

ENSINO DE INGLÊS COMO “LE” E CONTRIBUIÇÕES PEDAGÓGICAS DE UM GLOSSÁRIO BILÍNGUE DE COLOCAÇÕES

ADRIANE ORENHA OTTAIANO*
JEAN MICHEL PIMENTEL ROCHA**

RESUMO

Em conformidade com o arcabouço teórico-metodológico da Linguística de Corpus e da Fraseologia, este estudo visa discutir as contribuições pedagógicas de um glossário bilíngue de colocações da área médica para o ensino de inglês como Língua Estrangeira (doravante LE), compilado a partir do *corpus* paralelo constituído pelas legendas em inglês e em português do seriado *House M.D.* Com base nas colocações presentes no glossário, elaboramos propostas de atividades que poderão servir como subsídio para o ensino de inglês como LE, principalmente, quando se considera a relevância dessas unidades fraseológicas no enriquecimento da competência lexical de aprendizes. Além destas, também propomos outras atividades colocacionais baseadas em *corpus*.

PALAVRAS-CHAVE: Linguística de Corpus, colocações, Fraseologia, glossário bilíngue, ensino de colocações, aplicações pedagógicas.

1 INTRODUÇÃO

Ao considerar que o léxico mental de um indivíduo é muito mais amplo do que qualquer lista de palavras, sendo, desse modo, constituído de blocos pré-fabricados que, quando empregados, produzem textos coerentes e naturais (LEWIS, 2000; SINCLAIR, 1991), faz-se necessário auxiliar os aprendizes a desenvolver habilidades de observação do

* Professora Assistente Doutora do Departamento de Letras Modernas, da Universidade Júlio de Mesquita Filho (UNESP), Câmpus de São José do Rio Preto, SP, Brasil. Atua na Pós-Graduação em “Estudos Linguísticos” da UNESP/IBILCE. Email: adriane@ibilce.unesp.br.

** Discente do Programa de Pós-Graduação em Estudos Linguísticos (PPGEL) da Universidade Júlio de Mesquita Filho (UNESP), Câmpus de São José do Rio Preto, SP. Email: jeanpimentel_sp@yahoo.com.br.

potencial combinatório das palavras. Apesar de conhecerem um grande número de vocábulos isolados, os aprendizes de LE, de modo geral, não se atentam para as palavras com as quais esses vocábulos coocorrem, ou seja, para as palavras que os acompanham (FIRTH, 1957). Desse modo, neste trabalho, tendo como meta a expansão e o enriquecimento da competência lexical de aprendizes de LE, principalmente no que concerne à aprendizagem do fenômeno das colocações, objetivamos a aplicação pedagógica de um glossário bilíngue de colocações da língua geral e especializadas, ao propor atividades baseadas em *corpora*, a partir das colocações observadas no glossário, bem como outras atividades colocacionais também baseadas em *corpus*.

Para conduzir esta pesquisa, fomos amparados pelo arcabouço teórico-metodológico da Linguística de Corpus e da Fraseologia, áreas que têm contribuído fortemente não apenas para os estudos de descrição linguística baseados em *corpora*, mas também para os estudos da pedagogia do léxico, ao prezar pelo emprego da linguagem autêntica e contextualizada no ensino.

2 LINGUÍSTICA DE CORPUS E *CORPORA* NO ENSINO E APRENDIZAGEM DE LE

A Linguística de Corpus estuda e descreve a língua com base em exemplos reais de uso, evidenciados a partir da exploração de *corpora* eletrônicos, com auxílio do computador e de ferramentas computacionais (BERBER SARDINHA, 2004).

O advento do computador e o desenvolvimento de *softwares* para análise lexical de dados, como o *WordSmith Tools* (SCOTT, 2004), por exemplo, foram determinantes para o fortalecimento e a expansão da Linguística de Corpus, fazendo com que esta obtivesse projeção e prestígio no cenário das pesquisas linguísticas. Graças a essas tecnologias, a linguística de *corpus* (doravante LC) passou de um pseudo-procedimento a uma metodologia viável no estudo da língua (MCENERY; WILSON, 2001). As ferramentas computacionais permitiram a redução do trabalho árduo de compilação manual de grandes *corpora* para análise, permitiram ainda acesso a uma grande quantidade de dados em tempo recorde, com confiabilidade e precisão

estatística – o *Corpus of Contemporary American English*, por exemplo, possui 450 milhões de palavras.

Embora a utilização do termo LC para designar o estudo da língua baseado em exemplos reais de uso tenha surgido há apenas algumas décadas, os estudos com *corpora* já vinham sendo conduzidos desde o final do século XIX e início do século XX. De acordo com McEnery e Wilson (2001), tais estudos estão voltados para a aquisição da linguagem (STERN, 1924); a pedagogia da língua (THORNDIKE, 1921; PALMER, 1933, entre outros) etc. Nos Estados Unidos, a metodologia de pesquisa dominante nos estudos linguísticos se apoiava na observação da língua em uso. Os estudiosos desta época, a saber, linguistas da tradição estruturalista americana representada, principalmente, por Boas (1940), não se identificavam como linguistas de *corpus*, porém, se valiam de uma metodologia que certamente pode ser considerada baseada em *corpus*.

Na década de 1950, essa metodologia foi severamente criticada pela chamada revolução chomskyana. Ao postular o objeto de estudo da linguística como a competência linguística do falante, Chomsky reabriu um debate sobre a natureza dos dados no estudo da linguagem e retomou a oposição “racionalistas” e “empiricistas”. Os primeiros esposavam teorias que tomavam como base a introspecção para explicar o funcionamento da língua e, os segundos, teorias que se baseavam na observação dos dados para explicar o funcionamento das estruturas linguísticas. Dessa forma, Chomsky sugeriu que um *corpus* não seria uma ferramenta útil para o linguista, já que a linguística deveria se preocupar em estudar a competência, isto é, o conhecimento internalizado que os falantes têm da língua, e não a *performance*. Segundo ele, o *corpus* caracterizaria a *performance*, ou seja, a fala, que, permeada por influências externas ao sistema, não seria um guia útil para descrição da competência linguística do falante. Além disso, o *corpus*, dada sua finitude, não seria capaz de representar adequadamente a língua em sua infinidade; ele poderia dar conta de algumas, mas não de todas as sentenças válidas da língua. Outra crítica dizia respeito ao questionamento da adequação da metodologia baseada em *corpus*, sob a justificativa de que não era necessário se preocupar em buscar em um *corpus*, aquilo que poderia ser encontrado na mente e ser explicado

pela competência linguística do falante. Para Chomsky, era necessário se desvencilhar do empirismo e se valer do racionalismo para explicar a linguagem (MCENERY; WILSON, 2001).

Os defensores da LC, por sua vez, argumentavam que as pesquisas com *corpus* eram intrinsecamente verificáveis em relação a uma observação introspectiva, visto que, diferentemente daquela, os dados naturais teriam como principal benefício a observação e verificação por quem quer que os queira examiná-los, e não a partir de um ponto de vista privado, haja vista que o dado introspectivo pode ser artificial ou inadequado. Justificavam que, ao pensar em aspectos relacionados à frequência de determinadas expressões ou palavras, os falantes teriam apenas uma vaga noção, não sabendo precisar adequadamente aspectos de sua própria língua. Nesse sentido, observações provenientes de *corpus* seriam consideradas fontes confiáveis para a pesquisa de certas características e permitiriam análises sistemáticas dos fenômenos linguísticos (MCENERY; WILSON, 2001). Dessa maneira, embora reconheçamos a importância de Chomsky para a Linguística, sua visão mostra-se contrária aos preceitos da Linguística de Corpus.

Meyer (2004) aponta vários estudos que têm se apropriado do uso de *corpora* em pesquisas linguísticas, entre os quais se destacam: estudos focados na investigação e no uso de construções gramaticais de diversos tipos; utilização de *corpora* para a elaboração de gramáticas de referência; na Lexicografia, na produção de dicionários e estudos sobre a frequência de padrões vocabulares; no estudo da variação linguística na fala, na escrita e em diversos gêneros textuais; no estudo da variação diacrônica da língua; em investigações sobre aquisição de língua; em análise contrastiva na tradução; no processamento de linguagem natural; e também em trabalhos sobre pedagogia de línguas estrangeiras.

No que concerne aos processos de ensino e aprendizagem de LE, a LC trouxe e tem trazido grandes contribuições. Como examina Römer (2008), a LC e o ensino de LE mantêm entre si uma relação dinâmica em que ambas as áreas se beneficiam. O ensino de LE sofre influência das pesquisas, metodologias e abordagens derivadas da LC, e a LC, por sua vez, pode trabalhar em função das necessidades do ensino de LE, a partir do desenvolvimento de pesquisas e ferramentas que possam ser úteis às práticas pedagógicas.

No ensino, as aplicações de *corpora* podem ser observadas direta e indiretamente. O uso de aplicações pedagógicas indiretas de *corpora* gerais e também de *corpora* mais especializados, tais como *corpora* de línguas para fins específicos, *corpora* de aprendizes e *corpora* paralelos, têm fortes influências sob o currículo de ensino, sob a produção de materiais didáticos e a produção de materiais de referências (como gramáticas e dicionários). Dessa forma, colaboram na tomada de decisões sobre o que e quando ensinar. De acordo com Römer (2008), o currículo pedagógico baseado em *corpora* tem enfatizado a competência comunicativa e também o ensino de itens lexicais recorrentes, com maiores probabilidades de serem encontrados por aprendizes em situações comunicativas reais. Um trabalho pioneiro é o Projeto *Collins Cobuild*, cuja prioridade é o chamado “currículo lexical”, com conteúdo voltado para as combinações de palavras, considerando o alto grau de padronização da língua. A área de estudos fraseológicos, principalmente pelo ensino das colocações, também recebeu grandes contribuições a partir do uso de *corpora*. Nesse contexto, o léxico, muitas vezes, relegado, assumiu papel de destaque, principalmente com o surgimento da Abordagem Lexical (LEWIS, 1993).

A Abordagem Lexical pode ser compreendida como uma combinação entre Linguística Aplicada e metodologia no ensino de LE, cujo foco se volta para o léxico. Na concepção de Lewis (1997), “a Abordagem Lexical pode ser resumida em algumas palavras: a língua não é composta de vocabulário e gramática tradicionais, mas geralmente de blocos de palavras pré-fabricados”¹ (LEWIS, 1997, p. 03). Para ao autor, a combinação desses blocos de palavras armazenados no léxico mental do indivíduo permite a produção de textos coerentes.

O fenômeno das colocações assume papel fundamental quando se trata do ensino do léxico, e sua aplicação em sala de aula contribui fortemente para o desenvolvimento da fluência em uma LE. Podemos destacar, de acordo com Hill (2000), alguns fatores que evidenciam a relevância do ensino das colocações para aprendizes de uma LE: (i) o léxico não é arbitrário, pois os processos de fala e escrita não se constroem por itens vocabulares aleatórios pelo contrário, são previsíveis; (ii) as colocações permitem pensar de modo rápido e tornar a comunicação mais eficiente. A velocidade na fala dos nativos, por

exemplo, pode ser explicada pelo vasto repertório de língua pronta para uso, imediatamente disponível em suas memórias; (iii) as colocações tornam o pensamento mais fácil, uma vez que o falante não faz uso de uma nova língua o tempo todo, apenas resgata as palavras armazenadas em seu léxico mental.

Outra abordagem que pode ser utilizada para o ensino das colocações é a proposta por Johns (1991), conhecida como *Data-Driven Learning* (doravante DDL), ou seja, uma abordagem que trata do “Aprendizado Movido a Dados”. De acordo com Johns, por intermédio dessa abordagem, o aprendiz assumiria um papel importante, uma vez que, através da pesquisa dirigida por dados linguísticos, seu objetivo seria o de descobrir a LE. A tarefa do professor seria providenciar um contexto que possibilitasse ao aprendiz estratégias de descobrimento na língua e que ele pudesse *learn how to learn* – aprender a aprender (JOHNS, 1991, p. 01). O computador, ferramenta central para esta abordagem, funcionaria para busca de informações no *corpus*, e não substituiria o professor, cabendo a este orientá-lo na análise, na interpretação dos dados e nos questionamentos que surgissem.

Segundo Johns (1991), a ferramenta computacional mais importante para a abordagem DDL é o programa de concordância, capaz de recobrir no texto todo o contexto de um item particular, além de permitir uma análise rápida e comparativa. Com base em sua experiência com alunos, utilizando o programa de concordância, propõe algumas considerações: (i) o uso do programa de concordância pode ter uma considerável influência no processo de aprendizagem da língua, uma vez que estimula indagações e especulações por parte do aprendiz, e ajuda-o a desenvolver a habilidade de ver padrões no cerne da língua e a formar generalizações que deem conta dos padrões encontrados. A abordagem DDL diferencia-se da aprendizagem por processo de indução, porque parte do princípio da primazia dos dados, assim, o professor não sabe exatamente quais as regras ou padrões serão descobertos pelos estudantes; (ii) outra importante característica destacada pelo autor refere-se ao posicionamento do professor. Antes ele ditava as regras e o aluno recebia passivamente seus ensinamentos, mas, com a abordagem DDL, o professor tem que aprender como tornar-se um coordenador do

aluno iniciante na pesquisa; (iii) o terceiro importante efeito da DDL está na reavaliação do lugar da gramática no ensino e no aprendizado de línguas. Em vez do ensino da gramática tradicional, geralmente baseado em regras difíceis de serem compreendidas, e em muitas vezes, exemplos descontextualizados, a abordagem DDL torna possível uma nova consciência gramatical, já que o próprio aprendiz descobre o funcionamento da gramática, baseando-se em evidências autênticas da língua em uso.

Pelo exposto, vê-se que as combinações lexicais recorrentes assumem papel importante no ensino, haja vista que, ao internalizá-las, aprendizes de LE podem se comunicar de forma mais precisa e eficiente. Dessa forma, ambas as abordagens acima discutidas – a Lexical e a DDL – são válidas e se complementam, podendo ser aplicadas no ensino.

A disciplina dos estudos linguísticos que trata das combinações lexicais recorrentes é a Fraseologia, sobre a qual discorreremos a seguir.

3 FRASEOLOGIA E COLOCAÇÕES

Uma das formas de manifestação e construção da identidade cultural coletiva se dá pela linguagem. As normas culturais, implementadas através da linguagem e transmitidas ao longo das gerações, não são apenas reproduzidas, mas passam a ser obrigatórias para os falantes da língua pelas estruturas linguísticas utilizadas (TELIYA *et al.*, 1998). Há construções, por exemplo, que se explicam justamente pelo uso que determinada comunidade faz da língua, pela visão de mundo que compartilham, explicadas, portanto, pelas convenções sociais.

Saussure ([1916], 2012, p.112) afirma que “nenhuma sociedade conhece nem conheceu jamais a língua de outro modo que não fosse como um produto herdado de gerações anteriores e que cumpre receber como tal.” A língua, então, sendo parte social da linguagem, não pode existir a não ser por meio de um contrato estabelecido pelos membros da comunidade. Nela, nos deparamos com expressões, blocos de palavras e construções que aparecem com maior ou menor frequência, são arbitrárias e se consagram pelo uso (ORENHA-OTTAIANO,

2004). Essas combinações denominam-se fraseologismos e abrangem as colocações, as coligações, as expressões idiomáticas, os binômios, os provérbios, marcadores conversacionais etc.

No que concerne ao ensino de LE, o domínio dos fraseologismos é de suma importância, pois, quanto mais dessas unidades o aprendiz internalizar, mais próximo do falar nativo estará, considerando que os fraseologismos justificam-se pelo uso que determinada comunidade faz da língua. A tipologia dos fraseologismos é abrangente e, por esse motivo, priorizaremos, neste estudo, as colocações, em razão de se tratar da unidade fraseológica que compõe nosso glossário.

O termo colocação apareceu, primeiramente, nos trabalhos de Palmer (1933), em meados de 1920, quando ensinava inglês no Japão. Ao vislumbrar a necessidade de aprendizagem da combinatória das palavras no ensino de LE, definiu colocações como “a sucessão de duas ou mais palavras que devem ser aprendidas como um todo integrado e não reagrupadas pelas partes que a compõem”² (PALMER, 1933, *apud* COWIE, 1999, p. 54). Em sua publicação intitulada *Second Interim Report on English Collocations*, datada de 1933, justifica:

[...] Não é tanto as palavras ou a gramática que faz do inglês uma língua difícil, mas esse vago e indefinido obstáculo para progredir na aprendizagem consiste em sua grande parte na existência de muitas palavras que, estranhamente, aparecem juntas³ (PALMER, 1933, *apud* COWIE, 1999, p.52-53).

Como se pode notar pelas citações acima, e segundo afirma Cowie (1999), a definição de Palmer acerca das colocações não era linguística, mas pedagógica. Embora Palmer tenha sido o primeiro a criar o termo “colocação”, muitos pesquisadores consideram precursor o linguista Firth (1957), autor da máxima “uma palavra é conhecida pela companhia que mantém”⁴ (FIRTH, 1957 *apud* KENEDDY, 2003, p.468). Entretanto, diferentemente de Palmer, Firth enfatizou não apenas o contexto pedagógico, mas também o contexto linguístico, ao tratar das colocações (KENEDDY, 2003).

McIntosh, Francis e Poole (2009) entendem as colocações como a “combinação de palavras na língua para a produção de textos (falados e escritos) naturais”⁵. De acordo com os autores, pode-se dizer, em inglês,

strong wind e *heavy rain*, mas não seria comum dizer *heavy wind* e *strong rain*. Um falante nativo não teria problemas em produzi-las, mas não se pode esperar o mesmo de um aprendiz de inglês, o qual poderia até mesmo reconhecer o significado isolado de tais palavras, porém, provavelmente teria problemas na hora de combiná-las adequadamente. Daí a necessidade de encorajá-los a observar os padrões combinatórios mantidos pelas palavras, de modo que possam produzir textos mais coerentes e naturais.

Pode-se afirmar que ora as colocações são mais restritas, com menos possibilidades de combinações no eixo paradigmático da língua (*ulterior motives, mitigating circumstances etc.*), ora aproximam-se das combinações consideradas livres, como em *blue shirt e drink tea*, em que o número de possíveis combinações é maior.

4 A COMPILAÇÃO DO GLOSSÁRIO

Em um trabalho de pesquisa anterior (ROCHA, 2011; 2012), propusemos-nos a levantar e analisar os padrões colocacionais presente no seriado *House M.D.* Devido à grande quantidade de colocações extraídas do *corpus*, consideramos a possibilidade de reunir tais unidades fraseológicas em um glossário, visto que poderia servir como subsídio para o ensino de colocações temáticas da área médica. Dessa forma, a partir do glossário, várias propostas de atividades e exercícios poderiam ser desenvolvidas, a exemplo daquelas que apresentaremos nesse artigo.

Nosso *corpus* é composto pelas transcrições das legendas em inglês e em português da série de televisão *House M.D.*, divididas nos subcorpora: Corpus em Português e Corpus em Inglês. São oito temporadas, totalizando 176 episódios para cada subcorpus. Segundo a classificação de Berber Sardinha (2004, p. 26), o *corpus* paralelo de nossa pesquisa (isto é, é formado por um conjunto de textos originais e suas respectivas traduções) pode ser considerado de médio-grande porte, pois é extenso, possuindo mais de um milhão de palavras (1.337.395 em inglês e 3.195.227 em português), o que lhe garante maior representatividade. Pode também ser considerado específico (BERBER SARDINHA, 2004), uma vez que, embora

contenha situações da língua geral, tem como conteúdo predominante a temática médica, possuindo, dessa maneira, léxico e vocabulário especializados. Os episódios da série foram salvos em formato *txt* para levantamento das colocações pelo programa *WordSmith Tools* (SCOTT, 2004), por meio de suas três ferramentas básicas: *WordList*, *Concord* e *Keyword*.

Primeiramente, com a ferramenta *WordList*, identificamos as características gerais do *corpus* e criamos sua lista de palavras. Depois, utilizamos esta lista para geração da lista de palavras-chave (*KeyWords*), a partir de comparações entre listas de palavras do *corpus* de estudo e do *corpus* de referência (no caso da nossa pesquisa, o *British National Corpus - BNC*). Gerada a lista de palavras-chave, selecionamos aquelas pertencentes ao léxico da medicina que formariam as entradas de nosso glossário. Com a ferramenta *Concord*, observamos as linhas de concordância, ou seja, o contexto em que determinada palavra de busca se encontrava, para atestarmos se tal palavra se tratava ou não de uma colocação.

Após levantamento das colocações, da busca por exemplos contextualizados retirados do *Corpus em Inglês* e do *Corpus em Português*, e do levantamento de frequências de cada colocação nos *corpora* de referência COCA – *Corpus of Contemporary American English* e *Corpus do Português*, demos início à tarefa de delimitação da macro e microestrutura e compilação do glossário.

A macroestrutura do nosso glossário possui uma introdução, um índice alfabético e uma lista de entradas, como explicitados adiante:

- **Introdução:** Traz informações gerais e explicações sobre a organização interna e externa do glossário (macro e microestrutura), público-alvo, objetivos, e também sobre o domínio especializado abordado na obra.
- **Abreviações:** Apresenta as abreviações utilizadas no glossário. Em relação aos *corpora* de estudo, abreviamos: *Corpus em Português* para **Cport**; *Corpus em Inglês* para **Cing**. Em relação aos *corpora* de referência, mantivemos a abreviação original do *Corpus of Contemporary American English*, **COCA**; abreviamos o *Corpus do Português* para **CDP** e também mantivemos a grafia do *WebCorp*, cuja

frequência, caso necessário, aparecerá da seguinte maneira: Freq. **WebCorp**. Considerando a formação morfológica das colocações, abreviamos substantivo (**subst.**), adjetivo (**adj.**), locução adjetiva (**loc. adj.**), preposição (**prep.**), artigo (**art.**). Para as colocações em inglês, mantivemos a grafia.

- **Índice:** Funciona como um sistema de remissivas, em decorrência da classificação alfabética das entradas. Em nosso glossário, ordenamos a lista pela inserção das bases ou termos chaves em ordem alfabética, criando assim um índice alfabético, o qual remete aos verbetes e às colocações formadas por ele. Planejamos um índice nas direções tradutórias inglês → português e português → inglês.
- **Lista de entradas:** O conjunto de bases que compõe nossa lista de entradas descritas no verbete.

No que se refere à **microestrutura**, esta corresponderia à “organização dos dados contidos no verbete [...] o programa de informações sobre a entrada disposta no verbete” (BARROS, 2004, p. 157). Em nosso glossário, nos baseamos na proposta de Orenha-Ottaiano (2004) e, desse modo, a microestrutura compreende:

- **O verbete:** reúne os dados relativos à unidade lexical e as informações fornecidas sobre ela. No glossário, é encabeçada pelo registro de um único termo ou base, a partir do qual as colocações serão listadas. Nota-se que a base que encabeça o verbete possui um correspondente na Língua-alvo (LA). Este será traduzido conforme a classe gramatical do termo de origem. Assim, o substantivo **blood** será traduzido na LA pelo substantivo “**sangue**”, embora na listagem de colocações possa ser encontrado como adjetivo *blood clot* – coágulo (subst.) sanguíneo (adj.); *blood infection* – infecção (subst.) sanguínea (adj.) ou com outra tradução (*blood count* – hemograma, *blood culture* – hemocultura).
- **Colocações e equivalentes tradutórios:** Este campo é composto pelas colocações formadas pela entrada, sendo que as colocações em inglês organizam-se em ordem alfabética.

- **Informação gramatical na Língua Fonte (LF) e na Língua Alvo (LA):** Contém as informações morfológicas das colocações. A informação gramatical será inserida apenas uma vez, quando houver um padrão morfológico, ou seja, quando as colocações puderem ser agrupadas de acordo com uma mesma estrutura morfológica (*noun + noun*, *noun + adjective, verb + noun*), como se verifica nas colocações em inglês. Já nas colocações em português, excetuando-se as verbais, percebe-se que a informação gramatical é variável. Nesses casos, a estrutura morfológica será inserida ao lado da colocação correspondente.
- **Exemplos:** Foram retirados do Corpus em Inglês e seu equivalente retirado do Corpus em Português.
- **Frequências:** Ocorrências nos *corpora* de estudo (Corpus em Português e Corpus em Inglês) e *corpora* de referência *Corpus of Contemporary American English* (COCA) e Corpus do Português.
- **Observações gerais:** este campo é destinado à inserção de informações e comentários adicionais para a colocação, tais como: colocações que podem ser consideradas sinônimas (**colocações sinônimas**) em inglês ou português; **outras combinações** para a colocação e também frequência no *WebCorp* (Freq.*Webcorp*), em inglês ou português, para os casos que não encontramos ocorrências nos *corpora* de referência.

Para ilustrar a organização da microestrutura do glossário, selecionamos a entrada *disease* e as colocações formadas por ela:

QUADRO 1 - COLOCAÇÕES PARA A ENTRADA *DISEASE*.

DISEASE <i>NOUN + NOUN</i>	DOENÇA
autoimmune disease	doença autoimune (subst.+ adj.)
<i>Cing.26</i>	<i>Cport. 20</i>
<i>COCA158</i>	<i>CDP. 0</i>

DISEASE <i>NOUN + NOUN</i>	DOENÇA
His present symptoms could mean autoimmune disease, nasolacrimal tumor, an infection or just a cold.	Os sintomas dele podem significar doença autoimune, tumor nasolacrimal, uma infecção ou só um resfriado.
<i>Freq. WebCorp em português: 164</i>	
chronic disease <i>Cing. 01</i> <i>COCA 484</i>	doença crônica (subst. + adj.) <i>Cport. 01</i> <i>CDP. 08</i>
When you were a kid, you had a close relative with a chronic disease.	Quando você era criança, você teve um parente próximo que teve uma doença crônica.
genetic disease <i>Cing. 09</i> <i>COCA 109</i>	doença genética (subst. + adj.) <i>Cport. 10</i> <i>CDP. 03</i>
We just want to know if there's any diagnosed genetic disease in your family.	Só queremos saber se há alguma doença genética diagnosticada na sua família.
heart disease <i>Cing. 12</i> <i>COCA 5665</i>	doença cardíaca (subst. + adj.) <i>Cport. 12</i> <i>CDP. 22</i>
Do you have any history of heart disease in your family?	Você tem histórico de doença cardíaca na sua família?

Nota-se que as colocações a partir da entrada nominal *disease* (doença) foram bastante representativas tanto nos *corpora* de estudo como nos *corpora* de referência. A colocação *heart disease* (doença cardíaca), por exemplo, apresentou 12 ocorrências no *Cing*, impressionantes 5.665 ocorrências no *COCA*, 12 ocorrências no *Cport*, e 22 ocorrências no *CDP*. Não encontramos ocorrências para a colocação *doença autoimune* no *CDP*. No entanto, quando consultamos o *WebCorp*, tivemos um número expressivo: 164 ocorrências.

5 CONTRIBUIÇÕES PEDAGÓGICAS DO GLOSSÁRIO

Além da compilação do glossário, um dos objetivos deste artigo foi o de discutir sua aplicação pedagógica. Desse modo, desenvolvemos algumas propostas de atividades e exercícios baseados no glossário, além de outras atividades colocacionais a partir do *corpus*. Tais atividades foram elaboradas de acordo com os pressupostos teórico-metodológicos da LC, especificamente apoiadas na Abordagem Lexical – *Lexical Approach* (LEWIS, 1997, 2000) e na Abordagem movida a dados – *DDL Approach* (JOHNS, 1991). Para desenvolvê-las, utilizamos ainda sugestões do artigo *Integrating corpora with every day language teaching* (FRANKENBERG-GARCIA, 2012). Essas propostas foram apresentadas no Curso de Aperfeiçoamento “Linguística de Corpus e Fraseologia aplicadas à prática pedagógica de Professores de Língua Inglesa da Rede Pública”, coordenado pela Profa. Dra. Adriane Orenha-Ottaiano. O objetivo dessas atividades, além de testar a aplicabilidade do glossário em sala de aula, foi difundir a utilização do uso de *corpora* e do ensino lexical para além do ambiente acadêmico, isto é, tinham o propósito de desenvolver nos professores da rede pública consciência do potencial combinatório das palavras, de modo que eles pudessem transferir e adaptar esse conhecimento para a realidade da sala de aula, expondo os aprendizes a uma linguagem autêntica e contextualizada, conforme a proposta do referido curso e do Projeto de Extensão de mesmo título, aprovado pela Pró-Reitoria de Extensão da UNESP e executado em 2013. Entre as propostas de atividades desenvolvidas, selecionamos algumas, apresentadas a seguir:

5.1 ENSINO DE COLOCAÇÕES MÉDICAS A PARTIR DA SÉRIE *HOUSE M. D.* E DO GLOSSÁRIO DE COLOCAÇÕES

Como apontado anteriormente, o contato dos aprendizes com a LE estudada vai muito além do ambiente escolar. Na verdade, é no ambiente extraescolar que eles possuem maior exposição à língua, principalmente pelo intenso uso da *Internet*, a partir de canais como *youtube*, redes sociais, como *Facebook*, *Skype* e uma infinidade de

blogs e *websites*. Tais tecnologias encurtam as distâncias e aumentam as possibilidades de interação entre as pessoas, o que facilita as trocas linguísticas e, conseqüentemente, beneficia a aprendizagem de um novo idioma.

Aliada à influência da *Internet*, a aprendizagem de línguas, sobretudo, a de língua inglesa, também é determinada pela televisão, pela música, por filmes e pela literatura. No caso da televisão, umas das grandes influências provêm das séries de televisão. Os jovens, por exemplo, adoram as séries e recebem muita influência da cultura de um país por meio delas. Muitos assistem tais séries, pois acreditam que, por meio delas, se aproximam da língua em uso, falada no país que as originou e podem, assim, melhorar suas habilidades no idioma.

Nesse contexto, a utilização de trechos de séries de televisão na sala de aula traz grande contribuição para os alunos. Sugerimos, então, uma atividade a ser trabalhada que, além de focar em materiais que é do interesse do aluno (o que torna a aula mais atraente) e de promover contato com um aspecto cultural de um país, permite acesso ao falar muito próximo do natural e possibilita discussões orais, foca também na aquisição e internalização de colocações, na aprendizagem autônoma e no uso de glossários e dicionários.

Tomaremos como exemplo trechos da série de televisão *House M.D.*, da qual lançamos mão para a compilação do glossário bilingue de colocações da língua geral e especializadas na área médica. Primeiramente, selecionamos um trecho da série que apresentava colocações da língua geral e colocações especializadas. A partir da transcrição do trecho, copiamos a parte selecionada e a colamos em um *handout*. Uma vez escolhidas as colocações, apagamos uma parte da combinação, deixando aquela que mais fortemente sugere a parte faltante. Por exemplo, nas colocações **family history**, **radiation treatment**, **enviromental factors**, **brain tumor**, apagamos as partes em itálico, cabendo aos estudantes preencher a parte faltante com o colocado mais adequado. Para tanto, o professor passaria o trecho escolhido e pediria que os alunos completassem as lacunas. O trecho pode ser revisto algumas vezes, para que a classe tenha tempo de preenchê-las.

FIGURA 1 - PROPOSTA DE ATIVIDADE 02.

Wilson: Protein _____ for the three most prevalent brain _____ came up negative.

House: That's an HMO lab; you might as well have sent it to a high school kid with a chemistry _____.

Wilson: No family _____.

House: I thought your uncle died of cancer.

Wilson: Other side. No environmental _____.

House: That you know of.

Wilson: And she's not responding to radiation _____.



Na conferência das respostas está uma importante parte da aula, haja vista que o professor deve induzir os alunos a pensar nas possibilidades combinatórias de algumas colocações. Assim, no caso da colocação *brain tumor*, ele poderia perguntar: que outras palavras podem combinar com *brain* e com *tumor*? Conforme as dúvidas surgirem, principalmente no caso das colocações especializadas, os alunos poderão recorrer ao glossário. Lá, poderão encontrar diferentes colocados com os quais um nódulo pode ocorrer, além de seu contexto de uso e seu equivalente na LE. Se o glossário mostrar-se insuficiente para a pesquisa, o professor solicitará ao aluno que consulte os *corpora* de referência *on-line* ou até mesmo dicionário de colocações, e traga para compartilhar com a sala os resultados obtidos.

Essa proposta, ao ser apresentada no curso de extensão, foi considerada bem mais atraente e de fácil aplicabilidade. Os professores sugeriram que, a depender do nível da turma, algumas alterações precisam ser feitas. Assim, ao ser preparada, a atividade pode ser facilitada, por exemplo, para um aluno de nível intermediário, com sugestões, entre parênteses, de palavras que poderiam completar a colocação ou até mesmo uma box com as palavras da atividade. Sugeriram também que o vídeo fosse apresentando mais de uma vez, para que os alunos se sentissem mais preparados para fazer a atividade.

5.2 PRODUÇÃO TEXTUAL

Após a aplicação da atividade anterior, o professor pode desenvolver esta proposta, cujo objetivo é precisar a escrita dos alunos pela utilização de colocações. Para tanto, inicia uma discussão que aborde, por exemplo, uma consulta ao médico. Os alunos podem debater em pares, em grupo, ou discutir conjuntamente. Depois, são convidados a produzirem um texto em que relatam/descrevam uma visita ao médico. Eles podem falar sobre problemas de saúde, sintomas, descrever as dores etc. Para ajudá-los, algumas bases podem ser escritas na lousa: *needle, blood, heart, bleeding, test, dignosis, fever, patient, doctor, symptoms, treatment, urine, pain* etc. A redação poder ser feita na sala de aula, a fim de que o professor possa auxiliá-los. Estes também podem consultar o glossário para precisar suas produções. Ao fazer a correção das redações, o professor pode propor novas colocações (retiradas do glossário ou de *corpora* de referências). Depois de corrigidas, as redações são devolvidas para que os estudantes a reescrevam e vejam como seu texto ficou mais completo.

5.3 ASSOCIANDO IMAGENS A COLOCAÇÕES

Este exercício ajuda o aprendiz a formar colocações a partir de imagens. O aprendiz inferirá a que colocação a imagem se refere e criará uma sentença para cada uma delas.

FIGURA 2 - PROPOSTA DE ATIVIDADE 04.

ACTIVITY: THE IMAGES BELOW CAN BE REPLACED BY COLLOCATIONS. IDENTIFY THEM AND WRITE YOUR OWN SENTENCES ABOUT THEM.	
a) PAIN	b) BLOOD
	

c) HEART	d) SURGEON
	

Ao observar as imagens, o aluno consegue facilmente inferir a colocação na sua língua materna. No entanto, buscar a colocação equivalente em LE pode ser um pouco mais complicado. A imagem **b**, por exemplo, mostra um paciente cujo sangue está sendo coletado. O aluno teria que buscar pela colocação equivalente em inglês (*draw blood* ou *collect blood*). Caso não tenha a combinação internalizada em seu léxico mental, recorrerá ao glossário bilíngue. Para todas as imagens, o aluno fará a associação com uma colocação **a**) *back pain* – dor nas costas; **c**) *heart attack* – ataque cardíaco; **d**) *plastic surgeon* – cirurgião plástico), buscará o equivalente em inglês e produzirá uma sentença para cada uma delas. Lembrando que sempre poderá consultar, além do glossário, os *corpora* de referência.

Os professores aprovaram essa atividade, justificando que o uso de imagens ajuda bastante na aprendizagem de uma LE. Também sugeriram cuidado na escolha das imagens, pois algumas não exemplificavam claramente a colocação pretendida, levando a outras interpretações.

5.4 GLOSSÁRIO DE COLOCAÇÕES COMPILADO PELOS ALUNOS

Geralmente, nas aulas de línguas, muitos professores pedem aos alunos que mantenham um caderno de vocabulário para que possam consultá-lo em caso de dúvidas, de modo que possam ampliar

e fixar o vocabulário estudado. Entretanto, sabe-se que não é muito vantajoso manter tal caderno, pois o registro de palavras individuais e descontextualizadas pouco contribui para expansão lexical.

Nesse sentido, pensando na expansão do léxico mental de aprendizes de LE, principalmente no que concerne à aprendizagem das colocações, sugerimos o desenvolvimento de um trabalho que poderia ajudar os professores de LE desenvolver a competência colocacional (HILL, 2000) de seus alunos. Tal trabalho consistiria na adoção de um glossário de colocações, por parte dos alunos. Acreditamos que, em vez de manter somente um caderno de vocabulário com palavras individuais, é muito mais interessante para o estudante elaborar um glossário, cuja atenção recaia sobre o contexto combinatório em que uma palavra ocorre. Por meio deste, os aprendizes armazenariam as colocações com as quais se depararem ao longo das aulas de LE, e, especialmente, no ambiente extraescolar – considerando o vasto repertório linguístico a que estão expostos fora da escola. Dessa maneira, sempre que estivessem em contato com a LE em um contexto real de uso, poderiam observar tais unidades e registrá-las no glossário. Ao revisitá-lo e fazer novas inserções, haveria mais chances de internalizá-las.

Nesse caso, entraria em cena o que Hill (2000) chama de “o papel da memória”. As colocações já armazenadas no glossário tornam-se conhecidas pelo aprendiz, que as internaliza e recupera de sua mente quando sente necessidade de utilizá-las, da mesma maneira que consegue recuperar da memória um número de telefone ou um endereço. O glossário seria, então, uma fonte de pesquisa que ajudaria o estudante na produção oral e também escrita da LE, contribuindo para a ampliação do seu léxico mental.

A proposta ora apresentada se baseia nas sugestões de Woolard (2000) e seu caderno lexical (*lexical notebook*), a partir de sua experiência em sala de aula. Este seria um glossário simples, organizado alfabeticamente, com duas ou três folhas para cada letra. Os estudantes distribuiriam as colocações pelos nódulos ou base. Além da definição, da transcrição fonética e da tradução do nódulo, o aluno registraria um exemplo contextualizado da palavra colocada com o nódulo. Sempre que observasse um adjetivo, um substantivo, verbo ou advérbio que combinasse com o nódulo, ele faria uma inserção no glossário. Para

atestar a recorrência e o uso das unidades inseridas no glossário, poderiam fazer uso de *corpora on-line*, tais como: COCA (*Corpus of Contemporary American English*); *Corpus do Português* e *Corpus Brasileiro* (quando houver necessidade da busca de equivalentes em língua portuguesa), assim como o emulador de concordanciador *WebCorp*. Também poderiam utilizar os dicionários *Oxford Collocation Dictionary for Students of English* e *Cambridge Advanced Learners Dictionary*.

Além desse glossário, destinado à inserção de colocações mais gerais, sugerimos também que o professor trabalhe em um projeto paralelo, o qual consistiria no desenvolvimento de um glossário temático de colocações. Este seria compilado a partir da transcrição das legendas de uma série de televisão, de textos jornalísticos ou literários etc. Pelo desenvolvimento de um projeto como este, os alunos observariam o comportamento colocacional de uma área em particular. Nosso glossário bilíngue de colocações pode ser tomado com modelo. Tal projeto teria como suporte a Abordagem Lexical (LEWIS, 2000) e a abordagem movida a dados, *DDL - Data-driven Learning*, (JOHNS, 1991). O programa de concordância, como por exemplo, o *WordSmith Tools* (SCOTT, 2012), funcionaria como ferramenta central, pois recobriria no texto todo o contexto de um item particular e permitiria análise rápida e comparativa.

Esta atividade, ao ser apresentada no curso de extensão acima mencionado, foi bem recebida pelos professores; no entanto, muitos questionamentos surgiram, principalmente em relação à sua implementação. De acordo com os professores presentes, para desenvolvê-la seria necessário um projeto bem delineado, que leve em consideração desde o treinamento dos próprios professores e alunos na utilização do programa de concordância, bem como mudanças no planejamento das aulas, principalmente no currículo de ensino, haja vista que muitos professores da rede pública já possuem um plano de ensino a ser seguido, além da própria estruturação das salas de aula, pela necessidade de equipamentos eletrônicos para desenvolvimento da atividade. Embora vista de maneira positiva, os professores consideraram sua implementação trabalhosa. Nesse sentido, consideram o caderno de

colocações mais viável, visto que independe da utilização do programa de concordância.

6 CONSIDERAÇÕES FINAIS

Neste artigo, buscamos discutir as aplicações pedagógicas de um glossário bilíngue de colocações compilado a partir das legendas do seriado *House M.D.* Discorreremos, inicialmente, sobre o aporte teórico que deu embasamento para o desenvolvimento de nossa pesquisa, e também pontuamos brevemente as etapas que possibilitaram a compilação do glossário e a organização de sua macro e microestrutura. Por fim, apresentamos uma sequência de atividades elaboradas com base no glossário.

A nosso ver, tais sugestões podem oferecer subsídios teóricos e práticos aos professores de inglês das redes pública e particular de ensino para os quais apresentamos e aplicamos nossa pesquisa, de modo que pudessem ter uma ideia de como aplicar os conhecimentos da LC e da Fraseologia, áreas que têm fortemente contribuído para o ensino de LE centralizado no uso combinatórias lexicais autênticas, e adaptá-los à realidade de suas práticas pedagógicas. Além disso, acreditamos que o glossário compilado pode servir como fonte de consulta a estudantes, tradutores e demais interessados nas colocações do campo lexical da medicina.

TEACHING ENGLISH AS A FOREIGN LANGUAGE AND THE PEDAGOGICAL CONTRIBUTIONS OF A GLOSSARY OF COLLOCATIONS

ABSTRACT

In accordance with the theoretical and methodological background of Corpus Linguistic and Phraseology, this study aims to discuss the pedagogical contributions of a bilingual glossary of collocations in the field of medicine, which was compiled from a parallel corpus made up of English and Portuguese subtitles extracted from the TV show *House M.D.* Based on the glossary, we developed some activities that can be used in the classroom, considering that these phraseological units play an important role in the enrichment of EFL

learner's lexical competence. Moreover, we propose the development of other corpus-based collocational activities.

KEYWORDS: Corpus Linguistics, collocations, Phraseology, bilingual glossary, teaching collocations, pedagogical applications.

LA ENSEÑANZA DEL INGLÉS COMO LE Y CONTRIBUCIONES PEDAGÓGICAS DE UN GLOSARIO BILINGÜE DE COLOCACIONES

RESUMEN

De acuerdo con el referencial teórico y metodológico de la Lingüística de Corpus y de la Fraseología, el objetivo de este trabajo es discutir las aplicaciones pedagógicas de un glosario bilingüe de colocaciones en el área médica para la enseñanza de inglés como LE. La compilación del glosario fue posible a partir del corpus paralelo formado por subtítulos en inglés y portugués de la serie House M.D. Basadas en las colocaciones del glosario, elaboramos propuestas de actividades que pueden servir como un recurso para la enseñanza del inglés como LE, especialmente cuando se considera la relevancia de estas unidades fraseológicas en el enriquecimiento de la competencia léxica de los alumnos. Además, proponemos el desarrollo de otras actividades colocacionales basadas en corpus.

PALABRAS-CLAVE: Lingüística de Corpus, colocaciones, fraseología, glosario bilingüe, enseñanza de las colocaciones, aplicaciones pedagógicas.

NOTAS

- 1 "The Lexical Approach can be summarised in a few words: language consists not of traditional grammar and vocabulary but often of multi-word prefabricated chunks."
- 2 "A collocation is a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts."
- 3 "[...] it is not so much the words of English or the grammar of English that makes English difficult, but that vague and undefined obstacle to progress in the learning of English consists for the most part in the existence of so many odd-comings-together-of-words."

- 4 “you shall know a word by the company it keeps.”
- 5 “collocation is the way words combine in a language to produce natural-sounding speech and writing.”

REFERÊNCIAS

- BARROS, L. *Curso Básico de Terminologia*. São Paulo. Editora Universidade de São Paulo, 2004.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Manole, 2004.
- BOAS, F. *Race, Language and Culture*. New York: Macmillan, 1940.
- BRITISH NATIONAL CORPUS (BNC). Disponível em: <<http://www.natcorp.ox.ac.uk/>>. Acesso em: 20 jan. 2012.
- COWIE, A. P. *English Dictionary for Foreign Learners: A history*. Oxford: Oxford University Press, 1999.
- DAVIES, M. *The Corpus of Contemporary American English: 425 million words, 1990-present*. Disponível em: <<http://corpus.byu.edu/COCA/>>. Acesso em: 20 jan. 2012.
- _____; FERREIRA, M. (2006-) *Corpus do Português: 45 million words, 1300s-1900s*. Disponível em: <<http://www.corpusdoportugues.org>>. Acesso em: 20 jan. 2012.
- FIRTH, J. R. Modes of Meaning. In: FIRTH, J. R. (Ed.). *Papers in Linguistics – 1934-1951*. Oxford: Oxford University Press, 1957.
- FRANKENBERG-GARCIA, A. Integrating corpora with everyday language teaching. In: Thomas, J. and Boulton, A. (Ed.) *Input, Process and Product: developments in teaching and language corpora*. Brno: Masaryk University Press, 2012, p. 36-53.
- HILL, J. Revising priorities: from grammatical failure to collocational success. In: LEWIS, M. (Ed). *Teaching collocation: Further developments in the lexical approach*. London: Language Teaching Publications, 2000, p. 47-69.
- JOHNS, T. *Should you be persuaded: two examples of data-driven learning*. In: Johns, T.; KING, P. (Ed.), *Classroom Concordancing*. English Language. 1991.
- KENNEDY, G. *Amplifier Collocations in the British National Corpus: Implications for English Language Teaching*. TESOL Quaterly, v. 37, 2003, p. 467-487.

Lewis, M. *The Lexical Approach*. Hove: Language Teaching Publications, 1993.

_____. *Implementing the lexical approach: Putting theory into practice*. England: Language Teaching Publications, 1997.

_____. *Teaching Collocations*. London: Language Teaching Publications, 2000.

MCENERY, T; WILSON, A. *Early corpus linguistics and the Chomskyan revolution*. In: _____ (Ed). *Corpus Linguistics: An Introduction*. Edinburgh University Press: Edinburgh, 2001, p. 01-29.

MCINTOSH, C.; FRANCIS, B.; POOLE, R. (Org.). *Oxford Collocations Dictionary for Students of English.2.* ed. Oxford: Oxford University Press, 2009. 963 p.

MEYER, C. F. *English Corpus Linguistics – An Introduction*. Cambridge, 2004.

ORENHA-OTTAIANO, A. *A compilação de um glossário bilingue de colocações, na área de jornalismo de Negócios, baseado em corpus comparável*. 2004. 246 f. Dissertação (Mestrado em Estudos Linguísticos e Literários), FFLCH/USP, São Paulo, SP.

PALMER, H. E. *Second Interim Report on English Collocations* (Tokyo: Kaitakusha), 1933.

ROCHA, J. M. P. *Estudos fraseológicos com vista ao levantamento de colocações especializadas presentes no seriado House M. D.* Relatório de Estágio Básico. São José do Rio Preto (SP), IBILCE/UNESP, 2011.

_____. *A compilação de um glossário bilingue de colocações e colocações especializadas nas direções inglês-português/português-inglês, a partir do corpus House M. D.* Relatório de Iniciação Científica. São José do Rio Preto (SP), IBILCE/UNESP, 2012.

RÖMER, U. *Corpora and language teaching*. In: LÜDELING, A.; KYTÖ, M. (Ed.). *Corpus Linguistics: an International Handbook*, vol.1, p. 112-130. Berlin/New York: Walter de Gruyter, 2008.

SAUSSURE, F. *Curso de Linguística Geral*. São Paulo: Cultrix, 2012.

SCOTT, M. *WordSmith Tools: version 4.0*. Oxford: Oxford University Press, 2004. Disponível em: <<http://www.lexically.net/wordsmith/support/extras.html>>. Acesso em: 11 jul. 2013.

SINCLAIR, J. *Corpus, concordance and collocation*. Oxford: Oxford University Press, 1991.

SHORE, D. *House M. D. Estados Unidos: Fox/Universal Studios, 2005*.

STERN, W. *Psychology of Early Childhood up to six years of age*. New York: Holt, 1924.

TELIYA, V. et al. *Phraseology as a language of culture: its role in the representation of a collective mentality*. In: COWIE, A. P. (Ed.). *Phraseology: theory, analysis and applications*. Oxford: Oxford University Press, 1998. p. 55-75.

THORNDIKE, E. L. *Teacher's Wordbook*. New York: Columbia Teachers College, 1921.

Transcrições dos episódios da série *House M.D.* em inglês. Disponível em: <<http://clinic-duty.livejournal.com/12225.html>>. Acesso em: 11 jul. 2013.

Transcrições dos episódios da série *House M.D.* em português. Disponível em: <<http://www.tvsubtitles.net/tvshow-9-2.html>>. Acesso em: 11 jul. 2013.

WebCorp. Disponível em: <<http://webcorp.org.uk/>>. Acesso em: 21 jan. 2012.

WOOLARD, G. Collocation encouraging learner independence. In: LEWIS, M. (Ed.). *Teaching collocation: further developments in the lexical approach*. London: HeinleCengage Learning, 2000, p. 28-46.

Submetido em 1 de julho de 2014

Aceito em 6 de outubro de 2014

Publicado em 21 de dezembro de 2015
