

ENSINANDO INGLÊS COM *CORPORA* PARA PROFISSIONAIS DA ÁREA DE PUBLICIDADE

Cristina Mayer Acunzo (PUC-SP)

Resumo: Este artigo teve como objetivo principal apresentar uma atividade de ensino de Inglês como Língua Estrangeira com *corpora* para estudantes do idioma que atuam na área de Publicidade. Para tanto, encontramos suporte teórico principal na Linguística de *Corpus*, uma área que proporciona a pesquisa, o estudo e a exploração da língua em uso (BERBER SARDINHA, 2004; 2009; 2011). Particularmente, na sua vertente que trata da interface *corpus* e ensino de Língua Estrangeira, com base em abordagens desenvolvidas por Sinclair (1991; 2004a; 2004b), Johns (1991), Willis (1990) e Berber Sardinha (2009; 2011), que dizem respeito à exploração de *corpora* no ensino por meio da identificação e da análise de padrões lexicogramaticais da língua e ao ensino por meio da abordagem lexical. Os resultados da análise do *corpus* sugerem que existe uma grande variedade de padrões de linguagem relativos a Publicidade a serem estudados. Encontramos diversos padrões e colocados que, se bem explorados por professores e alunos, podem auxiliar na comunicação dos profissionais dessa área. Propomos, assim, uma atividade de ensino de Língua Inglesa com alguns desses padrões. Nossa pesquisa, bem como este artigo, pretende contribuir para o desenvolvimento de atividades de ensino baseadas em *corpora*, bem como para estudantes que atuam na área de Publicidade, para a qual existe pouco material didático no mercado.

Palavras-chave: Linguística de *Corpus*; Ensino de Língua Inglesa como Língua Estrangeira.

Abstract: This paper aimed at presenting an activity for the teaching of English as a foreign language with *corpora* to students who work in Advertising. In order to achieve this objective, the main theoretical framework for the research is provided by Corpus Linguistics, which is an area that supports the research and the study of the language in use (BERBER SARDINHA, 2004; 2009, 2011). Particularly, this study is mainly concerned with the interface between corpus and the teaching of foreign languages, based on the approaches developed by Sinclair (1991; 2004a; 2004b), Johns (1991), Willis (1990) and Berber Sardinha

(2009; 2011), which refer to the exploration of corpora in language teaching through the identification and analysis of lexicogramatical patterns and the teaching based on a lexical approach. The results of the analysis of the corpus showed that there is a great variety of language patterns related to Advertising to be studied. We have found a range of patterns and collocates which, if adequately explored by both teachers and students, can help the professionals in this field to communicate effectively. We propose, then, an activity for the teaching of English with some of those patterns. Both the study and the present paper intend to contribute to the development of classroom activities based on corpora, as well as to professionals in Advertising, for whom there is few teaching materials available.

Keywords: Corpus Linguistics; Teaching of English as a Foreign Language.

1. INTRODUÇÃO

O ensino de Línguas Estrangeiras para profissionais de áreas específicas é um desafio, pois existe pouco material disponível no mercado que atenda às necessidades dos alunos de comunicar-se em seu meio profissional. A minha experiência profissional, como professora de Inglês como Língua Estrangeira, em aulas particulares e em empresas, especialmente em agências de Publicidade, levou-me a observar que esse contexto de ensino requer um conteúdo e um material adequado às necessidades dos alunos de utilizarem o idioma em seu meio profissional. Em nossa experiência, pudemos verificar que os alunos normalmente conhecem os termos usados em sua área de atuação, porém apresentam dificuldades no uso desse vocabulário específico. É por meio de materiais autênticos¹, como textos

1 Consideramos material autêntico o que não foi produzido ou criado para fins pedagógicos.

e vídeos, que os alunos podem ter contato com essa língua em uso; observamos que os materiais didáticos disponíveis no mercado proporcionam muito pouco contato com esse tipo de material, bem como com o vocabulário utilizado em seu meio profissional.

Diante desse panorama e de nosso contexto de ensino de Inglês como Língua Estrangeira, doravante ILE, em agências de Publicidade, desenvolvemos uma pesquisa² para a criação de atividades de ensino específicas, a qual descrevemos neste artigo, a fim de que professores que lecionam em contextos semelhantes possam aproveitar-se desta metodologia para criar suas próprias atividades.

Em síntese, nossa pesquisa consistiu em dois objetivos principais: o de extrair e analisar o uso de alguns dos padrões lexicogramaticais³ mais frequentes no meio publicitário e, em seguida, o de utilizar esses padrões na preparação de atividades⁴ de ensino.

Para atingir os objetivos propostos, desenvolvemos a seguinte metodologia de pesquisa: (1) coleta de um *corpus*⁵

2 Pesquisa em nível de Mestrado desenvolvida pela autora de 2009 a 2012 pela PUC-SP com auxílio bolsa Capes.

3 Entendemos por padrão lexicogramatical a combinação de duas ou mais palavras que ocorre com frequência com um significado associado a ela (BERBER SARDINHA, 2004).

4 Entendemos atividade como “uma sequência de trabalhos feitos pelos alunos na sala de aula, individualmente ou não, veiculados em um material de ensino” (BERBER SARDINHA, 2011, p. 307).

5 Corpus refere-se à coletânea de textos em formato eletrônico coletados de forma criteriosa

de mais de 1 milhão de palavras composto por artigos escritos e transcrição de vídeos com textos lidos e entrevistas de uma revista digital de Publicidade, que chamamos de *ad corpus*; (2) análise do *corpus* e sua comparação com um *corpus* de referência, o BNC (*British National Corpus*) para a identificação dos padrões lexicogramaticais distintivos do *corpus* de estudo, por meio do programa *Wordsmith Tools* (SCOTT, 2008); (3) transposição dos achados para a atividade de ensino.

Encontramos suportes teórico e metodológico principais na Linguística de *Corpus*, doravante LC, que, dedicando-se à coleta criteriosa e exploração de conjuntos de dados linguísticos textuais, denominados *corpora*, proporciona a pesquisa, o estudo e a exploração da língua em uso, de forma empírica e por meio do uso de computador (BERBER SARDINHA, 2004). Mais especificamente, nossa proposta fundamenta-se na área de pesquisa baseada em *corpus* que se preocupa com o ensino de Línguas Estrangeiras. Uma das questões principais da LC nesta área é como incorporar o estudo e a análise da língua por meio de *corpora* no ensino-aprendizagem de línguas.

Sinclair (1991; 2004a; 2004b) e Johns (1991) apresentam

a fim de servir para a pesquisa de uma língua ou variedade linguística (BERBER SARDINHA, 2004). O plural de *corpus* é *corpora*.

formas de se trabalhar os conceitos e as ferramentas da LC no ensino por meio de concordâncias (ver um exemplo na Metodologia), que são “listagens de uma palavra específica (o ‘nódulo’, *node word* ou *search word*) juntamente com parte do texto onde ocorreu” (BERBER SARDINHA, 2009, p.9). Johns é o criador do *Data Driven Learning* (DDL), abordagem na qual o aluno é incentivado a estudar os padrões lexicogramaticais da língua por meio de concordâncias, desenvolvendo, assim a habilidade de um pesquisador. Cabe ao professor guiar o aluno, facilitando esse trabalho de formular hipóteses sobre a língua, buscar respostas para suas questões, identificar padrões recorrentes e fazer generalizações a respeito dos achados.

Berber Sardinha (2009; 2011) também apresenta maneiras de se usar *corpora* em sala de aula propondo três tipos de atividades para o ensino de ILE, quais sejam: centradas na concordância, centradas no texto e multimídia/multigênero. As atividades centradas na concordância focam na identificação dos padrões lexicogramaticais e são sugeridos exercícios acompanhando as concordâncias, o que propomos em nossa atividade. O segundo tipo de atividade parte de um texto e trabalha as questões que o envolvem, como tema, gênero, ideias principais, além de trazer concordâncias

com palavras do texto. Da mesma maneira, propomos que os alunos usem o texto, não só concordâncias. As atividades multimídia/multigênero envolvem diferentes mídias e gêneros e incentivam o aluno a entender as diferenças entre eles e seus padrões frequentes ou típicos.

Em contato com professores de ILE em contextos diversos, bem como na área de Publicidade, observamos que são utilizados materiais autênticos, geralmente baseados em textos da *Internet*, de *sites* das áreas em que os alunos atuam. Entretanto, as atividades preparadas pelo professor, assim como a escolha dos itens a serem explorados nestes textos são comumente feitas sem acesso a recursos da LC e os alunos acabam por estudar somente itens gramaticais retirados dos textos ou então extraem o vocabulário que acreditam ser usualmente empregado sem analisar sua frequência ou o seu uso.

Pretendemos, aqui, mostrar como utilizar esses recursos, contribuindo, assim, para que o ensino de ILE possa modernizar-se e que tecnologias sejam usadas para o desenvolvimento de aulas mais significativas e customizadas às necessidades dos alunos em seu meio profissional.

2. LINGUÍSTICA DE *CORPUS*

Este trabalho tem como fundamentação teórica principal a Linguística de *Corpus* que, segundo Berber Sardinha (2004, p.3), “ocupa-se da coleta e 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”.

Ainda segundo o autor, um *corpus* pode ser definido como:

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

É importante destacar que um *corpus*, peça-chave da pesquisa em LC, deve ser coletado para servir a um propósito específico. De acordo com Berber Sardinha (2004), alguns dos principais critérios para a formação de um *corpus* computadorizado são: (1) deve ser formado por textos autênticos, ou seja, não pode ter sido produzido para servir

de objeto de pesquisa; (2) deve ter um conteúdo selecionado criteriosamente, de acordo com as características e objetivos estabelecidos pelo criador e (3) deve ser um conjunto representativo de uma variedade linguística ou de uma língua.

Firth (1957) é considerado um dos estudiosos mais importantes para a fundamentação da Linguística de *Corpus*, especialmente no que diz respeito ao estudo de uma língua em função do seu contexto de uso (McENERY; WILSON, 2001). Os estudos de Firth nos anos 50 contribuíram para o surgimento de uma nova tradição de linguistas, chamados neo-firthianos, como John Sinclair e Michael Halliday. No final dos anos 50, o empirismo, e conseqüentemente as pesquisas com *corpora*, perderam força com a chegada da teoria racionalista de Chomsky, que rejeitava a pesquisa com dados autênticos e criticava a confiabilidade dos *corpora* coletados manualmente. As pesquisas com *corpora* reapareceram e consolidaram-se nos anos 80, com o advento de uma tecnologia que possibilitou a análise de grandes quantidades de dados linguísticos por meio do computador. Na década de 90, suítes de *software* como o *Wordsmith Tools*, de Mike Scott, facilitaram ainda mais o trabalho na área e atualmente existem diversos programas e ferramentas à disposição

de linguistas de *corpus*, que atuam em universidades em diversos países. No Brasil, a LC vem ganhando força, com diversas publicações (livros, revistas, jornais, dissertações e teses) e encontros entre pesquisadores da área em diferentes pontos do país (Cf. BERBER SARDINHA e ALMEIDA, 2008).

A LC não faz a separação rígida entre léxico e sintaxe, ao contrário, aponta a existência de um nível do sistema linguístico conhecido por lexicogramática, que engloba o vocabulário e a gramática (BERBER SARDINHA, 2004). Um dos conceitos que justificam o estudo de padrões lexicogramaticais é o de linguagem enquanto sistema probabilístico. Segundo essa visão, a linguagem é padronizada, isto é, muitos traços linguísticos teoricamente possíveis não ocorrem na língua em uso com a mesma frequência (BERBER SARDINHA, 2004). Segundo Berber Sardinha, “um padrão léxico-gramatical é uma sequência recorrente de (pelo menos duas) palavras, dentro de um espaço delimitado (normalmente equivalente a até quatro palavras de distância), que possui um sentido específico” (2009, p.12). Em nossa língua nativa, usamos esses padrões de forma automática, ou seja, não selecionamos palavras isoladas para a construção dos discursos e é esse uso de unidades já padronizadas que nos permite usar a língua com fluência (SINCLAIR, 1991).

Os padrões lexicogramaticais podem ser formalizados em três conceitos principais, que são: colocação, coligação e prosódia semântica (BERBER SARDINHA, 2004). A colocação é considerada “a tendência de duas palavras de coocorrerem, ou como a tendência de uma palavra atrair outra”⁶ (HUNSTON, 2002, p.68), ou seja, uma associação de itens lexicais. Segundo Sinclair (1991), a colocação ilustra o princípio idiomático (*idiom principle*) em contraposição ao princípio de escolha aberta (*open-choice*), pois “palavras não ocorrem de forma aleatória em um texto”⁷ (SINCLAIR, 1991, p.110). Um exemplo de colocação é *shed tears*; alguns outros colocados de *shed* são: *light, jobs* e *blood* (HUNSTON, 2002). A coligação é, segundo Firth, “a inter-relação das categorias sintáticas”⁸ (PALMER, 1968, p.23), ou seja, está no nível gramatical e concerne à associação entre itens lexicais e gramaticais. Um exemplo bastante claro é o selecionado por Berber Sardinha (2004, p.40): “*start* é mais comum com sintagmas nominais e orações *-ing*, enquanto *begin* é mais usado com um complemento *to*”. A prosódia semântica é a associação entre itens lexicais e conotação (negativa ou positiva) ou instância avaliativa (BERBER SARDINHA, 2004).

6 “the tendency of two words to co-occur, or as the tendency of one word to attract another” (HUNSTON, 2002, p.68).

7 “words do not occur at random in a text” (SINCLAIR, 1991, p.110).

8 “the interrelations of the syntactical categories” (PALMER, 1968, p.23).

Um dos exemplos de prosódia semântica mais citados é o dado por Stubbs (1996): a palavra *cause* tem uma prosódia semântica negativa por coocorrer com palavras de conotação negativa como *problem(s), illness, disaster, damage* e *concern*.

As principais áreas de interesse da Linguística de *Corpus* de acordo com Kennedy (1998) são, de forma sucinta, a compilação e a construção de *corpora*, o desenvolvimento de ferramentas para análise de *corpora*, a descrição da linguagem e o ensino-aprendizagem de línguas. Como já apontado, nosso foco é no uso de *corpora* no ensino-aprendizagem de língua estrangeira, o que detalhamos a seguir.

2.1 Ensino de Língua Inglesa por meio da Linguística de *Corpus*

O interesse de profissionais da área de ensino e aprendizagem de línguas no trabalho com *corpora* advém do reconhecimento da necessidade de priorizar o ensino de padrões lexicais e de que esses padrões não poderiam ser selecionados apenas pela introspecção (SINCLAIR, 2004b). Esse interesse e o uso de *corpora* por esses profissionais ocorrem há mais de 20 anos, como podemos ver pelas pesquisas de Sinclair (1987, 1991, 2004a), e tem sido crescente, especialmente pela disponibilidade de programas e ferramentas de fácil utilização e da maior facilidade de

acesso a *corpora*, que estão disponíveis *online* para serem baixados e/ou comprados (RÖMER, 2006, 2009; BERNARDINI, 2004).

Com relação ao uso de *corpora* em sala de aula de língua estrangeira, Aijmer (2009) afirma que ainda há certa resistência ou falta de habilidade por parte de professores, mas prevê um futuro promissor. Römer (2009) sustenta que os linguistas de *corpus* têm um trabalho árduo para enfrentar o desafio de professores e alunos que estão acostumados a métodos mais tradicionais e sugere que uma tarefa importante é divulgar a LC e informar os professores sobre o que existe disponível para ajudá-los em sua prática pedagógica, como *corpora online*, materiais baseados em *corpora*, materiais de referência sobre *corpora*, entre outros.

Para Sinclair (2004a) e Tim Johns (1991), em sua proposta de Aprendizado Movido por Dados (*Data Driven Learning 9^o*, doravante DDL, o aluno deve ter acesso direto a um *corpus* e explorar a linguagem respondendo às suas próprias questões e hipóteses, desenvolvendo a habilidade de um pesquisador. Já Aston (1997) afirma que o professor pode simplificar os resultados das buscas no *corpus*, selecionando e filtrando os textos de acordo com o nível de compreensão do aluno,

9 Aprendizagem movida a dados (Data Driven Learning – DDL) é uma abordagem cuja ênfase é desenvolver no estudante da língua a habilidade de pesquisa e de descoberta, colocando à sua disposição dados de um corpus.

especialmente no caso de iniciantes. Segundo o autor, “talvez o uso mais óbvio de *corpora* seja tratá-los como fontes de materiais que o professor pode selecionar e adaptar de acordo com o que for necessário”¹⁰ (ASTON, 1997, p.52).

As grandes vantagens do DDL são (BERNARDINI, 2004; BERBER SARDINHA, 2004; O’SULLIVAN, 2007; CHAMBERS, 2010; GILQUIN e GRANGER, 2010):

1) Uso e exploração de linguagem autêntica na sala de aula e grandes quantidades de exemplos, não para a simples repetição (*drills*), mas sim para o que Ellis (2002, p.177) chama de repetição consciente (*mindful repetition*), ou seja, a leitura das mesmas instâncias visando ao entendimento dos padrões e itens linguísticos. Esse acesso à linguagem autêntica e a uma quantidade de exemplos maior do que pode ser encontrado em livros de referência usados por alunos e professores, como dicionários, gramáticas e livros didáticos, leva à expansão do vocabulário e à percepção e conscientização de padrões lexicogramaticais.

2) A modificação do papel do professor de “Fonte

10 “perhaps the most obvious pedagogic use of corpora is to treat them as sources of classroom materials which the teacher can select from and adapt according to requirements.” (ASTON, 1997, p.52)

de conhecimento” para facilitador, auxiliando os alunos a interpretar os dados e a fazer suas buscas, com acesso a exemplos de língua em uso.

3) O fato de o aluno desenvolver maior autonomia, independência e responsabilidade sobre sua aprendizagem, além de habilidades como fazer previsões e observações, pensar, notar, analisar, raciocinar, interpretar, deduzir, comparar, entre outras.

4) A possibilidade de o aluno corrigir-se comparando seu uso da língua ao uso de nativos ou ótimos usuários (*expert*) e de refletir sobre as questões da língua; o elemento da descoberta faz a aprendizagem ser mais motivadora e divertida.

Bernardini (2004, p.22), seguindo as premissas do DDL, propõe o DL (*Discovery Learning*), uma abordagem na qual os alunos “são guiados a observar grandes coletâneas de textos variados de uma maneira aberta e exploratória”¹¹. A autora opta pelo acesso direto dos alunos a *corpora*; entretanto, diferentemente de Johns (1991), considera que os alunos devam assumir o papel de descobridores, e não de pesquisadores, considerando que os alunos podem

11 “are guided to browse large and varied text collections in open-ended, exploratory ways” (BERNARDINI, 2004, p.22)

não ter as competências e capacidades que professores e/ou linguistas possuem enquanto pesquisadores. Para a autora, a Aprendizagem como Descoberta (*Discovery Learning*) incentiva os alunos a seguir seus interesses e desenvolver capacidades de forma que “suas buscas tornem-se mais focadas, suas interpretações mais precisas e seu entendimento mais afiado sobre o uso de *corpus* e sobre a língua”¹² (BERNARDINI, 2004, p.23).

Conforme aponta Chambers (2010), não se defende a ideia do DDL como o principal ou único componente no ensino de línguas, mas sim como um recurso para encorajar a observação da língua em uso, a descoberta, a formulação de hipóteses e, conseqüentemente, a autonomia, não fugindo às principais premissas da abordagem, que são o princípio idiomático da língua e a descoberta de padrões lexicogramaticais, bem como o uso de concordâncias como instrumento para a observação desses padrões. A maneira como o professor irá usar o *corpus* e as concordâncias irá depender do contexto de aprendizagem, do nível dos alunos, da disponibilidade de materiais, entre outros fatores.

12 “their searches become better focused, their interpretation of results more precise, their understanding of corpus use and their language awareness sharper” (BERNARDINI, 2004, p.23)

2.2 Concordâncias na sala de aula

Como exposto anteriormente, uma das formas de se utilizar textos autênticos e *corpora* em sala de aula é por meio das concordâncias. Segundo Berber Sardinha (2004)), “a concordância é empregada para exemplificar o uso (*usage*) de traços linguísticos e as situações nas quais ocorre” (p.273). Usando o programa *Wordsmith Tools*, de Mike Scott (2008), por exemplo, podemos obter uma listagem das ocorrências de um item selecionado no conteúdo de vários textos ao mesmo tempo. Assim, a palavra estudada fica centralizada na tela e é possível fazer uma leitura hipertextual das palavras que ocorreram junto a ela no *corpus*. Dessa forma, a palavra está em seu contexto original e o usuário tem acesso a uma porção de texto maior da ocorrência a ser estudada no próprio programa.

Berber Sardinha (2011) lista o que deve ser levado em conta ao preparar atividades baseadas em concordâncias, sendo: (a) escolha da palavra de busca, que é guiada pelos objetivos da atividade, da aula, da unidade e do curso; (b) modo de apresentação, ou seja, se a aula será em um laboratório ou na sala de aula, o que implica na necessidade de o professor trazer a concordância pronta impressa ou para projetar; (c) tipo de concordância: geral (com todas as

ocorrências), selecionada, editada (sem símbolos ou palavras cortadas, por exemplo), preenchida (com a palavra de busca) ou lacunada (com lacunas no lugar da palavra de busca); (d) quantidade de concordâncias e (e) exercícios, como perguntas, completar frases, ligar colunas, entre outros.

O autor então explica os ingredientes que o professor precisa para preparar a atividade, o *corpus* e um *software* para produção de concordâncias e indica que o professor deve ensinar ao aluno como ler uma concordância e como observar os padrões lexicogramaticais. O professor pode selecionar as concordâncias previamente, atentando a questões como número, seleção e simplificação das linhas; outra opção é o aluno produzir as concordâncias, atividade que requer que o professor ensine o aluno a usar o *software*, formular buscas, selecionar linhas, etc.

Nas atividades centradas no texto, propostas por Berber Sardinha (2011), a concordância está presente, mas assume um papel secundário, visto que o foco é no texto. Nesse tipo de atividade, não há a desvantagem de o aluno sentir falta do contexto ou de uma situação concreta de uso da língua. Dentre suas vantagens, citamos a contextualização dos padrões enfocados e o tempo de preparação que, se feito com o uso do *software*, é relativamente curto.

3. METODOLOGIA DE COLETA E ANÁLISE DOS DADOS

A compilação do *corpus* de estudo, denominado *ad corpus*, foi feita segundo os seguintes critérios levantados por Berber Sardinha (2004): (a) composição, por textos autênticos utilizando linguagem natural, (b) autenticidade, textos produzidos por falantes nativos e em situações reais e (c) seleção criteriosa de acordo com os objetivos da pesquisa.

O *ad corpus* é composto por 1.118 textos em formato eletrônico e tem 1.004.733 palavras (*tokens*). Trata-se de um *corpus* sincrônico de porte médio-grande e que consideramos ser especializado, por ser composto por textos de uma área específica para a exploração da linguagem nela usada. Os textos foram coletados de fonte autêntica, da *Internet*, de uma revista digital que publica artigos sobre a atuação das agências de Publicidade na área. O *site* selecionado é o *AdAge – Advertising Agency*¹³, que fornece informações, notícias e análise de dados sobre *marketing* e mídia, possibilitando a comunicação para as comunidades de Publicidade e *Marketing*. A revista teve início em 1930, em formato de jornal e atualmente, em seu formato impresso, é semanal e distribuída em diversos países do mundo, incluindo o Brasil. A revista digital traz publicações em forma de artigos, colunas, vídeos, *blogs*, *podcasts*, entre outros.

13 Disponível em <http://adage.com>. Acesso de julho de 2009 a fevereiro de 2012.

A escolha por essa fonte para o *corpus* deve-se a diversos fatores. O primeiro é a questão da disponibilidade e variedade de tipos de textos, de autores e de informações. Em segundo lugar, verificamos que a maior parte dos profissionais da área a consideram uma importante fonte de informações e a utilizam diariamente. Por último, em nossa experiência com aulas nesse contexto, observamos que diversos professores de ILE que trabalham com esses profissionais utilizam essa revista em suas aulas e essa prática é aprovada e até mesmo solicitada pelos próprios alunos, que geralmente desejam ter aulas com temas e linguagem relacionados à área em que atuam. Textos dessa revista e de diversas outras fontes foram comumente utilizados em nossas aulas e, de acordo com os alunos, em conversas informais, a *Adage* é a mais completa e eles sentiram que o uso dela em sala de aula foi bastante proveitoso para sua aprendizagem do idioma.

Fizemos a seleção dos textos no período de setembro de 2009 a março de 2011 e os textos selecionados datam de maio de 2009 a março de 2011. Selecionamos textos gratuitamente disponíveis e salvos em arquivos do processador de texto *Microsoft Word* em formato tipo texto sem formatação (.txt). Selecionamos os seguintes textos da revista para a coleta: artigos informativos, denominados

stories, que contêm notícias sobre a atuação das agências de Publicidade ao redor do mundo, novas campanhas, eventos de premiação, dentre outros temas; artigos informativos e opinativos, chamados de *blogs*, que apresentam algum tema relacionado à Publicidade juntamente com o comentário e/ou opinião do autor; transcrições de vídeos chamados *3 minutes Adage*, que são vídeos de 3 minutos que apresentam algum tema ou campanha realizada por alguma agência de Publicidade. Eles podem ser assistidos no próprio *site* ou pode ser feito o *download* (baixar) gratuito por meio do *iTunes*¹⁴.

A coleta dos textos foi feita da seguinte forma: com relação aos textos escritos, utilizamos o processo de selecionar, copiar e colar cada texto em um arquivo do *Microsoft Word*, tanto em formato de documento do *Word* como em formato texto sem formatação (.txt). Salvamos nos dois formatos para ter a possibilidade de acessar os textos com as imagens e outros itens que são eliminados no formato .txt com mais facilidade. Os vídeos foram assistidos e transcritos por nós e salvos da mesma maneira que os textos escritos. Nomeamos os arquivos da seguinte forma:

14 iTunes é um reprodutor de áudio e vídeo desenvolvido pela Apple, para reproduzir e organizar música digital, arquivos de vídeo e para a compra de arquivos de mídia digital.

1) Artigos (*stories*): *st_* + numeração. Exemplo: *st_0001*, *st_0002*, etc.

2) Artigos opinativos (*blogs*): *bl_* + numeração. Exemplo: *bl_0001*, *bl_0002*, etc.

3) Vídeos: *vi_* + numeração. Exemplo: *vi_0001*, *vi_0002*, etc.

O *corpus* de referência serve de contraponto ao *corpus* de estudo, para fins de comparação e de extração de seus padrões lexicogramaticais mais típicos, que emergem a partir do contraste dos dois *corpora*. Segundo Berber Sardinha (2009, p.226-227), um *corpus* de referência, também denominado *corpus* de controle, normalmente possui as seguintes características: (1) é pelo menos 5 vezes maior do que o *corpus* de estudo; (2) não deve conter o *corpus* de referência para que não seja feita uma comparação do *corpus* de estudo com ele mesmo; e (3) deve ser selecionado com um propósito específico para que não sejam comparadas listas de palavras que não tenham relação alguma uma com a outra.

O *corpus* de referência selecionado foi o *British National Corpus*, doravante *BNC*, composto por inglês escrito e falado, com cerca de 100 milhões de palavras (*tokens*). A parte escrita inclui textos de jornais, livros acadêmicos e de ficção,

cartas, redações, entre outros. A parte falada consiste em transcrições de conversas informais, de idade, regiões e classes sociais diferentes, reuniões governamentais, programas de rádio, entre outros. Essa escolha deveu-se ao fato desse *corpus* ser referência da Língua Inglesa em uso. Além disso, no momento da pesquisa, era um dos únicos *corpora* disponíveis gratuitamente para uso em programas computacionais. A lista de palavras apresenta 99.465.296 ocorrências (*tokens*) e 512.588 itens (*types*)¹⁵.

Os *corpora* foram analisados por meio da ferramenta computacional *WordSmith Tools*, versão 5.0 (SCOTT, 2008), que oferece meios de analisar *corpora* em busca da padronização lexicogramatical.

O primeiro passo para a análise foi o uso da ferramenta *WordList*, que faz uma lista de palavras de um *corpus*. Ao fazer uma lista de palavras, o programa produz três abas principais: (1) dados estatísticos do *corpus* (*statistics*); (2) uma lista de palavras por ordem alfabética (*alphabetical*); (3) uma lista de palavras por frequência (*frequency*), que é a que mostramos aqui (Figura 1). Na lista de palavras por frequência, encontramos as seguintes informações: (1) *Word*: palavra ou item; (2) *Freq.*: frequência das palavras no *corpus*

15 Em “Paulo viajou com o seu pai e com o seu irmão”, por exemplo, temos 11 ocorrências (*tokens*) e 8 itens (*types*).

de estudo; (4) %: porcentagem da frequência das palavras considerando a porcentagem do total de ocorrências das palavras; (5) *Texts*: quantidade de textos nos quais a palavra ocorre; e (6) %: porcentagem da frequência de ocorrência das palavras nos textos.

O *corpus* de estudo é constituído de 1.004.733 *tokens* ou ocorrências, e 32.166 *types*, que chamamos de itens. A proporção item/ocorrência (*Type/Token Ratio*) é 3,27. Essa proporção é feita dividindo-se o número de itens pelo número de ocorrências; o resultado é multiplicado por 100 para transformar o valor em porcentagem. Isso significa que quanto maior o valor, mais palavras diferentes e quanto menor o valor, maior o número de repetições; esse valor pode ser interpretado para verificar a riqueza lexical de um texto ou *corpus*. (BERBER SARDINHA, 2009)¹⁶.

16 As demais informações apresentadas na janela de estatística não serão aqui abordadas; para explicações detalhadas sobre elas, vide Berber Sardinha (2009).

The screenshot shows the WordList tool interface with a table of word frequencies. The table has columns for N, Word, Freq, %, Texts, %_emmas, and Set. The word 'THE' is highlighted in row 1. At the bottom, there are tabs for 'frequency', 'alphabetical', 'statistics', 'filenames', and 'notes', with 'frequency' selected. A search bar at the bottom left contains '32,166' and 'Type-in'.

N	Word	Freq	%	Texts	%_emmas	Set
1	THE	48.322	4.81	1,118	100.00	
2	TO	28.187	2.81	1,116	99.82	
3	AND	26.029	2.59	1,113	99.55	
4	A	23.789	2.37	1,114	99.64	
5	OF	23.670	2.36	1,117	99.91	
6	#	19.779	1.97	1,103	98.66	
7	IN	17.757	1.77	1,115	99.73	
8	THAT	11.428	1.14	1,105	98.84	
9	FOR	10.955	1.09	1,107	99.02	
10	IS	10.042	1.00	1,089	97.41	
11	ON	8.178	0.81	1,111	99.37	
12	WITH	7.556	0.75	1,086	97.14	
13	IT	7.395	0.74	1,080	96.60	
14	AS	6.986	0.70	1,063	95.08	
15	ARE	5.718	0.57	1,037	92.75	
16	SAID	5.417	0.54	818	73.17	
17	BUT	5.243	0.52	1,073	95.97	
18	AT	5.188	0.52	1,052	94.10	
19	BE	4.824	0.48	1,030	92.13	
20	HAVE	4.521	0.45	1,016	90.88	
21	FROM	4.328	0.43	1,024	91.59	
22	HAS	4.179	0.42	980	87.66	
23	ITS	4.127	0.41	903	80.77	
24	MORE	4.115	0.41	969	86.67	
25	BY	4.014	0.40	1,052	94.10	
26	AN	3.907	0.39	1,012	90.52	
27	THIS	3.706	0.37	999	89.36	
28	WAS	3.680	0.37	886	79.25	
29	OR	3.633	0.36	944	84.44	
30	YOU	3.599	0.36	850	76.03	
31	NEW	3.517	0.35	931	83.27	
32	NOT	3.372	0.34	980	87.66	
33	MEDIA	3.350	0.33	764	68.34	

Figura 1: Tela do WordList do Wordsmith Tools com resultado parcial de lista de palavras do *corpus* de estudo por frequência.

Podemos observar que na lista das palavras mais frequentes do *corpus* de estudo encontramos artigos, preposições, pronomes e verbos, sendo que o primeiro item relacionado à área de Publicidade, *media*, está na 33ª posição. Assim, para a observação e seleção das palavras mais representativas no *corpus* de estudo, fizemos a comparação entre a lista de palavras dele e a do *corpus* de referência (*BNC*). Para tanto, utilizamos a ferramenta *Keywords*, “que permite a seleção de itens de uma lista de palavras (ou mais)

por meio da comparação de suas frequências com uma lista de referências” (BERBER SARDINHA, 2004, p.96). A Figura 2 mostra uma janela com uma parte dos resultados.

N	Key word	Freq	%	RC. Freq	RC. %	Keyness	F	Lemmas
1	AD	2,987	0.30	3,078	19,175.07	0.0000000000		
2	MEDIA	3,350	0.33	7,862	17,348.12	0.0000000000		
3	MARKETING	2,821	0.28	5,198	15,692.83	0.0000000000		
4	BRAND	2,283	0.23	2,032	15,105.70	0.0000000000		
5	MARKETERS	1,390	0.14	61	12,299.38	0.0000000000		
6	FACEBOOK	1,139	0.11	0	10,491.74	0.0000000000		
7	COM	1,212	0.12	281	9,725.86	0.0000000000		
8	BRANDS	1,355	0.13	770	9,714.46	0.0000000000		
9	ADVERTISING	1,872	0.19	4,400	9,687.03	0.0000000000		
10	CONSUMERS	1,525	0.15	2,290	8,959.66	0.0000000000		
11	ADS	1,220	0.12	599	8,944.69	0.0000000000		
12	ONLINE	1,185	0.12	597	8,654.87	0.0000000000		
13	AGENCY	1,804	0.18	5,645	8,484.58	0.0000000000		
14	GOOGLE	905	0.09	0	8,336.11	0.0000000000		
15	CEO	886	0.09	93	7,548.24	0.0000000000		
16	DIGITAL	1,279	0.13	1,915	7,519.67	0.0000000000		
17	TV	1,610	0.16	6,294	6,966.77	0.0000000000		
18	TWITTER	727	0.07	11	6,582.24	0.0000000000		
19	ADAGE	692	0.07	115	5,715.39	0.0000000000		
20	U	1,210	0.12	4,344	5,410.42	0.0000000000		
21	ADVERTISERS	644	0.06	310	4,734.98	0.0000000000		

Figura 2: Tela do KeyWords do *Wordsmith Tools* com resultado parcial de chavidade (*keyness*) da comparação entre a lista de palavras do *corpus* de estudo e da lista de palavras do *corpus* de referência.

As palavras-chave aparecem em ordem de chavidade, de forma que no topo da lista estão as positivas, ou seja, as que possuem frequência significativamente mais alta no *corpus* de estudo do que no de referência. Quando o contrário acontece, as palavras-chave são denominadas negativas.

A partir da lista de palavras-chave, selecionamos as mais frequentes, tendo em mente a transposição dos achados

para atividades didáticas. A análise das palavras-chave e seus colocados foi feita com a ferramenta *Concord* do *Wordsmith Tools*. A Figura 3 mostra a aba *Patterns*, na qual podemos verificar os colocados de *ad*, ou seja, as palavras que mais ocorrem juntamente com *ad*.

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	THE	THE	THE	THE	THE	AD	AGE	AND	THE	THE	THE
2	TO	A	A	A	AN	A	SPENDING	THE	A	TO	TO
3	OF	AND	TO	OF	OF		AGES	IN	TO	AND	OF
4	A	OF	OF	IN	TO		AGENCY	FOR	IN	OF	AND
5	AND	TO	AND	TO	IN		NETWORKS	IS	OF	A	A
6	IN	IN	IN	AND	AND		REVENUE	TO	AND	IN	IN
7	FOR	FOR	FOR	FOR	TOLD		DOLLARS	A	ON	THAT	ON
8	IS	AS	ON	SUPER	ONLINE		CAMPAIGN	THAT	IT	ON	THAT
9	IT	ON	THAT	THAT	WITH		AGENCIES	WITH	IS	FOR	WITH
10	WITH	IT	AS	ACCORDING	BOWL		NETWORK	ON	THAT	WITH	FOR
11	AS	WITH	AN	WITH	THAT		INDUSTRY	BUT	YOU	YOU	YEAR
12	THAT	IS	IS	U	TV		SALES	ARE	FOR	ARE	BY
13	ON	AD	ITS	ON	MOBILE		FOR	HAS	ARE	AD	IS
14	WAS	THAT	WITH	FROM	THIS		AND	FROM	THIS	IS	FROM
15	BE	ARE	YEAR	30	FOR		SUPPORTED	WHAT	THEY	AN	IT
16	HAS	WILL	ARE	AN	ITS		IN	BY	AS	IT	AS
17	FROM	HAVE	AD	IT	THEIR		CAMPAIGNS	TACENTER	HAS	FROM	YOU
18	ITS	AT	FIRST	AS	NEW		BUSINESS	HOW	AN	ITS	WHO
19	ARE	NOT	BE	ITS	BY		THAT	AT	WITH	WAS	AD
20	YOU	MORE	ACCORDING	AD	S		IS	YOU	BY	HAS	AT
21	AN	AN	IT	THEIR	DISPLAY		INVENTORY	IT	WAS	MORE	BUT
22	AD	WAS	AT	NEW	MORE		BREAKS	SO	HAVE	MEDIA	ARE
23	MR	YOU	THIS	AT	SECOND		BUYERS	WILL	SAID	AS	THIS
24	HAVE	ITS	MORE	BILLION	ON		MARKET	LAST	FROM	BY	WHICH
25	MEDIA	BUT	HAVE	WHEN	VIDEO		PAGES	HAVE	PAPER	BE	WILL

Figura 3: Tela do *Concord*, aba *Patterns*, do *Wordsmith Tools* com os padrões de *ad*.

Ainda dentro da ferramenta *Concord*, analisamos os padrões lexicogramaticais típicos com as palavras selecionadas por meio de cálculos estatísticos para medir o grau de associação entre os itens, ou seja, o grau de coocorrência entre um nóculo e seus colocados. Dentre as medidas de associação disponíveis, utilizamos a Informação Mútua (*Mutual Information*) e o Escore T (*T Score*). Ambas

consideram a razão O/E, ou seja, Observado/Esperado, levando em conta, segundo Berber Sardinha (2004, p.201), “quantas vezes de fato duas palavras ocorreram juntas dentro do horizonte delimitado (o valor observado) e quantas vezes seria esperado que elas ocorressem (o valor esperado), dado o tamanho do *corpus* e a frequência de cada palavra”. O Escore T, por sua vez, identifica as associações não tão constantes, porém mais frequentes na língua (BERBER SARDINHA, 2004). Um exemplo do nosso *corpus* de estudo é *ad to*. Segundo Berber Sardinha (2004), podemos afirmar, com base nos cálculos de IM, que uma associação entre duas palavras é, de fato, uma colocação, se o resultado for superior a 3. Já no caso do Escore T, uma palavra é considerada colocada de um nóculo quando o resultado for superior a 2. Levando em conta esses cálculos, selecionamos em nossa análise os colocados que tivessem resultado de IM (Informação Mútua) acima de 3 e de Escore T acima de 2.

Observamos na Tabela 1 a seguir as 20 primeiras palavras-chave do *corpus* de estudo em ordem de chavicidade.

	Palavra	Frequência no <i>corpus</i> de estudo	% estatística da frequência no <i>corpus</i> de estudo	Frequência no <i>corpus</i> de referência	% estatística da frequência no <i>corpus</i> de referência	Chavicidade (<i>log-likelihood</i>)
1	ad	2.987	0,30	3.078	0,0030945	19.175,07
2	media	3.350	0,33	7.868	0,0079102	17.348,12
3	marketing	2.821	0,28	5.198	0,0052259	15.692,83
4	brand	2.283	0,23	2.032	0,0020392	15.105,70
5	marketers	1.390	0,14	61	0,0000613	12.299,38
6	facebook	1.139	0,11	0	0	10.491,78

7	com	1.212	0,12	281	0,0002825	9.725,86
8	brands	1.355	0,13	770	0,0007741	9.714,46
9	advertising	1.872	0,19	4.400	0,0044236	9.687,03
10	consumers	1.525	0,15	2.290	0,0023023	8.959,66
11	ads	1.220	0,12	599	0,0006022	8.944,69
12	online	1.185	0,12	597	0,0006002	8.654,87
13	agency	1.804	0,18	5.645	0,0056753	8.484,58
14	google	905	0,09	0	0	8.336,11
15	ceo	886	0,09	93	0,0000934	7.548,24
16	digital	1.279	0,13	1.915	0,0019252	7.519,67
17	tv	1.610	0,16	6.294	0,0063278	6.966,77
18	twitter	727	0,07	11	0,000011	6.582,24
19	adage	692	0,07	115	0,0001156	5.715,39
20	u	1.210	0,12	4.344	0,0043673	5.410,42

Tabela 1: Resultado da extração de palavras-chave - 25 primeiras em ordem de chavidade

(keyness).

O valor p usado em testes estatísticos varia de 0 a 1. O valor 0,01, por exemplo, indica que existe 1% de possibilidade de erro em se afirmar a existência de uma relação, ou seja, pode ter ocorrido “por acaso”. Todas as palavras da Tabela 1 possuem valor menor do que 0,000001, o que mostra que elas possuem grande significância, visto que a probabilidade de a diferença ter ocorrido por acaso é muito baixa.

A seleção dos itens foi feita de acordo com o nosso objetivo, que foi o de proporcionar o estudo do uso do vocabulário da área de Publicidade. Assim, optamos por eliminar as palavras que nomeiam *sites* da *Internet* e redes de relacionamentos, como *Facebook*, *Google* e *Twitter*. Apesar de relevantes para o ensino da língua, preferimos desenvolver, neste trabalho, atividades com palavras mais específicas da área. A palavra “*com*” no *corpus* de estudo refere-se, em geral, ao domínio que compõe um endereço da *Internet*; *adage* refere-se ao nome da revista e “*u*” refere-se a “*U.S.*” – Estados Unidos

(*United States*) ou “U.K.” – Reino Unido (*United Kingdom*). Em nossa pesquisa, foi feita uma análise mais extensa; no entanto, aqui apresentamos a análise apenas da primeira palavra-chave (*ad*) e como ela pode ser estudada pelos alunos.

3.1 Análise de *ad*

A palavra-chave selecionada para o desenvolvimento da atividade aqui apresentada foi *ad*, que aparece na primeira posição na lista de palavras-chave quando da comparação do *ad corpus* com o *corpus* de referência. Usamos a ferramenta *Concord*, do *Wordsmith Tools* para observar seus colocados, conforme mencionado anteriormente (ver Figura 3). Dentre eles, observamos que os que mais ocorrem na posição L1, ou seja, logo à esquerda de *ad*, são palavras gramaticais, isto é, determinantes e preposições, como *the*, *to*, *an* e *of*. À direita, encontramos *age* com maior frequência, o que pode ser explicado pelo nome da revista (*Adage*). A seguir, vêm as palavras *spending*, *agency* e *networks* (ver Tabela 2).

Como nossa proposta foi analisar os colocados que apresentam forte associação, optamos por analisar os que tivessem valores de IM altos. Esses colocados não são tão frequentes na língua, mas possuem ocorrência constante;

consideramos também que eles tivessem Escore T maior do que 2, o que indica que não deixam de ser frequentes.

Dentre esses colocados, encontramos somente uma palavra gramatical, que é *an*; entendemos que esse resultado corrobora nossa ideia inicial de estudar as palavras de conteúdo, por terem uma relação mais direta com o nóculo. Observando as linhas de concordância, vimos que os colocados selecionados possuem frequência alta na posição R1, à direita do nóculo. Fizemos, portanto, uma busca dos colocados nessa posição e encontramos, dentre as 10 primeiras, em ordem de frequência:

	COLOCADO	FREQUÊNCIA NO CORPUS	FREQUÊNCIA NA POSIÇÃO R1	IM	ESCORE T
1	age	553	540	9,700	23,448
2	spending	140	130	9,422	11,815
3	Age's	98	95	18,135	9,899
4	agency	131	79	9,551	11,430
5	networks	106	76	10,819	10,290
6	revenue	100	74	9,656	9,988
7	campaign	76	63	8,032	8,684
8	dollars	67	63	10,374	8,179
9	agencies	67	55	9,325	8,472
10	industry	60	49	6,655	7,699

Tabela 2: Colocados principais de *ad* com frequência na posição R1.

Encontramos também *campaigns*, no plural, em 16º lugar e *network*, no singular, na 11ª posição. Observando novamente as linhas de concordância, pudemos concluir que esses padrões apresentam um sentido específico, sendo, portanto, relevantes para estudo e, conseqüentemente, para o ensino. O próximo passo foi, então, selecioná-los para uma observação de seu uso. Os padrões *ad age*, *ad age's*

e *ad agency* estavam, em sua grande maioria, relacionados ao nome da revista, *AdAge*. Dessa forma, selecionamos os seguintes para a análise: *ad spending*, *ad network(s)*, *ad revenue* e *ad campaign*.

Para analisar o uso desses padrões, utilizamos a ferramenta *Concordance*, que traz as linhas de concordância e os colocados do item selecionado. Neste artigo apresentamos a análise de *ad campaign(s)*, por esse padrão estar diretamente relacionado com a atividade apresentada. Analisamos os 5 itens que ocorrem com ele com mais frequência e depois discutiremos os que podem ser relevantes para serem ensinados com os padrões.

Com o padrão *ad campaign(s)*, encontramos uma coocorrência de preposições, como se pode ver nas Tabelas 3 e 4.

	COLOCADO	FREQUÊNCIA	IM	ESCORE T
1	the	27	6,793	5,149
2	an	22	10,657	4,688
3	a	20	7,833	4,453
4	of	10	6,349	3,123
5	new	10	10,965	3,161
6	with	10	8,558	3,154
7	to	10	6,580	3,129
8	from	9	9,037	2,994
9	and	8	6,244	2,791
10	that	8	7,563	2,813
11	in	7	6,483	2,616
12	for	6	7,404	2,435
13	by	5	7,920	2,227

Tabela 3: colocados principais de *ad campaign*.

	COLOCADO	FREQUÊNCIA	IM	ESCORE T
1	and	17	7,332	4,098
2	of	12	6,612	3,429
3	the	11	5,497	3,243
4	for	7	7,627	2,632
5	a	5	5,833	2,197

Tabela 4: Colocados principais de *ad campaigns*.

Observamos a seguir as linhas de concordância com o nódulo *ad campaign(s)* e os colocados *to, for, from e by*:

- 1 a sweet deal in which LBL will create a new **ad campaign** to stimulate sales at 50% off th
- 2 nized the need for an organized, strategic **ad campaign** to drive the ongoing demobilizat
- 3 ng, pricing strategies, retail promotions and **ad campaigns** to reach consumers effectively
- 4 the recession, earning itself a new national **ad campaign** from Publicis' Saatchi & Saatchi
- 5 f course, to the software giant's ongoing **ad campaign** from agency Crispin Porter & Bog
- 6 mployees. Chrysler breaks the "New Day" **ad campaign** from BBDO. The spots use third
- 7 Fox somewhat, but he largely credits an **ad campaign** from Publicis Worldwide, New Y
- 8 y a just-launched umbrella marketing and **ad campaign** from WPP's Y&R that includes s
- 9 mployees. Chrysler breaks the "New Day" **ad campaign** from BBDO. The spots use third
- 10 rm details about the upcoming national **advertising campaign** for Via, other than the offi
- 11 t's not to say there aren't marketing and **advertising campaigns** for the Beatles' selected p
- 12 n the course of 2010, Kraft has launched new **ad campaigns** for 15 of its "power brands,"
- 13 new standard of the world," and forthcoming **ad campaigns** for the brand will center on a
- 14 is year." The cross-platform **ad campaign** by the Atlanta shop Lawler Ballard Va
- 15 better known as FARC. An **ad campaign** by Bogota-based Lowe SSP3 is helping
- 16 is being promoted with an **ad campaign** by WPP's Y&R with a "Heaven on Earth
- 17 OO by the end of the year. An **ad campaign** by Sancho BBDO highlights that the s

Analizamos as linhas de concordância e vimos que o uso de *ad campaign(s)* com as preposições, *to, for e from* à direita mostrou-se interessante. O primeiro bloco de linhas de concordância nos mostra o uso de adjetivo + *ad campaign*

+ *to* + verbo. Esse uso refere-se ao propósito da campanha publicitária. Em segundo lugar, verificamos o padrão adjetivo/nome da campanha + *ad campaign* + *from/by* + nome da agência; vemos aqui que esse padrão é usado em geral para referir-se a uma certa campanha e que agência a criou ou desenvolveu. Por último, encontramos *ad campaign* + *for* + marca /empresa, relacionando a marca ou empresa para a qual a campanha foi criada ou desenvolvida.

Como podemos observar nas concordâncias, o adjetivo *new* é um colocado típico à esquerda de *ad campaign(s)*, provavelmente pela natureza dos negócios feitos nas agências, que é a criação de campanhas.

Essas concordâncias nos permitem observar como expandir as ideias relacionadas às campanhas e quais os verbos comumente usados para isso, como *use*, *break* e *leverage*.

Conforme podemos observar nas concordâncias, *in* é uma preposição que coocorre com *ad campaign*.

Os padrões que pudemos encontrar com o nódulo estudado são: *ad campaign* + *to* / *for* / *from* / *by*, *new* + *ad campaign*, *in* + *a/an* + (adjetivo) + *ad campaign* e *ad campaign* + *that*.

4. DESENVOLVIMENTO DA ATIVIDADE

Nossa proposta visa desenvolver no aluno a capacidade de observação do vocabulário na língua inglesa em uso em seu meio profissional a fim de detectar os padrões que o envolvem para poder usá-los em sua comunicação. Os alunos também entram em contato com os diversos itens gramaticais apresentados nos livros didáticos de maneira mais livre e significativa, em ocasiões diversas e sem um sequenciamento rígido. Conforme Willis, “considerar o léxico como ponto inicial permite a identificação dos significados e padrões mais comuns e oferece aos alunos uma imagem do que é típico na maneira como o inglês é usado”¹⁷ (WILLIS, 1990, p.129), o que, a nosso ver, desperta o interesse dos alunos.

Essa proposta condiz com as abordagens aqui discutidas no que concerne à exposição do aluno a linhas de concordância com questões formuladas pelo professor que irão guiá-lo a encontrar os padrões mais frequentes (JOHNS, 1991; SINCLAIR, 1991; HUNSTON, 2002; BERNARDINI, 2004; BERBER SARDINHA, 2004, 2009, 2011). Concordamos com a ideia de Bernardini (2004) ao tratar o aluno como um descobridor e não como pesquisador, como sugere Johns

17 “taking lexis as a starting point enables us to identify the commonest meanings and patterns and to offer students a picture which is typical of the way English is used” (WILLIS, 1990, p.129)

(1991); entretanto, usamos como base a análise da linguagem por meio de concordâncias, premissa do DDL.

De acordo com as propostas de ensino de ILE baseado em *corpora*, elencamos os seguintes pressupostos para o desenvolvimento de nossa atividade:

- 1) Uso de textos autênticos, não criados com um propósito pedagógico, mas sim para uma comunicação real;
- 2) Criação de oportunidades para os alunos usarem a língua para comunicação real;
- 3) Uso de linhas de concordância para a exploração da língua em uso;
- 4) Foco na lexicogramática, nos padrões lexicogramaticais e nas colocações;
- 5) Desenvolvimento das habilidades de exploração e descoberta dos alunos;
- 6) Papel do professor de observador, mediador e facilitador, guiando os alunos para as descobertas.

Procuramos desenvolver uma atividade em que os alunos pudessem lidar com o texto e com o *corpus*, pois acreditamos que “texto e *corpus* são duas unidades de língua que se completam naturalmente, até porque a segunda

é uma coletânea da primeira” (BERBER SARDINHA, 2009). Conforme explica o autor, um *corpus* pode ser interessante, porém não pode ser lido de forma linear como um texto, mas sim convida a uma leitura hipertextual; um *corpus* explicita a intertextualidade, enquanto o texto a supõe. A leitura e discussão dos textos da área são importantes para que os alunos possam comunicar-se sobre assuntos de sua realidade, o que, em nossa experiência, mostrou-se ser um fator de motivação.

Observa-se que é possível desenvolver a atividade com ferramentas e recursos variados, de forma que o professor pode preparar os exemplos previamente, se não tiver os recursos em sala de aula, assim como usá-los juntamente com os alunos, se houver a possibilidade. O tempo da atividade varia de acordo a disponibilidade de um computador em sala de aula ou não, com o número de alunos e seu rendimento, entre outros fatores. A princípio, no entanto, calculamos que ela possa ser realizada em 2 aulas de 1 hora e trinta minutos cada, lembrando que esse tempo depende do contexto de ensino, do rendimento dos alunos, dos recursos disponíveis para o professor em sala de aula, entre outros fatores.

Cabe ressaltar que essa proposta pode ser usada em aulas individuais ou em grupos. No caso de aulas individuais,

os alunos podem refletir sozinhos e reportar suas ideias ao professor ou então discutir as questões juntamente com ele, que terá um papel de participante da discussão, e depois reportar a discussão por escrito, em forma de redação, por exemplo. Além disso, sugerimos em alguns momentos que o aluno busque mais exemplos em linhas de concordância do item que está estudando. Isso pode ser feito pelo professor antecipadamente, de forma impressa, ou, se houver a possibilidade, com o uso de um ou mais computadores em sala de aula.

A atividade aqui apresentada parte de um padrão a ser explorado, que é o *ad campaign*. Desenvolvemos a atividade utilizando o *corpus* de estudo e uma das ferramentas do programa *AntConc*¹⁸, por ser de mais fácil uso, de acordo com pareceres de professores que já utilizaram o *Wordsmith Tools* e o *AntConc*. Sugerimos também que o professor explore o uso dos padrões usando o COCA (*Corpus of Contemporary American English*), mencionado no exercício (1). Da mesma maneira, é possível preparar as linhas de concordância previamente e utilizá-las em sala de aula em formato impresso. A atividade é apresentada em formato de slides do programa *Microsoft PowerPoint*.

A princípio, sugerimos que os alunos reflitam e discutam

18 <http://www.antlab.sci.waseda.ac.jp/software.html>

sobre uma campanha publicitária (*ad campaign*). A primeira parte da atividade traz linhas de concordância para o estudo do padrão e seus colocados. Os alunos observam as linhas do *corpus* de estudo e associam os usos aos propósitos (Ver Figuras 4, 5 e 6).

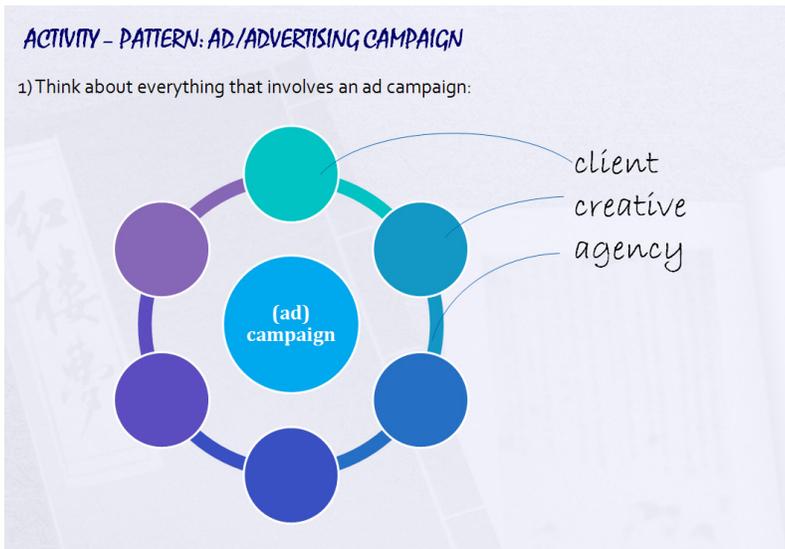


Figura 4: Slide 1 da atividade.

2) Look at the right of 'ad campaign' and mark:

a → for the ones that are about the agency (that created the campaign)

b → for the ones that are about the product / brand

c → for the ones that are about the purpose of the campaign

- () than Fox somewhat, but he largely credits an **ad campaign** from Publicis Worldwide, New York, th
 () 75 employees. Chrysler breaks the "New Day" **ad campaign** from BBDO. The spots use third-party
 () or a sweet deal in which LBL will create a new **ad campaign** to stimulate sales at 50% off their nor
 () sen, CMO of the clinic, which started a national **ad campaign** earlier this year from Adworks, Washi
 () cognized the need for an organized, strategic **ad campaign** to drive the ongoing demobilization eff
 () new standard of the world," and forthcoming **ad campaigns** for the brand will center on a theme
 () h new storytelling techniques and high-impact **ad campaigns** like the great ads for Apple. But ev
 () rs to attend "the unveiling of its first global **advertising campaign**" from TBWA/Chiat/Day tomo
 () nfirm details about the upcoming national **advertising campaign** for Via, other than the offi
 () at's not to say there aren't marketing and **advertising campaigns** for the Beatles' selected p

2) Underline the words that common to each one. What pattern can you see?



Figura 5: Slide 2 da atividade.

3) Read the excerpts below and identify:

- ♦ which / what kind of ad campaign they are about;
 - ♦ in which ones you know about the company or agency involved in the campaign.
- a) Several years ago, the Syracuse University football team did an **ad campaign** in New York City...
- b) "The Beatles: Rock Band," for instance, launched last week with an estimated \$20 million **ad campaign**...
- c) Sergio Jaramillo, Colombia's vice minister for defense, recognized the need for an organized, strategic **ad campaign** to drive the ongoing demobilization effort...
- d) ...LBL will create a new **ad campaign** to stimulate sales at 50% off their normal sticker price for creative services.
- e) Chrysler breaks the "New Day" **ad campaign** from BBDO.
- f) It was a nod, of course, to the software giant's ongoing **ad campaign** from agency Crispin Porter & Boguysy.
- g) The company wouldn't confirm details about the upcoming national **advertising campaign** for Via...
- h) New York-based journalists have gotten invites from the French tire maker and awarder of restaurant stars to attend "the unveiling of its first global **advertising campaign**" from TBWA/Chiat/Day tomorrow...

Figura 6: Slide 3 da atividade.

Em seguida, sugerimos que os alunos pesquisem no COCA as palavras que coocorrem com o padrão, comparando esse uso com o uso encontrado no *corpus* de estudo (Ver Figura 7). Conforme explicamos, os alunos podem conhecer o padrão, mas geralmente encontram dificuldades ou limitações quanto ao seu uso; com esse exercício, eles terão a possibilidade de expandir seu conhecimento e utilizar outros ou colocados frequentemente empregados com ele.

Para realizar um exercício com um texto com mais exemplos do padrão, selecionamos um texto do *corpus* de estudo por meio da ferramenta *Concordance Plot* do programa *AntConc* (Ver Figura 8). Selecionado o texto, desenvolvemos um exercício em que retiramos algumas palavras que coocorrem com *ad campaign* para que os alunos as reposicionem no texto, como uma tarefa de preencher lacunas (Ver Figura 9).

3) Do some research on the corpus collected or on COCA (*Corpus of Contemporary American English*) and find more instances of:

- what collocates with ad campaign
- verbs that collocate with ad campaign(s)
- adjectives that collocate with ad campaign(s)

COCA – * + ad campaign*		CONTEXT
1	<input type="checkbox"/>	AN AD CAMPAIGN
2	<input type="checkbox"/>	NEW AD CAMPAIGN
3	<input type="checkbox"/>	THE AD CAMPAIGN
4	<input type="checkbox"/>	MILLION AD CAMPAIGN
5	<input type="checkbox"/>	" AD CAMPAIGN
6	<input type="checkbox"/>	'S AD CAMPAIGN
7	<input type="checkbox"/>	TELEVISION AD CAMPAIGN

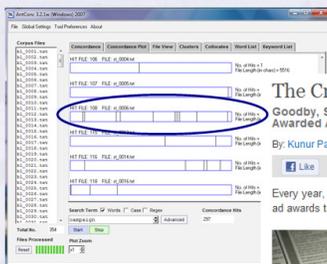
COCA – verb + an ad campaign		CONTEXT
1	<input type="checkbox"/>	LAUNCHED AN AD CAMPAIGN
2	<input type="checkbox"/>	DO AN AD CAMPAIGN
3	<input type="checkbox"/>	DEVELOPING AN AD CAMPAIGN
4	<input type="checkbox"/>	DESIGNING AN AD CAMPAIGN
5	<input type="checkbox"/>	DESIGNED AN AD CAMPAIGN
6	<input type="checkbox"/>	LAUNCHING AN AD CAMPAIGN
7	<input type="checkbox"/>	BEGAN AN AD CAMPAIGN

<http://corpus.byu.edu/coca/>

4) Are these collocations similar or different from the ones found in the Advertising corpus?

Figura 7: Slide 4 da atividade.

FURTHER PRACTICE WITH A TEXT



Finding a text for further practice using the corpus collected and *AntConc* (Concordance plot)

The Creativity Awards Report

Goodby, Silverstein & Partners Tops Agency Winners, While BBDO Remains the Most-Awarded Ad Network

By: Kunur Patel

Published: September 28, 2009

Every year, once the gold dust has settled, Ad Age's *Creativity Combs* through the results of the leading global ad awards to determine which people, companies and campaigns claimed the most hardware.



This Columbus, Ohio-based agency's campaign for Toyota's Camry was the most-awarded campaign of 2009.

This year, top agencies earned distinction largely due to interactive work. But it's not just interactive know-how that sets an agency apart to awards juries. The most-awarded agency of 2009 is Goodby & Silverstein & Partners, a shop that simultaneously was named Cannes' digital agency of the year and, according to our count, had the most-awarded TV spot this season with "Rabbit" for Comcast.

What's more, the San Francisco-based Omnicom agency didn't earn this distinction for hitting one home run that won everything — last year's top winners could chalk success up to standard campaigns like HBO's "Voyeur" and "Halo 3." A broad range of award-winning work and strength across media proved fruitful for Goodby, which won for work such as the YouTube-quaking "Warriorland Shake It" for Nintendo, the split-screen "There Can Only Be One" spots for the NBA that inspired spoofs on the cover of Time magazine and on SNL, the "Save the Honey Bees" campaign for Honey Bunches of Oats and the virtual haunted "Hotel 626" for Doritos. Goodby also won for the first augmented-reality campaign for a major advertiser, the GE Plug into the Smart Grid website, as well as the data-crunching Sprint Now widget.

See all the winning work. Last year, Goodby was the eighth agency on our awards list. Agency co-

Figura 8: Slide 5 da atividade - Exemplo de como encontrar um texto que contenha as palavras ou padrões para desenvolver a atividade usando a ferramenta *Concordance Plot* do programa *AntConc*.

FURTHER PRACTICE WITH A TEXT

5) Read the first part of the text below and decide in which of the black spaces the following words fit:

standout
for
"Save the honey bees"
most-awarded
for

The Creativity Awards Report
Goodby, Silverstein & Partners Tops Agency Winners, While BBDO Remains the Most-Awarded Ad Network

By: Kunur Patel | Published: September 28, 2009

Like Tweet 0 +1 0 [+] Share | Email | Reprints | Print | S

Every year, once the gold dust has settled, Ad Age's Creativity **combs through** the results of the leading global ad awards to determine which people, companies and campaigns claimed the most hardware.



This year, top agencies earned distinction largely due to interactive work. But it's not just interactive know-how that sets an agency apart to awards juries. The most-awarded agency of 2009 is **Goodby**, Silverstein & Partners, a shop that simultaneously was named Cannes' digital agency of the year and, according to our count, had the most-awarded TV spot this season with "Rabbit" for Comcast.

What's more, the San Francisco-based Omnicom agency didn't earn this distinction for hitting one home run that won everything -- last year's top winners could chalk success up to **campaigns** like HBO's "Voyeur" and "Halo 3." A broad range of award-winning work and strength across media proved fruitful for Goodby, which won for work such as the **Yogi The Baskin "Wainland Shake II" for Nintendo** the edit.

This **campaign** for Tourism Queensland was the **campaign** of 2009.

<http://adage.com/article/agency-news/creativity-awards-report-2009/139270/>

6) Talk about:

- ♦ the campaigns mentioned in the text;
- ♦ ongoing / forthcoming campaigns.

Figura 9: Slide 6 da atividade.

5. CONSIDERAÇÕES FINAIS

Nossa prática de ensino nos permitiu verificar as necessidades dos alunos que atuam na área de Publicidade na aprendizagem de ILE. Observamos que os alunos possuíam conhecimento do vocabulário do meio, mas apresentavam dificuldades em utilizá-lo. Devido à falta de materiais didáticos voltados para esse público, que possibilitam aos alunos o contato com a linguagem da área, professores optam por criar suas próprias atividades com textos retirados da *Internet*; entretanto, essas atividades comumente focam primordialmente questões gramaticais e de vocabulário separadamente, o que não auxilia o aluno a entender o funcionamento da língua em uso. Com a LC, que vê a padronização da língua e considera a lexicogramática, professores e alunos podem enxergar as relações entre os itens da língua, ou seja, as palavras frequentemente usadas com o vocabulário da área, formando padrões que auxiliam os alunos a comunicar-se com mais fluência (SINCLAIR, 1991).

Neste artigo mostramos o desenvolvimento de uma atividade para o ensino de ILE com base no estudo dos padrões lexicogramaticais distintivos do *ad corpus*, um *corpus* compilado com o objetivo de servir para a pesquisa dos padrões distintivos da área de Publicidade. Encontramos

diversos padrões com as palavras-chave resultantes da comparação entre o *ad corpus* e o BNC, que usamos como *corpus* de referência. Pudemos analisar os diversos padrões colocados em nossa trajetória; por motivos de espaço, apresentamos a análise dos padrões com o nódulo *ad* e *ad campaign*.

Selecionamos alguns dos padrões encontrados para mostrar o desenvolvimento da atividade proposta não seguindo somente uma abordagem de ensino, mas envolvendo diversos pressupostos, de diferentes abordagens de ensino em geral e de ensino com *corpora*, como o DDL (JOHNS, 1991), Lexical Syllabus (WILLIS, 1990) e as propostas de Berber Sardinha (2009; 2011) de atividades centradas na concordância e no texto.

Com o uso da LC em sala de aula, objetivamos levar o aluno à reflexão sobre a linguagem e ao uso da língua com eficiência, dando-lhe voz, de forma que o ensino possa partir de seus questionamentos sobre esse uso da língua e o professor possa agir como um guia e facilitador, ajudando-o a encontrar suas respostas e customizando o ensino às suas necessidades. Acreditamos, assim, que esta proposta poderá servir de incentivo para que professores possam mudar seus paradigmas, questionar e refletir sobre suas práticas,

livrando-se da resistência ao novo e do conformismo (CELANI, 1998).

Segundo Moita Lopes, muitas pesquisas em Linguística Aplicada estão “orientadas *para* a sala de aula” e não “*na* sala de aula” (2002, p.86 - grifo meu), ou seja, o pesquisador e os envolvidos no processo, professores e alunos, estão distantes entre si. No entanto, o autor aponta para uma tendência de pesquisa, que procuramos fazer, relacionada ao movimento do professor-pesquisador, a pesquisa-ação, na qual o pesquisador está envolvido diretamente com a investigação crítica de sua prática.

Essa é uma das questões mais discutidas hoje com relação ao ensino de LC com *corpora*. Conforme O’Keefe et al. (2007, p.246), para que a LC possa “informar” para o ensino, professores precisam “informar” à LC, visto que os professores são os melhores mediadores entre as descobertas com *corpora* e a prática. Os autores fazem algumas sugestões para ajudar professores a sentirem-se confortáveis em usar *corpora* no ensino, sendo elas maior disponibilidade de *corpora* e ferramentas *online* e uma maior diversidade de *corpora*. Mostramos aqui algumas ferramentas e *corpora* disponíveis e como o professor pode compilar um *corpus* e analisar o uso da língua, transpondo seus achados para atividades de ensino.

Uma das limitações desse trabalho corresponde à compilação do *corpus*, que pode ser trabalhosa e demandar tempo. A análise e seleção dos itens lexicogramaticais requer certa prática e a mudança da visão da língua como estruturas a serem preenchidas com o léxico para a lexicogramática. O desenvolvimento das atividades também demanda tempo e criatividade para que as atividades sejam variadas e compreendam assuntos variados e interessantes ao aluno. Esses são alguns dos “contras” assinalados por Berber Sardinha (2011) e por professores e pesquisadores que usam a LC no ensino em conversas informais. Acreditamos, no entanto, que quando o professor adquire certa prática, a preparação de aulas e de atividades com *corpora* torna-se menos trabalhosa e é feita em menos tempo, visto que o professor pode reutilizar algumas estruturas de atividades, somente modificando o texto. Em nossa opinião, esse trabalho compensa, pois não é mais possível conceber o ensino da língua sem considerar seu uso e os padrões lexicogramaticais que a envolvem; as descobertas que diversos pesquisadores já fizeram e fazem e que nós estamos começando a fazer são inestimáveis. Acreditamos que ensinar o idioma dessa maneira traz mudanças importantes para os alunos, tanto com relação à visão da linguagem no que diz respeito à padronização, quanto ao desenvolvimento de habilidades

e capacidades de uso de tecnologias, de manipulação de dados, de descoberta, reflexão e conclusões.

Infelizmente, em nosso país ainda são poucas e limitadas as propostas de LC para o ensino de ILE (BERBER SARDINHA, 2004). Se por um lado podemos tirar proveito do que já foi feito, por outro essa situação abre-nos a oportunidade de ir à busca da construção de novos conhecimentos. Por esses motivos e pela realidade com a qual nos deparamos em sala de aula é que nos propomos a desenvolver esta pesquisa e que apresentamos parte dela no presente artigo.

REFERÊNCIAS

AIJMER, Karin (2009). "Introduction". In AIJMER, Karin (Ed.). *Corpora and language teaching*. Amsterdam and Philadelphia: John Benjamins.

ASTON, Guy (1997). "Enriching the learning environment: corpora in ELT". In: WICHMANN, Anne; et al. *Teaching and language corpora*. London/New York: Longman.

BERBER SARDINHA, Tony (2004). *Linguística de Corpus*. São Paulo: Manole.
_____. (2006). "Preparação de material didático para Aprendizagem Baseada em Tarefas com Wordsmith Tools e corpora". *Calidoscópio*, 4(3), 148-155. In www.unisinos.br/publicacoes_cientificas/images/stories/pdfs_calidoscopio/vol4n3/art03_berber.pdf. Acesso em Mai.2008.

_____. (2007). Corpus linguistics: readings in a widening discipline. *DELTA* [online], 23 (1), 127-137. In http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502007000100006. Acesso em Jul.2009.

_____. (2009). *Pesquisa em Linguística de Corpus com Wordsmith Tools*. Campinas: Mercado de Letras.

_____. (2011). "Como usar a Linguística de Corpus no ensino de língua

estrangeira. Por uma Linguística de Corpus educacional brasileira". In: VIANA, Vander.; TAGNIN, Stella. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial.

_____; ALMEIDA, Gladis Maria de Barcellos (2008). "A Linguística de Corpus no Brasil". In: Tagnin, Stella Esther Ortweiler; Vale, Oto Araujo (Orgs.). *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas.

BERNARDINI, Silvia (2004). "Corpora in the classroom: an overview and some reflections on future developments". In: SINCLAIR, John (Ed.) *How to use corpora in language teaching*. Amsterdam and Philadelphia: Routledge.

BROWN, H. Douglas (2007). *Teaching by principles: an interactive approach to language pedagogy*. NY: Pearson.

CELANI, Maria Antonieta (1998). "Transdisciplinaridade na Linguística Aplicada no Brasil". In: SIGNORINI, Inês; CAVALCANTI, Marilda (Orgs.). *Linguística Aplicada e Transdisciplinaridade*. Campinas: Mercado de Letras.

CHAMBERS, Angela (2005). "Integrating Corpus Consultation in Language Studies". *Language Learning and Technology*, 9 (2), 111–125.

_____. (2010). "What is data-driven learning?" In: MCARTHY, Michael e O'KEEFE, Anne (Eds.) *The Routledge handbook of Corpus Linguistics*. New York: Routledge.

CHENG, Winnie (2010). "What can a corpus tell us about language teaching?" In: MCARTHY, Martin e O'KEEFE, Anne (Eds.) *The Routledge handbook of Corpus Linguistics*. London e New York: Routledge.

_____; WARREN, Martin; XU, Xun-feng (2003). "The Language Learner as Language Researcher: Corpus Linguistics on the Timetable", *System*, 31 (2), 173–186.

ELLIS, Nick (2002). "Frequency Effects in Language Processing. A Review with Implications for Theories of Implicit and Explicit Language Acquisition". *Studies in Second Language Acquisition*, 24, 143–188.

FIRTH, John Rupert (1957) *Papers in linguistics – 1934-1951*. Oxford: Oxford University Press.

GAVIOLI, Laura; ASTON, Guy (2001). "Enriching Reality: Language Corpora in Language Pedagogy". *ELT Journal*, 55 (3) 238–246.

GILQUIN, Gaëtanelle; GRANGER, Sylviane (2010) "How can data-driven learning be used in language teaching?" In *MCARTHY, Martin; O'KEEFE, Anne (Eds.). The Routledge handbook of Corpus Linguistics. London e New York: Routledge.*

HALLIDAY, Michael Alexander Kirkwood (1994). *An Introduction to Functional Grammar*. 2.ed. London: Edward Arnold.

HUNSTON, Susan (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

_____ (1991). "Should you be persuaded: two examples of data-driven learning". *ELT Journal*. Birmingham: University of Birmingham. 4, 1-16.

KENNEDY, Graeme (1998) *An introduction to corpus linguistics*. New York: Longman.

McENERY, Tony e WILSON, Andrew (1997). "Teaching and language corpora". *ReCall*, 9 (1), 5-14.

MOITA LOPES, Luiz Paulo da (2002). *Oficina de Lingüística aplicada – A natureza social e educacional dos processos de ensino-aprendizagem de línguas*. Campinas: Mercado de Letras.

O'KEEFE, Anne; McCARTHY, Martin; CARTER, Ronald (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.

O'SULLIVAN, Íde (2007). "Enhancing a Process-Oriented Approach to Literacy and Language Learning: The Role of Corpus Consultation Literacy". *ReCALL*, 19 (3), 269–86.

PALMER, Firth Rupert (Ed.) (1968). *Selected Papers of J.R. Firth 1952-1959*. Longman.

RÖMER, Ute (2006). "Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments". *Zeitschrift für Anglistik und Amerikanistik*, 54 (2), 121-134.

_____ (2009). "Corpus research and practice: what help do teachers need and what can we offer?" In *AIJMER, Karin (Ed.). Corpora and*

- language teaching*. Amsterdam and Philadelphia: John Benjamins.
- SCOTT, Mike (2008). *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.
- SINCLAIR, John et al. (1987). *Cobuild English Dictionary*. London and Birmingham: Collins COBLUID.
- _____ (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- _____ (2004a). *Trust the text: Language, corpus and discourse*. London and New York: Routledge.
- _____ (2004b). "Introduction". In: SINCLAIR, John (Ed.) *How to use corpora in language teaching*. Amsterdam and Philadelphia: Routledge.
- STUBBS, Michael (1996). *Text and corpus analysis: computer-assisted studies of language and culture*. Oxford: Blackwell.
- WIDDOWSON, Henry George (1991). "The description and prescription of language". In: ALATIS, James (Orgs.) *Georgetown University Round Table on Languages and Linguistics*. Washington, D. C.: Georgetown University Press.
- WILLIS, Dave (1990). *The Lexical Syllabus*. London: Collins ELT.
- _____; WILLIS, Jane (2007). *Doing Task-based Teaching*. Oxford: Oxford University Press.

Cristina Mayer Acunzo é Doutoranda em Linguística Aplicada e Estudos da Linguagem na Pontifícia Universidade Católica de São Paulo – PUC-SP, Mestre em Linguística Aplicada e Estudos da Linguagem pela Pontifícia Universidade Católica de São Paulo – PUC-SP. PUC-SP. CAPES. Berber Sardinha, T., Kauffmann, C., & Acunzo, C. M. A multidimensional analysis of register variation in Brazilian Portuguese. *Corpora*, 9(2), 239-271, 2014. GELC – Grupo de Estudos em Linguística de Corpus. cristinaacunzo@hotmail.com.

Recebido em 08 de junho de 2017.

Aprovado em 28 de julho de 2017