

**A COMPILAÇÃO DE ATIVIDADES DIDÁTICAS A PARTIR DE COLOCAÇÕES
ESPECIALIZADAS DA ÁREA CRIMINAL EXTRAÍDAS DO CORPUS *CSI – CRIME
SCENE INVESTIGATION***

Ariane Donizete Delgado Ribeiro CALDAS

Adriane ORENHA-OTTAIANO

Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)

Resumo: Por meio do arcabouço teórico-metodológico da Linguística de Corpus e da Fraseologia, apresentamos uma análise de colocações especializadas em inglês da área criminal a partir de um corpus paralelo, composto pelas transcrições do seriado criminal *CSI – Crime Scene Investigation*. Com base nesta análise, propomos atividades que favoreçam o aprimoramento do conhecimento léxico-fraseológico de aprendizes de inglês como língua estrangeira, no que diz respeito ao uso e à compreensão das colocações especializadas da área criminal. Pretende-se, com estas atividades, ressaltar a importância da Linguística de Corpus para a Fraseologia, e a relevância das duas áreas para o desenvolvimento de pesquisas e a elaboração de materiais didáticos, contribuindo, dessa maneira, para a área da Pedagogia do Léxico Baseada em Corpus. **Palavras-chave:** Linguística de Corpus. Pedagogia do Léxico Baseada em Corpus. Fraseologia. Colocação. Colocações Especializadas.

**THE COMPILATION OF PEDAGOGICAL ACTIVITIES BASED ON SPECIALIZED
COLLOCATIONS FROM THE CRIMINAL AREA EXTRACTED FROM *CSI – CRIME
SCENE INVESTIGATION* CORPUS**

Abstract: Based on the Corpus Linguistic and Phraseology theory and methods, we present the analysis of specialized collocations of the criminal area extracted from a parallel corpus composed of the transcriptions of the TV series *CSI – Crime Scene Investigation*. According to the results of the analysis, we will propose activities that may contribute to improve the phraseological lexicon knowledge of learners of English as a foreign language, regarding the use and comprehension of specialized collocations of the criminal area. With these activities, it is intended to highlight the importance of Corpus Linguistic to Phraseology, and the relevance of both areas to the development of research and the preparation of teaching materials, thus contributing to the area of Corpus-based Lexicon Pedagogy.

Keywords: Corpus Linguistic. Corpus-based Lexicon Pedagogy. Phraseology. Collocations. Specialized Collocations.

LA COMPILACIÓN DE ACTIVIDADES DIDÁCTICAS BASADAS EN COLOCACIONES ESPECIALIZADAS DEL ÁREA CRIMINAL EXTRAÍDAS DEL CORPUS CSI – *CRIME SCENE INVESTIGATION*

Resumen: A través del aporte teórico y metodológico de la Lingüística de Corpus y de la Fraseología, presentamos un análisis de las colocaciones especializadas en inglés en el área criminal, a partir de un corpus paralelo compuesto pelas transcripciones de la serie criminal CSI – *Crime Scene Investigation*. Sobre la base de este análisis, se proponen actividades que promuevan la mejora del conocimiento léxico-fraseológico de los estudiantes de inglés como lengua extranjera, en relación con el uso y la comprensión de las colocaciones especializadas en el área criminal. Se pretende, con estas actividades, hacer resaltar la importancia de la Lingüística de Corpus a la Fraseología y la pertinencia de las dos áreas para el desarrollo de la investigación y de materiales de enseñanza, contribuyendo de esta manera a la área de la Pedagogía del Léxico Basado en Corpus.

Palabras-clave: Lingüística de Corpus. Pedagogía del Léxico Basado en Corpus. Fraseología. Colocaciones. Colocaciones Especializadas.

INTRODUÇÃO

Devido à crescente procura por cursos de idiomas, mais especificamente cursos de língua inglesa, pesquisas que tratam do desenvolvimento e aprimoramento do ensino e da aprendizagem da língua inglesa, bem como de línguas estrangeiras de modo geral, também têm crescido muito nos últimos anos (WALKER, 2011; LAUFER; WALDMAN, 2011; DURRANT, 2015). Por meio dessas pesquisas, percebe-se o interesse crescente por parte de pesquisadores e professores pelo uso da Linguística de Corpus como auxiliar nesse processo.

Por ser uma disciplina que trabalha com amostras de usos reais da língua, a Linguística de Corpus tem despertado o interesse dos estudiosos para questões que, até então, eram pouco ou quase nada discutidas, e que, atualmente, tem se mostrado de grande importância para o trabalho em sala de aula, principalmente no que diz respeito ao ensino do léxico. A própria frequência de uso das unidades lexicais passou a ser considerada como algo importante a ser observado e apresentado aos alunos. Isso fica bastante visível ao analisarmos os materiais de ensino de idiomas que têm sido disponibilizados no mercado como, por exemplo, a coleção *Touchstone* (MCCARTHY; MCCARTEN; SANDIFORD, 2005), da editora *Cambridge University Press*, que apresenta indicações e notas sobre o léxico baseadas em *corpora*, ou seja, amostras

reais do uso da língua. Também podemos citar vários dicionários para aprendizes de inglês como língua estrangeira, tais como: *Oxford Student's Dictionary of English*, *Cambridge Dictionary of American English*, *Collins Cobuild Learner's Dictionary*, *Longman Dictionary of Contemporary English*, entre outros.

Estas novas obras, assim como o crescimento de pesquisas na área, chamam a atenção para aspectos lexicais e, principalmente, aspectos fraseológicos, que até então não tinham tanto destaque, tais como: as expressões idiomáticas (*Once in a blue moon*), as colocações (*blood pressure*, *blood evidence*) e os binômios (*step by step*; *safe and sound*). No que diz respeito às colocações, observamos um grande número de investigações que ultimamente tratam dessas unidades fraseológicas no ensino, na Lexicografia e na Tradução (ORENHA-OTTAIANO, 2009, 2015; FRITZING; HEID, 2009; WALKER, 2011; LAUFER; WALDMAN, 2011; ALMELA, 2011, entre outros), dada sua relevância na produção escrita e oral. Algumas pesquisas (ORENHA-OTTAIANO, 2012, 2015; TAGNIN, 2013) já mostraram que as colocações não apresentam, para o aprendiz de língua inglesa, grande desafio no que tange à compreensão, seja ela escrita ou oral, pois, a partir do momento em que ele conhece uma palavra ou mais, daquelas que compõem uma dada colocação, ele consegue, por meio do contexto em que está inserida, entender seu significado. O desafio no ensino das colocações está centrado na produção destas pelos estudantes, uma vez que eles tendem a produzi-las da mesma forma em que pensam na língua materna, ou seja, haverá uma grande interferência da língua materna na produção das colocações em uma língua estrangeira (ORENHA-OTTAIANO, 2004). Por exemplo, se um aluno brasileiro de nível A1, A2 ou, até mesmo B1, tivesse que produzir em inglês, de forma oral ou escrita, a colocação “fazer um elogio”, tenderia a dizer *make or do a compliment*, ao invés de usar a colocação adequada *pay a compliment*, influenciado por sua língua materna (Bom, se é “fazer”, deve ser *make or do!*). No entanto, ele não teria dificuldades em entender a colocação *pay a compliment* ao ouvi-la ou lê-la.

Pensando neste desafio, e considerando os estudos com foco nas colocações, a presente investigação, vinculada ao Grupo de Pesquisa “Pedagogia do Léxico, da Tradução e Linguística de Corpus”, cadastrado no Diretório de Grupos de Pesquisa do CNPq, e parte integrante do projeto de pesquisa guarda-chuva “A compilação de materiais didáticos e glossários especializados baseados em corpora e sua contribuição para uma Pedagogia do

Léxico e da Tradução”, apresenta a análise de colocações especializadas da área criminal. O referido campo foi escolhido para a presente pesquisa, em razão de se tratar de uma área que desperta o interesse de pessoas das mais variadas idades ao redor de todo mundo, graças à crescente criação e divulgação de seriados que abordam o cotidiano de investigadores e cientistas forenses, demandando, assim, tradutores que conheçam o léxico da área, bem como as colocações especializadas que fazem parte deste léxico.

As colocações que compõem esta pesquisa foram extraídas de um corpus paralelo composto pelas transcrições do seriado norte-americano *CSI – Crime Scene Investigation*, de forma a apresentar aos alunos de tradução as colocações com as quais poderão vir a trabalhar, preparando-os, desse modo, para um futuro mercado de trabalho. O objetivo final deste estudo é a proposta de atividades didático-pedagógicas que foquem o ensino e a aprendizagem das colocações da língua especializada da área criminal, considerando que estas atividades farão parte do *Online Corpus-based English Collocational Workbook*¹, em fase de compilação pela orientadora do projeto mencionado anteriormente.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, faremos uma breve revisão das teorias que embasam nossa pesquisa, ressaltando seus aspectos mais importantes. Trataremos, primeiramente, da Linguística de Corpus e da Terminologia. Em seguida, daremos enfoque às colocações da língua geral e especializadas, objeto de estudo de nosso trabalho. Trataremos, também, da Abordagem Lexical, que servirá de base para o desenvolvimento das atividades que serão propostas.

2.1 LINGUÍSTICA DE CORPUS E TERMINOLOGIA

Ao refletirmos acerca do *status* da Linguística de Corpus, é comum pensá-la, apenas, como uma metodologia, ou abordagem. No entanto, para delimitá-la como uma ou outra, é preciso “perceber que a Linguística de Corpus é um campo heterogêneo”² (MCENERY; HARDIE,

¹ Disponível em <http://www.gbd.institucional.ibilce.unesp.br/workbook/index>, conforme referências bibliográficas.

² “[...] realize that corpus linguistics is a heterogeneous field”.

2012, p. 1) e, dessa maneira, seu *status* poderá variar conforme o tipo de pesquisa a ser realizada. Os autores se referem à Linguística de Corpus (doravante LC) como um campo heterogêneo, devido ao fato de não haver um consenso entre os linguistas de corpus quanto à sua classificação, já que alguns a consideram como uma abordagem, outros como uma metodologia e, os Neo-Firthianos, conforme ressaltado pelos autores, a vêem como uma disciplina, os quais afirmam que “o próprio corpus incorpora a sua própria teoria da linguagem”³ (MCENERY; HARDIE, 2012, p. 6).

McEnery e Hardie (2012) não concordam com a ideia dos Neo-Firthianos, não considerando, assim, a LC como uma disciplina autônoma. Do mesmo modo, Biber (1998) trata a LC como uma abordagem, e Tognini-Bonelli (2002), ao abordar tal discussão, categoriza a LC como “uma nova abordagem filosófica para a investigação linguística”⁴ (TOGNINI-BONELLI, 2002, p. 1).

Assim, em nossa pesquisa, compartilhamos que a visão de que a LC é:

uma abordagem empírica, na qual, como todos os tipos de investigação científica, o ponto de partida são dados reais autênticos. O procedimento para descrever os dados que fazem uso de corpus é, portanto, indutivo, já que são afirmações de caráter teórico sobre a língua ou a cultura, as quais se chegaram a partir de observações das instâncias reais. A observação dos fatos de linguagem conduz à formulação de uma hipótese para explicar esses fatos; esta, por sua vez, leva a uma generalização baseada na evidência dos padrões utilizados na concordância; a última etapa é a unificação destas observações em uma declaração teórica.⁵ (TOGNINI-BONELLI, 2002, p. 2).

Com essa definição de Tognini-Bonelli, conseguimos relacionar o uso da LC em pesquisas de diversas áreas, como Ensino e Aprendizagem de Línguas, Estudos da Tradução,

³ “[...] that the corpus itself embodies its own theory of language”.

⁴ “[...] a new research enterprise and a new philosophical approach to linguistic enquiry”.

⁵ “[...] as an *empirical approach* in that, like all types of scientific enquiry, the starting point is actual authentic data. The procedure to describe the data that makes use of corpus is therefore *inductive* in that it is statements of a theoretical nature about the language or the culture which are arrived at from observations of the actual instances. The observation of language facts leads to the formulation of a hypothesis to account for these facts; this in turn leads to a generalization based on the evidence of the repeated patterns in the concordance; the last step is the unification of these observations in a theoretical statement.”

Fraseologia e Terminologia, por exemplo, uma vez que podemos analisar dados reais autênticos da língua com o auxílio desta abordagem empírica, em conjunto com os pressupostos teóricos das áreas com as quais associamos nossas pesquisas.

Os dados reais autênticos mencionados por Tognini-Bonelli (2002) são os dados que compõem o corpus escolhido para cada pesquisa. Entendemos o corpus como sendo uma

[...] coleção de textos considerados representativos de uma dada língua, reunidos de forma que possam ser usados para análise linguística. Normalmente, pressupõe-se que a língua armazenada em um corpus é de ocorrência natural, que é compilada de acordo com os critérios do projeto, com um propósito específico em mente, e com a pretensão de representar os ‘pedaços maiores do idioma’ selecionados de acordo com uma tipologia específica.⁶ (TOGNINI-BONELLI, 2002, p.2)

A partir da definição de corpus dada por Tognini-Bonelli, ressaltamos que, além de primar pelo uso de dados reais autênticos, conforme mencionado anteriormente, é preciso levar em consideração os objetivos que se pretendem alcançar na pesquisa, para que o corpus seja condizente com estes fins. Desse modo, no caso de nossa pesquisa, ao escolhermos o corpus *CSI – Crime Scene Investigation*, levamos em consideração o fato de ele ser representativo da área criminal, fato que nos permitirá a extração de colocações especializadas, as quais são consideradas, no caso de nossa pesquisa, as “porções da língua” (*language chunks*), além do fato de nosso corpus ter sido selecionado de acordo com uma tipologia específica, a saber, as legendas de um seriado da área criminal, atendendo, assim, a todas as exigências defendidas por Tognini-Bonelli.

Uma vez que nosso corpus é composto por textos extraídos de uma área específica, convém ressaltarmos alguns pontos importantes da Terminologia, disciplina que se ocupa do estudo dos termos, que serão considerados para o desenvolvimento de nossa análise.

⁶ “[...] a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis. Usually the assumption is that the language stored in a corpus is naturally-occurring, that is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology.”

De acordo com a Terminologia, os nódulos de buscas extraídos da lista de palavras-chave a partir de nosso corpus de estudo, gerada por uma ferramenta criada no âmbito da LC, são classificados como termos, e não como palavras, ou unidades lexicais simples, haja vista que “palavra e termo dissociam-se na medida em que a primeira possui significado e o termo evoca conceito” (KRIEGER, 2013, p. 31). Conforme Krieger (2013) explica, enquanto o significado da palavra pode gerar a polissemia lexical por ser variável, o conceito evocado pelo termo, por ser invariável, é independente do contexto enunciativo. Ou seja, os nódulos de busca por nós levantados, são termos, pois, independente do contexto no qual ocorrerem, estarão relacionados diretamente à área da investigação criminal, fazendo com que o leitor/ouvinte faça a relação entre o termo e a área privilegiada nesta pesquisa.

Convém ressaltar aqui que, em nossa pesquisa, entendemos *termo* como sendo “uma unidade lexical com um conteúdo específico dentro de um domínio específico”, assim como o define Barros (2004, p. 40)

Assim, em nossos estudos observamos haver uma interdisciplinaridade entre Terminologia e LC, uma vez que o suporte teórico da Terminologia nos auxiliará na extração de termos da área da investigação criminal, os quais serão analisados e descritos, também, pelo aporte teórico-metodológico proveniente da LC.

2.2 COLOCAÇÕES NA LÍNGUA GERAL E ESPECIALIZADA

O termo “colocação” tem sido utilizado por diversos autores, tais como Hausmann (1985), Sinclair (1991), Nesselhauf (2005), Fritzingler e Heid (2009), fato esse que pode ter contribuído para que diferentes definições acerca do termo “colocação” surgissem a partir da visão de cada um deles.

Para Hausmann (1984), por exemplo, as colocações são “produtos semi-acabados da língua”, sendo o aspecto mais importante das mesmas “[...] seu estado de disponibilidade mental como um todo, não como uma criação produzida para uma tal finalidade pelo falante.” (HAUSMANN, 1984). Hausmann acredita que apenas reutilizamos estes “produtos semi-

acabados da língua” ao nos comunicarmos, ou seja, que as colocações são pré-determinadas, semi-fixas, em nosso vocabulário, e que as reutilizamos conforme necessário.

Já Fritzingler e Heid (2009) afirmam que o termo “colocação” tem sido usado para se referir a uma vasta gama de fenômenos, sendo tratado com distinção pelos linguistas computacionais, e por aqueles que defendem o contexto. No entanto, os autores ressaltam que, para eles, colocação é entendida de acordo com a definição proposta por Bartsch (2004), a qual afirma que as “colocações são lexicalmente e/ou pragmaticamente coocorrências restritas recorrentes de pelo menos dois itens, os quais estão em uma relação sintática direta um com o outro”⁷ (BARTSCH, 2004, p.76). De acordo com os autores, esta definição baseia-se em três critérios, a saber: coseleção lexical, significância estatística de coocorrência e padrões sintáticos.

Pode-se dizer que o critério de coseleção lexical apontado por Fritzingler e Heid (2009) está intimamente ligado à distinção terminológica dos termos que compõem uma colocação, a saber, a *base* e o *colocado*, feita por Hausmann (1984), haja vista que, segundo os autores, a base é que seleciona seus colocados. Hausmann, por sua vez, entende que a base é determinante dentro da colocação, ou seja, é semanticamente autônoma, enquanto, o colocado, é o determinado na colocação, uma vez que é escolhido por uma dada base para formar a colocação.

A significância estatística de coocorrência está relacionada, por sua vez, a uma das características apontadas por Tagnin (1999) que devem ser observadas para a qualificação de uma colocação, a recorrência. Segundo a autora, uma combinação lexical deve apresentar frequência superior a 1 para ser considerada uma colocação, o que é aferido pelo critério de significância de coocorrência estatística mencionado por Fritzingler e Heid (2009), na definição de Bartsch (2004).

⁷ “Collocations are lexically and/or pragmatically constrained recurrent cooccurrences of at least two items which are in a direct syntactic relation with each other.”

Com relação à tipologia das colocações, ORENHA-OTTAIANO (2004) propõe uma classificação tipológica, a partir do que foi proposto por Hausmann (1985), conforme mostramos abaixo, com exemplos já retirados de nosso corpus de estudo:

Verbais

- Substantivo **base** + Verbo **colocado** = *ballistics confirms*
- Verbo **base** + Determinante + Substantivo **colocado** = *verify an alibi*
- Verbo **colocado** + Pronome + Substantivo **base** = *treats his body*

Nominais

- Substantivo **base** + Substantivo **colocado** = *arrest warrant*
- Substantivo **base** + Preposição + Substantivo **colocado** = *marks on the body*

Adjetiva

1. Adjetivo **colocado** + Substantivo **base** = *full arrest*

Adverbial

2. Verbo **base** + Advérbio **colocado** = *shot directly*
3. Advérbio **colocado** + Verbo **base** = *easy kill*

Até este momento, definimos as colocações da língua geral, entretanto, tendo em vista a especificidade de nosso corpus de estudo, trabalharemos com as colocações especializadas, o que nos leva a fazer aqui uma pequena diferenciação entre estas e as colocações da língua geral.

Bevilacqua (2004, 2005) dedica-se ao estudo das unidades fraseológicas especializadas (UFE). Em seu estudo, a autora afirma que há uma variedade na denominação dada a estas UFEs, visto que alguns autores as tratam como unidades fraseológicas, outros como colocações. Contudo, opta por denominá-las unidades fraseológicas especializadas, as quais são definidas como:

Unidades formadas por um núcleo eventivo, considerado como tal por ser de base verbal ou derivada de verbo (nominalização ou particípio), e por um núcleo terminológico (termo). Entre estes dois núcleos se

estabelecem relações sintáticas, mas especialmente semânticas, determinadas pelas propriedades do texto em que são utilizadas. Portanto, são unidades que se conformam no e pelo texto em que são utilizadas. Cumprem, tal como os termos, a função de representar e transmitir conhecimento especializado. (BEVILACQUA, 2005, p. 83).

Dois pontos chamam a atenção na definição apresentada por Bevilacqua (2005): a presença de um termo na formação da unidade fraseológica, a qual denominamos colocação especializada em nossa pesquisa; e o fato de tal colocação ser usada para representar e transmitir um conhecimento especializado, em um texto específico.

Levando em consideração o destaque dado ao termo presente na UFE, proposto por Bevilacqua (2005), retomamos ORENHA-OTTAIANO (2009, 2012a), que nos diz que a principal diferença entre as colocações especializadas e as colocações da língua geral consiste no fato de a base das colocações da língua geral ser constituída por uma unidade lexical pertencente à própria língua geral, enquanto que a base das colocações especializadas é composta por uma unidade terminológica de uma língua de especialidade.

ORENHA-OTTAIANO (2012a), embasada em L'Homme (2000), afirma que, assim como as colocações da língua geral são definidas por meio da convenção estabelecida pelos falantes de uma dada língua, o mesmo se observa no caso das colocações especializadas, as quais são estabelecidas pela convenção estabelecida por um determinado grupo de falantes de uma dada especialidade, fato que corrobora a presença de um termo de uma área específica, na formação da colocação especializada.

Além disso, ORENHA-OTTAIANO (2012a), retomando L'Homme (2000) afirma que:

as colocações especializadas podem ser melhor descritas em termos de coocorrência lexical livre (L'HOMME, 2000, p. 90), e isso pode ser verificado por meio de diferentes propriedades: 1) pelo fato de a não composicionalidade não ser um critério predominante para a identificação de uma combinação lexical especializada; e 2) em razão de ser altamente produtiva, a definição dos grupos de termos semanticamente relacionados está associada a coocorrentes. (ORENHA-OTTAIANO, 2012a, p. 161)

Dessa forma, em nossa pesquisa, consideraremos colocações especializadas aquelas que contenham um termo em sua formação, e procuraremos descrever as tipologias das colocações encontradas de acordo com a tipologia proposta por ORENHA-OTTAIANO (2004).

2.3 A ABORDAGEM LEXICAL

Ao nos propormos a trabalhar com o ensino-aprendizagem das colocações, parece-nos bastante apropriado tomarmos os postulados teórico-metodológicos da Abordagem Lexical como base para nosso projeto, visto que tal abordagem, amplamente defendida por Lewis (1993; 2000), tem como foco não somente o léxico, mas também, e principalmente, as colocações.

Como o próprio nome sugere, a Abordagem Lexical privilegia o ensino-aprendizagem baseado no léxico, ou seja, tem como foco o ensino do vocabulário, uma vez que entende como desnecessária a dicotomia gramática/vocabulário. Segundo o autor, tendemos a simplificar um dos mais complexos e importantes aspectos da língua ao promovermos essa dicotomia (LEWIS, 1993,), visto que, ao analisarmos a língua que falamos, percebemos que esta divisão não acontece; ao contrário, esta divisão apenas ocorre nas aulas de língua estrangeira.

Lewis (1993) faz uma relação das principais polaridades e dicotomias que se percebe, no que diz respeito ao desenvolvimento do ensino de língua, como, por exemplo, a dicotomia que existe entre língua falada e língua escrita; a já citada, entre vocabulário e gramática; entre habilidades receptivas e produtivas; entre competência e desempenho; entre outras. Para o autor,

O tipo de pensamento binário encorajado por este tipo de terminologia é quase sempre uma simplificação, que destaca algumas características do assunto, mas apenas à custa de não enfatizar, ou suprimir positivamente, outros aspectos. Devo concordar que muitos dos termos introduzidos, [...], tiveram um sério efeito inibidor sobre o desenvolvimento do pensamento sobre o currículo e os métodos adequados para o ensino de línguas. (LEWIS, 1993, p. 7)⁸

⁸ “The kind of binary thinking encouraged by this type of terminology is almost always a simplification, which highlights some features of the subject matter, but only at the expense of de-emphasizing, or

Percebe-se, assim, que o autor não acredita, nem compartilha das dicotomias e polaridades por ele elencadas, o que pode ser comprovado com as sugestões que ele propõe aos professores de língua inglesa, de se ajustarem aos valores por ele sugeridos (LEWIS, 1993, p. 32-35), o que faria com que deixassem de pensar e agir em suas aulas de maneira binária.

O autor propõe que, ao invés de pensarmos de maneira binária, pensemos em *espectros*, ou *conceitos não-lineares* (LEWIS, 1993), para que, de fato, seja possível entender as polaridades, já que a dicotomia feita entre vocabulário e gramática só é bem definida quando trabalhamos com palavras muito específicas. Contudo, ao analisarmos palavras que são de nosso uso cotidiano, essa diferenciação se torna mais complicada, visto que estas palavras são situadas em uma posição mais central no âmbito desta dicotomia, ou em um espectro (LEWIS, 1993, p.7), o que faz com que deixemos de pensar em tal dicotomia.

Assim, Lewis passa a abordar a questão das colocações, ao afirmar que “de um lado, estão as palavras semanticamente fortes, as quais possuem um alcance colocacional limitado, [...]”⁹ (1993, p. 37). Com isso, já é possível dizer que a Abordagem Lexical se ocupará, além do ensino de palavras com diferentes cargas semânticas, das colocações, visto que estas são combinações lexicais que contêm elevada carga semântica, e são formadas tanto por palavras classificadas como gramaticais quanto por palavras que, normalmente, são tratadas como vocabulário.

No que diz respeito ao ensino das colocações mais especificamente, Lewis (2000) conta com o apoio de outros pesquisadores, como Hill (2000), Conzett (2000) e Woolard (2000), para ajudar os professores a incorporar o ensino das colocações em suas aulas de língua inglesa, já que acredita que nosso vocabulário é, em grande parte, composto por tipos diferentes de unidades pré-fabricadas, sendo as colocações consideradas o tipo mais importante dessas unidades. (LEWIS, 2000, p. 8).

positively suppressing, other aspects. I shall agree that many of the terms introduced, [...], have had a seriously inhibiting effect on the development of thinking about the syllabus and methods appropriate to language teaching”.

⁹ “At one end, are semantically strong words which have limited collocational range, [...]”

Hill (2000) defende que é necessário dedicar mais tempo ao ensino de colocações desde o nível mais básico do ensino de língua inglesa, visto que, ao focar o ensino de colocações, o professor estará permitindo que um aluno, com um conhecimento restrito a 2000 palavras do idioma, ao ser apresentado às várias colocações que são formadas por essas palavras, poderá ter sua competência colocacional desenvolvida, sua competência comunicativa ampliada (p. 62), e ser comunicativamente competente, objetivo da grande maioria dos alunos que buscam aprender um novo idioma.

Fica evidente que, para Hill (2000), a maneira mais apropriada de desenvolver a competência comunicativa dos alunos de língua inglesa é por meio do trabalho voltado para o ensino de colocações em sala de aula. Dessa forma, ele propõe algumas sugestões de como o professor pode introduzir as colocações em sua prática didática, e de como tornar este ensino relevante aos alunos. O autor sugere que os professores ensinem as colocações individualmente, tal como fazem ao ensinar um novo léxico, uma vez que é melhor ensiná-las aos alunos do que esperar que eles sejam capazes de agrupá-las sozinhos (HILL, 2000, p. 61).

Além disso, os professores devem fazer com que os alunos se conscientizem da importância das colocações como algo vital para o aprendizado de uma língua, já que, conforme mencionado anteriormente, a autora acredita que o uso das colocações amplia a competência comunicativa dos alunos (HILL, 2000, p. 61). Para tal, Hill propõe que os professores incentivem seus alunos a manterem um caderno lexical no qual anotem as colocações conforme forem conhecendo-as e aprendendo-as (p. 62).

Uma das sugestões dadas por Hill, Lewis e Lewis (2000, p. 99) é que os professores desenvolvam atividades de memorização das colocações com os seus alunos, e para tal apresentam um modelo de atividade intitulado *5-1 box*, no qual os alunos memorizam os colocados que ocorrem com cada nóculo em destaque.

Rocha e Orenha-Ottaiano (2015), por sua vez, discutem 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. A partir das colocações compiladas no

glossário, propõem uma série de atividades a fim de serem utilizadas em sala de aula 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

Nesse sentido, espera-se que a proposta de elaboração de atividades colocacionais que irão fazer parte de um *workbook* de colocações on-line contribua para o ensino de aprendizagem de inglês como LE. A seguir, passaremos a discorrer sobre a metodologia adotada para o desenvolvimento de nossa pesquisa, levando em consideração o disposto nesta seção.

3 METODOLOGIA

Para melhor explicarmos a metodologia utilizada para a realização da nossa pesquisa, trataremos, primeiramente, da coleta dos dados, para, em seguida, focarmos os procedimentos para a elaboração das atividades colocacionais.

3.1 COLETA DOS DADOS

Conforme mencionado previamente, nosso corpus de estudo é constituído pelas legendas, em inglês e português, de doze temporadas do seriado *CSI: Crime Scene Investigation*. Para compilá-lo, levamos em consideração alguns aspectos importantes, tais como: representatividade da língua, propósito de pesquisa linguística, e ocorrência natural. Dessa forma, seguimos aqui a definição de corpus dada por Tognini-Bonelli (2002, p.2), já mencionada anteriormente.

O seriado citado acima foi escolhido para compor o corpus de estudo, em razão de se tratar de um programa de TV norte-americano voltado para o público jovem e adulto, e por colaborar quanto ao aspecto de ocorrência natural dos textos, uma vez que o seriado retrata a rotina de trabalho das pessoas envolvidas em investigações criminais, enfatizando o vocabulário específico da área criminal. Dessa maneira, nosso corpus de estudo atende a outro aspecto importante proposto por Tognini-Bonelli (2002), o propósito de pesquisa linguística, ou

seja, o corpus nos fornece amostras reais de uso da língua especializada, possibilitando a busca e análise das colocações especializadas, alvo de nossa pesquisa.

Após a coleta das transcrições e seu armazenamento em formato *.txt*, o corpus foi analisado com a ajuda do programa *WordSmith Tools*, versão 6.0 (SCOTT, 2012), o qual nos possibilita a busca e análise das colocações. O referido programa é composto por algumas ferramentas, a saber, *Wordlist*, *KeyWord* e *Concord*. Em nossa pesquisa, nos valem das três ferramentas mencionadas acima.

No primeiro momento da pesquisa, após a compilação dos *corpora*, utilizamos a ferramenta *Wordlist*, que nos permitiu gerar uma lista de palavras, evidenciando desde a palavra mais frequente até a menos frequente em nosso corpus. Após obtermos a *Wordlist* (lista de palavras), esta foi utilizada para a geração da lista de palavras-chave. Para tal, utilizamos uma segunda *Wordlist* proveniente de um corpus de referência, o qual, para nossa pesquisa, foi o BNC (*British National Corpus*). Assim, de posse das duas *Wordlists*, podemos, por meio da ferramenta *KeyWord*, gerar a lista com as palavras-chave do nosso corpus. Somente a partir desta etapa, de posse das palavras-chave, passamos à utilização da terceira ferramenta do programa, a *Concord*, gerador de linhas de concordância que permite ao pesquisador visualizar as palavras que ocorrem próximas da palavra de busca, verificando, assim, a ocorrência das colocações e, posteriormente, possibilitar a análise das colocações extraídas de nosso corpus de estudo. Conforme mencionamos, após a busca pelas colocações em nosso corpus, passamos à análise das mesmas. Para tal, contamos com as considerações de Hausman (1984) para a identificação da base e do colocado das colocações, além de nos basear na tipologia proposta por ORENHA-OTTAIANO (2004) para classificação das colocações encontradas.

3.2 ELABORAÇÃO DO MATERIAL DIDÁTICO

Após a busca e a análise das colocações especializadas, passamos à elaboração das atividades que promoverão o ensino/aprendizagem destas colocações. Para tal, além de contarmos com os pressupostos teóricos da abordagem lexical, defendida por Lewis (1993; 2000), nos basearemos nas pesquisas desenvolvidas por ORENHA-OTTAIANO (2012b; 2013;

2015) e seguiremos os passos teórico-metodológicos empregados na compilação do *Online English Collocations Workbook*, de modo mais específico, na alimentação dos jogos que o compõem o *Memory Game* e o *Gap Fill*, que passarão a contar com uma categoria a mais, a de Investigação (Investigation), a qual irá conter colocações extraídas de nosso corpus de estudo. Todos os jogos e atividades podem ser acessados on-line, após cadastro dos alunos interessados.

No *Memory Game* (jogo de memória), os aprendizes terão que juntar os elementos da combinatória (a base e o colocado), a fim de formar as colocações especializadas. Convém ressaltar que os jogos de memória estão sendo propostos de forma a trabalhar, individualmente, com as diferentes tipologias das colocações especializadas que o corpus de estudo nos permitir encontrar, sendo, a princípio, as colocações nominais, verbais, adjetivas e adverbiais. Já no *Gap Fill*, os aprendizes escolherão, entre as colocações apresentadas, aquela que julgar correta para preencher a lacuna de cada frase apresentada na atividade. Assim como no *memory game*, as atividades do *gap fill* são separadas de acordo com as tipologias das colocações especializadas.

Conforme proposta do *Online English Collocations Workbook*, uma vez elaboradas as atividades, estas são classificadas de acordo com três diferentes níveis de dificuldade, a saber: fácil, médio e difícil. A fim de verificar se os níveis de dificuldade das atividades estão coerentes, serão aplicados questionários aos alunos dos cursos de Licenciatura em Letras e Bacharelado em Tradução. Esse questionário já tem sido aplicado nos últimos dois anos para as atividades que já estão on-line, conforme proposta do projeto de pesquisa guarda-chuva, no qual nosso projeto se insere.

A título de ilustração, exibimos abaixo imagens das atividades do tipo *memory game* e *gap fill* que já se encontram inseridas no *Online English Collocations Workbook*.

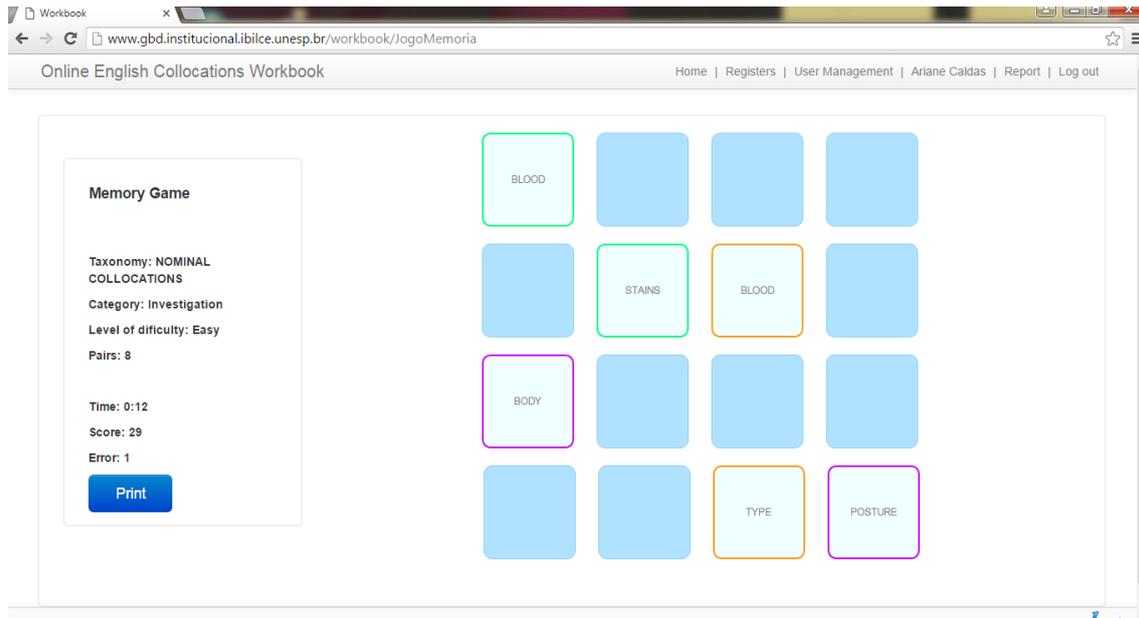


Figura1: Tela do Memory Game no Online English Collocations Workbook.

Fonte: Online English Collocations Workbook. Disponível em:
<http://www.gbd.institucional.ibilce.unesp.br/workbook/index>.

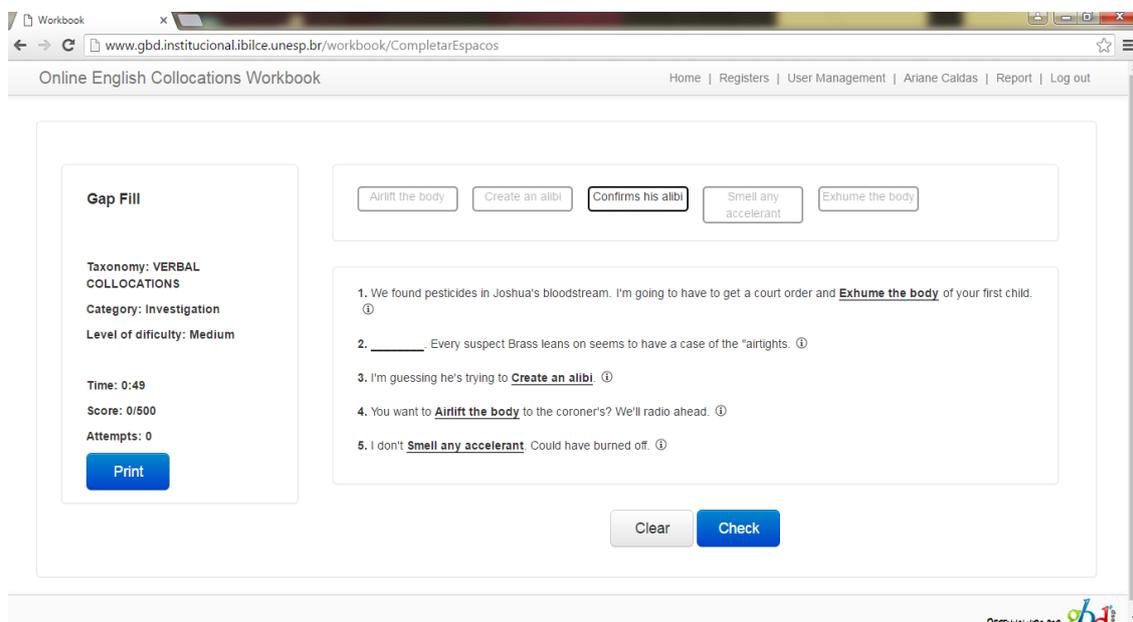


Figura2: Tela do Gap Fill no Online English Collocations Workbook

Fonte: Online English Collocations Workbook. Disponível em:
<http://www.gbd.institucional.ibilce.unesp.br/workbook/index>.

A seguir, apresentamos as colocações especializadas encontradas em nosso corpus de estudo a partir de um dos nódulos de busca selecionados.

4 ANÁLISE DOS DADOS

Para a análise de nossos dados levamos em consideração o objetivo a ser alcançado ao concluirmos a pesquisa, o de levantarmos e classificarmos as colocações especializadas presentes em nosso corpus, de forma que estas viessem a ser inseridas no *Online English Collocations Workbook*, compondo as atividades do tipo *Memory Game* e *Gap Fill*. Assim, para obtermos tal resultado, foi necessário realizar uma análise da formação das colocações levantadas, objetivando a sua classificação taxonômica, uma vez que a inserção no referido *Workbook* é feita a partir da tipologia de cada colocação.

Dessa forma, optamos por trazer, neste artigo, a análise do nódulo *body*, um dos nódulos que se mostraram mais representativos em nosso levantamento, nos retornando um total de 128 colocações especializadas. Deste total, foi possível a identificação de 41 colocações nominais, 71 colocações verbais e 16 colocações adjetivas. Desse modo, apresentamos a seguir alguns exemplos das colocações especializadas levantadas de nosso corpus de estudo, a partir do nódulo *body*.

Quadro 1: Colocações especializadas a partir do nódulo *body* levantadas.

Colocações especializadas formadas a partir do nódulo <i>body</i>	
Colocações	Classificação Taxonômica
<i>Body count</i>	Nominal
<i>Desecration of the body</i>	Nominal
<i>A body bag</i>	Nominal
<i>A body arm</i>	Nominal
<i>Body fluid</i>	Nominal
<i>To exhume a body</i>	Verbal
<i>To spot a body</i>	Verbal
<i>To tie up a body</i>	Verbal
<i>To propel a body</i>	Verbal
<i>To hoist a body</i>	Verbal
<i>A fresh body</i>	Adjetiva
<i>An unconscious body</i>	Adjetiva
<i>A foreign body</i>	Adjetiva
<i>A warm body</i>	Adjetiva
<i>An entire body</i>	Adjetiva

Para chegarmos à classificação taxonômica apresentada no quadro acima, utilizamos a tipologia proposta por ORENHA-OTTAIANO (2004), a qual foi proposta pela autora com base nos pressupostos teóricos de Hausmann (1985). Assim, entendemos que cada componente de cada colocação especializada é classificado da seguinte maneira, no que diz respeito às colocações nominais:

Body_{s subs. colocado}+ count_{s subst. base}

Desecration_{s subst. base} + of_{prep.} the_{art det.}+body_{s subst. colocado}

A body_{s subst. colocado} + bag_{s subst. base}

A body_{s subst. colocado} + farm_{s subst. base}

Body_{s subst. colocado}+ fluid_{s subst. base}

Conforme podemos perceber, *body* nem sempre é a base de cada colocação. No caso de *body fluid*, a base é *fluid* e *body* é um dos colocados que podem ocorrer com esta base. No entanto, optamos por manter todas as colocações levantadas sob o mesmo nódulo, *body*, uma vez que este foi o termo determinado como nódulo de busca para o levantamento das colocações especializadas.

Já as colocações especializadas verbais, tiveram seus componentes classificados da seguinte maneira:

To exhume_{verbo colocado}+ a_{det.}+body_{s subst. base}

To spot_{verbo colocado}+ a_{det.}+ body_{s subs. base}

To spot_{verbo colocado}+ a_{det.}+ body_{s subs. base}

To tie up_{verbo colocado}+ a_{det.}+ body_{s subs. base}

To hoist_{verbo colocado}+ a_{det.}+ body_{s subs. base}

No caso das colocações verbais, é possível perceber a predominância de *body* como base da colocação, constituindo-se, assim, um exemplo clássico de colocação especializada conforme a definição dada por ORENHA-OTTAIANO (2009) na qual a base da colocação especializada é sempre um termo, neste caso *body*.

Com as colocações especializadas do tipo adjetivas, obtivemos as seguintes classificações para cada um de seus componentes:

A fresh_{adj. colocado} +body_{subst. base}

An unconscious_{adj. colocado} +body_{subs. base}

A foreign_{adj. colocado} +body_{subs. base}

A warn_{adj. colocado} + body_{subs. base}

An entire_{adj. colocado} + body_{subs. base}

As colocações especializadas adjetivas mostraram, assim, a prevalência de *body* enquanto base da colocação, sendo os adjetivos seus colocados, fato explicado graças à convenção de nível gramatical da língua inglesa que prescreve o uso dos adjetivos antepostos aos substantivos.

De forma geral, podemos dizer que a análise se mostrou produtiva, tanto para a classificação da tipologia das colocações especializadas a serem inseridas no *Online English Collocations Workbook*, quanto para a observação e determinação de suas bases. Podemos perceber, como no caso do nódulo de busca *body*, que este é um termo que tem variação de ocorrência nas colocações especializadas, sendo usado ora como base, ora como colocado.

CONSIDERAÇÕES FINAIS

Por meio desta pesquisa, esperamos ressaltar a importância da Linguística de *Corpus* para a Fraseologia, e a relevância das duas áreas para o desenvolvimento de pesquisas na área, principalmente, no que tange à elaboração de materiais didáticos para o ensino de colocações, fornecendo, dessa maneira, subsídios teórico-práticos para a área da Pedagogia do Léxico Baseada em *Corpus*.

Além disso, ao inserirmos mais uma categoria (*Investigation*) no *Online English Collocations Workbook*, esperamos contribuir para o desenvolvimento do ensino-aprendizagem sobre colocações especializadas da área criminal para os alunos do curso de Tradução, haja vista que, conforme dito anteriormente, os alunos tendem a apresentar uma maior dificuldade

ao fazer uso de uma colocação do que de entendê-la. Além disso, os tradutores aprendizes e futuros tradutores precisam estar familiarizados com o maior número possível de colocações, de diversas áreas, já que este conhecimento poderá facilitar o trabalho que desempenharão futuramente.

Embora as colocações levantadas até aqui tenham sido utilizadas para o desenvolvimento das atividades já constantes do *English Online Collocations Workbook*, futuramente, esperamos expandir o uso das colocações aqui levantadas e analisadas, de forma a abranger não apenas alunos da tradução, como também, alunos de inglês como língua estrangeira, também público-alvo do *Collocations Workbook*, além de utilizarmos estas colocações em novas atividades que poderão ser desenvolvidas.

Outrossim, os professores de língua inglesa poderão acessar as atividades on-line em sala de aula, além de desenvolver com seus alunos atividades de memorização, por exemplo, tomando por base a proposta apresentada aqui.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMELA, M. The case for verb-adjective collocations: corpus-based analysis and lexicographical treatment. *Revista de Linguística y Lenguas Aplicadas*. v.6. 2011.

BARROS, L. A. *Curso Básico de Terminologia*. São Paulo: Editora da Universidade de São Paulo, 2004.

BARTSCH, S. Structural and functional properties of collocations in English. In: *A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Gunther NarrVerlag, 2004.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus Linguistics: investigating language structure and use*. Cambridge: Cambridge University Press, 1998.

British National Corpus (BNC). Disponível em: <http://www.lexicaly.net/wordsmith/support/extras.html>. Acesso em: 25 abr. 2014.

CASSIDY, C. J. *Cambridge Dictionary of American English*. Cambridge University Press. 2007.

COLLINS. *Collins Cobuild Learner's Dictionary*. Heinle ELT. 2nd ed. 2005.

Corpus of Contemporary American English (COCA). Disponível em: <http://corpus.byu.edu/coca/>. Acesso em: 08 set. 2015

CSI: Crime Scene Investigation. Disponível em: <http://www.tvsubtitles.net>. Acesso em: 01 jul. 2014.

DURRANT, P. *Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories*. *Applied Linguistics*, 2015: 1-30.

FRITZINGER, F.; HEID, U. Automatic Grouping of Morphologically Related Collocations. In: *Proceedings of the Corpus Linguistics 2009 Conference Liverpool/UK*.

KRIEGER, M.G. *A heterogeneidade do léxico especializado e perfis terminológicos*. In: MURAKAWA, C. A. A.; NADIN, O. L. *Terminologia: uma ciência interdisciplinar*. São Paulo: Cultura Acadêmica, 2013.

LAUFER, B.; WALDMAN, T. *Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English*. *Language Learning* 61:2, June 2011, pp. 647-672.

LEWIS, M. *The Lexical Approach: The State of ELT and a Way Forward*. Language Teaching Publications. 1993.

_____. *Teaching Collocation: Further Developments in the Lexical Approach*. Language Teaching Publications. 2000.

L'HOMME, M. C. Understanding Specialized Lexical Combinations. In: *Terminology*, v. 6, n. 1, 2000, p. 89-110.

MCCARTHY, M.; MCCARTEN, J.; SANDIFORD, H. *Touchstone*. Cambridge University Press. 2005.

MCENERY, T.; HARDIE, A. *Corpus linguistics: method, theory and practice*. Cambridge University Press.

ORENHA-OTTAIANO, A. *Collocations workbook: um material de apoio pedagógico on-line e baseado em corpus para o ensino de colocações em inglês*. In *Revista de Estudos da Linguagem*, 2015, p. 833-881.

_____. *An English collocations E-workbook designed to Brazilian Portuguese speakers*. In: *Learner Corpus Research Abstract Book*, Bergen, Noruega, p. 116-117, 2013.

ORENHA-OTTAIANO, A. *Collocations and the design of teaching materials for second language learners*. In: *TALC10 Proceedings of the 10th International Conference on Teaching and*

Language Corpora 2012b, Varsóvia, Polônia. Warsaw: Institute of Applied Linguistics, p. 93-103, 2013a.

_____. English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. In: *Acta Scientiarum. Language and Culture*. V.34, 2012, p. 241-251.

_____. Semelhanças e diferenças entre colocações e colocações especializadas. In: ORTIZ-ALVAREZ, M. L. (Org.). *Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia*. 1 ed. Campinas: Editora Pontes, 2012a, v.2, p.147- 163.

_____. *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.

_____. *A compilação de um glossário bilíngüe 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.

_____. *Online English Collocations Workbook*. Disponível em: <http://www.gbd.institucional.ibilce.unesp.br/workbook/index>. Acesso em: 28 de jun. 2016.

OXFORD. *Oxford Student's Dictionary of English*. Oxford. 3rd Ed. 2013.

ROCHA, J. M. P.; [ORENHA-OTTAIANO, A.](#) Ensino de inglês como LE e contribuições pedagógicas de um glossário bilíngüe de colocações. *Revista Signótica*, v. 27, p. 485-509, 2015.

SCOTT, M. *WordSmith Tools*. Versão 6.0. Liverpool: Lexical Analysis Software, 2012.

SUMMERS, D. *Longman Dictionary of Contemporary English*. 2003.

TAGNIN, S. E. O. *O jeito que a gente diz: combinações consagradas em inglês e português*. Barueri, SP: DISAL, 2013.

TOGNINI-BONELLI, E. *Corpus Linguistics at work*. John Benjamins Publishing, 2001.

WALKER, C. P. A Corpus-Based Study of the Linguistic Features and Processes Which Influence the Way Collocations Are Formed: Some Implications for the Learning of Collocations. *Tesol Quarterly*, vol. 45, No. 2, June 2011.

WebCorp. Disponível em: <http://www.webcorp.org.uk/live/>. Acesso em: 08 set. 2015.

Ariane Donizete Delgado Ribeiro CALDAS

Mestranda no Programa de Pós-Graduação em Estudos Linguísticos da UNESP-IBILCE. Possui licenciatura em Letras pelo Centro Universitário do Norte Paulista - UNORP (2008) e Especialização em Formação de Professores de Língua Estrangeira pelo Instituto de Biociências, Letras e Ciências Exatas - IBILCE-UNESP (2014). Atua na área de Linguística Aplicada com ênfase em Linguística de Corpus, Fraseologia e Ensino de Inglês como língua estrangeira

Adriane ORENHA-OTTAIANO

Possui Bacharelado em Letras com Habilitação para Tradutor, pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Mestrado em Estudos Linguísticos e Literários em Inglês, pela Universidade de São Paulo (USP), e Doutorado em Estudos Linguísticos, pela UNESP. É líder do Grupo de Pesquisa “Pedagogia do Léxico, da Tradução e Linguística de Corpus”. É Professora Assistente Doutora do Departamento de Letras Modernas, da Universidade Júlio de Mesquita Filho, Câmpus de São José do Rio Preto e atua na Pós-Graduação em “Estudos Linguísticos” da UNESP/IBILCE, nas linhas “Estudos da Tradução” e “Pedagogia do léxico e da tradução a partir de corpora”. Possui experiência em Estudos da Tradução Baseados em Corpus, Ensino de Inglês como LE, Linguística de Corpus e Fraseologia, atuando principalmente nos seguintes temas: corpus, corpus de aprendizes, ensino de LE, fraseologia (colocações, colocações estendidas e unidades fraseológicas especializadas), pedagogia da tradução e pedagogia do léxico.

Recebido em junho/2016 - Aceito em março/2017