

A utilização de *corpora* em atividades de língua inglesa voltadas para a compreensão escrita de vestibulandos

DOI: <http://dx.doi.org/10.21165/el.v48i2.2286>

Luana Aparecida Nazzi Laranja¹

Resumo

O presente trabalho objetiva aplicar a abordagem teórico-metodológica da Linguística de *Corpus* (LC) no contexto de ensino aprendizagem de língua inglesa (McCARTHY, 2001; SARDINHA, 2010; VIANA, 2011) por meio do desenvolvimento de atividades de compreensão escrita voltadas para o vestibular. Para tanto, foi compilado um *corpus* de textos jornalísticos autênticos advindos dos *sites* CNN® e BBC® a fim de gerar as palavras-chave encontradas nos mesmos e, a partir de então, extrair os agrupamentos lexicais dessas palavras para posterior criação das atividades de compreensão escrita. Em função disso, buscou-se responder às seguintes perguntas: 01. Como o uso de um *corpus* criado a partir de textos jornalísticos autênticos pode auxiliar no levantamento de agrupamentos lexicais para o vestibular? 02. Que impacto a utilização de atividades desenvolvidas por meio desse *corpus* terá em relação ao desenvolvimento da compreensão escrita dos alunos? As atividades desenvolvidas foram aplicadas nas aulas de língua inglesa do terceiro ano do ensino médio em uma escola de ensino regular do interior de São Paulo. Por meio de um questionário entregue aos alunos ao final da atividade, foi possível observar que o efeito das atividades com base em *corpora* foi positivo, visto que grande parte dos estudantes avaliaram as atividades de compreensão escrita como auxiliares no entendimento dos textos propostos nas atividades, especialmente pelo vocabulário apresentado e pelas técnicas de leitura abordadas.

Palavras-chave: ensino e aprendizagem de língua estrangeira; Linguística de Corpus; atividades de compreensão escrita em língua inglesa.

¹ Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), São José do Rio Preto, São Paulo, Brasil. luananzzi@gmail.com; <https://orcid.org/0000-0003-3546-1707>

The use of corpora in writing comprehension activities aimed at the entrance examination in Brazilian universities

Abstract

The present work aims to apply the theoretical-methodological approach of Corpus Linguistics (LC) in the context of English language learning (MCCARTHY, 2001; SARDINHA, 2009; VIANA, 2011) through the development of written comprehension activities aimed at the entrance exam. In order to do so, a corpus of authentic journalistic texts was compiled from the CNN® and BBC® websites in order to generate the keywords found in them and, then, to extract the clusters of these words for later development of the written comprehension activities. As a result, we aimed to answer the following questions: 01. How can the use of a corpus created from authentic journalistic texts help in the collection of lexical clusters for the entrance exam? 02. What impact will the use of activities developed through this corpus have in relation to the development of students' written comprehension? The developed activities were applied to the English classes of the senior year in a regular school in the interior of São Paulo. Through a questionnaire given to the students at the end of the activity, it was possible to observe that the effect of the activities based on corpora was positive, since a great part of the students evaluated the activities of written comprehension as auxiliary in the understanding of the texts proposed in the activities, especially by the vocabulary presented and the approached reading techniques.

Keywords: foreign language teaching and learning; Corpus Linguistics; written comprehension activities in English.

Introdução

Sendo o inglês considerado um pré-requisito crucial para qualquer área profissional ou acadêmica (CRYSTAL, 2003; SEIDLHOFER, 2005), a busca pela sua aprendizagem é de igual importância. O ensino de língua inglesa (LI) para fins específicos (IFE) é uma das ramificações da abordagem LinFE², Línguas para Fins Específicos, que tem como foco os objetivos que o aprendiz necessita atingir em situações de usos reais da LI. Além disso, tal abordagem exige dos professores novos mapas mentais, posturas, maneiras de encarar o ensino e o processo de formação de profissionais conscientes linguisticamente (DIAS, 2011).

O IFE surgiu a partir da Segunda Guerra Mundial e começou a se expandir, tornando-se, assim, uma língua internacional para a melhor comunicação entre pessoas de diferentes

2 Nova denominação para a abordagem até então denominada de Inglês Instrumental, que em seu início no Brasil era vista como sinônimo de ensino de leitura (OCHIUCCI, 2015).

locais. A partir de então, as pessoas começaram a mostrar um enorme interesse e, conseqüentemente, uma maior procura para o aprendizado dessa língua, não mais por prestígio e sim por interesse no comércio e na tecnologia internacional. (HUTCHINSON; WATERS, 1987). Dessa forma, o aprendizado passou a surgir de necessidades particulares e houve, então, a necessidade de cursos voltados para pessoas com objetivos já determinados, fossem eles acadêmicos, profissionais ou pessoais, fazendo o IFE ganhar uma abrangência cada vez maior.

A primeira iniciativa de aplicação da abordagem IFE, o Projeto Nacional Ensino de Inglês Instrumental em Universidades Brasileiras, teve a sua fase experimental no Brasil em 1978 e sua oficialização em 1980, sendo coordenado pela Prof^a Dr^a Maria A. Celani (PUC - SP). A habilidade em foco do projeto era a compreensão escrita de alunos universitários que necessitavam compreender textos acadêmicos em inglês. No entanto, hoje se sabe que, em um curso IFE, o professor pode abordar tanto as habilidades de produção (*writing* e *speaking*) quanto as habilidades de compreensão (*listening* e *reading*), tudo dependerá das necessidades de seu aluno.

É fato que, em contexto de língua estrangeira geral, as quatro habilidades deveriam ser trabalhadas em sala de aula para um pleno domínio da língua-alvo, porém, devido a inúmeros fatores que não serão discutidos aqui, as escolas brasileiras ainda mantêm o foco no ensino-aprendizagem da leitura, como é mostrado nos 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. (BRASIL, 1998, p. 20).

Sabendo, portanto, que a compreensão escrita é a habilidade exigida nos exames que abrem portas para o ensino superior e devido à relevância de se ter um curso voltado para as necessidades específicas de cada aluno, o presente trabalho tem como *objetivo geral* aplicar recursos da Linguística de *Corpus* (LC) no levantamento de agrupamentos lexicais para a criação de material de língua inglesa para vestibular por meio de aulas utilizando a abordagem IFE. Como *objetivos específicos*: 1. Compilar um *corpus* de textos jornalísticos autênticos advindos dos *sites* CNN[®] e BBC[®], visto que são considerados confiáveis e uma rica fonte de textos de diferentes temas; 2. Desenvolver atividades de compreensão escrita com o uso de agrupamentos lexicais, voltando-se para o vestibular; 3. Trabalhar técnicas de compreensão escrita, a partir dos agrupamentos lexicais levantados, a fim de auxiliar os alunos no momento de prestar o exame.

Ainda que muitos trabalhos já tenham sido desenvolvidos com foco na compreensão escrita em língua inglesa (BRAGA, 1997; SOLÉ, 1998; TOMITCH, 2009), poucos são os que se voltam para a criação de atividades que se baseiam no levantamento lexical por meio de *corpora*.

Dessa forma, as atividades produzidas foram aplicadas em aulas de língua inglesa de uma escola de ensino regular situada no interior de São Paulo, no primeiro semestre de 2019, voltando-se para os estudantes do terceiro ano do ensino médio. Tanto as atividades desenvolvidas bem como as aulas ministradas tiveram aparato teórico em Inglês para Fins Específicos com os estudiosos Hutchinson e Waters (1987), Dudley-Evans e St. Johns (1998), Tahir (2009) e Celani (1984), que também abordam a Linguística Aplicada; ensino de leitura baseado nas concepções de Kato (1985); Scaramucci (1995, 1999); Schmitt e McCarthy (1997); Kleiman (1985); preparação de materiais com Leffa (1996), ensino de segunda língua com Brown (2000) e Linguística de Corpus com Berber Sardinha (2009), Viana e Tagnin (2011) e O’Keefe, McCarthy e Carter (2007).

Fundamentação teórica

A importância da abordagem IFE

Segundo Tahir (2009), há características que diferenciam a abordagem ESP (English for Specific Purposes), em português IFE (Inglês para Fins Específicos), da abordagem EGP (English for General Purposes), em português, IFG (Inglês para fins gerais):

IFE é contrastado com IFG, ou Inglês para fins gerais. Se o inglês é ensinado como uma segunda língua juntamente com outras disciplinas para fins educacionais como algum assunto útil para os alunos no futuro, então este é o EGP. Neste tipo de aprendizado, geralmente não há exigência imediata para os alunos usarem o inglês para quaisquer propósitos comunicativos reais. Em contraste, se o inglês é ensinado para aprendizes especializados com algum propósito vocacional e educacional específico em mente, então isto é ESP. (TAHIR, 2009, p. 98, tradução nossa)³.

3 No original: “ESP is contrasted with EGP, or English for General Purposes. If English is taught as a second language along with other subjects for educational purposes as some useful subject to the learners in the future, then this is EGP. In this type of learning, there is generally no immediate requirement for the learners to use English for any real communicative purposes. In contrast, if English is taught for specialized learners with some specific vocational and educational purpose in mind, then this is ESP” .

Desse modo, pode-se constatar que, diferentemente de cursos gerais de línguas, que abordam a língua em contextos comuns de uso, a abordagem IFE se volta para o objetivo de cada aluno, sendo que essa consciência linguística tem implicação direta na aprendizagem da língua-alvo. Conforme afirmado por Hutchinson e Waters (1987, p. 53): “se professor e aluno sabem o porquê estão estudando o idioma, essa consciência terá uma influência positiva ao selecionar conteúdos que serão pertinentes à aprendizagem, e quais habilidades serão exploradas em sala de aula”.

Em consonância com os autores citados, Holmes (1981) e Dudley-Evans e St. John (1998) mencionam algumas vantagens do LinFE (e, conseqüentemente, IFE) em relação aos cursos gerais de línguas:

1. alunos aprendem o que realmente precisam de acordo com suas necessidades;
2. alunos tornam-se mais motivados, pois percebem que estão em um *curso feito sob medida (tailored course)* e que suas necessidades são o foco norteador das aulas;
3. professores têm liberdade para preparar aulas específicas a alunos específicos, isso reflete a consciência que o professor possui sobre a importância das necessidades de cada aluno.

Nesse contexto de ensino-aprendizagem de línguas, é importante levar em consideração como se dá o processo tanto de ensinar quanto de aprender, especialmente quando diz respeito à abordagem IFE. Para Hutchinson e Waters (1987, p. 59, tradução nossa),

Uma abordagem centrada no aluno para a aprendizagem e o ensino vê a aprendizagem como a construção ativa de significado e o ensino como o ato de guiar, montar e facilitar o aprendizado. Esta abordagem considera o conhecimento como sendo um processo em constante mudança, construído com base na experiência prévia dos alunos.⁴

Assim, o professor deixa de ser o centro detentor do saber e passa a ser facilitador no processo de ensino, guiando o aluno na solidificação do próprio conhecimento por meio do *background* já existente do estudante, ou seja, aquilo que este já conhece em relação à língua. É dessa maneira que, na presente pesquisa, os alunos entram em contato com

4 No original: “A learner-centered approach to learning and teaching sees learning as the active construction of meaning, and teaching as the act of guiding, scaffolding and facilitating learning. This approach considers knowledge as being an ever-changing process, which is built upon the learners prior experience”.

o inglês, ativando, especialmente, o que eles já têm de informação e acrescentando o que há de novo, sendo abordado nas atividades, como as técnicas de leitura, que serão discutidas no próximo subcapítulo, bem como o vocabulário apresentado durante os exercícios.

Nesse contexto, conforme mencionado por Dutra e Silero (2010, p. 911), “a Linguística de Corpus (LC) pode colaborar com a Linguística Aplicada”, visto que o uso de *corpus* em sala de aula (SA) fornece descrições da língua em uso, o que se torna uma excelente ferramenta no ensino de línguas, em especial, para contextos específicos, bem como pode levantar textos apropriados a serem selecionados como materiais pertinentes para a aula.

Em função das vantagens acima mencionadas, é esperado que o professor transite com relativa segurança por temas diversificados e específicos, sendo que tanto professor quanto aluno aproveitam melhor o tempo em sala de aula, visto que, como mencionado pelos autores acima, os alunos aprendem o que condiz com suas necessidades, e pelo mesmo motivo se tornam mais motivados. Isso faz com que as aulas sejam mais interessantes e a relação entre professor e aluno se solidifique de forma mais colaborativa.

O ensino de leitura como objeto de estudo

Com o início do projeto nacional Ensino de Inglês Instrumental em Universidades Brasileiras, constatou-se que grande parte das dificuldades dos alunos não era com a língua estrangeira, mas sim com a incapacidade de interagir com o texto escrito na própria língua materna.

A partir de então, professores e pesquisadores de língua materna e estrangeira passaram a ter um objetivo em comum: compreender melhor os processos subjacentes à leitura e à formação do bom leitor. Esta, porém, não é uma tarefa a ser iniciada no âmbito da universidade. Se o aluno universitário não sabe ler os textos acadêmicos nem em língua materna nem em língua estrangeira, é preciso desencadear um trabalho preventivo e formativo e não apenas corretivo (KATO, 1985).

Visto isso, é evidente que o problema no ensino de leitura é algo que deve ser tratado desde a formação inicial dos alunos, porém, o objetivo do presente trabalho não é discutir tal situação, mas sim ilustrar brevemente algumas técnicas de leitura em LE, e como elas refletem a preparação de materiais com foco na compreensão escrita.

Carvalho (2005), Farrell (2003) e Kato (1985), em suas obras, descrevem três modelos populares de como um leitor interage com um texto. Tais modelos são descritos pela teoria cognitivista como *bottom-up*, ou modelo ascendente, *top-down*, ou modelo descendente e

um terceiro modelo chamado de modelo interativo de leitura, esse último é considerado como uma interação entre os modelos *bottom-up* e *top-down*.

De uma maneira simples, pode-se dizer que, no modelo *top-down*, o leitor traz consigo seus conhecimentos e experiências prévias para compreender o texto, e continua a ler enquanto o texto confirme suas expectativas, no modelo *bottom-up* ocorre o inverso, o leitor lê as palavras e as frases e olha para a organização do texto, sem relacioná-lo com o conhecimento e a experiência anteriores. Desse modo, ele consegue construir o significado do que foi escrito no texto pela junção do conhecimento do vocabulário com a estrutura sintática. Farrell (2003) explica que a oposição entre os dois modelos se dá pelo fato de que, na abordagem *top-down*, o leitor trabalha do significado para o texto, já no modelo *bottom-up*, o leitor parte do texto para o significado (com o foco na palavra e no nível da frase).

O modelo interativo de leitura é explicado por Farrell (2003), que ilustra a interação entre os modelos *bottom-up* e *top-down* de modo que descreve o processo de leitura feito “por um leitor maduro e fluente” (FARRELL, 2003, p. 4), bem como apresenta quatro passos que podem ser seguidos por esses leitores na abordagem interativa. Esses passos são:

01. Formular hipóteses ou fazer previsões: o leitor inicia o texto prevendo o assunto que será abordado, inferindo, por meio de seu *background* (conhecimento prévio), se o texto será interessante para o prosseguimento da leitura.

02. Experimentar o texto: o leitor não utiliza muito de seu tempo com a gramaticalidade do texto, mas seleciona palavras que se encaixam no momento da leitura.

03. Continuar o ciclo: como em um ciclo, o leitor continua examinando sempre o texto e deduzindo seu sentido, ignorando aquilo que não é de seu entendimento até esse tipo de problema se tornar recorrente e perceber que sua previsão estava equivocada.

04. Sendo a compreensão impossível, o leitor formula outra hipótese e reinicia o ciclo de previsão.

Vale lembrar que não era esperado que os alunos alvo desta pesquisa possuísem as características de um leitor “maduro e fluente”, conforme descrito por Farrell (2003), porém essas técnicas que levam o aprendiz a alcançar esse nível de compreensão escrita foram trabalhadas durante as aulas de língua inglesa, para que os alunos se tornassem leitores mais independentes, e se sentissem mais preparados para o vestibular ou quaisquer eventuais leituras em inglês que eles precisem fazer durante suas vidas.

A partir da ideia de que o léxico é parte essencial para a compreensão de textos em uma segunda língua (KLEIMAN, 1985; LAUFER, 1997; SCHMITT; McCARTHY, 1997), outra estratégia que foi abordada durante as aulas é a inferência, à luz dos estudos de Kleiman (1985), que discorre acerca dessa técnica no estudo de línguas com o uso de textos autênticos, como foi feito no presente trabalho. Para a autora, o conhecimento lexical dos alunos é, muitas vezes, limitado, e é por isso que uma estratégia como a inferência se faz necessária para a leitura de textos em uma língua estrangeira. Kleiman (1985, p. 73) afirma que “se o conhecimento linguístico do aluno é limitado, ele passará a usar outros tipos de conhecimentos, de caráter textual, discursivo, enciclopédico, a fim de suprir as limitações”.

Apesar das críticas feitas acerca desse tipo de técnica (SCHMITT; McCARTHY, 1997), a inferência será utilizada, nesse trabalho, de forma contextualizada, ou seja, não serão abordadas palavras soltas que podem, com certeza, ter mais de um significado dependendo do contexto em que são inseridas, mas, como mencionado, serão trabalhados os agrupamentos lexicais, e esses serão apresentados dentro do texto, de forma que a inferência será muito útil para a análise textual. Schmitt (2000) afirma que inferir o significado de palavras desconhecidas através do contexto é uma estratégia essencial no que concerne ao aprendizado lexical, e é corroborando com o autor que, nesse estudo, deve-se levar em consideração que os vestibulares tratam de assuntos diversos, que, muito provavelmente, trarão palavras novas para os alunos. Sendo assim, é primordial que eles conheçam e apliquem técnicas de leitura, como a inferência.

De acordo com Brown (2001), há várias técnicas que são muito úteis no momento da leitura de um texto escrito em língua estrangeira, dentre elas, *skimming* e *scanning*. A primeira faz referência ao ato de passar os olhos rapidamente pelo texto a fim de encontrar características chave que mostraram do que se trata o texto de modo mais geral. Essa técnica é muito utilizada, por exemplo, quando se lê um jornal, em que o leitor busca rapidamente por aquilo que lhe é de interesse. A segunda estratégia se baseia no fato de examinar o texto em busca de informações mais específicas, como palavras-chave, por exemplo. Para o autor, esses tipos de técnicas que introduzem um tema são muito importantes durante a compreensão textual.

O estudioso Françoise Grellet (1981) assegura que os alunos nunca vão ler de maneira eficiente a menos que adaptem sua velocidade de leitura e técnica a seu objetivo quando estiver lendo. Ao lerem todos os textos da mesma maneira, os alunos perderiam tempo e deixariam de lembrar pontos importantes para eles porque estariam absorvendo muita informação desnecessária (GRELLET, 1981).

Levando em consideração que um vestibulando não dispõe de muito tempo para realizar as questões da prova, faz-se necessário que as técnicas mencionadas acima sejam ensinadas aos alunos a fim de que eles as apliquem quando forem realizar os exercícios de compreensão textual, visto que, assim como a inferência, o *skimming* e o *scanning*

auxiliarão os alunos no que diz respeito à introdução ao assunto do texto, assim como o entendimento de determinadas questões. Aliando as técnicas já mencionadas ao uso da abordagem teórico-metodológica oferecida pela LC, duas outras estratégias serão incluídas no ensino de compreensão escrita da presente pesquisa, o uso de palavras-chave (*Keywords*) e de agrupamentos lexicais (*Clusters*), que serão elucidados em um próximo tópico.

Uso de *corpora* no ensino de LE

Para muitos professores, um dos fatores que implica muito trabalho quando utilizada a abordagem LinFE é a necessidade de se organizarem previamente, além de uma grande quantidade de tempo para se dedicarem à preparação e seleção de materiais específicos utilizados em suas aulas, já que cada aluno possui um objetivo diferente. Por isso, pesquisadores da área de Linguística Aplicada passaram a utilizar *corpora* eletrônicos como fornecedores de contribuições importantes para o desenvolvimento de suas pesquisas, visto 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. 03).

Dentro dos estudos de *corpora* há duas abordagens que sustentam a pesquisa por ângulos diferentes, sendo a abordagem “dirigida pelo *corpus*” e a “baseada no *corpus*” (BIBER, 1989). A primeira abordagem parte do princípio de que as hipóteses ou teorias levantadas em um estudo devem surgir a partir do *corpus*, ou seja, este norteará o pesquisador mostrando qual caminho seguirá. Por outro lado, na segunda abordagem, a pesquisa já é baseada em uma teoria acreditada pelo pesquisador e então esse utiliza o *corpus* para corroborar ou não sua hipótese. O presente trabalho faz uso da primeira abordagem, a “dirigida por *corpus*”, pois serão utilizados os agrupamentos lexicais extraídos do *corpus* compilado para o desenvolvimento das atividades, ou seja, a pesquisa será “dirigida” pelos dados encontrados no *corpus*.

Segundo Berber Sardinha (2004), a utilização de *corpora* no ensino se divide em quatro áreas principais, sendo elas: Descrição da linguagem nativa; Descrição da linguagem do aprendiz; Transposição de metodologias de pesquisa acadêmica para a sala de aula e Desenvolvimento de materiais de ensino, currículos e abordagens. Complementando isso, Viana e Tagnin (2011) também abordam o ensino de línguas por meio de *corpora*, afirmando que há duas maneiras de se obter mais informações a respeito de uma língua e seu funcionamento. O primeiro modo diz respeito à intuição, ou seja, julga-se quando uma regra ou determinado uso da língua é adequado ou não, obtendo, dessa

maneira, uma ideia do que é/seria a língua em si, o que, segundo o autor, não corresponde exatamente ao seu uso real. Esse tipo de situação pode, muitas vezes, deixar o professor em uma situação um tanto delicada, já que, para se ter certeza se determinados usos da língua estariam apropriados ou não a certas ocasiões, eles teriam de consultar um falante nativo.

A segunda maneira de se estudar uma língua é por meio da observação de dados autênticos da língua, ou seja, como ela é realmente utilizada em um contexto que não se preocupa em utilizar conhecimentos intuitivos na comunicação, mas, ao contrário disso, tem-se a língua em sua espontaneidade. Nesse caso, o estudo é baseado em informações empíricas, por meio do uso da língua em interações que ocorrem naturalmente entre falantes nativos (VIANA, 2011).

É importante ressaltar também que, como afirma McCarthy (2001), a LC representa uma grande mudança no que diz respeito à educação em geral, tanto em relação aos métodos de ensino quanto ao papel do professor em sala de aula, além de fazer um importante intermédio entre teoria e prática no contexto educacional.

Porém, no ensino ainda existem alguns mitos que também são citados por Sardinha (2004), por exemplo, a crença de que a sintaxe é superior ao léxico, sendo este apenas um modo de preencher as lacunas sintáticas que impedem um melhor desenvolvimento em sala de aula, enquanto o tempo dos alunos e do professor é “desperdiçado” com informações que poderiam ser mais relevantes para o ensino-aprendizagem de línguas.

Em relação à implementação da LC no ensino, Sardinha (2004, p. 261) afirma:

Não devemos sobrecarregar os alunos com vastas quantidades de informação sintática [“gramática”] de um lado e informação lexical [“vocabulário”] de outro, que eles terão de juntar de acordo com princípios não disponíveis naturalmente para eles na condição de falantes não-nativos.

É dessa forma que a LC pode contribuir para a criação de materiais desse projeto piloto, visto que o curso será voltado para objetivos específicos de alunos que irão prestar o vestibular, não sendo necessário expandir o ensino da língua inglesa de modo com que o aluno tenha de começar a estudar no nível básico até alcançar o avançado. Portanto, a compilação de textos jornalísticos (uma vez que há uma grande demanda desse tipo de texto nos vestibulares) auxiliará os alunos a compreender técnicas de leitura que são necessárias no momento do vestibular, utilizando material autêntico, que aparece com frequência nesse tipo de exame. Assim, haverá um melhor aproveitamento do conteúdo e do tempo que serão disponíveis tanto para o professor quanto para o aluno em sala de aula.

Palavras-chave e agrupamentos lexicais – *Keywords e clusters*

Como mencionado, nesta pesquisa, foram extraídas as palavras-chave do *corpus* de estudo – aquele compilado por meio de textos jornalísticos – em comparação com um *corpus* de referência, EnglishTenTen® – de inglês geral. Sardinha (2009, p. 193) afirma que as palavras-chave são “aquelas cujas frequências são diferentes, de modo estaticamente significativo, das frequências encontradas em um *corpus* de referência”. O autor também explica que essas palavras-chave não têm de ser necessariamente uma palavra individual, como, por exemplo, “*café*”, mas também, o que é nomeado de *clusters*, ou seja, agrupamentos lexicais, como “*café com leite*”.

Outro ponto relevante no que diz respeito às palavras-chave refere-se à classificação entre positivas e negativas. As primeiras se referem àquelas cuja frequência é mais alta no *corpus* compilado pelo pesquisador do que no *corpus* de referência – esse que representará a língua de maneira geral. As negativas são exatamente o contrário, há uma frequência de palavras-chave muito mais relevante no *corpus* de referência em relação ao de estudo.

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

Sardinha (2009) também cita as finalidades de se utilizar palavras-chave em uma pesquisa, sendo elas: 1. identificar a temática (“*aboutness*”) de um *corpus* ou de texto; 2. descrever a organização interna dos textos; 3. localizar marcas indicativas de posicionamento ideológico; e 4. traçar um perfil lexical de um autor ou de outros elementos básicos.

É essencial esclarecer que, no presente trabalho, as palavras-chave foram extraídas com um único objetivo, que é o primeiro citado pelo autor, ou seja, identificar o tema do *corpus* compilado, pois, como mencionado, é a partir dele que os agrupamentos lexicais foram levantados para a criação das atividades de compreensão escrita, como será mais detalhadamente apresentado na metodologia.

Os *clusters*, ou agrupamentos lexicais, são parte do que a LC oferece para descrever o funcionamento de determinada língua, assim, pode-se fazer uma pesquisa de agrupamentos lexicais de duas, três ou quatro palavras que se agrupam. Em outras palavras, pode-se dizer que, nesses agrupamentos, há certa preferência lexical por uma ou outra palavra que, geralmente, estão juntas. Para exemplificar, foi coletado por meio do *corpus* de estudo, que abrange 3 temas: política, tecnologia e mundo, três

agrupamentos lexicais com maior chavicidade (maior chance de ser chave no contexto em que é encontrada) de duas ou três palavras. Em primeiro lugar, o *corpus* apresenta o agrupamento *bread roll* e, em seguida, *fake news*. É importante mencionar que os resultados podem mudar em concordância com o aumento do *corpus* de estudo.

De acordo com O'Keeffe, McCarthy e Carter (2007), esses agrupamentos são, muitas vezes, tão frequentes quanto palavras "soltas" quando não, mais encontrados no vocabulário em geral. O que isso sugere é que o conteúdo programático de vocabulário para o nível básico fica incompleto, sem a devida atenção voltada para os agrupamentos lexicais mais frequentes, uma vez que muitos deles são tão frequentes quanto ou ainda mais do que palavras isoladas, que todo mundo concordaria que devem ser ensinadas (O'KEEFE; MCCARTHY; CARTER, 2007).

Não apenas para o nível básico, mas considerando que o presente trabalho se volta para um público específico, é de extrema relevância o levantamento de agrupamentos lexicais dos textos compilados para os alunos que prestam vestibulares, visto que, a partir deles, um tema será enfatizado e o aluno terá a possibilidade de compreender melhor o texto. Além de que "[...] um *corpus* também pode fornecer mais informações sobre uma forma-chave e destacar padrões léxico-gramaticais e semânticos associados à forma⁵" (O'KEEFE; MCCARTHY; CARTER, 2007, p. 100, tradução nossa). É também corroborando a ideia dos estudiosos citados que se pode afirmar que os agrupamentos lexicais deveriam ser considerados como qualquer outra unidade da língua e ainda, trabalhados não isoladamente, mas, para uma prática mais sustentada, seria interessante abordá-los dentro de um contexto.

Como é possível verificar, os agrupamentos lexicais são muito válidos na compreensão de um texto, visto que fazem parte de seu vocabulário e, assim, podem auxiliar os alunos no entendimento tanto do tema geral de determinado texto quanto de sentenças mais específicas, além de remeterem também à parte gramatical, ou seja, à estrutura que o texto apresenta, como tempo verbal, por exemplo, o que, novamente, vai ajudar o aluno a compreender melhor a mensagem que está sendo transmitida.

Metodologia e análise

Para a presente pesquisa, um *corpus* de artigos jornalísticos foi compilado, contendo textos autênticos retirados dos *sites* CNN®, e BBC®. Os artigos abrangem três temas diferentes, sendo eles: *politics* (política), *tech* (tecnologia) e *world* (mundo). Foram compilados, no período de 10 semanas, 3 textos de cada um dos assuntos semanalmente, totalizando 90 artigos jornalísticos. O *corpus* de estudo foi compilado por meio do *software* Sketch

5 No original: "[...] a corpus can also provide more information about a key form and underline lexico-grammatical and semantic patterns associated with the form".

Engine⁶ e possui 88,261 itens e 75,474 formas⁶, sendo importante ressaltar que a compilação do *corpus* ocorreu durante os três últimos meses do último semestre de 2018.

Com base nos critérios concebidos por Sardinha (2000), o conteúdo e o propósito do *corpus* do presente projeto foram organizados da seguinte forma:

Quadro 1. Conteúdo e propósito do corpus

Modo	Escrito
Tempo	Contemporâneo.
Seleção	De amostragem.
Conteúdo	O conteúdo do <i>corpus</i> será de caráter especializado, composto por textos jornalísticos autênticos advindos dos sites CNN [®] e BBC [®] .
Autoria	Os autores serão falantes nativos e não-nativos fluentes em LI.
Finalidade	O <i>corpus</i> será utilizado na preparação de atividades com foco na compreensão escrita.

Fonte: Elaboração própria.

Em relação ao material desenvolvido nesta pesquisa, é de extrema importância salientar que apenas textos autênticos foram selecionados, ou seja, textos que não foram criados com objetivo pedagógico – para serem trabalhados em sala de aula – pelo contrário, eles são retirados do contexto real de uso da língua inglesa, *sites* de grande repercussão mundial. Os autores O’Keeffe, McCarthy e Carter (2007) enfatizam a relevância da autenticidade de material para o ensino de línguas juntamente com o uso de *corpora* afirmando que:

A disponibilidade de exemplos de *corpus* produziu uma perspectiva diferente uma vez que nós podemos encontrar em *corpora* diversos exemplos de textos que são livres, na medida em que são independentes de qualquer função voltada para o ensino de línguas. Eles estão em seu próprio contexto autêntico, e são compostos para um público particular (que tende a ser diferente do aprendiz de idiomas). Portanto, quando eles são apresentados com exemplos de

⁶ Item (*tokens*) diz respeito ao total de palavras contido no *corpus*, já a forma (*types*) é o número de palavras repetidas. Para mais informações, vide Sardinha (2009).

corpus, os aprendizes encontram a língua real como é na verdade utilizada, e nesse sentido é “autêntica”. (O’KEEFFE; MCCARTY; CARTER. 2007, p. 26, tradução nossa)⁷.

Para a elaboração das atividades, foram utilizadas as ferramentas palavras-chave (*keywords*) e agrupamentos lexicais (*clusters*). Dessa forma, ao utilizar a primeira ferramenta, o Sketch Engine® gera automaticamente duas listas de palavras-chave – uma referente às palavras soltas, denominada *single words*, e outra com os agrupamentos lexicais mais frequentes, denominada *multi-words*. A segunda lista foi utilizada como base para o desenvolvimento dos exercícios de vocabulário.

Foram extraídos os 10 primeiros agrupamentos lexicais, como é mostrado na figura abaixo, retirada do *corpus* de estudo.

1	bread roll	14	13
2	fake news	11	21
3	military parade	10	14
4	trade war	9	24
5	warning system	10	74
6	trade deal	9	51
7	no-deal brexit	7	0
8	crown prince	8	34
9	coal plant	8	53
10	saudi government	7	25

Imagem 1. Agrupamentos lexicais do corpus

Fonte: *Corpus* de estudo.

Assim, no desenvolvimento das atividades, os 10 blocos lexicais exibidos acima foram utilizados, de modo contextualizado, como meio de enriquecer o léxico dos alunos com palavras encontradas em textos autênticos e atuais.

7 No original: “The availability of corpus examples has produced a different perspective since we can find in corpora numerous examples of texts that are free-standing, in so far as they are independent of any language learning task. They are in their own authentic context, and they are composed for a particular audience (which tends to be different to that of the language learner). Thus, when they are presented with corpus examples, learners encounter real language as it is actually used, and in this sense, it is ‘authentic’”.

Para exemplificar, pode-se verificar na figura abaixo como *military parade* foi utilizado em um exercício. Por meio do *corpus*, foi possível encontrar as combinações verbais mais frequentes com esse agrupamento, assim, os alunos tiveram a oportunidade não apenas de entrar em contato com uma expressão nova, bem como aprender a utilizar os verbos que são frequentemente utilizados nesse contexto.

2. Read the sentences below and then fill in the blanks with the expressions in the box:

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.

Imagem 2. Atividade com o agrupamento *military parade*

Fonte: *Corpus* de estudo.

Após a elaboração de atividades com base no vocabulário encontrado no *corpus*, foram também desenvolvidos exercícios de compreensão escrita e, para esses, foi escolhido, pela própria pesquisadora, o agrupamento *fake news*, o segundo da lista de *multi-words* exibida anteriormente. O *software* Sketch Engine® permite que o pesquisador clique sobre um determinado agrupamento lexical e encontre tanto o contexto em que este está inserido quanto o texto fonte, ou seja, de onde foi retirado o agrupamento. Seguindo esse passo, um dos textos que continha o agrupamento *fake news* foi “Facebook touts fight on fake news, but struggles to explain why InfoWars isn’t banned” encontrado no *site* CNN®. Desse modo, as atividades de compreensão escrita foram baseadas no texto mencionado.

Os alunos foram, a todo momento, instruídos a como utilizar técnicas de leitura durante a aula, como *skimming* e *scanning*, análise de palavras cognatas e repetição de palavras-chave. As aulas foram também baseadas no modelo interativo de leitura, combinando *bottom-up* e *top-down*, que são relevantes na compreensão escrita dos textos, como explicado anteriormente.

Em um último momento, foi entregue aos alunos participantes um questionário com três questões: 1. O que você achou mais interessante na aula de hoje? 2. Qual a dificuldade das atividades (aqui o aluno deveria escolher entre 1= muito fácil e 5=muito difícil), e 3. As atividades de vocabulário te ajudaram a compreender o texto? (três opções foram dadas para assinalar: Sim, em partes ou não).

No total, foram 16 participantes que responderam o questionário. Em relação à primeira pergunta, 4 se referiram ao vocabulário que ajudou na compreensão, 9 disseram ter gostado dos temas abordados e 7 mencionaram alguma técnica de leitura, sendo que algumas, como a identificação de cognatos e sinônimos, diziam respeito ao léxico abordado. Em relação à dificuldade das atividades, 3 alunos acharam as atividades de compreensão escrita muito fáceis, assinalando o número 1, 10 colocaram o número 2 e 3 alunos assinalaram o número 3, ou seja, acharam as atividades com nível médio de dificuldade. 15 participantes disseram que as atividades de vocabulário ajudaram na compreensão do texto e apenas 1 respondeu que os exercícios o ajudaram em partes.

Deste modo, a investigação se posiciona nas discussões teóricas de métodos e técnicas baseadas em um ensino prático reflexivo, juntamente com a questão teórica da formação de professores de línguas (ALMEIDA FILHO, 1993, 2004; WIDDOWSON, 2005). Os instrumentos de pesquisa foram assim organizados:

Quadro 2. Organização do corpus

Objetivos	Instrumentos
Coleta de dados	Compilação do <i>corpus</i> de textos retirados dos <i>sites</i> BBC® e CNN®.
Análise de dados	Uso do <i>software Sketch Engine</i> ® para análise do <i>corpus</i> . Questionário qualitativo como forma de obter um <i>feedback</i> das atividades aplicadas, bem como da aprendizagem dos alunos.
Apresentação dos resultados	Apresentação dos dados obtidos pela análise do <i>corpus</i> . Comparação qualitativa e quantitativa da melhora da habilidade de leitura dos alunos no decorrer da pesquisa.

Fonte: Elaboração própria.

A presente pesquisa utilizou o método quantitativo a partir do levantamento e análise dos dados coletados por meio da aplicação das atividades, além da utilização da abordagem teórico-metodológica da LC por meio do concordanciador Sketch Engine® para obtenção das palavras-chave e agrupamentos lexicais para a criação dos exercícios.

Simultaneamente, será feito o uso do método qualitativo, tendo em vista a relação colaborativa entre professor e aluno, bem como a oportunidade de compartilhamento de diferentes perspectivas acerca da elaboração e aplicação das atividades em sala de aula.

É importante ressaltar que o uso dos dois métodos já foi defendido por vários pesquisadores (REICHARDT; COOK, 1979; COHEN, 1989), visto que tendem a se complementar diante de suas limitações. Deve-se levar em consideração que ambos os métodos contribuem para o melhor entendimento e análise dos dados durante o processo (subjetivo) e resultado (objetivo) da pesquisa (ALLWRIGHT; BAILEY, 1991), oferecendo mais validade à mesma.

Considerações finais

É relevante mencionar que esta pesquisa se deu no contexto de uma escola de ensino regular particular do interior de São Paulo, onde os alunos participantes já tinham conhecimento de algumas das técnicas de leitura mencionadas aqui, o que pode ter facilitado o entendimento das atividades de compreensão escrita. Entretanto, os alunos tiveram a oportunidade de entrar em contato com outras técnicas desconhecidas por eles até então, além de conhecerem um novo vocabulário, o que foi primordial para a compreensão dos textos, segundo a análise do questionário entregue ao final da atividade.

Portanto, por meio da presente pesquisa, foi possível verificar que a utilização de *corpus* na composição de materiais para aulas cuja abordagem é IFE, mais especificamente atividades voltadas para vestibulandos, deu-se de forma positiva, visto que o *corpus* fornece tanto as palavras-chave quanto os agrupamentos lexicais de forma precisa, por meio de dados estatísticos. Desse modo, o pesquisador e/ou profissional da educação não necessita confiar apenas em sua intuição, passando aos alunos um vocabulário relevante e confiável em relação aos textos que são lidos em SA.

REFERÊNCIAS

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

ALMEIDA FILHO, J. C. P. *O Professor de Língua(s) Profissional, Reflexivo e Comunicacional*. Campinas: UNICAMP, 2004.

ALMEIDA FILHO, J. C. P. *Dimensões comunicativas no ensino de línguas*. Campinas: Pontes, 1993.

BERBER SARDINHA, T. *Pesquisa em linguística de Corpus com WordSmith Tools*. Campinas: Mercado de Letras, 2009.

BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Editora Manole, 2004.

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.

BIBER, D. A Typology of English texts. *Linguistics*, Berlim, De Gruyter, v. 27, n. 1, p. 3-43, 1989.

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.

BRAGA, D. Ensino de Língua via Leitura: uma Reflexão sobre a Elaboração de Material Didático para Auto-Instrução. *Trabalhos em Lingüística Aplicada*, IEL/UNICAMP, v. 30, p. 5-16, jul./dez. 1997.

BROWN, D. H. *Principles of Language Teaching and Learning*. New York: Longman, 2000.

BROWN, D. H. *Teaching by principles: and interactive approach to language pedagogy*. Second edition. New York: Longman, 2001.

CARVALHO, S. N. O enfoque instrumental na leitura. *Soletas*, São Gonçalo: UERJ, ano V, n. 10, p. 116-129, jul./dez. 2005.

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

COHEN, A. D. Attrition in the productive lexicon of two Portuguese third-language speakers. *Studies in Second Language Acquisition*, v. 11, p. 135-149, 1989.

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

DIAS, E. *Avaliação: meio para intervenção pedagógica na formação do professor de língua estrangeira e (re)construção da sala de aula*. 2011. 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-EVANS, T.; JOHN, M. J. *Developments in English for Specific Purposes. A Multidisciplinary approach*. Cambridge: 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.

FARRELL, T. S. C. *Planejamento de atividades de leitura para aulas de idiomas*. São Paulo: SBS, 2003.

GRELLET, F. *Developing Reading Skills: A Practical Guide to Reading Comprehension Exercises*. Cambridge: Cambridge University Press, 1981.

HOLMES, J. What do you mean by ESP? *Working Papers*. n. 2. CEPRIL. São Paulo: PUC-SP, 1981.

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

KATO, M. *O aprendizado da leitura*. São Paulo: Martins Fontes Editora, 1985.

KLEIMAN, A. B. Estratégias de inferência lexical na leitura de segunda língua. *Ilha do Desterro*, Campinas: UNICAMP, p. 67-82, 1985.

LAUFER, B. *The lexical plight in second language reading*. In: COADY, J.; T. N. HUCKIN. *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, 1997. p. 20-34.

LEFFA, V. J. *Aspectos da leitura: uma perspectiva psicolinguística*. Porto Alegre: Sagra/Luzzatto, 1996.

LEFFA, V. J. Metodologia do ensino de línguas. In: BOHN, H. I.; VANDRESEN, P. *Tópicos em linguística aplicada: o ensino de línguas estrangeiras*. Florianópolis: Ed. da UFSC, 1988.

OCHIUCCI, M. S. M. Do projeto nacional ensino de inglês instrumental ao LinFE: a contribuição dos estudos de Rosinda de Castro Guerra Ramos ao ensino de línguas para fins específicos – a aventura continua.... In: LIMA-LOPES, R. E.; FISCHER, C. R.; GAZOTTI-VALLIM, M. A. (org.). *Perspectivas em língua para fins específicos: festschrift* para Rosinda Ramos. Campinas: Pontes Editores, 2015.

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

REICHARDT, C. S.; COOK T. D. Beyond qualitative versus quantitative methods. In COOK, T. D.; REICHARDT, L. S. (org.). *Qualitative and Quantitative Methods in Evaluation Research*. Sage: Londres, 1979.

SCARAMUCCI, M. V. R. Vestibular e ensino de língua estrangeira (inglês) em uma escola pública. *Trab. Ling. Apl.*, Campinas, p. 07-20, 1999.

SCARAMUCCI, M. V. R. *O Papel do Léxico na Compreensão em Leitura em Língua Estrangeira: Foco no Produto e no Processo*. 1995. Tese (Doutorado em Linguística Aplicada) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1995.

SEIDLHOFER, B. Key concepts in ELT. English as a lingua franca. *ELT Journal*, v. 59, n. 4, doi:10.1093/elt/cci06, 2005.

SCHMITT, M.; MCCARTHY, M. *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press, 1997.

SCOTT, M. *WordSmith Tools*. Version 3. Oxford: Oxford University Press, 1997.

SOLÉ, I. *Estratégias de leitura*. Porto Alegre: Artes médicas, 1998.

SCHMITT, N. *Vocabulary in Language Teaching*. Cambridge University Press, Cambridge, 2000.

TAHIR, M. M. English for specific purposes (ESP) and syllabus design (2009). Disponível em: <https://www.iasj.net/iasj?func=fulltext&ald=45953>. Acesso em: 20 mar. 2019.

TAGNIN, S. *A Identificação de equivalentes tradutórios em corpora comparáveis*. 2007. Disponível em: <http://bit.ly/2XGw6Tt>. Acesso em: 19 abr. 2018.

TOMITCH, L. M. B. Aquisição de leitura em língua inglesa. *In*: LIMA, D. C. (org.). *Ensino e aprendizagem de língua inglesa: conversas com especialistas*. São Paulo: Parábola Editorial, 2009. p. 191-201.

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

VILAÇA, M. L. C. Pesquisa e Ensino: considerações e reflexões. *Revista E-scrita*, Nilópolis, v. 1, n. 2, mai./ago. 2010.

WIDDOWSON, H. G. *O ensino de línguas para a comunicação*. Tradução José Carlos P. de Almeida Filho. 2. ed. Campinas: Pontes, 2005.

WIDDOWSON, H. G. *Text, Context, Pretext*. Oxford: Blackwell, 2004.