
LINGUÍSTICA DE *CORPUS* E ENSINO DE LÍNGUA ESTRANGEIRA: O USO DE *CORPORA* NA ELABORAÇÃO DE MATERIAIS DIDÁTICOS E ENSINO DE VOCABULÁRIO

CARNEIRO, Raphael Marco Oliveira¹

Recebido em: 2014.12.10

Aprovado em: 2014.07.18

ISSUE DOI: 10.3738/1982.2278.1025

RESUMO: O objetivo deste artigo é apresentar uma análise comparativa de dois materiais didáticos de ensino de língua inglesa, no que diz respeito ao ensino de vocabulário. Tal comparação se justifica pelo fato de que um dos materiais analisados foi elaborado com base em *corpus* enquanto o outro não. Nesse sentido, pretendemos verificar as contribuições da Linguística de Corpus para o ensino de língua inglesa no que concerne a elaboração de materiais didáticos e ensino de vocabulário. Para o embasamento teórico da pesquisa partimos principalmente dos estudos de Berber Sardinha (2004, 2010) sobre Linguística de *Corpus*, de McCarten (2007) e McCarthy (2004) sobre o uso de *corpus* na elaboração de materiais didáticos e Pikulski e Templeton (2004) sobre ensino de vocabulário. A análise realizada aponta que os materiais informados por *corpus* apresentam vantagens significativas em relação aos não informados, uma vez que são elaborados tendo como base o uso real da língua em contextos autênticos.

Palavras-chave: Linguística de *Corpus*. Linguística aplicada. Ensino de vocabulário. Análise de material Didático.

SUMMARY: This paper aims to present a comparative analysis of two course books designed for teaching English on what concerns vocabulary teaching. This comparison is justified insofar as one of the course books is based on corpus and the other is not. Thus, we intend to verify the contributions of Corpus Linguistics to English Language Teaching concerning course books design. For the theoretical framework, this research is based on the Corpus Linguistics studies of Berber Sardinha (2004, 2010), on McCarten (2007) and McCarthy (2004) studies on the use of corpus for course books design and Pikulski and Templeton (2004) studies on vocabulary teaching. The analysis made points out that corpus-based course books have advantages in relation to traditional course books, once they are designed according to real language usage in authentic contexts.

Keywords: Corpus Linguistics. Applied Linguistics. Vocabulary Teaching. Coursebook Analysis.

INTRODUÇÃO

Pesquisas no ensino de língua estrangeira têm acarretado diversas mudanças em como ensinar e aprender uma língua estrangeira. A natureza da linguagem, o papel do aprendiz no processo de ensino-aprendizagem e novas perspectivas sobre a aquisição de uma segunda língua levaram a uma crescente e sofisticada visão a respeito da aprendizagem e do ensino de uma segunda língua. Segundo Nunan (1999) nosso grande desafio agora é pensar em como incorporar novas maneiras de se ensinar na prática. Pela conexão de desenvolvimentos no ensino de línguas às principais correntes educacionais e pelo teste crítico de novas ideias poderemos contribuir para a evolução e o desenvolvimento do ensino de línguas.

Nessa perspectiva, a Linguística de *Corpus* (LC) tem fornecido importantes contribuições para o ensino de línguas estrangeiras a partir das evidências empíricas de uso da língua em contextos naturais observados por meio dos chamados *corpora*. Segundo Tagnin (2005):

¹ Graduando do Curso de Licenciatura Plena em Letras – Habilitação em Inglês e Literaturas de Língua Inglesa no âmbito do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia (UFU)

Para a LC, um corpus é uma coletânea de textos, necessariamente em formato eletrônico, compilados e organizados segundo critérios ditados pelo objetivo da pesquisa a que se destina. O formato eletrônico permite que esses textos sejam investigados e analisados automaticamente, com o uso de ferramentas computacionais específicas (TAGNIN, 2005, p. 21).

Ainda sobre a definição de *corpus*, McCarten (2007) acrescenta uma dimensão um pouco mais metodológica de como os *corpora* são coletados e analisados.

Um corpus é basicamente uma coleção de textos armazenada em um computador. Os textos podem ser escritos ou orais. Textos escritos como jornais ou revistas, para que sejam armazenados em formato eletrônico podem ser digitalizados, inseridos no computador por um CD, ou pela Internet. Textos orais, como diálogos, são gravados e transcritos, ou seja, são transcritos palavra por palavra para que os textos desses diálogos possam alimentar o banco de dados do computador. Assim, é possível analisar a amostra linguística de um corpus por meio de ferramentas computacionais, a fim de se investigar como as pessoas realmente falam ou escrevem² (MCCARTEN, 2007, p. 2; tradução minha).

Segundo Berber Sardinha (2004), a LC tem feito incursões não apenas em áreas já consagradas pelo uso de *corpora* como a Tradução e a Lexicografia, mas no campo da Linguística Aplicada também. Nesse sentido, face à descrição linguística proporcionada pela LC podemos repensar o tipo de instrução oferecida em aulas de língua estrangeira.

O que torna a proposta da linguística de *corpus* particularmente interessante [...] é que, antes de qualquer outra coisa, ela questiona um dos postulados da linguística tradicional (teórica), com base em uma perspectiva de ordem eminentemente prática (RAJAGOPALAN, 2006, p.160).

Assim, o uso de *corpus* para efeitos de análise e descrição linguística, descrição de padrões e estudo de frequências e de tradução (BERBER SARDINHA, 2004) tem contribuído sobremaneira para o desenvolvimento de estudos e pesquisas nessas áreas, dentre muitas outras. Porém no ensino, apesar de já haver pesquisas (BERBER SARDINHA, 2004, 2010) nota-se certa timidez.

Segundo Berber Sardinha (2010), a área de avaliação de material didático elaborado com base em *corpora* tem tido uma produção importante, tendo em vista o objetivo de comparar os textos e/ou explicações contidas em livros didáticos baseados em *corpora* com a finalidade de verificar até que ponto o conteúdo dos livros é fidedigno à linguagem real, não inventada para fins educacionais. Por isso, justifica-se a realização desta pesquisa.

Em vista das considerações acima, pretendemos verificar a contribuição da LC para o ensino de língua inglesa por meio da análise comparativa entre materiais didáticos informados e não informados por *corpora*. Dessa maneira partiremos dos seguintes questionamentos:

1. Como a Linguística de *Corpus* pode contribuir para a elaboração de materiais didáticos e o ensino de vocabulário?
2. Que diferenças significativas para o ensino de línguas podemos encontrar ao compararmos materiais didáticos elaborados a partir dos estudos de *corpora* com materiais convencionais?

² Texto original: A corpus is basically a collection of texts which is stored in a computer. The texts can be written or spoken language. Written texts like newspapers and magazines can be entered into the computer from a scanner, a CD, or the Internet. Spoken texts, like conversations, are recorded and then the recordings are transcribed; that is, they are written down word for word, so that the texts of these conversations can be fed into the computer database. It is then possible to analyze the language in the corpus with corpus software tools to see how people really speak or write (MCCARTEN, 2007, p. 2).

Assim, conduzimos nossa investigação, tendo em mente os objetivos, o referencial teórico e a metodologia, apresentados a seguir.

OBJETIVOS

O objetivo geral da pesquisa aqui descrita é realizar uma análise de materiais didáticos baseados em *corpora* e compará-los com materiais que não os utilizam. Para tanto, comparamos dois materiais didáticos: *Touchstone 3* (baseado em *corpus*; doravante MD1) e *New American Inside Out* (não baseado em *corpus*; doravante MD 2), investigando como as informações dos *corpora* são incorporadas pelo livro didático para o ensino de vocabulário. Analisamos também as diferenças entre as informações fornecidas pelos materiais didáticos em relação aos conteúdos léxico-gramaticais.

REFERENCIAL TEÓRICO

A presente pesquisa fundamenta-se tanto nos construtos teóricos da Linguística de *Corpus* (doravante LC) quanto no ensino e aprendizagem de vocabulário. Inicialmente, traçamos algumas considerações em relação ao ensino e aprendizagem de vocabulário, e em sequência sobre a natureza da LC e suas incursões no ensino de línguas.

Quando falamos de ensino de vocabulário, uma das primeiras questões que surgem é a seguinte: o que precisamos ensinar sobre vocabulário? Essa questão nos leva a outra: o que é aprender uma palavra? É possível saber quando aprendemos uma palavra de fato?

Segundo McCarten (2007) aprender vocabulário é um desafio para os aprendizes, principalmente porque é uma tarefa de enormes proporções. Visto que o léxico de uma língua é um sistema aberto que sofre constantes modificações, há uma variedade muito grande de palavras e combinações de palavras a serem aprendidas, como colocações, vocabulário estratégico, bem como padrões gramaticais, expressões idiomáticas e expressões fixas.

Conforme destacado por McCarten (op.cit.), Richards (1976) e Nation (2001) listam algumas informações que os aprendizes precisam saber antes que possam dizer que realmente aprenderam uma palavra, são elas: o(s) significado(s) da palavra, a sua forma falada (pronúncia) e escrita, os componentes da palavra (por exemplo, prefixo, sufixo, radical), comportamento gramatical (por exemplo, classe da palavra, padrões gramaticais típicos), colocações, registro, associações (por exemplo, palavras com significado similar ou oposto), conotações da palavra e frequência.

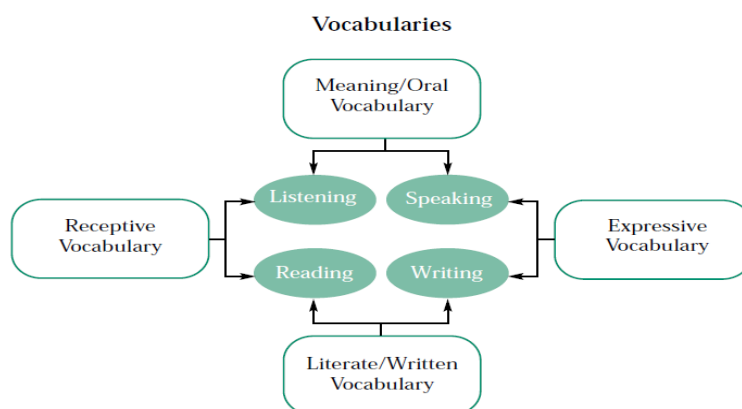
De forma similar, Thornbury (2006) afirma que:

Tipicamente [lições de vocabulário] são conduzidas por meio do ensino de conjuntos lexicais, ou seja, grupos de palavras tematicamente relacionadas) usando uma variedade de meios, incluindo auxílio visual, demonstração, situações, textos, e atividades que demandam o uso de dicionário. Não só o significado das palavras, como também a forma, tanto falada (pronúncia) quanto escrita (ortografia), precisam ser trabalhados, principalmente se as palavras estão sendo ensinadas visando à produção linguística. Outros aspectos referentes ao conhecimento das palavras que precisam ser destacados incluem a conotação e o estilo, colocação, formas derivadas e características gramaticais, como as classes de palavras. Vocabulário também é ensinado como uma preparação para compreensão oral e leitura (*pre-teaching vocabulary*) ou como um subproduto dessas habilidades³ (THORNBURY, 2006, p. xxxiv; tradução minha).

³ Texto original: Typically these take the form of teaching *lexical sets* of words (i.e. groups of thematically linked words) using a variety of means, including visual **aids**, demonstration, situations, texts, and dictionary work. As well as the **meaning** of the items, the **form**, both spoken (i.e. **pronunciation**), and written (i.e. **spelling**), needs to be dealt

Outra questão a se considerar no ensino de vocabulário, como Thornbury (2006) menciona no excerto supracitado, refere-se ao propósito de ensiná-lo, ou seja, quando ensinamos vocabulário é preciso pensar em nossos objetivos, se queremos que os alunos utilizem o vocabulário para falar ou escrever ou para facilitar a compreensão de um texto, por exemplo. Nesse sentido, faz-se a distinção entre vocabulário receptivo e vocabulário ativo. Pikulski e Templeton (2004) fazem uma distinção em relação aos vocabulários que constituem o repertório linguístico de um falante. Cada tipo de vocabulário está relacionado a uma das habilidades linguísticas como ilustra a Figura 1:

Figura 1: Integração dos vocabulários às diferentes habilidades linguísticas



Depreende-se do esquema acima que o vocabulário é uma peça fundamental na aprendizagem de uma língua, estando presente e integrado a todas as habilidades. Nesse sentido, o vocabulário expressivo refere-se aos itens lexicais que o falante usa produtivamente, tanto na modalidade escrita quanto na falada. O vocabulário receptivo é aquele que o falante se utiliza para compreender textos, sejam eles escritos ou falados.

Dada a complexidade da empreitada de se aprender o vocabulário de uma língua, é importante que o professor trabalhe com os alunos de forma a auxiliá-los nessa aprendizagem. Ainda segundo McCarten (2007), materiais didáticos podem auxiliar os aprendizes em duas áreas:

Primeiro, eles [os materiais] precisam apresentar e praticar em contextos naturais o vocabulário que é frequente e apropriado para as necessidades dos aprendizes. Segundo, materiais didáticos devem ajudar os estudantes a se tornarem melhores aprendizes de vocabulário, ensinando diferentes técnicas e estratégias que eles possam usar para continuar aprendendo fora da sala de aula⁴ (MCCARTEN, 2007, p.19-20; tradução minha).

Percebe-se então que materiais didáticos têm dupla função em relação ao ensino de vocabulário, tendo que, não só apresentar itens lexicais em contextos naturais e apropriados de uso, como também

with, especially if the words are being taught for productive use. Other aspects of word knowledge that may need to be highlighted include **connotation** and **style**, **collocation**, derived forms, and grammatical features, such as the word's **word class**. Vocabulary is also taught as preparation for listening or reading (*pre-teaching vocabulary*) or as a by-product of these skills (THORBURY, 2006, p. xxxiv).

⁴ Texto original: First, they need to present and practice in natural contexts the vocabulary that is frequent, current, and appropriate to learner's needs. Second, materials should help students become better learners of vocabulary by teaching different techniques and strategies they can use to continue learning outside the classroom (MACCARTEN, 2007, p. 19-20).

fornecer estratégias e meios para que o aluno desenvolva a sua autonomia, para que possa dar continuidade ao aprendizado de vocabulário fora da sala de aula.

Nesse sentido, Thornbury (2006) afirma que seria impossível ensinar em sala de aula todas as palavras que os aprendizes precisam. Desse modo, eles necessitam de oportunidades para que haja a aprendizagem accidental de vocabulário fora de sala de aula, que pode ser estimulada por meio do que se chama de leitura extensiva (*extensive reading*).

Tendo essa discussão em vista nos perguntamos: o que um *corpus* teria a nos dizer em relação ao vocabulário de uma língua?

Segundo McCarten (2007) um *corpus* pode fornecer as seguintes informações sobre vocabulário:

Frequência: palavras e expressões mais frequentes e raras da língua;
Diferenças na fala e na escrita: palavras mais comuns na fala e na escrita;
Contextos de uso: as situações nas quais as pessoas usam determinadas palavras;
Colocações: palavras que frequentemente ocorrem juntas;
Padrões gramaticais: como as palavras se combinam em frases na formação de padrões;
Uso estratégico de vocabulário: palavras e expressões usadas para organizar o discurso⁵ (MCCARTEN, 2007, p. 3; tradução minha).

Desse modo, por meio da observação de um *corpus* eletrônico é possível descrever as línguas em vários aspectos. Owen (1993 apud RAJAGOPALAN, 2006, p. 161) afirma que a LC se apresenta como “uma nova abordagem da descrição gramatical, que penetra nas partes da linguagem inacessíveis a outras gramáticas”. Dessa forma, achamos mais adequado referir-se ao nível léxico-gramatical de uma língua, visto que a LC desfaz a crença de que há dois níveis independentes de organização da linguagem (sintaxe e léxico).

É inegável que a Linguística de *Corpus* tem trazido contribuições importantes para os estudos da linguagem. No ensino, a influência dos *corpora* pode ser resumida em quatro grandes áreas de investigação: (a) descrição da linguagem nativa; (b) descrição da linguagem do aprendiz; (c) transposição de pesquisa acadêmica para a sala de aula, (d) desenvolvimento de materiais de ensino, currículos e abordagens (BERBER SARDINHA, 2004). Segundo Berber Sardinha (2004):

A influência da Linguística de Corpus no ensino ocorre de modo indireto, apenas quando os resultados da pesquisa são absorvidos e incorporados, em geral parcialmente, nos materiais de ensino. A entrada no ambiente pedagógico ocorre, primordialmente, pelos livros didáticos e muito pouco por intermédio do professor. Ainda é raro, em sala de aula, o professor usar a descrição da linguagem baseada em corpus, e a razão principal é o distanciamento entre o profissional de ensino e a academia (BERBER SARDINHA, 2004, p.255).

Segundo Berber Sardinha (op. cit.) há duas controvérsias em relação ao uso de *corpora* no ensino. A primeira refere-se ao espaço no ensino de línguas para o uso de *corpus* e da descrição proveniente dele. Alguns linguistas de *corpus* defendem que a utilização de *corpora* no ensino deve ser direta, ou seja, a descrição feita pelos linguistas de *corpus* é incorporada diretamente pelos professores, os quais devem respeitar o princípio de usar somente linguagem autêntica. “Apresente somente exemplos reais. Esta talvez seja a primeira lição a ser aprendida por meio do estudo de um *corpus*. A linguagem não pode ser

⁵ Texto original: **Frequency:** Which words and expressions are most frequent and which are rare; **Differences in speaking and writing:** Which vocabulary is more often spoken and which is more often written; **Contexts of use:** The situations in which people use certain vocabulary; **Collocation:** Which words are often used together; **Grammatical patterns:** How words and grammar combine to form patterns; **Strategic use of vocabulary:** Which words and expressions are used to organize and manage discourse (MACARTEN, 2007, p. 3).

inventada; ela pode ser apenas capturada”⁶ (SINCLAIR, 1988 apud BERBER SARDINHA, 2004, p. 256, tradução minha).

Henry Widdowson critica essa posição dizendo que o ensino e a descrição de língua são atividades independentes ainda que relacionadas, e assim “não só as duas práticas são distintas, mas para ele, a descrição proporcionada pelo lingüista não deve ter precedência à pedagogia levada a cabo pelo lingüista aplicado” (BERBER SARDINHA, 2004, p. 256). Segundo essa posição, a descrição fornece fatos baseados em dados que não apresentam garantia de relevância pedagógica.

A segunda controvérsia seria a de que o contato entre o *corpus* e o aluno deveria ser mediado. Nesse sentido, questiona-se “se o aluno deve ter acesso ao *corpus* ou concordâncias brutas ou se o acesso deve ser simplificado” (op. cit., p. 257).

Nesse sentido, Berber Sardinha destaca que,

aquilo que o professor quiser ensinar não precisa ser relacionado ao que o lingüista de *corpus* já descreveu. O professor tem liberdade para fazer suas próprias análises, enfocando o que achar mais relevante e apresentar os resultados como achar mais apropriado, filtrando e simplificando-os, se necessário (BERBER SARDINHA, 2004, p. 257).

A partir dessas considerações, o livro didático se constitui como um instrumento de mediação entre a descrição linguística proveniente dos *corpora* e o aluno. E é nessa relação que o professor será responsável por tomar as decisões que forem mais adequadas.

METODOLOGIA

A partir das perguntas de investigação propostas (Como a Linguística de *Corpus* pode contribuir para a elaboração de materiais didáticos e o ensino de vocabulário? Que diferenças significativas para o ensino de línguas podemos encontrar ao compararmos materiais didáticos elaborados a partir dos estudos de *corpora* com materiais convencionais?) seguimos algumas etapas para sistematizar e organizar o processo de condução da coleta e análise dos dados.

Para tanto, realizamos leituras e resenhas dos referencias teóricos para um embasamento da pesquisa. A escolha do material analisado se deu pautada nos seguintes motivos: o livro *New American Inside Out* (MD 2, não baseado em *corpus*) é o livro adotado em um instituto de idiomas⁷ conceituado na cidade de Uberlândia, e o livro *Touchstone 3* (MD 1, baseado em *corpus*) também é utilizado em outra escola de Uberlândia. O primeiro livro mencionado é classificado pelos seus autores como *Pre-Intermediate* e o segundo como *Low-Intermediate* os quais, segundo Harmer (2001) são níveis equivalentes. Priorizamos o mesmo nível, a fim de que a comparação pudesse ser feita de forma a revelar as diferenças em níveis iguais.

Em uma análise qualitativa, observamos as unidades dos livros manualmente, ou seja, não compilamos um *corpus* eletrônico dos livros para ser analisado por um *software* de análise linguística. Nessa análise observamos o modo como exemplos e explicações léxico-gramaticais são apresentados e o modo de apresentação e prática de exercícios de vocabulário. A seguir, apresentamos a análise feita e a discussão dos resultados.

⁶ Texto original: Present real examples only. This is perhaps the first lesson to be learned from corpus study. Language cannot be invented; it can only be captured (SINCLAIR, 1988 apud BERBER SARDINHA, 2004, p. 256).

⁷ Os nomes dos institutos de idiomas não serão mencionados, a fim de preservar a identidade dos mesmos.

ANÁLISE, RESULTADOS E DISCUSSÃO

Nessa seção serão contempladas algumas considerações feitas em relação a uma análise geral do material sem nos atermos a uma unidade específica dos livros mencionados anteriormente. Salientamos também que, as considerações apresentadas a seguir referem-se a trechos específicos dos materiais didáticos em análise, não representando a totalidade dos mesmos. Mas antes de apresentarmos a análise vale apresentar algumas considerações em relação aos materiais analisados.

Primeiramente, o MD 1 é uma coleção de quatro livros, os quais, segundo os autores são informados pelo *Cambridge International Corpus* (CIC).

*O Cambridge International Corpus é uma base de dados de mais de 700 milhões de palavras nas modalidades faladas e escritas em inglês que foram coletadas a partir de uma variedade de fontes autênticas. Desenvolvido pela Cambridge University Press como uma ferramenta de auxílio na elaboração de materiais para aprendizes de inglês, o Corpus pode ser analisado para determinar como a língua é realmente usada*⁸ (texto apresentado na contracapa do MD 1; tradução minha).

O MD 2, por sua vez, é uma coleção de seis livros, os quais representam seis níveis respectivamente, *beginner, elementary, pre-intermediate, intermediate, upper-intermediate e advanced*. Não há nenhuma menção ao fato do livro ser informado ou não por *corpora* o que entendemos como sendo não informado.

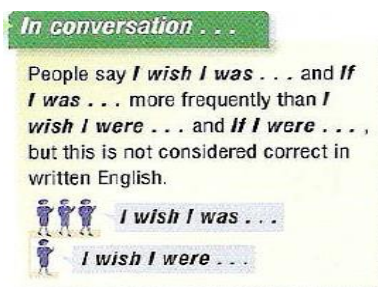
De modo geral, o MD 2 apresenta o seu conteúdo distribuído em 12 unidades. Cada unidade é composta por diferentes seções denominadas de acordo com o conteúdo focado: *Listening, Vocabulary, Reading, Grammar, Pronunciation, Speaking: anecdote*, ou a combinação desses conteúdos. Ao final de cada unidade há uma página destinada ao que é chamado de *Useful Phrases* e uma chamada de *Vocabulary Extra*.

Assim como o MD 2, o MD 1 apresenta o seu conteúdo em 12 unidades. Cada unidade é subdividida em quatro lições (A, B, C, D). Cada lição é também subdividida de forma variada sem um padrão aparente. As subdivisões das lições são semelhantes às das unidades do MD 2 com algumas variações: *Getting Started, Speaking naturally, Building vocabulary and Grammar, Conversation strategy, Strategy plus, Talk about it, Free talk*. E ao final de cada unidade há uma página chamada *Vocabulary Notebook*.

Os títulos das seções nos permite afirmar que o MD 1 apresenta maior variedade na organização dos conteúdos. Além disso, é possível observar que o MD 1 oferece mais informações sobre a língua do que o MD 2. Tais informações assumem a forma de pequenos quadros intitulados *In Conversation* como ilustra a figura seguinte:

⁸ Texto original: The Cambridge International Corpus is a database of over 700 million words of spoken and written English that have been collected from a variety of authentic sources. Developed by Cambridge University Press as a tool to help in writing materials for learners of English, the Corpus can be analyzed to determine how the language is actually used (Texto apresentado na contracapa do MD 1).

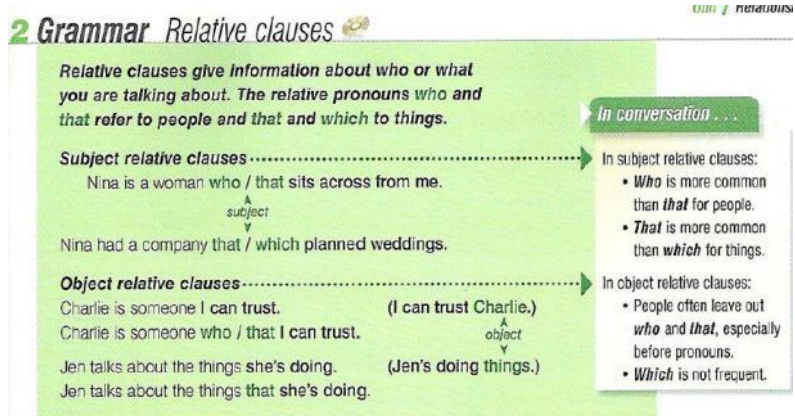
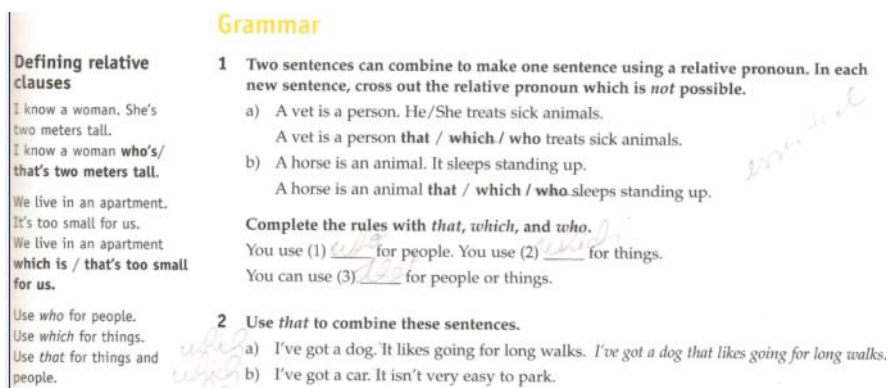
Figura 2: Excerto do MD 1, p. 77



Percebe-se que nesse exemplo é feita uma distinção dos usos da língua em relação à fala e à escrita. O MD 2 não traz esse tipo de informação e nem mesmo apresenta esse conteúdo. Desse modo, os livros são referentes ao mesmo nível de aprendizagem (*pre-intermediate* ou *low-intermediate*), mas mesmo assim os conteúdos divergem. Talvez, uma possível explicação para essa divergência seja a de que a partir do *corpus* esse tipo de estrutura tenha sido considerado frequente o suficiente para ser inserido nesse nível de aprendizagem. Mesmo assim, uma busca mais detalhada por uma explicação para essa divergência está além do escopo da presente investigação.

Observando a apresentação de conteúdos similares, fizemos o seguinte recorte:

Figura 3: Excerto de MD 2, p.99 e MD 1, p. 67, respectivamente



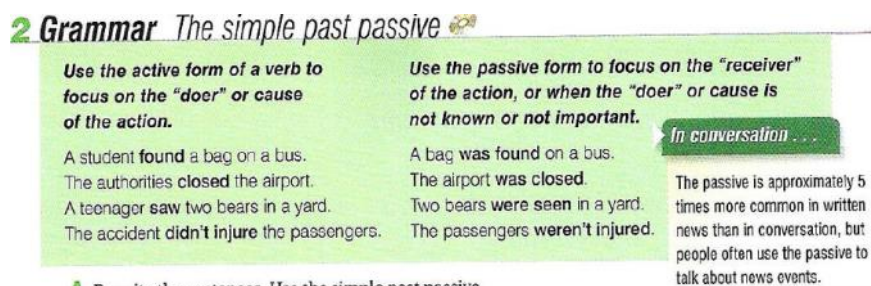
Nos exemplos acima observamos como foi realizada a explicação do uso dos pronomes relativos *who*, *that* e *which*. No primeiro exemplo, do MD 2, vemos alguns exemplos na margem da página seguidos pela regra gramatical. No MD 1 também temos exemplos e a mesma regra gramatical: *who* e *that* se usa para pessoas e *that* e *which* se usa para coisas. A diferença está no fato de que o MD 1 vai além da mera regra gramatical. À direita, no quadro *In conversation* apresentam-se informações relevantes para o uso. Assim, percebe-se que no MD 1 há maior preocupação com o uso desses itens lexicais em um contexto autêntico de comunicação, enquanto no MD 2 não. Portanto, percebemos que no MD 2 a explicação gramatical se esgota no nível da abstração de regras. O MD 1, por sua vez, vai além disso, fornecendo informações de uso, o que certamente auxilia o aprendiz na adequação de sua produção linguística.

Figura 4: Excertos do MD 1, p. 55 e p. 5, respectivamente



Nos exemplos acima, vemos informações relacionadas à frequência do uso dos itens *should*, *ought to* e *had better*. Esses três itens por terem uma função semelhante (dar um conselho ou recomendação) poderiam causar confusão no momento de usá-los. Nesse caso, percebe-se que, na fala, *should* é mais comum do que as outras formas. No exemplo ao lado, temos a comparação entre *really* e *pretty* que são muito mais usados na fala do que na escrita. Mais uma vez, observamos a explicitação de diferenças entre fala e escrita na explicação de um conteúdo gramatical conforme a Figura 5:

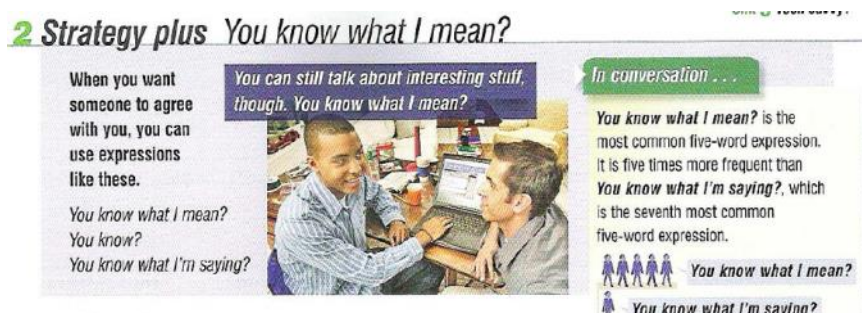
Figura 5: Excerto do MD 1, p. 119



Na figura abaixo, observamos a explicação do uso estratégico de vocabulário usado na organização do discurso por meio da expressão *you know what I mean?* É possível perceber que há uma

preocupação por parte do material em relação ao tipo de vocabulário que será mais útil e apropriado para ser usado em uma interação, a fim de se organizar o discurso típico da fala.

Figura 6: Excerto do MD 1, p. 91



McCarten (2007) afirma que conjuntos lexicais como *I mean*, *I don't know* e *or something* são muito mais frequentes do que *woman*, *six* e *black*. No entanto, essas palavras simples são mais encontradas em cursos básicos, mesmo sendo menos frequentes. O quadro seguinte apresenta as expressões mais frequentes de conversas norte-americanas de listas do *Cambridge International Corpus*:

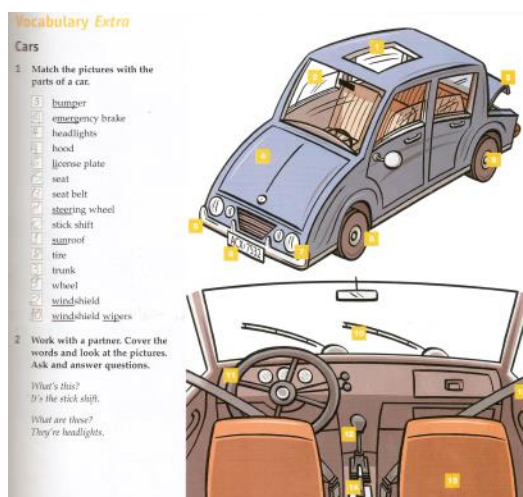
Figura 7: Expressões das listas de frequência do *Cambridge International Corpus*

No. of words in phrase	Examples
two	<i>you know, I mean, I guess, or something</i>
three	<i>a little bit, and all that</i>
four	<i>or something like that, and things like that</i>
five	<i>you know what I mean, as a matter of fact</i>
six	<i>it was nice talking to you; and all that kind of stuff</i>
seven+ words	<i>a lot of it has to do (with) . . .</i>

O ensino das expressões acima é importante, pois elas auxiliam os alunos a soarem mais fluentes no idioma, porém, em muitos casos, elas recebem pouca atenção nas aulas de língua inglesa e nos materiais didáticos. O MD 2, por exemplo, não explora esse tipo de uso do vocabulário estratégico.

Na figura seguinte, observamos uma tendência do MD 2, que é a de apresentar vocabulário dentro de um determinado campo semântico. Neste exemplo específico, vemos as palavras que se referem aos componentes de um carro. Apesar de ser um conjunto lexical interessante de se ter conhecimento, acreditamos que esse tipo de vocabulário seja de pouca relevância para os alunos em um nível pré-intermediário, pelo menos para fins de expressão. Não temos dados para comprovar isso, mas podemos supor que as unidades lexicais que nomeiam as partes de um carro são itens pouco frequentes em um *corpus* geral de língua, tendo pouca, ou nenhuma relevância para o nível em questão. É claro que a relevância ou não de determinado vocabulário depende dos objetivos do aprendiz. Por isso, não podemos ser categóricos ao afirmar se dado conjunto vocabular é relevante ou não para o aprendiz.

Figura 8: Excerto do MD 2, p. 95

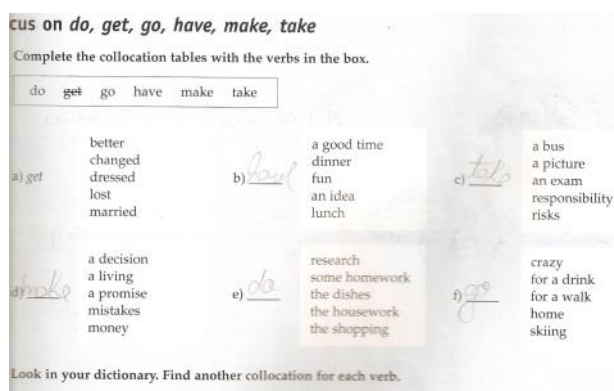


Quando analisamos o tipo de exercício proposto logo a seguir, vemos como as palavras foram trabalhadas de modo irrelevante para o aluno por meio de perguntas como, *What's this?* e *What are those?* Esse é o tipo de pergunta que dificilmente os alunos seriam questionados em uma situação autêntica. O que se priorizou nas atividades foi a memorização por meio de auxílio visual e uso de perguntas. A nosso ver seria muito mais relevante o ensino do uso estratégico de vocabulário, que, aliás, é muito mais frequente e fornece subsídios para que os alunos estejam cientes das estratégias que podem facilitar a interação e a expressão de significados em sua produção oral.

A apresentação de expressões idiomáticas no MD 2 é feita por meio de diálogos curtos entre personagens fictícios. Não podemos afirmar com certeza, mas visto que esse material não parte de uma análise de *corpus*, supomos que esses diálogos sejam tão inventados quanto as personagens.

A seguir temos um exemplo de como as colocações⁹ foram trabalhadas no MD 2:

Figura 9: Excerto de exercícios de colocação no MD 2, p. 111.



Percebe-se que o exercício proposto visa apenas a relacionar apropriadamente um verbo com um respectivo grupo de colocados¹⁰. Não há nenhum espaço para que o aluno possa personalizar a atividade atribuindo significados às colocações. Elas são apresentadas completamente fora de um contexto.

⁹ De acordo com Tagnin (2005, p. 37) “o termo *collocation* [colocação] foi introduzido pelo linguista britânico J. R. Firth para designar casos de co-ocorrência léxico-sintática, ou seja, palavras que usualmente ‘andam juntas’”.

Figura 10: Excerto de um exercício sobre colocações no MD 1, p. 52

Learning tip Collocations – words that go together
When you learn new words, learn them in combination with other words. For example, learn adjectives that can go before a particular noun.
boiled eggs, fried eggs, raw eggs

Talk about eating
The most common collocations conversation with these 6 cooking words are:
1. *fried* chicken 4. *grilled* chicken
2. *boiled* eggs 5. *smoked* fish
3. *baked* potatoes 6. *raw* fish

1 Cross out the adjective that doesn't go well with each noun.

a. fried } onions
pickled }
~~smoked~~ }
raw }

b. barbecued } rice
steamed }
fried }
boiled }

c. stir-fried } noodles
spicy }
boiled }
pickled }

d. smoked }
fresh }
canned }
dried }

2 How many cooking or taste words can you put before these foods? List them from your favorite to your least favorite ways of eating them.

least favorite	----->	favorite
grilled		fish
		chicken
		eggs
		red peppers
		pineapple
		carrots

Já na figura acima, percebe-se que além de se relacionar uma palavra-chave com os seus colocados, o exercício dois permite ao aluno fazer uso das colocações de forma que estas signifiquem para ele como uma comida favorita e uma menos favorita. Assim, as colocações assumem uma relevância mais significativa para os alunos, uma vez que eles a relacionarão a um gosto pessoal.

Tendo feito um breve comentário em relação às diferenças observadas em alguns traços linguísticos, observamos também se os materiais se preocupam com o *como* aprender vocabulário. Nesse sentido, o MD 2 não apresenta nenhum tipo de contribuição para que o aluno desenvolva estratégias de aprendizagem. Não há nenhuma menção a como aprender melhor determinado traço linguístico. Já o MD 1 apresenta várias dicas, como as seguintes:

Figura 11: Excertos do MD 1, P. 126 e P. 20 respectivamente, sobre dicas de aprendizagem

Learning tip Collocations
When you learn a new word, use a dictionary to find out what other words are typically used with it. For example, you can say *flash floods*, but not usually *quick floods*. Or you can say *seriously injured*, but not usually *completely injured*.

Learning tip Linking new words to your experiences
When you learn a new word or expression, link it to something you have recently seen or done.

Essas dicas são encontradas ao final de cada unidade em uma página chamada de *vocabular notebook*, que não só objetiva expandir o vocabulário dos aprendizes como também fornecer estratégias para melhor aprender.

De um modo geral, percebemos que no md 2, a apresentação de vocabulário é feita de modo isolado, sem a inserção do vocabulário em um contexto autêntico, seja escrito ou falado. Por outro lado, no md 1, ao longo de todo o livro, o vocabulário apresentado está integrado à determinada habilidade linguística dentro de um contexto possível e provável de ser vivenciado fora da sala de aula

¹⁰ *Colocado* é o termo utilizado para designar as palavras que se combinam em colocações. Tendo uma palavra-chave como referência, as palavras que se combinam frequentemente com ela são os seus colocados.

Tendo feito essas considerações, podemos dizer que o uso de *corpora* no ensino tem contribuído sobremaneira para que a linguagem seja descrita para os aprendizes em toda a sua multiplicidade de usos. Por meio do livro didático informado por *corpus*, o aprendiz terá acesso a amostras autênticas da língua, e uma riqueza de informações que o auxiliarão na adequação de sua produção linguística, bem como na compreensão de textos autênticos.

Um ponto importante em relação ao uso da língua autêntica fornecida pelos *corpora* é o fato de que a posição do falante nativo como fonte de língua autêntica é relativizada, ou seja, não é necessário que o aluno viaje para o exterior ou tenha contato com um falante nativo para estudar a língua como ela é usada em contextos reais de comunicação.

CONSIDERAÇÕES FINAIS

Ao realizar esta pesquisa pretendemos investigar como os *corpora* são usados na elaboração de materiais didáticos e principalmente a influência que eles exercem no ensino de vocabulário. Para isso, comparamos dois materiais didáticos, um informado por *corpus* e outro não informado por *corpus*. Ao longo da investigação tornou-se evidente que materiais baseados em *corpus* são muito mais ricos e fornecem ao aprendiz mais informações em relação ao modo como a língua é usada em situações autênticas. Assim, compartilhamos da opinião de McCarthy (2004) de que materiais informados por *corpus* são genuinamente especiais porque:

- São baseados no uso real;
- Os exemplos, apesar de que às vezes são editados ou adaptados, são um reflexo do uso real; eles não são inventados;
- O conteúdo do curso (tanto os itens a ser ensinados quanto a sequência na qual são apresentados) é informado pela frequência: por exemplo, pode-se priorizar a gramática e o vocabulário que é mais frequente e mais útil;
- Os contextos em que as palavras e as estruturas gramaticais ocorrem são autênticos, baseados nos contextos que ocorrem em *corpora*;
- A apresentação do conteúdo e as atividades podem focar em diferenças importantes entre a língua falada e escrita;
- Os materiais podem incluir traços linguísticos que foram ignorados ou despercebidos no passado, mas que têm importância central na comunicação. Por exemplo, quando dizemos indiretamente as palavras de outras pessoas, nós dizemos *Mary said* ou *Daniel told me*, mas o *corpus* mostra que frequentemente dizemos *Mary was saying* ou *Daniel was telling me*, principalmente quando temos informações interessantes para passar. A formação *was + ing* para a discurso indireto foi despercebida no passado mesmo em grandes gramáticas;
- *Corpora* especializados podem ser analisados para atender às necessidades de um grupo específico de aprendizes. Por exemplo, podemos usar um *corpus* acadêmico coletado em contextos universitários para ajudar aprendizes que irão estudar fora, ou um *corpus* sobre negócios para elaborar materiais para empresários e administradores que precisam trabalhar com uma segunda língua;
- Os autores de materiais informados por *corpus* podem antecipar erros mais comuns observando *corpora* de aprendizes que apresentam línguas maternas distintas;
- Estudantes não precisam viver no ambiente da língua alvo para ter contato com linguagem autêntica – ela está bem aqui, em seus livros didáticos e dicionários¹¹ (MCCARTHY, 2004, p. 15-16; tradução minha).

¹¹ Texto original: They are based on actual usage. The examples used in them, although they may sometimes be edited or adapted, are a reflection of real usage; they are not invented. The syllabus (the items to be taught as well as the sequence in which they will be presented) is informed by frequency information: For instance, we can prioritize grammar and vocabulary that is most frequent and most useful. The contexts in which words and grammar structures are used are authentic ones, based on the contexts that occur in corpora. The presentation and activities can focus on

É importante lembrar que é muito improvável que os professores consigam cobrir em sala de aula o grande número de palavras que os estudantes precisarão para produzir ou compreender uma língua estrangeira, mesmo com o auxílio dos *corpora*. Nesse sentido, não basta apenas auxiliar os alunos em relação ao conteúdo a ser aprendido, mas também em *como* aprender vocabulário. Nesse sentido, também observamos que o material analisado baseado em *corpus* demonstrou interesse em como os alunos aprendem, fornecendo dicas e estratégias que poderão ser usadas em seu aprendizado fora da sala de aula, o que contribui para o desenvolvimento e a consolidação de uma aprendizagem calcada na autonomia.

Gostaríamos de finalizar frisando que o ensino de vocabulário, apesar do auxílio dos *corpora*, não se esgota e não pode ser simplesmente restrito a eles. Nas palavras de McCarten (2007):

[um corpus] não pode nos dizer exatamente o que ensinar ou como ensinar, e não tem nada a nos dizer sobre como os estudantes aprendem melhor. ele não pode substituir a expertise dos professores, ou dos próprios estudantes, sobre como ensinar e aprender melhor vocabulário. ele é *uma* ferramenta. não é a *única*¹² (MCCARTEN, 2007, p.3, tradução minha).

REFERÊNCIAS

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

BERBER SARDINHA, T. Como usar a linguística de corpus no ensino de língua estrangeira: por uma linguística de corpus educacional brasileira. In: VIANA, V.; TAGNIN, S.E.O. (Org.) **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010.

CARTER, R.; MCCARTHY, M.; SANDIFORD, H. **Touchstone 3**. Cambridge: Cambridge University Press, 2006.

HARMER, J. **The practice of English language teaching**. England: Longman, 2001.

JONES, V.; KAY, S. **New American Inside Out Student's Book Pre-Intermediate**. Macmillan Publishers Ltd., 2009.

MCCARTEN, J. **Teaching Vocabulary: lessons from the corpus, lessons for the classroom**. Cambridge: Cambridge University Press, 2007.

MCCARTHY, M. **Touchstone: from corpus to course book**. Cambridge: Cambridge University Press, 2004.

the important differences between spoken and written language. The materials can include language that was ignored or not noticed in the past but that is at the heart of real communication. For example, when we report other people's words, we often say *Mary said* or *Daniel told me*, but the corpus shows we also frequently say *Mary was saying* or *Daniel was telling me*, especially when we have some interesting news to pass on. The *was + -ing* construction for reporting speech has gone unnoticed in the past, even in comprehensive grammar reference books. Specialized corpora can be analyzed to meet the needs of particular groups of learners. For example, we can use an academic corpus collected in university and college contexts to help learners who are going to study abroad, or a business corpus to construct materials for businesspeople who need to work in a second language. The writers of corpus-informed materials can anticipate common errors by looking at corpora of learners' work from a wide variety of language backgrounds. Students don't have to live in the target language environment to experience authentic language – it's right there, in their course books and dictionaries (MCCARTHY, 2004, p. 15-16).

¹² Texto original: [a corpus] cannot tell us exactly what to teach or how to teach, and it has nothing to tell us with respect to how students learn best. It cannot replace the expertise of teachers, or of students themselves, on how best to teach and learn vocabulary. It is *a* tool. It is not the *only* tool (MACARTEN, 2007, p. 3).

NUNAN, D. From the traditional to the contemporary in second language teaching and learning. In: _____. **Second language teaching & learning**. Boston: Heinle & Heinle Publishers, 1999.

PIKULSKI, J. J.; TEMPLETON, S. Teaching and Developing Vocabulary: key to long-term reading success. In: **Houghton Mifflin Reading**, 2004.

RAJAGOPALAN, K. Repensar o papel da lingüística aplicada. In: MOITA LOPES, L. P. da. **Por uma lingüística aplicada indisciplinar**. São Paulo: Parábola Editorial, 2006, p.151-168.

TAGNIN, S. E. O. **O jeito que a gente diz**. São Paulo: Disal, 2005.

THORNBURY, S. Excerpts from: An A-Z of ELT (2006). In: KAY, S.; JONES, V.; GOMM, H. **New American Inside Out Teacher's Edition Intermediate**. Macmillan Publishers Ltd., 2009.

