

GLOSSÁRIO DE COLOCAÇÕES DA LÍNGUA GERAL
E ESPECIALIZADAS BASEADO EM *CORPUS*: UMA CONTRIBUIÇÃO
PARA O ENSINO DE LE E TRADUÇÃO

Adriane Orenha-Ottaiano
Universidade Estadual Paulista
adriane@ibilce.unesp.br

Roberta Fiel
Universidade Estadual Paulista
roh_fiel@hotmail.com

RESUMO:

Esta investigação, fundamentada na Linguística de Corpus e na Fraseologia, tem como propósito o levantamento e a análise das colocações da língua geral e das colocações especializadas na área médica, extraídas do *corpus* paralelo composto pelas transcrições do seriado de TV *Grey's Anatomy*. A partir desse levantamento, propomos a compilação de um glossário bilíngue, a fim de que esse material possa ser utilizado por tradutores aprendizes, bem como em sala de aula por professores de língua inglesa.

PALAVRAS-CHAVE: colocações da língua geral; colocações especializadas; glossário bilíngue; *corpus* paralelo; Fraseologia.

ABSTRACT:

This research, theoretically founded on Corpus Linguistics and Phraseology, has the purpose of extracting and analyzing general language and specialized collocations in the medical field, taken from a parallel corpus comprised of transcriptions of the TV serial *Grey's Anatomy*. Based on this extraction, it is proposed a compilation of a bilingual glossary, so that the referred material can be used by learner translators as well as English language teachers.

KEYWORDS: general language collocations; specialized collocations; bilingual glossary; parallel corpus; Phraseology.

Introdução.

Neste trabalho, realizamos um estudo fraseológico, com foco nas colocações da língua geral e da língua de especialidade na área médica, presentes em um *corpus* formado pelas transcrições do seriado de TV *Grey's Anatomy*. Para isso, nos baseamos na Linguística de Corpus, área que tem como um dos objetivos o estudo do léxico, e que tem colaborado e fornecido subsídios para a realização de pesquisas em ensino e aprendizagem de línguas; na Terminologia, área responsável pelo estudo dos conceitos e termos usados nas línguas de especialidade; e na Fraseologia, área voltada para a investigação de combinações lexicais recorrentes, que abrangem as expressões idiomáticas, as coligações, os binômios, os provérbios e as colocações, alvo de nosso estudo. As colocações, inseridas no âmbito da Fraseologia, são uma das categorias mais relevantes, também consideradas a maior classe de unidades multipalavras, e representam grande dificuldade para um falante não nativo, pois ele sofre influência da língua-mãe no momento da enunciação. Orenha-Ottaiano (2004) observa que algumas colocações são de fácil compreensão para um aprendiz, como a combinatória *place an order* (= *fazer um pedido*). No entanto, na construção do enunciado, o aprendiz encontra grandes dificuldades e, quase sempre influenciado pela língua-mãe, acaba cometendo erros, como em *make an order*, que, para ele, faz mais sentido, mas que para um falante nativo soa estranho. Segundo Béjoint (1994), “as palavras têm ‘características embutidas’ e, assim, a escolha de uma palavra, ou até um sentido específico de uma palavra, acarreta necessariamente a escolha obrigatória ou preferencial de outras palavras, ou de alguma construção sintática”.

O estudo fraseológico citado a cima, consiste de um levantamento das colocações e colocações especializadas mais frequentemente empregadas no seriado TV *Grey's Anatomy*, a partir desse levantamento compilamos um *corpus* paralelo formado pelas transcrições em inglês e em português de todas as temporadas da série, o que possibilitou a elaboração de um glossário bilíngue nas direções tradutórias inglês-português/português-inglês de colocações especializadas da área Médica e na geral, as quais poderão servir de material de apoio para aulas ministradas em língua inglesa, mostrando a relevância da Linguística de Corpus e da Fraseologia para o ensino de LE, já que, baseado em nossa experiência como aprendiz de uma língua estrangeira (doravante LE), notamos que a inter-relação entre léxico e gramática é muito importante para o aprendizado de uma LE. Segundo Orenha-Ottaiano (2004, p. 01), “a gramática, sem o potencial combinatório, sintagmático e paradigmático do léxico, desempenha poucas funções no sistema. Apenas a gramática não daria conta

de explicar certas escolhas linguísticas do falante nativo”. Com base na autora, podemos afirmar que um dos motivos pelos quais os aprendizes de uma LE encontram tantas dificuldades para se comunicar se deve ao fato de aprenderem tal LE de uma maneira segmentada e artificial, muitas vezes baseada apenas na gramática da língua, tornando, assim, sua comunicação lenta e ineficiente.

Este estudo mostra a relevância da Linguística de Corpus, e da Fraseologia para o ensino e aprendizagem de uma LE, já que aponta casos reais de uso, e aproxima o aprendiz cada vez mais da LE estudada.

Para a realização desse trabalho, contamos com o arcabouço teórico-metodológico da Linguística de Corpus (BERBER SARDINHA, 2004; BIBER; CONRAD; REPPEN, 1998; MEYER, 2004), da Terminologia (CARVALHO, 2007; PAVEL; NOLET, 2002; AUBERT, 2001; BARROS, 2004) e da Fraseologia (FIRTH, 1957, HILL, 2000; NESSELHAUF, 2005; ORENHA-OTTAIANO, 2004, 2009; SINCLAIR, 1991).

1. A Linguística de Corpus.

Conforme já mencionado, nossa pesquisa se baseia na Linguística de Corpus que, de acordo com Berber Sardinha (2004, p. 3) “ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística”. Com o advento da tecnologia da informática, a Linguística de Corpus ganhou grande espaço nas pesquisas, no ensino de LE e exerce grande influência na linguística.

Berber Sardinha (2004) afirma que para um *corpus* ser representativo, esse deve ser extenso, ou seja, o maior possível a fim de que ele se aproxime ao máximo da população do qual deriva.

Neste aspecto, Berber Sardinha (2004, p. 26) sugere uma classificação quanto à extensão de um *corpus*, como mostra a tabela abaixo:

Tamanho em palavras	Classificação
Menos de 80 mil	Pequeno
80 a 250 mil	Pequeno-médio
250 mil a 1 milhão	Médio
1 milhão a 10 milhões	Médio-Grande
10 milhões ou mais	Grande

Tabela 1: Extensão de um *corpus*, segundo Berber Sardinha (2004, p. 27).

O *corpus* compilado para esta presente pesquisa é composto por sete temporadas do seriado *Grey's Anatomy*, e pode ser classificado como um *corpus* de médio-grande porte, posto que as transcrições de todas as temporadas contabilizaram 2.200.305 palavras.

Uma área muito importante na qual a Linguística de Corpus atua é a Linguística Aplicada ao ensino de LE. Segundo Berber Sardinha (2004), já há algum tempo existe uma influência de estudos baseados em *corpora* no ensino de línguas, antes mesmo do advento do *corpus* computadorizado. Há cerca de 80 anos, Edward Thorndike lançava a primeira de sua obra baseada em um *corpus* de 4,5 milhões de palavras, o *Teacher's Workbook* (1921), composto por um levantamento monumental das frequências de palavras da língua inglesa. Quando foi publicado, o *Workbook* impulsionou mudanças no ensino de língua materna e estrangeira, por mostrar a língua real falada pelos nativos.

A exploração de *corpora* no ensino de LE pode ser resumida, de acordo com Berber Sardinha (2004), em quatro áreas de concentração majoritárias:

- Descrição da linguagem nativa;
- Descrição da linguagem do aprendiz;
- Transposição de metodologias de pesquisa acadêmica para sala de aula;
- Desenvolvimento de materiais de ensino, currículos e abordagens.

A primeira área diz respeito à descrição da linguagem de falantes nativos, tanto escrita quanto falada. A influência da Linguística de Corpus no ensino ocorre, normalmente, de modo indireto, em geral, no conteúdo de materiais de ensino.

A segunda área trata da descrição da linguagem, mas, diferentemente da outra, esta trata da fala e da escrita de estudantes de LE, chamada de *corpora* de aprendizes (*learner corpora*). A descrição de *corpora* de aprendizes é uma atividade nova, que tem ganhado impulso nos últimos anos.

A terceira área é ligada ao uso instrumental analítico da Linguística de Corpus na sala de aula, como concordâncias de listas de palavras. A concordância é atividade mais empregada, servindo de base para matérias de ensino.

Por fim, a última área refere-se à criação de metodologias de ensino inspiradas na Linguística de Corpus. Nesse sentido, Berber Sardinha (2004) diz que há pelo menos três propostas de utilização de *corpora* no ensino: o Currículo Lexical, a Abordagem Lexical e o Ensino Movido a Dados (DDL). Neste trabalho, daremos alguns detalhes desta última, dada a sua importância para pesquisas no ensino de LE, que pretendemos realizar no futuro.

A proposta DDL (*Data Driven Learning*) é defendida por Tim Johns (2000), originalmente foi criada para ensinar gramática no inglês, mas que, entretanto, têm se expandido para outras áreas de ensino. A ênfase do DDL é desenvolver no aluno a habilidade de descoberta. O professor tem o papel de proporcionar meios para que o aluno adquira estratégias de descoberta e o computador tem, meramente, o papel de informante, sem jamais tomar o lugar do professor, o qual assume o papel de orientador de pesquisa.

A pesquisa que está sendo apresentada se pauta, de certa forma, no método DDL, pois se baseia na busca de evidências reais na língua e proporciona vantagens ao aluno, que assume o papel de aluno-pesquisador, desenvolvendo, assim, a habilidade de identificar regularidades e de fazer generalizações para explicá-las.

De modo geral, a Linguística de Corpus, entendida como investigação criteriosa de *corpora* de linguagem natural, tem sido influente no ensino de línguas há pelo menos 80 anos, desde a publicação do primeiro *Workbook*, de Thorndike, em 1921, até os dias de hoje, principalmente no ensino de LE, pois possibilita ao aluno entrar em contato com o falar real de uma língua.

No que concerne às diferentes tarefas de ensino baseados em *corpora*, estas podem ser feitos diretamente pelo professor com a descrição da linguagem de seus próprios alunos, pela descrição da linguagem de nativos, pelo método DDL, entre outros, como foi explicado anteriormente, ou indiretamente por meio de materiais, livros e dicionários que se baseiem em *corpora*. Vários estudiosos como Biber et al. (1999), Lewis (2000) e Sinclair (1991) enfatizam a importância de combinações de palavras recorrentes e expressões em um contexto pedagógico, devido seu grande potencial na promoção da fluência.

Vale enfatizar que somente o ensino de gramática não é suficiente para aprendizes que desejem uma fluência boa em uma língua. É preciso, dessa maneira, obter vocabulário de uso diário de nativos, e, o ensino baseado em *corpora*, proporciona isso aos alunos. Berber Sardinha (2009) nos mostra a existência de uma padronização léxico-gramatical, também anteriormente exposta por Sinclair (1991). A padronização léxico-gramatical é uma sequência recorrente de palavras. Entre os vários padrões existentes, o que é mais saliente são os agrupamentos (*clusters*), ou seja, sequências fixas de palavras, como “bom dia”. Esses padrões são bastante salientes, devido à sua rigidez e, assim, se tornam padrões importantes para se ensinar em aulas de LE, pois são facilmente notados. As palavras tendem a se associar em grupos padronizados, recorrentes e essas associações não são aleatórias: as palavras se combinam umas às outras com graus diferentes de probabilidade.

Do ponto de vista da sala de aula, há várias razões que salientam a importância de padrões no aprendizado de LE. Os padrões são um tipo de “unidade pré-fabricada” da língua, que parece estar na memória em forma pronta, isto é, sem estar decomposto em unidades individuais. Dessa maneira, um agrupamento como “bom dia” parece estar disponível por inteiro na memória, para produção e recepção, dos falantes nativos. Quando falamos, possivelmente acessamos nosso estoque lexical mental e retiramos de lá tal unidade semi-pronta e a utilizamos na fala e na escrita. Do mesmo modo, ao processar a língua falada ou escrita, é possível decompor a sequência de fala ou escrita em unidades compostas, muitas vezes em forma de padrões. Esse acesso a unidades pré-fabricadas permite uma maior rapidez no processamento, o que, por sua vez, sustenta nossa capacidade natural de falar e escrever fluentemente, de ouvir e entender, de modo síncrono, nosso interlocutor e de muitas vezes poder ler com rapidez, sem nos deter a cada palavra individual do texto. Embora essas sejam habilidades naturais que todo falante nativo possui, em maior ou menor grau, elas não são automaticamente transferidas para a LE. Normalmente, os aprendizes de LE tendem a ver a linguagem como um conjunto de palavras individuais sustentadas por regras gramaticais. Desse modo, a compreensão e a produção se dão, em muitos casos, com base na junção de palavras individuais.

O autor ainda evidencia que o trabalho com *corpora* em sala de aula pode ajudar a melhorar essa situação, pois ele naturalmente faz saltar aos olhos essas unidades “pré-fabricadas”, que são os padrões léxico-gramaticais. Ou seja, começando por um trabalho de conscientização, o aluno passa a ter uma visão diferente do que é o vocabulário de uma língua, deixando de lado a ideia de que se trata de um conjunto de palavras isoladas que se juntam por meio de regras gramaticais. Ele passa a perceber que as palavras se juntam por meio de atração e assim criando os agrupamentos, pacotes *chunks* que, por sua vez, se juntam e formam o “tecido da linguagem”.

O uso de *corpora* na sala de aula de LE pode trazer vários benefícios para os alunos, pois eles entram em contato com muitas ocorrências de padrões que se repetem em outros textos, permitindo, desse modo, desenvolver a consciência a respeito da natureza probabilística e associativa da língua, visando ao desenvolvimento da fluência e ao enriquecimento do vocabulário.

2. Convencionalidade e Fraseologia.

O conceito de Convencionalidade na língua, segundo Orenha-Ottaiano(2004), está relacionado ao uso que fazemos, ao interagir, de uma série de

convenções, expressões e blocos de palavras já preestabelecidos e consagrados em nossa comunidade. Todos os dias usamos conjuntos de palavras na nossa língua materna que, segundo nossa concepção, parecem-nos óbvias.

O termo convencionalidade foi empregado por Fillmore (1979, *apud* Tagnin, 1998) para designar “o conjunto dos elementos linguísticos, cuja coocorrência não é explicada sintática ou semanticamente, mas sim pelo uso”. O autor também chama a atenção para o fato de que a Convencionalidade está intimamente relacionada à fluência em uma língua, ou seja, o desconhecimento dessas unidades convencionais faz com que o aprendiz não se comunique de maneira eficiente.

Desse modo, o termo Convencionalidade também está relacionado à Fraseologia, área Linguística, voltada à investigação de combinações lexicais recorrentes. Entre elas, estão as colocações, as coligações, as expressões idiomáticas, os binômios, os provérbios, entre outros, ou seja, as expressões fixas que usamos no dia a dia.

Portanto, segundo Orenha-Ottaiano (2004), a Fraseologia é um recorte da Convencionalidade, como mostra a figura abaixo:



Figura 1: Esquema referente à convencionalidade e à fraseologia (ORENHA-OTTAIANO, 2004, p. 13)

Como já foi mencionado, a Fraseologia abrange muito mais do que apenas as coligações, as expressões idiomáticas, os provérbios. As colocações, um dos tipos de fraseologismos, são o alvo de nossa pesquisa e, por esse motivo, nos concentraremos especificamente nelas.

2.1 As colocações.

As colocações estão inseridas no âmbito da Fraseologia e, por conseguinte, da convencionalidade. Esse termo foi pela primeira vez empregado por J. R. Firth (1957) para designar casos de coocorrência léxico-sintática, ou seja,

combinações que já estão consagradas em uma dada língua. Nesselhauf (2005) caracteriza as colocações como um grupo de expressões cuja importância na linguagem tem sido cada vez mais reconhecida nos últimos anos. Este grupo de expressões tem sido chamado de unidades pré-fabricadas, casas pré-fabricadas, unidades fraseológicas, pedaços, multi-unidades, ou uma sequência de fórmulas.

A esse respeito, Sinclair (1987) estabelece dois princípios organizadores da língua: o “princípio de livre escolha” (*the open-choice principle*) e o “princípio idiomático” (*the idiom principle*). O princípio da livre escolha é uma maneira de ver o texto como um grande número de escolhas complexas. Sempre quando uma frase está para ser completada, abre-se uma vasta gama de escolhas e a única restrição é a gramaticalidade. Isto é provavelmente a maneira normal de ver e descrever a língua, frequentemente chamada *slot-and-filler model*, ou seja, o texto é uma série de lacunas que têm que ser preenchidas. Em cada lacuna pode ocorrer praticamente qualquer palavra, dado que a língua supostamente opera simultaneamente em vários níveis, como o nível fonológico, morfológico, sintático, semântico e pragmático (TAGNIN, 1989).

Os níveis que mais nos interessam, nesta pesquisa, são o nível sintático e o nível semântico. O sintático é aquele que trata da ordem das palavras e, conforme exemplo citado por Orenha-Ottaiano (2004), na língua inglesa é comum usar o adjetivo anteposto ao termo que ele dá característica, enquanto que, na língua portuguesa, é possível usarmos o adjetivo anteposto ou posposto ao termo caracterizado, mas que, no entanto, a posição desse adjetivo pode acarretar uma mudança de significado. Como exemplo, a autora cita as colocações em inglês *a big business* e *a large business*; e em português, as colocações “um grande negócio” e um “negócio grande, amplo”. Observar que, em português, a posição do adjetivo “grande” mudou o significado do substantivo “negócio”, o primeiro trata-se de um negócio que lucrara muito dinheiro, e o segundo um negócio que gastará muito dinheiro, por se tratar de um negócio amplo. Já, em inglês, para dar essas características ao termo “*business*” é necessário trocar o adjetivo “*big*”, que dá o significado de um negócio rentável, por “*large*”, que mostra uma ideia de um negócio amplo. O nível semântico, por sua vez, mostra que não há nenhuma relação semântica que motive uma colocação a ter determinados elementos combinados, a única explicação é a convenção na língua.

As palavras não ocorrem aleatoriamente e o princípio da livre escolha não é responsável por todas as restrições. Não seríamos capazes de produzir um texto apenas por procedermos de acordo com o princípio da livre escolha. Neste ponto é que entra em cena o princípio idiomático: o falante tem à sua disposição um grande número de grupos de palavras pré-construídos,

apesar de estes poderem apresentar alguma variação da ordem das palavras. Nesse contexto se inserem as colocações e outros fraseologismos que foram mencionados.

As colocações apresentam uma grande dificuldade para um falante não nativo. Apesar de esse falante conseguir compreender uma colocação, para ele é complicadíssimo criá-la, pois as colocações, como foi exposto, já estão consagradas na língua e, se o aprendiz não tem o conhecimento dessas colocações, não conseguirá usá-las e acabará cometendo erros influenciado pela língua mãe, como em português a combinatória “fazer um cheque”. O aluno que pode crer que sua colocação correspondente em inglês seja *make a cheque*, o que soaria estranho para um falante nativo, pois usariam, na realidade, a combinatória *write a cheque*.

Orenha-Ottaiano (2004) discute alguns requisitos necessários, propostos por Tagnin (1999), para que uma combinação de palavras possa ser considerada uma colocação. O primeiro desses é a recorrência, ou seja, é necessário que a combinação seja recorrente, aparecendo mais de uma vez; o segundo requisito é a não-idiomaticidade, enfatizando que a combinação deve ter um significado composicional e não idiomático, ou seja, o sentido da combinação pode ser deduzido do significado de cada um de seus elementos; o terceiro é coesão, mostrando que é necessário que haja uma ligação muito forte entre seus elementos, muito mais forte do que se esperaria de uma combinação qualquer; a quarta é a restrição contextual, referindo-se à probabilidade de que a combinação ocorra dentro de um conceito específico; e, por fim, a coocorrência arbitrária entre seus elementos, ou seja, não há uma razão semântica que explique tal combinação.

Quanto à composição das colocações, elas são constituídas de uma base e de um colocado, de acordo com a terminologia de Hausmann (1985, *apud* ORENHA-OTTAIANO 2004). A base e o colocado não possuem o mesmo valor semântico na combinação, pois há uma hierarquia entre esses elementos, pois um determina e o outro é determinado. Aquele que determina é chamado de base, e o determinado de colocado.

Como já foi mencionado, a produção das colocações representa um problema para os aprendizes de uma LE, e estes problemas são maiores em relação àqueles que apresentam com o uso do vocabulário em geral e, por essa razão, acabam usando menos colocações do que os falantes nativos. Segundo Nesselhauf (2005), o uso das colocações pelos aprendizes é de apenas 25%, isso porque os aprendizes não têm consciência nem das restrições colocacionais, nem do potencial combinatório dos itens lexicais. Neste contexto, Fillmore (1997) chama os aprendizes de “falantes ingênuos” e, nessa “ingênuo” con-

cepção de linguagem, o aprendiz acredita que pode combinar as palavras de maneira livre e aleatória.

Pelos motivos expostos acima, fica evidente não só a importância, mas a necessidade de se abordar colocações em sala de aula, pois quanto maior for o repertório de colocações aprendidas, mais a comunicação pode ser agilizada. Esse domínio facilita a comunicação já que, uma vez adquiridas e armazenadas, o discurso dá-se com maior rapidez e fluência.

Esta pesquisa, futuramente, poderá ser aplicada em sala de aula de língua inglesa, proporcionando aos alunos contato não somente com as colocações, mas com os fraseologismos em geral.

3. Metodologia de pesquisa.

Para a construção do *corpus* de estudo (doravante CE), utilizamos as transcrições do seriado de TV *Grey's Anatomy*, um drama médico norte-americano exibido no horário nobre da rede ABC. A partir dele, foi possível extrair várias colocações na área médica, além das colocações da língua geral. Para verificar a frequência de coocorrência na língua, utilizamos o *COCA (Corpus of Contemporary American English)*, corpus on-line que traz um vasto número de textos do inglês americano. Os dados do CE foram armazenados em arquivos, convertidos em formato *txt*, e armazenados em pastas, para que pudessem ser utilizados pelo programa computacional *WordSmith Tools* (SCOTT, 2004).

Ter em mãos um *corpus* representativo já é um ponto importante para uma pesquisa linguística, contudo, de nada vale se não tivermos uma boa ferramenta de investigação para realizar as buscas, fazer o levantamento das frequências etc. Dessa forma, a fim de auxiliar na seleção das colocações, utilizamos o programa acima mencionado, que atendeu nossas necessidades no âmbito desta pesquisa. É preciso notar que esse programa não interpreta os dados, já que essa função cabe ao pesquisador. O programa simplesmente disponibiliza os dados, para que sejam selecionados, analisados e interpretados pelo pesquisador.

Desenvolvido por Mike Scott (1997), o *WordSmith Tools* é um programa que contém um conjunto de ferramentas integradas que possibilita a observação das palavras dentro de um contexto, a saber: a) *Wordlist* (Lista de Palavras), b) *Concord* (Concordância) e c) *Keywords* (Palavras-Chave). A título de ilustração, exibimos abaixo as figuras geradas pelas ferramentas do *WordSmith Tools*, a partir dos dados de nosso CE:

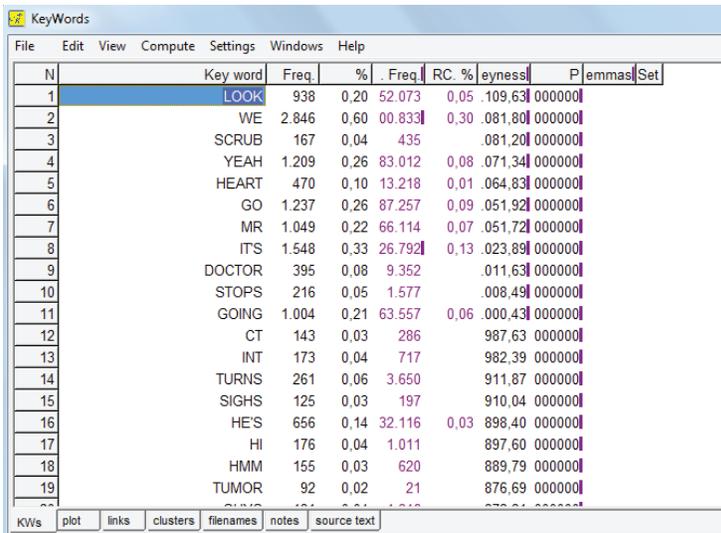
a) *WordList* (Figura 2): apresenta o número de palavras presentes nas sete temporadas do seriado de TV *Grey's Anatomy*, com 2.200.305 palavras, em destaque:

N	Overall	1	2	3	4	5	6	7
text file	Overall	2.txt	3.txt	4.txt	5.txt	6.txt	7.txt	8.txt
file size	7.923.663	50.676	45.798	35.968	38.548	37.238	40.636	35.990
tokens (running words) in text	2.200.305	9.035	8.178	6.055	6.768	6.378	6.865	6.127
tokens used for word list	1.567.581	8.975	8.163	6.012	6.729	6.346	6.837	6.104
types (distinct words)	19.568	1.433	1.378	1.191	1.225	1.167	1.349	1.349
type/token ratio (TTR)	1	16	17	20	18	18	20	22
standardised TTR	26,24	37,14	36,97	38,80	37,08	37,55	39,88	41,87
standardised TTR std. dev.	67,95	53,91	54,05	50,99	51,97	52,15	50,94	48,88
standardised TTR basis	1.000,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
mean word length (in characters)	3	4	4	4	4	4	4	4
word length std. dev.	1,82	2,17	2,10	2,05	2,16	2,02	2,06	2,08
sentences	143.313,00	148,00	928,00	966,00	962,00	992,00	053,00	942,00
mean (in words)	11	8	9	6	7	6	6	6
std. dev.	10,25	6,43	7,31	4,72	5,63	5,05	4,84	4,57
paragraphs	146,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
mean (in words)	10.737	8.975	8.163	6.012	6.729	6.346	6.837	6.104
std. dev.	4.437,41							
headings								
mean (in words)	0	0	0	0	0	0	0	0
std. dev.								
sections	146,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
mean (in words)	10.737	8.975	8.163	6.012	6.729	6.346	6.837	6.104
std. dev.	4.437,41							
numbers removed	632.724,00	60,00	15,00	43,00	39,00	32,00	28,00	23,00

Figura 2: Lista de Palavras do CE com dados estatísticos.

b) *Keywords* (figura 3): lista de palavras-chave obtida a partir do contraste da lista de palavras do *corpus* de referência (doravante CR), para contrastar com o CE. No caso de nosso estudo, usamos a lista de palavras do BNC¹ (*British National Corpus*), por se tratar de um *corpus* de grande porte, com 2.530.849 palavras, e a lista de palavras do CE, gerada pelo *WordList*. O resultado do contraste é uma lista de palavras-chave, ou seja, palavras cujas frequências são significativamente diferentes, em termos estatísticos, no CE e no CR:

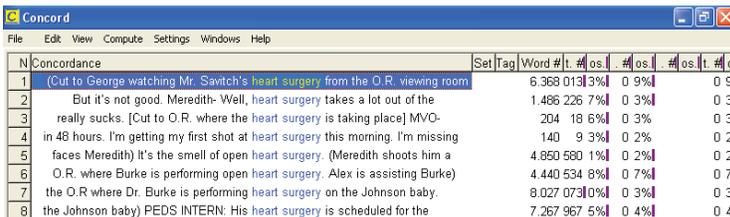
1 Disponível para download em <http://www.lexically.net/wordsmith/support/extras.html>.



N	Key word	Freq.	%	. Freq.	RC. %	eyness	P	emmas	Set
1	LOOK	938	0,20	52.073	0,05	.109,63	0,000000		
2	WE	2.846	0,60	00.833	0,30	.081,80	0,000000		
3	SCRUB	167	0,04	435		.081,20	0,000000		
4	YEAH	1.209	0,26	83.012	0,08	.071,34	0,000000		
5	HEART	470	0,10	13.218	0,01	.064,83	0,000000		
6	GO	1.237	0,26	87.257	0,09	.051,92	0,000000		
7	MR	1.049	0,22	66.114	0,07	.051,72	0,000000		
8	ITS	1.548	0,33	26.792	0,13	.023,89	0,000000		
9	DOCTOR	395	0,08	9.352		.011,63	0,000000		
10	STOPS	216	0,05	1.577		.008,49	0,000000		
11	GOING	1.004	0,21	63.557	0,06	.000,43	0,000000		
12	CT	143	0,03	286		.987,63	0,000000		
13	INT	173	0,04	717		.982,39	0,000000		
14	TURNS	261	0,06	3.650		.911,87	0,000000		
15	SIGHS	125	0,03	197		.910,04	0,000000		
16	HE'S	656	0,14	32.116	0,03	.898,40	0,000000		
17	HI	176	0,04	1.011		.897,60	0,000000		
18	HMM	155	0,03	620		.889,79	0,000000		
19	TUMOR	92	0,02	21		.876,69	0,000000		

Figura 3: Lista de Palavras-Chave do CE.

c) *Concord* (figura 4): gera uma lista de concordâncias (*concordance*), que possibilitam ao pesquisador buscar a palavra ou expressão, acompanhada das palavras que ocorrem junto com a palavra de busca no *corpus*, para que, assim, o pesquisador possa fazer uma análise mais rápida e, ao mesmo tempo, detalhada. Na figura a seguir, observamos as linhas de concordância para a colocação *heart surgery* e o contexto em que ela está situada. Vemos que colocação pesquisada aparece centrada, destacada e alinhada:



N	Concordance	Set	Tag	Word #	t.	#	os	#	os	#	os	t.	#	c	
1	(Cut to George watching Mr. Sawitch's heart surgery from the O.R. viewing room			6.368	013	3%	0	9%						0	9
2	But it's not good. Meredith- Well, heart surgery takes a lot out of the			1.486	226	7%	0	3%						0	3
3	really sucks. [Cut to O.R. where the heart surgery is taking place] MVO-			204	18	6%	0	3%						0	3
4	in 48 hours. I'm getting my first shot at heart surgery this morning. I'm missing			140	9	3%	0	2%						0	2
5	faces Meredith) It's the smell of open heart surgery. (Meredith shoots him a			4.850	580	1%	0	2%						0	2
6	O.R. where Burke is performing open heart surgery. Alex is assisting Burke)			4.440	534	8%	0	7%						0	7
7	the O.R. where Dr. Burke is performing heart surgery on the Johnson baby.			8.027	073	0%	0	3%						0	3
8	the Johnson baby) PEDS INTERN: His heart surgery is scheduled for the			7.267	967	5%	0	4%						0	4

Figura 4: Linhas de concordância para a expressão *heart surgery*.

Para esta investigação, também utilizamos a lista de colocados (*collocates*), gerada na aba da própria lista de concordâncias da ferramenta *Concord*, ao clicar no ícone *collocates*. Essa ferramenta teve grande importância neste trabalho, uma vez que nos possibilitou ter acesso a uma lista de colocados, ou seja, uma lista de palavras que aparecem ao redor da palavra de busca, em

posições determinadas (esquerda, direita, por exemplo), permitindo examinar os padrões fixos ou unidades fraseológicas. A figura abaixo mostra uma lista de colocados, gerada a partir do termo *heart* (coração). A coluna *centre* (centro) indica a posição da palavra *heart*, e é possível notar, na primeira linha, que sua frequência no CE, que é de 470 vezes, conforme destaque:

N	Word	With	elation	Total	tal	Left	Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	HEART	Heart	0.000	474	2	2	1	0	0	1	0	91	470	0	1	0	0	1
2	THE	Heart	0.000	158	120	28	11	11	4	13	9	0	0	0	7	9	1	11
3	A	Heart	0.000	140	122	18	9	9	9	26	69	0	1	7	7	3	1	
4	TO	Heart	0.000	89	52	37	4	9	29	8	2	0	3	4	12	13	5	
5	AND	Heart	0.000	69	35	34	9	5	9	11	1	0	17	4	8	4	1	
6	YOU	Heart	0.000	63	47	16	14	18	10	3	2	0	4	3	6	2	1	
7	IS	Heart	0.000	55	16	39	4	8	2	2	0	0	17	18	1	3	0	
8	MY	Heart	0.000	54	45	9	4	3	2	6	30	0	1	3	1	3	1	
9	HIS	Heart	0.000	52	47	5	0	1	0	2	44	0	1	2	1	1	0	
10	S	Heart	0.000	50	40	10	2	4	3	1	30	0	5	3	2	0	0	
11	I	Heart	0.000	44	35	9	20	7	8	0	0	0	3	2	1	2	1	
12	IN	Heart	0.000	36	20	16	1	3	2	13	1	0	5	4	2	3	2	
13	HER	Heart	0.000	35	35	0	2	1	3	1	28	0	0	0	0	0	0	
14	OF	Heart	0.000	35	26	9	2	9	1	11	3	0	1	0	0	6	2	
15	FOR	Heart	0.000	32	19	13	2	3	7	7	0	0	8	2	0	3	0	
16	GET	Heart	0.000	30	24	6	0	0	3	21	0	0	0	0	1	1	4	
17	THIS	Heart	0.000	30	23	7	0	1	3	1	18	0	4	2	0	1	0	
18	ON	Heart	0.000	28	17	11	2	1	3	11	0	0	1	6	4	0	0	
19	YOUR	Heart	0.000	28	25	3	0	0	0	4	21	0	0	1	0	1	1	
20	HAVE	Heart	0.000	27	22	5	6	6	5	5	0	0	0	2	1	1	1	
21	MONITOR	Heart	0.000	27	0	27	0	0	0	0	0	0	27	0	0	0	0	
22	SURGERY	Heart	0.000	25	2	23	1	0	1	0	0	0	21	1	0	1	0	
23	THAT	Heart	0.000	25	20	5	1	5	3	5	6	0	3	0	0	2	0	
24	WITH	Heart	0.000	21	14	7	2	0	7	5	0	0	1	3	1	2	0	
25	ITZIE	Heart	0.000	21	17	4	6	5	2	2	2	0	2	1	0	1	0	

Figura 5: Lista, parcial, de collocates para o termo *heart*.

Segundo a figura, as colunas “L1”, “L2”,... “L5” (L = *left*), mostram a posição das palavras que estão à esquerda do termo de busca. Notamos que na segunda linha, o artigo *the*, ocorre 91 vezes à esquerda da palavra *heart*, formando *the heart*.

As colunas “R1”, “R2”, ..., “R5” (R = *right*), estão relacionadas às palavras que estão a direita da palavra de busca. No item 22, notamos que o termo *surgery* (cirurgia) ocorre 21 vezes à direita da palavra de busca, formando, assim, a colocação *heart surgery*. Desse modo, a lista de colocados pode auxiliar-nos obter uma melhor visualização e extração dos padrões colocacionais de um dado termo de busca.

Para melhor organizar nossa pesquisa, elaboramos uma lista de colocados em uma tabela, organizadas em blocos de frequência:

1	Word	With	Relation	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1
2	YOU	heart	0	63	47	16	14	18	10	3	2	0	4
3	IS	heart	0	55	16	39	4	8	2	2	0	0	17
4	MY	heart	0	54	45	9	4	3	2	6	30	0	1
5	HIS	heart	0	52	47	5	0	1	0	2	44	0	1
6	S	heart	0	50	40	10	2	4	3	1	30	0	5
7	I	heart	0	44	35	9	20	7	8	0	0	0	3
8	IN	heart	0	36	20	16	1	3	2	13	1	0	5
9	HER	heart	0	35	35	0	2	1	3	1	28	0	0
10	OF	heart	0	35	26	9	2	9	1	11	3	0	1
11	FOR	heart	0	32	19	13	2	3	7	7	0	0	8
12	GET	heart	0	30	24	6	0	0	3	21	0	0	0
13	THIS	heart	0	30	23	7	0	1	3	1	18	0	4
14	ON	heart	0	28	17	11	2	1	3	11	0	0	1
15	YOUR	heart	0	28	25	3	0	0	0	4	21	0	0
16	HAVE	heart	0	27	22	5	6	6	5	5	0	0	0
17	MONITOR	heart	0	27	0	27	0	0	0	0	0	0	27
18	SURGERY	heart	0	25	2	23	1	0	1	0	0	0	21
19	THAT	heart	0	25	20	5	1	5	3	5	6	0	3
20	WITH	heart	0	21	14	7	2	0	7	5	0	0	1
21	IZZIE	heart	0	21	17	4	6	5	2	2	2	0	2
22	HEART	heart	0	474	2	2	1	0	0	1	0	470	0
23	THE	heart	0	158	130	28	11	11	4	13	91	0	0
24	A	heart	0	140	122	18	9	9	9	26	69	0	1
25	TO	heart	0	89	52	37	4	9	29	8	2	0	3

Figura 6: Lista colocado, para o termo *heart*.

Na figura acima, podemos notar a mesma estrutura exibida anteriormente. A coluna *centre* mostra onde está situada o termo *heart*, e nas colunas laterais onde estão situadas as palavras, à direita “R1”, “R2”, ..., “R5” (R = right), ou à esquerda (“L1”, “L2”, ..., “L5” (L = left), do termo em questão. Nas linhas 17 e 18, do lado direito (R1), encontramos os substantivos *monitor* (monitor) e *surgery* (cirurgia), que ocorrem, respectivamente, 27 e 21 vezes, combinado com o termo *heart*, formando, assim, as colocações, *heart monitor* e *heart surgery*.

Na figura 7, observamos que o verbo *beat* (bater), na linha 30, ocorre 3 vezes à esquerda e 5 vezes à direita do termo *heart*, formando as colocações *beat heart* e *heart beat*, respectivamente.

Ainda é possível notar, nas linhas 36, 40, 45 e 47, os substantivos *rate* (frequência), *attack* (ataque), *patient* (paciente) e *transplant* (transplante), que ocorrem, 11, 13, 5 e 12, respectivamente, formando as colocações *heart rate*, *heart attack*, *heart patient* e *heart transplant*.

26	AND	heart	0	69	35	34	9	5	9	11	1	0	17
27	HE	heart	0	20	18	2	4	8	6	0	0	0	0
28	IT	heart	0	19	7	12	1	4	1	1	0	0	0
29	BURKE	heart	0	19	17	2	5	6	2	4	0	0	0
30	BEATING	heart	0	17	3	14	0	0	0	0	3	0	5
31	T	heart	0	17	9	8	1	1	7	0	0	0	0
32	GEORGE	heart	0	16	10	6	3	2	3	2	0	0	2
33	NEW	heart	0	16	16	0	0	0	0	0	16	0	0
34	NOT	heart	0	16	11	5	3	0	7	1	0	0	0
35	HAD	heart	0	15	12	3	4	1	3	4	0	0	2
36	RATE	heart	0	15	0	15	0	0	0	0	0	0	15
37	OPEN	heart	0	15	12	3	0	0	1	0	11	0	1
38	STARTS	heart	0	14	3	11	0	1	2	0	0	0	0
39	GETTING	heart	0	14	13	1	2	0	3	8	0	0	0
40	ATTACK	heart	0	13	0	13	0	0	0	0	0	0	13
41	WE	heart	0	13	12	1	3	6	3	0	0	0	0
42	WILL	heart	0	12	5	7	1	1	3	0	0	0	6
43	BEEPING	heart	0	12	0	12	0	0	0	0	0	0	0
44	CAN	heart	0	12	7	5	0	2	5	0	0	0	2
45	PATIENT	heart	0	12	3	9	0	1	0	2	0	0	5
46	OUT	heart	0	12	4	8	0	1	1	2	0	0	1
47	TRANSPLANT	heart	0	12	0	12	0	0	0	0	0	0	12
48	DENNY	heart	0	11	8	3	1	2	2	3	0	0	2
49	FROM	heart	0	11	3	8	0	1	0	2	0	0	3

Figura 7: Lista colocados, para o termo *heart*.

Na figura 11, podemos notar, nas linhas 85 e 97, que os substantivos *valve* (válvula) e *failure* (insuficiência) ocorrem 4 e 6 vezes, respectivamente e formam, dessa maneira, as colocações *heartvalve* e *heartfailure*.

74	THEY	heart	0	8	6	2	3	2	1	0	0	0	0
75	STOP	heart	0	7	4	3	0	0	0	4	0	0	1
76	LOOKING	heart	0	7	6	1	0	2	3	0	1	0	1
77	BEAT	heart	0	7	0	7	0	0	0	0	0	0	6
78	WANT	heart	0	7	4	3	2	0	0	2	0	0	0
79	HIM	heart	0	7	7	0	1	1	3	2	0	0	0
80	I'M	heart	0	7	6	1	1	0	5	0	0	0	0
81	STILL	heart	0	6	1	5	0	0	1	0	0	0	0
82	THERE	heart	0	6	3	3	3	0	0	0	0	0	0
83	TODAY	heart	0	6	1	5	0	0	1	0	0	0	2
84	START	heart	0	6	2	4	0	1	0	1	0	0	0
85	VALVE	heart	0	6	0	6	0	0	0	0	0	0	4
86	THEN	heart	0	6	4	2	1	0	1	1	1	0	1
87	WHAT	heart	0	6	4	2	1	2	1	0	0	0	0
88	RE	heart	0	6	4	2	0	2	0	2	0	0	0
89	DIDN	heart	0	6	6	0	1	5	0	0	0	0	0
90	CHEST	heart	0	6	3	3	0	2	1	0	0	0	0
91	CRISTINA	heart	0	6	6	0	3	2	1	0	0	0	0
92	KNOW	heart	0	6	4	2	0	1	3	0	0	0	0
93	IT'S	heart	0	6	3	3	0	1	2	0	0	0	0
94	BY	heart	0	6	1	5	0	0	0	0	1	0	3
95	BLOOD	heart	0	6	2	4	0	1	1	0	0	0	0
96	GETS	heart	0	6	5	1	0	0	1	4	0	0	1
97	FAILURE	heart	0	6	0	6	0	0	0	0	0	0	6

Figura 8: Lista colocados, para o termo *heart*.

Notamos, na figura 12, que as linhas 108 e 114 apresentam o substantivo, *disease* (doença) coocorrendo 5 vezes à direita do termo *heart*, e o substantivo- *muscle* (músculo) coocorrendo 1 vez à direita e 1 outra vez à esquerda, gerando as colocações *heartdisease*, *muscleheart* e *heartmuscle*.

98	ONE	heart	0	6	6	0	0	3	1	2	0	0	0
99	TOO	heart	0	5	1	4	1	0	0	0	0	0	0
100	EMILY	heart	0	5	5	0	0	0	0	5	0	0	0
101	FIRE	heart	0	5	0	5	0	0	0	0	0	0	0
102	DONOR	heart	0	5	5	0	1	0	0	0	4	0	0
103	DOWN	heart	0	5	1	4	0	1	0	0	0	0	0
104	WOULD	heart	0	5	3	2	1	2	0	0	0	1	0
105	BAD	heart	0	5	5	0	0	0	0	0	5	0	0
106	BABY	heart	0	5	2	3	0	1	1	0	0	0	0
107	COULD	heart	0	5	3	2	0	0	3	0	0	0	0
108	DISEASE	heart	0	5	0	5	0	0	0	0	0	5	0
109	DENNY'S	heart	0	5	5	0	1	0	0	0	4	0	0
110	DECANNULATED	heart	0	5	5	0	0	0	0	5	0	0	0
111	NEEDS	heart	0	5	2	3	0	1	0	0	1	0	3
112	RIGHT	heart	0	5	0	5	0	0	0	0	0	1	0
113	MUCH	heart	0	5	2	3	0	2	0	0	0	0	0
114	MUSCLE	heart	0	5	3	2	0	0	2	0	1	0	1
115	RICHARD	heart	0	5	5	0	1	1	2	1	0	0	0
116	PUMP	heart	0	5	2	3	0	1	0	1	0	0	2
117	PADDLES	heart	0	5	4	1	1	2	0	1	0	0	1
118	NO	heart	0	5	3	2	2	0	1	0	0	0	0
119	OR	heart	0	5	1	4	0	1	0	0	0	2	0
120	TAKE	heart	0	5	2	3	0	0	2	0	0	0	0
121	GLASS	heart	0	5	5	0	0	1	4	0	0	0	0
122	FLAT	heart	0	5	0	5	0	0	0	0	0	0	0

Figura 9: Lista colocados, para o termo *heart*

Seguindo os mesmos padrões acima apresentados, passamos a armazenar as colocações extraídas, para que pudessem compor o glossário.

4. O glossário.

Carvalho (2007) define glossário como um repertório de unidades lexicais de uma dada especialidade, que não possui pretensão de exaustividade. Dessa maneira, o glossário compilado nesta pesquisa é bilíngue, composto por colocações e colocações especializadas, retiradas de nosso CE, nas direções tradutórias inglês-português.

O glossário dispõe de quatro colunas, organizadas em ordem alfabética, sendo uma com a entrada em inglês (colocação + exemplo retirado do CE), outra com sua respectiva tradução (correspondente em português + exemplo em português). Cabe ressaltar, que os exemplos em português foram retirados do *corpus* paralelo, ou seja, todos são traduções dos exemplos em inglês; outra com a sua frequência no CE e, a última, com a sua frequência no COCA. Vale lembrar, que todos os exemplos, tanto em português quanto em inglês, foram retirados do CE.

Para este estudo, utilizamos o dicionário *The Oxford Dictionary of American English for learners*; e os sites de tradução <http://www.linguee.com/>, e www.tradukka.com onde também verificamos outras possibilidades tradutórias e as comparamos com as legendas em português do CE. Em seguida, apresentamos uma mostra de nosso glossário de colocações especializadas na área médica e na língua geral.

Artery		Freq. CE	Freq. COCA
Carotid artery Eg: Your dad has a 90% blockage in his right carotid artery.	Artéria carótida Ex: Seu pai está com 90% da artéria carótida direita entupida.	2	258
Cerebral artery Eg: Is that the middle cerebral artery?	Artéria cerebral Ex: Esta é a artéria cerebral média?	2	12
Coronary artery Eg: Okay everyone. It's actually a coronary artery aneurysm.	Artéria coronária Ex: Ok todos. Isto é realmente um aneurisma na artéria coronária.	14	441
Femoral artery Eg: Laceration to the superficial femoral artery.	Artéria femoral Ex: Laceração da artéria femoral superficial.	4	54
Pulmonary artery Eg: The coronary artery is attached to the pulmonary artery not the aorta.	Artéria pulmonária Ex: A artéria coronária é ligada à artéria pulmonar não a aorta.	9	56
Renal artery Eg: The kid had a renal artery thrombosis.	Artéria renal Ex: O garoto teve uma trombose da artéria renal.	6	14
Uterine artery Eg: Come on. Let's get in there. Did you check the uterine artery?	Artéria uterina Ex: Vamos. Vamos chegar lá. - O senhor checkou a artéria uterina?	2	11

Blood		Freq. CE	Freq. COCA
Blood bank Eg: We need more O-negative. I'll call the blood bank.	Banco de sangue Ex: Precisamos de mais O-negativo. Vou ligar para o banco de sangue.	8	138
Blood pressure I'm just gonna take your blood pressure now, okay?	Pressão arterial Ex: Eu só vou medir a sua pressão arterial, ok?	46	6015
Blood test Eg: I ordered a blood test and a bio-physical test on the baby.	Exame de sangue Ex: Eu pedi um exame de sangue e um teste bio-físico no bebê.	10	726

Blood type Eg: Well, Jackson's blood type is O-positive.	Tipo sanguíneo Ex: Bem, o tipo sanguíneo de Jackson é O-positivo.	11	203
Blood vessel Eg: I wanna see every blood vessel and nerve I'm gonna screw up if I ever smoke again.	Vaso sanguíneo Ex: Eu quero ver todos os vasos sanguíneos e nervos que eu vou estragar se eu fumar novamente.	18	323

Brain		Freq. CE	Freq. COCA
Brain activity Eg: Okay, now we just wait for the brain activity in that area to stop.	Atividade cerebral Ex: Ok, agora só esperar a atividade cerebral naquela área parar.	7	298
Brain damage Eg: Clearly, he has brain damage.	Lesão cerebral Ex: Sem dúvida, ele tem uma lesão cerebral.	9	821
Brain dead Eg: Uh, my patient, Kendra Thomas. She's brain dead. Her parents want to keep her alive to have the baby.	Morte cerebral Ex: Uh, minha paciente, Kendra Thomas. Ela teve morte cerebral. Seus pais querem mantê-la viva para ter o bebê.	24	138
Brain surgery Eg: I hate surgery, especially brain surgery.	Cirurgia no cérebro Ex: Eu odeio cirurgia, especialmente cirurgia cerebral.	24	345
Brain tumor Eg: Benjamin O'Leary, 32. In for the removal of a brain tumor that's pushing on his frontal temporal lobe.	Tumor cerebral Ex: Benjamin O'Leary, 32 anos. Está pronto para a remoção de um tumor cerebral que está empurrando o seu lóbulo frontal temporal.	23	537

Chest		Freq. CE	Freq. COCA
Chest pain Eg: I need to know what time your chest pain started.	Dor no peito Ex: Eu preciso saber a que horas sua dor no peito começou.	20	478
Chest tube Eg: No breath sounds on the right. Set up a chest tube.	Dreno de tórax Ex: Não há nenhuma respiração à direita. Inicie um dreno de tórax.	45	22

Exam		Freq. CE	Freq. COCA
Pelvic exam Eg: Actually it's not a pap smear. It's a pelvic exam.	Exame pélvico Ex: Na verdade não é um exame de Papanicolaou. É um exame pélvico.	3	41
Physical exam Eg: Dr. Burke is doing a physical exam on her. George is there as well.	Exame físico Ex: Dr. Burke está fazendo um exame físico com ela.	5	213
Rectal exam Eg: I'm gonna have to give him a rectal exam.	Exame retal Ex: Eu vou ter que dar a ele um exame retal.	5	75
Vaginal exam Eg: No you don't. If someone doesn't want to submit to a vaginal exam they don't have to. It's the law.	Exame vaginal Ex: Se alguém não quer se submeter a um exame vaginal não é obrigado. É a lei.	2	5

Fracture		Freq. CE	Freq. COCA
Skull fracture Eg: I've got a patient with a depressed skull fracture and probably an inter-cranial bleed.	Fratura de crânio Ex: Eu tenho uma paciente com uma fratura de crânio e, provavelmente, um sangramento inter-cranial.	12	44
Tibial plateau fracture Eg: Meg whealon, age 28, Tibial plateau fracture caused by a bicycle accident.	Fratura do platô tibial Ex: Meg whealon, de 28 anos, fratura de planalto tibial causada por um acidente de bicicleta.	3	785

Spinal		Freq. CE	Freq. COCA
Spinal canal Eg: It's a tumor in her spinal canal. But the good news is we can operate. You have a 95% chance of fully recovering if we get you into surgery as soon as possible.	Canal medular Ex: É um tumor em seu canal vertebral. Mas a boa notícia é que podemos operar. Você tem uma chance de 95% da plena recuperação, se você entrar em cirurgia o quanto antes.	2	52

<p>Spinal cord Eg: We have been at this for four hours. Maybe he just injured his spinal cord and there's nothing to fix.</p>	<p>Medula Espinhal Ex: Nós estivemos nisso por quatro horas. Talvez ele apenas tenha machucado a medula espinhal assim não há nada para corrigir</p>	19	1213
<p>Spinal curvature Eg: Can I ask...I know I'm just here to remove a few ribs but... have you ever tried anything more radical to treat the spinal curvature?</p>	<p>Curvatura da coluna Ex: Posso perguntar... eu sei que estou aqui apenas para remover algumas costelas, mas ... você já tentou algo mais radical para o tratamento da coluna vertebral?</p>	2	14
<p>Spinal fusion Eg: Did you cut her heart open in your spinal fusion?</p>	<p>Fusão espinhal Ex: Você cortou o coração na sua fusão espinhal?</p>	3	51
<p>Spinal injury Eg: Sir, you could have a spinal injury. You need to stop fighting me and stay calm, okay?</p>	<p>Lesão medular Ex: Senhor, você pode ter uma lesão na coluna vertebral. Você precisa parar de lutar e manter a calma, ok?</p>	3	41
<p>Spinal shock Eg: Um, on my exam, she looks fine. No signs of spinal shock.</p>	<p>Choque medular Ex: Hum, no meu exame, ela parece bem. Sem sinais de choque medular.</p>	2	2
<p>Spinal fractures Eg: He has a few spinal fractures, but it looks like he's out of the woods</p>	<p>Fraturas da coluna vertebral Ex: Ele tem algumas fraturas na coluna vertebral, mas parece que ele está fora de perigo</p>	2	4
<p>Spinal surgery Eg: There's no way that this baby is going to survive spinal surgery.</p>	<p>Cirurgia da coluna vertebral Ex: Não tem nenhum jeito deste bebê sobreviver a uma cirurgia da coluna vertebral.</p>	4	45

Spinal tumor Eg: So um I'm looking for a neurosurgeon to consult on a fetal spinal tumor. You know anybody good?	Tumor espinhal Ex: Então, eu estou procurando um neurocirurgião para consultar sobre um tumor espinhal do feto. Você conhece alguém bom?	2	5
---	---	---	---

Traumatic		Freq. CE	Freq. COCA
Traumatic brain Eg: Traumatic brain injury could indicate anything from a concussion to intracranial hemorrhaging.	Trauma cerebral Ex: Lesão cerebral traumática pode indicar um abalo de hemorragia intracraniana.	2	276
Post-traumatic stress Eg: Post-traumatic stress is a real injury, Owen.	Estresse pós-traumático Ex: Estresse pós-traumático é uma lesão real, Owen.	3	581

Tube		Freq. CE	Freq. COCA
Endotracheal tube Eg: Now explain why you chose that size endotracheal tube.	Tubo endotraqueal Ex: Agora explique por que você escolheu esse tamanho de tubo endotraqueal.	4	67
Fallopian Tube Eg: You had an extra uterine pregnancy. Your left fallopian tube burst.	Trompa de falópio Ex: Você teve uma gravidez extrauterina. Sua trompa de falópio esquerda estourou.	2	57

Considerações finais.

Acreditamos que esta pesquisa possa colaborar com o desenvolvimento dos estudos em Linguística de Corpus, uma vez que se debruça sobre a análise do léxico, mais especificamente das colocações na língua inglesa, com base em um *corpus* paralelo. Ao compilar um CE, analisá-lo e interpretá-lo, à procura de combinatórias recorrentes na língua inglesa, o aluno aprendiz muda o foco do seu aprendizado, antes baseado nos compêndios gramaticais, e passa a visualizar uma forma mais eficaz de aprender uma língua. Ao fazer isso, ele se aproxima cada vez mais do falar nativo, pois está em contato com exemplos que aparecem com frequência na língua em análise e, desse modo, o aprendizado se torna mais prazeroso e eficaz.

Descobrimos que, ao introduzir na sala de aula resultados extraídos a partir de *corpora* eletrônicos, o professor muda o foco do seu ensino, antes baseado na gramática, e passa a visualizar uma forma mais eficiente e inovadora de ensinar uma LE.

É necessário ressaltar aqui a importância do uso de um glossário, tanto para alunos e professores quanto para tradutores em formações e poucos experientes, já que ele pode nos auxiliar no aprendizado de um vocabulário novo e de seus diferentes empregos, que muitas vezes não é conhecido por todos como, por exemplo, os termos da área médica, mostrados neste trabalho. Além disso, conforme já mencionamos, esta pesquisa aponta casos reais de uso da língua, aproximando o aprendiz cada vez mais da LE estudada.

Referências.

- AllSubs*. Disponível em: <<http://www.allsubs.org//search-subtitle/grey%27s+anatomy/20>>. Acesso em 03 de jul. de 2011.
- AUBERT, Henrik Aubert. *Introdução à metodologia da pesquisa terminológica bilingue*. São Paulo: Humanitas, 1996.
- BARROS, Lidia Almeida. *Curso Básico de Terminologia*. São Paulo. Editora da Universidade de São Paulo, 2004.
- BERBER SARDINHA, Tony. *Linguística de Corpus*. Barueri, Manole, 2004.
- _____. *Pesquisa em Linguística de Corpus com WordSmith Tools*. Campinas: Mercado de Letras, 2009.
- BIBER, Douglas; Conrad, Susan; Reppen, Randi. *Corpus Linguistics: investigating language structure and use*. Cambridge: Cambridge University Press, 1998.
- British National Corpus (BNC)*. Disponível em: <<http://www.lexically.net/wordsmith/support/extras.html>>. Acesso em 20 jan. 2011.
- CARVALHO, Estela Maria Faustino de. *Metodologia de construção de um glossário bilingüe com base em um corpus de domínio técnico*. 2007. 81. Dissertação (mestrado Estudos de Tradução)-Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina. Florianópolis, 2007.
- Corpus Brasileiro*. Disponível em: <<http://corpusbrasileiro.pucsp.br/x/>>. Acesso em: 19 jan. 2012.
- DAVIES, Mark. *The Corpus of Contemporary American English: 425 million words, 1990-present*. Disponível em: <<http://corpus.byu.edu/coca/>>. Acesso em: 20 jan. 2012.

- FILLMORE, Charles J. On fluency. In: _____ et al. *Individual differences in language ability and language behavior*. New York: Academic Press, 1979, p. 85-99.
- FIRTH, John Rupert. *Papers in Linguistics*. Oxford: Oxford University Press. 1957.
- HILL, Jimmy. Revising priorities: from grammatical failure to collocational success. In:
- LEWIS, Michael (Ed.). *Teaching collocations: further developments in the lexical approach*. London: Language Teaching Publications, 2000, p. 47-69.
- Hypnoweb*. Disponível em: <<http://grey-s-anatomy.hypnoweb.net/episodes/saison-1.62.3/>>. Acesso em: 03 de jul. de 2011.
- JOHNS, Tim. Should you be persuaded: Two samples of data-driven learning. *English Language Research Journal*, 2000.
- LEWIS, Michael. *Teaching Collocation: further developments in the Lexical Approach*. London: Language Teaching Publications, 2000.
- Linguee*. Disponível em: <<http://www.linguee.com/>> Acesso em: 23 jan. de 2012.
- MEYER, Charles. F. *English Corpus Linguistics: an introduction*. Cambridge: Cambridge, 2004.
- NESSSELHAUF, Nadja. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins Publishing Company, 2005.
- ORENHA-OTTAIANO, Adriane. *A compilação de um glossáriobilingüe de colocações, na área de jornalismo de Negócios, baseado em corpus comparável*. 2004. 246. Dissertação (Mestrado em Estudos Linguísticos e Literários)-FFLCH/USP, São Paulo, 2004.
- _____. *Unidades fraseológicas especializadas: colocações e colocações estendidas em contratos sociais e estatutos sociais traduzidos no modo juramentado e não-juramentado*. 2009. 282. Tese (Doutorado em Estudos Linguísticos)- IBILCE/UNESP, São José do Rio Preto, 2009.
- Oxford Collocations Dictionary For students of English*. Oxford. Oxford University Press, 2002.
- PAVEL, Silvia; NOLET, Diane. *Manual de Terminologia*. Disponível em: <http://www.translationbureau.gc.ca/pwgsc_internet/en/publications/documents/presport.pdf>. Acesso em 11 de jun de 2012.
- ROMER, Ute. Corpora and Language Teaching. In: LUDELING, Anke et al. *Corpus Linguistics: An International Handbook*. Alemanha: Walter de Gruyter, 2008, p. 112-131.

SCOTT, Michael. *WordSmith Tools*: version 4.0. Oxford: Oxford University Press, 2004.

SINCLAIR, John. *Corpus, concordance, collocation: Describing English language*. Oxford: Oxford University Press. 1991.

Tradukka. Disponível em: <www.tradukka.com> Acesso em: 23 jan. 2012

Webcorp. Disponível em: <<http://www.webcorp.org.uk/live/wlse.jsp>> Acesso em: 21 jan. de 2012.

Recebido em 4 de março de 2013.

Aceito em 5 de maio de 2013.