-EVENS, T.; John, M.J. *Developments in English for Specific Purposes. A Multidisciplinary approach*. Cambridge University Press, 1998.

DUTRA D. P.; SILERO, R. P. Descobertas linguísticas para pesquisadores e aprendizes: a linguística de *corpus* e o ensino de gramática. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 10 n. 4, p. 909-930, 2010.

HARVEY, A.; HORSELLA, M. *Technical Communication via Computational Abstracts: Implications for ESP*. Paper presented at the Second International Eindhoven LSP Conference, Eindhoven, Netherlands, 1988.

HUTCHINSON, T.; Waters, A. *English for Specific Purposes: a learning-centred approach*. Cambridge University Press, 1987.

HYLAND, K. *Academic clusters: text patterning in published and postgraduate writing*. *International Journal of Applied Linguistics*.18 (1): 41-62, 2008.

JENKINS, J. *Current perspectives on teaching world Englishes as English as a lingua franca*. *Tesol quarterly*, v. 40, n. 1, 2006.

JENKINS, J. 2009. English as a lingua franca: interpretations and attitudes. In: *World Englishes*. 2009, v. 28, No. 2, p. 200–207.

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. *From Corpus to Classroom*. Language Use and Language Teaching. Cambridge: Cambridge University Press, 2007.

RÖMER, U. *Corpus Research Applications in Second Language Teaching*. *Annual Review of Applied Linguistics*, 31, 205-225, 2011. doi:10.1017/S0267190511000055.

SCOTT, M. *WordSmith Tools*. Version 3. Oxford, Oxford, 1997.

_____. In memory of Tim Johns. *International Journal of Corpus Linguistics* 14(3): 271-274, 2009.

SILVA, L. F.; PINTO, P. T.; DIAS, E. Atividades de compreensão oral com base em corpora de TED Talks: um estudo piloto. In: FINATTO, M. J. B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (Orgs.) *Linguística de corpus: perspectivas*. Porto Alegre: Instituto de Letras - UFRGS, 2018, p. 527 – 554. Disponível em: < http://www.ufrgs.br/elc-ebralc2017/livro-ebralc-elc-2017/Linguisticadecorpus_NOVO_FINAL_ogmai2018.pdf>. Acesso em: 10 out. 2019.

TOGNINI-BONELLI, A. *Theoretical overview of the evolution of corpus linguistics*. In: *The Routledge Handbook of Corpus Linguistics* by Anne O'Keeffe, Michael McCarthy, Routledge: New York, 2010, p. 14-28.

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

VIANA V.; TAGNIN S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2011.

Data de submissão: 26/08/2019. Data de aprovação: 30/10/2019.