



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Campus de São José do Rio Preto

ARIANE DONIZETE DELGADO RIBEIRO CALDAS

A identificação de colocações especializadas extraídas do Corpus CSI e do Corpus Comparável Criminal para a elaboração de atividades didáticas.

SÃO JOSÉ DO RIO PRETO
2017

ARIANE DONIZETE DELGADO RIBEIRO CALDAS

A identificação de colocações especializadas extraídas do Corpus CSI e do Corpus Comparável Criminal para a elaboração de atividades didáticas.

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto. (Área de concentração: Linguística Aplicada)

Orientadora: Prof^a. Dr^a. Adriane Orenha-Ottaiano

SÃO JOSÉ DO RIO PRETO
2017

Caldas, Ariane Donizete Delgado Ribeiro.

A identificação de colocações especializadas extraídas do Corpus CSI e do Corpus Comparável Criminal para a elaboração de atividades didáticas. / Ariane Donizete Delgado Ribeiro Caldas. -- São José do Rio Preto, 2017
137 f. : il.

Orientador: Adriane Orenha-Ottaiano

Dissertação (mestrado) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas

1. Linguística. 2. Linguística de corpus. 3. Fraseologia. 4. Análise baseada em corpus (Linguística) 5. Palavras e expressões. 6. Lexicologia. I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. II. Título.

CDU – 413

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

ARIANE DONIZETE DELGADO RIBEIRO CALDAS

A identificação de colocações especializadas extraídas do Corpus CSI e do Corpus Comparável Criminal para a elaboração de atividades didáticas.

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto. (Área de concentração: Linguística Aplicada)

Comissão Examinadora

Prof^ª. Dr^ª. Adriane Orenha-Ottaiano
UNESP – São José do Rio Preto
Orientadora

Prof^ª. Dr^ª. Marilei Amadeu Sabino
UNESP – São José do Rio Preto

Prof. Dr^ª. Ieda Maria Alves
USP – São Paulo

SÃO JOSÉ DO RIO PRETO
01 de Setembro de 2017

Aos meus pais, meu alicerce, minha base!

AGRADECIMENTOS

Agradeço infinitamente a meus pais por sempre me apoiarem e incentivarem a lutar pelos meus sonhos e objetivos e por todo carinho e suporte durante esta caminhada.

Muito obrigada a minha orientadora, Prof^ª. Dr^ª. Adriane Orenha-Ottaiano, por ter acreditado e confiado em mim e por todas as orientações e ensinamentos ao decorrer deste trabalho.

Muito obrigada, também, à Prof^ª Dr^ª Maria José Bocorny Finatto, por ter se disponibilizado a debater este trabalho no VIII Seminário de Estudos Linguísticos da Unesp, pelas contribuições enriquecedoras e pelos elogios.

Ao Prof. Dr. Maurizio Babini por ter aceitado o convite para a banca de qualificação e pelas contribuições para o enriquecimento do trabalho.

À Prof^ª. Dra. Marilei Amadeu Sabino pelas preciosas contribuições na banca de qualificação e por ter aceitado o convite para fazer parte da banca de defesa, tendo igualmente contribuído para o aprimoramento deste trabalho.

Muito obrigada à Prof^ª. Dr^ª Ieda Maria Alves por ter aceitado o convite para fazer parte da banca de defesa, pelos elogios ao trabalho e pelas valiosas sugestões para o enriquecimento deste.

Agradeço aos professores das disciplinas cursadas durante o Mestrado, os quais foram responsáveis por ampliar a minha visão sobre o léxico, a fraseologia, corpus e linguística em geral.

Meu muito obrigada aos amigos que fiz durante as disciplinas e com os quais compartilhei angústias, dúvidas, certezas e vitórias, em especial a Jean Pimentel, Janaína Camoleze, Mariele Seco e Beatriz Curti. Vocês fizeram toda a diferença durante esta etapa da minha vida e os levarei sempre comigo.

Por fim, mas não menos importante, agradeço a Deus por ter me capacitado e me fortalecido durante toda minha vida, em especial, por ter me sido meu sustentáculo durante a realização desta pesquisa.

RESUMO

A presente pesquisa está vinculada ao Grupo de Pesquisa “Pedagogia do Léxico, da Tradução e Linguística de Corpus”, cadastrado no Diretório de Grupos de Pesquisa do CNPq, bem como faz parte do projeto guarda-chuva “A compilação de materiais didáticos e glossários especializados baseados em corpora e sua contribuição para uma Pedagogia do Léxico e da Tradução”, desenvolvido na UNESP. Tem como objetivo extrair e analisar, do ponto de vista sintático-morfológico e léxico-semântico, as colocações da área de investigação criminal presentes no corpus constituído pelas legendas da série de TV norte-americana *CSI – Crime Scene Investigation* e por um corpus comparável compilado com o auxílio da ferramenta *Boot Cat Front End*, versão 0,71 (ZANCHETTA; BARONI; BERNARDINI, 2013), o qual recebeu a denominação de Corpus Comparável Criminal. Como resultado prático desta pesquisa, as colocações especializadas levantadas e analisadas foram utilizadas para a elaboração das atividades que compõem o *Online English Collocations Workbook* (ORENHA-OTTAIANO, 2015), na categoria *Investigation*. Para isso, este trabalho fundamenta-se no arcabouço teórico-metodológico da Linguística de Corpus (BIBER, 1998; MEYER, 2004; MCENERY; HARDIE, 2012; TOGNINI-BONELLI, 2001), para proceder à compilação, extração e análise dos corpora de estudo, além de contar com os construtos da Fraseologia, de forma mais específica, aqueles que dizem respeito às colocações e às colocações especializadas (BEVILACQUA, 1998, 2005; L’HOMME; BERTRAND, 2000; ORENHA-OTTAIANO, 2004, 2009, 2012; SINCLAIR, 2001; TAGNIN, 2013), permitindo-nos, assim, analisar as colocações especializadas extraídas de ambos os *corpora*. Para a elaboração das atividades que são propostas como resultado final, contamos com o apoio teórico da Abordagem Lexical, proposta por Lewis (1993, 2000) e a proposta de Orenha-Ottaiano (2015), concernente à compilação de materiais didáticos para o ensino de colocações. No que diz respeito à metodologia, a pesquisa conta com dois corpora, o Corpus CSI e com o Corpus Comparável Criminal, compilado a fim de ser uma fonte de comprovação da utilização e da frequência de coocorrência das colocações especializadas encontradas no Corpus CSI fora do ambiente da série. Além disso, contamos com o emulador de concordanciador *WebCorp* como mais uma fonte de verificação da frequência de coocorrência das colocações especializadas levantadas. A partir desta pesquisa, foi possível a identificação e extração de 148 colocações especializadas a partir do nóculo *evidence* no Corpus CSI e 167 colocações deste mesmo nóculo no Corpus Comparável Criminal, sendo que 90 destas colocações foram selecionadas para compor as atividades do *Online English Collocations Workbook*. Espera-se que estas atividades possam contribuir para o aprimoramento do conhecimento lexical de alunos de inglês como língua estrangeira, bem como de alunos do curso de Tradução, evidenciando, assim, a formação, o uso e a compreensão das colocações especializadas na área enfocada nesta pesquisa. Espera-se, ainda, que esta investigação possa vir a ressaltar a importância da Linguística de Corpus para o estudo da Fraseologia e a importância de ambas as áreas para o desenvolvimento de estudos e elaboração de materiais didáticos na área da Pedagogia do Léxico Baseada em Corpus.

Palavras-chave: Linguística de Corpus. Pedagogia do Léxico. Fraseologia. Colocações. Colocações Especializadas.

ABSTRACT

The present research is linked to the Research Group Corpus-based Lexicon and Translation Pedagogy, registered in the Directory of Research Groups of CNPq, as well as it is part of the umbrella project "The compilation of didactic materials and specialized glossaries based on corpora and its contribution to a Teaching of the Lexicon and of Translation ", developed in UNESP. It aims to extract and analyze, from the syntactic-morphological and lexical-semantic point of view, the specialized collocations present in a corpus constituted by the subtitles of North American TV series CSI – Crime Scene Investigation and by a comparable corpus compiled with the aid of the Boot Cat Front End tool, version 0.71 (ZANCHETTA; BARONI; BERNARDINI, 2013), which received the name of Criminal Comparable Corpus. As a practical result of this research, the specialized collocations extracted and analyzed were used to elaborate the activities that integrate the Online English Collocations Workbook (ORENHA-OTTAIANO, 2015), in the Investigation category. To do this, this work is based on the theoretical-methodological framework of Corpus Linguistics (BIBER, 1998; MEYER, 2004; MCENERY; HARDIE, 2012; TOGNINI-BONELLI, 2001) to compile, extract and analyze the study corpora, besides counting with the constructs of Phraseology, more specifically, those related to collocations and specialized collocations (BEVILACQUA, 1998, 2005; L'HOMME; BERTRAND, 2000; ORENHA-OTTAIANO, 2004, 2009, 2012; SINCLAIR, 2001; TAGNIN, 2013), thus allowing us to proceed with the analysis of specialized collocations extracted from both corpora. In order to elaborate the activities that are proposed as final results, we count on the theoretical support of the Lexical Approach proposed by Lewis (1993, 2000) and on Orenha-Ottaiano (2015) proposal concerning the compilation of didactic materials for the teaching of collocations. Regarding the methodology, the research has two corpora, the CSI Corpus and the Criminal Comparable Corpus, compiled in order to be a source of proof of the use and frequency of co-occurrence of specialized collocations found in CSI Corpus outside the series environment. In addition, we have the WebCorp concordance emulator as another source of verification of the frequency of co-occurrence of the specialized collocations raised. From this research it was possible to identify and extract 148 specialized collocations from the evidence node in the CSI Corpus and 167 collocations of this same node in the Criminal Comparable Corpus, and 90 of these collocations were selected to compose the activities of Online English Collocations Workbook. We hope that these activities may contribute to the improvement of the lexical knowledge of students of English as a foreign language, as well as students of the Translation course, thus evidencing the training, use and understanding of specialized collocations in the area focused in this research. It's hoped that this research may highlight the importance of Corpus Linguistics for the study of Phraseology and the importance of both areas for the development of studies and elaboration of didactic materials in the area of Corpus-Based Lexicon Pedagogy.

Keywords: *Corpus Linguistics. Lexicon Pedagogy. Phraseology. Collocations. Specialized Collocations.*

LISTA DE FIGURAS

Figura 1 – Alguns personagens da primeira temporada da série <i>CSI</i>	18
Figura 2 – Tela com o resultado obtido pela ferramenta <i>WordList</i> do programa <i>WordSmith Tools</i>	47
Figura 3 – Tela com resultado obtido pela ferramenta <i>KeyWords</i> do programa <i>WordSmith Tools</i>	48
Figura 4 – Tela de definição do projeto.....	50
Figura 5 – Tela com as <i>seeds</i> (sementes) inseridas na ferramenta.....	51
Figura 6 – <i>Tuples</i> apresentadas pela ferramenta.....	52
Figura 7 – Levantamento de <i>URLs</i>	53
Figura 8 – Compilação do <i>corpus</i>	54
Figura 9 – Tela da ferramenta <i>Concord</i> do programa <i>WordSmith Tools</i>	55
Figura 10 – Tela inicial do <i>Online English Collocations Workbook</i>	58
Figura 11 – Tela de cadastro de novos usuários no <i>Online English Collocations Workbook</i>	59
Figura 12 – Tela inicial do <i>Memory Game</i> no <i>Online English Collocations Workbook</i>	60
Figura 13 – Tela inicial do <i>Gap Fill</i> no <i>Online English Collocations Workbook</i>	60
Figura 14 – Tela do <i>Memory Game</i> no <i>Online English Collocations Workbook</i>	61
Figura 15 – Tela do <i>Gap Fill</i> no <i>Online English Collocations Workbook</i>	61
Figura 16 – Tela de registro de categorias no <i>Online English Collocations Workbook</i>	62
Figura 17 – Tela de cadastro das colocações especializadas no <i>Online English Collocations Workbook</i>	63
Figura 18 – Tela de seleção das colocações para o <i>Memory Game</i> no <i>Online English Collocations Workbook</i>	64
Figura 19 – Tela do <i>Memory Game</i> no <i>Online English Collocations Workbook</i>	64
Figura 20 – Tela de seleção das colocações para o <i>Gap Fill</i> no <i>Online English Collocations Workbook</i>	65
Figura 21 – Tela do <i>Gap Fill</i> no <i>Online English Collocations Workbook</i>	66
Figura 22 – Tela da verificação das respostas do <i>Gap Fill</i>	66
Figura 23 – Tela do desempenho do aluno no <i>Gap Fill</i>	67
Figura 24 – Tela de memorização dos pares das colocações especializadas.....	124
Figura 25 – Tela da atividade de <i>Memory Game</i> com colocações verbais de nível fácil.....	124
Figura 26 – Tela da atividade <i>Gap Fill</i> com colocações especializadas verbais de nível fácil.....	125
Figura 28 – Tela de correção da atividade <i>Gap Fill</i> com colocações especializadas verbais de nível fácil.....	126

LISTA DE QUADROS

Quadro 1 - Estrutura das colocações especializadas nominais: <i>Evidence</i> como base.....	73
Quadro 2 - Estrutura das colocações especializadas nominais: <i>Evidence</i> como colocado.....	74
Quadro 3 - Estrutura das colocações especializadas adjetivas.....	76
Quadro 4 - Estrutura das colocações especializadas verbais constituídas por verbos frasais...	77
Quadro 5 - Estrutura das colocações especializadas verbais: base como sujeito.....	78
Quadro 6 - Estrutura das colocações especializadas verbais: base como objeto.....	79
Quadro 7 - Estrutura das colocações especializadas adverbiais.....	81
Quadro 8 - Estrutura das colocações especializadas nominais: <i>Evidence</i> como base.....	82
Quadro 9 - Estrutura das colocações especializadas nominais: <i>Evidence</i> como colocado.....	83
Quadro 10 - Estrutura das colocações especializadas adjetivas.....	84
Quadro 11 - Estrutura das colocações especializadas verbais.....	85
Quadro 12 - Estrutura das colocações especializadas adverbiais.....	86
Quadro 13 - Colocações nominais do nóduo <i>Evidence</i> incluídas no <i>Online English Collocations Workbook</i>	117
Quadro 14 - Colocações verbais do nóduo <i>Evidence</i> incluídas no <i>Online English Collocations Workbook</i>	118
Quadro 15 - Colocações adjetivas do nóduo <i>Evidence</i> incluídas no <i>Online English Collocations Workbook</i>	119
Quadro 16 - Colocações adverbiais inseridas no <i>Online English Collocations Workbook</i>	119
Quadro 17 - Colocações nominais classificadas de acordo com seu nível de dificuldade.....	120
Quadro 18 - Colocações verbais classificadas de acordo com seu nível de dificuldade.....	121
Quadro 19 - Colocações adjetivas classificadas de acordo com seu nível de dificuldade.....	122
Quadro 20 - Colocações adverbiais classificadas de acordo com seu nível de dificuldade....	123

LISTA DE TABELAS

Tabela 1 – <i>KeyWords</i> selecionadas como núdulos de busca.....	48
Tabela 2 – Exemplo de análise das colocações obtidas a partir da palavra-chave <i>blood</i>	57
Tabela 3 – Frequências das palavras-chave em ambos os corpora de pesquisa.....	69
Tabela 4 – Frequência das colocações nominais: <i>Evidence</i> como base.....	88
Tabela 5 – Frequência das colocações nominais: <i>Evidence</i> como colocado.....	90
Tabela 6 – Frequência das colocações verbais: <i>Evidence</i> como sujeito.....	91
Tabela 7 – Frequência das colocações verbais: <i>Evidence</i> como objeto.....	93
Tabela 8 – Frequência das colocações verbais formadas por verbos frasais.....	94
Tabela 9 – Frequência das colocações adjetivas.....	96
Tabela 10 – Frequência das colocações adverbiais.....	98
Tabela 11 - Resumo dos dados obtidos em ambos os corpora e no WebCorp.....	100

LISTA DE ABREVIATURAS E SIGLAS

Adj.	Adjetivo
Adv.	Advérbio
CSI	Crime Scene Investigation
CCSI	Corpus CSI
CCC	Corpus Comparável Criminal
Det.	Determinante
Prep.	Preposição
Part.	Partícula
Subst.	Substantivo

SUMÁRIO

INTRODUÇÃO.....	14
1 A SÉRIE CSI – CRIME SCENE INVESTIGATION.....	18
2 FUNDAMENTAÇÃO TEÓRICA.....	22
2.1 Fraseologia.....	22
2.2 Fraseologia Baseada em Corpus.....	26
2.3 Colocações da língua geral e colocações especializadas.....	29
2.4 Linguística de Corpus.....	37
2.5 Abordagem Lexical.....	41
3 METODOLOGIA DA PESQUISA.....	45
3.1 Compilação e tratamento do <i>Corpus CSI</i>	45
3.2 Compilação e tratamento do <i>corpus</i> comparável.....	49
3.3 Procedimento de extração e análise das colocações especializadas.....	55
3.4 O <i>Online English Collocations Workbook</i>	57
3.5 Da elaboração das atividades didáticas.....	62
4 ANÁLISE DOS DADOS	68
4.1 Análise quantitativa dos dados em ambos os corpora de estudo.....	68
4.2 Análise morfossintática das colocações a partir do <i>Corpus CSI</i>.....	71
4.2.1 Colocações especializadas nominais.....	73
4.2.2 Colocações especializadas adjetivas.....	75
4.2.3 Colocações especializadas verbais.....	77
4.2.4 Colocações especializadas adverbiais.....	80
4.3 Análise morfossintática das colocações a partir do <i>Corpus Comparável Criminal</i>.....	81
4.3.1 As colocações especializadas nominais.....	81
4.3.2 As colocações especializadas adjetivas.....	83
4.3.3 As colocações especializadas verbais.....	84
4.3.4 As colocações especializadas adverbiais.....	85
4.4 Análise contrastiva das colocações especializadas nos corpora e no <i>WebCorp</i>.....	87
4. 5 Análise léxico semântica das colocações especializadas.....	101
4.5.1 <i>To mishandle an evidence</i>	102
4.5.2 <i>To jeopardize an evidence</i>	103
4.5.3 <i>To tamper with evidence</i>	104
4.5.4 <i>To flush evidence</i>	105
4.5.5 <i>To ditch evidence</i>	106

4.5.6 <i>Stitch of evidence</i>	107
4.5.7 <i>An evidence ledger</i>	109
4.5.8 <i>Crime Spree</i>	109
4.5.9 <i>Duty weapon</i>	110
4.5.10 <i>Allegedly fire</i>	111
4.5.11 <i>Reportedly fire</i>	112
4.5.12 <i>A body evidence X body of evidence</i>	113
4.5.13 <i>Examination of evidence X evidence examination</i>	114
5 AS ATIVIDADES DO <i>ONLINE ENGLISH COLLOCATIONS WORKBOOK</i>	116
CONSIDERAÇÕES FINAIS	128
REFERÊNCIAS	132

INTRODUÇÃO

A área de investigação criminal tem despertado o interesse de várias pessoas nos últimos tempos, abrangendo desde crianças e adolescentes, até mesmo adultos, com as mais variadas formações. Isso se deve ao fato de as séries que abordam este tema estarem ganhando cada vez mais destaque entre os gêneros de séries produzidas pelos Estados Unidos, principal país criador e exportador deste tipo de programa televisivo.

Dentre as séries que abordam este tema em sua trama, a série *CSI – Crime Scene Investigation* é apontada como a responsável por impulsionar a expansão desse gênero, uma vez que, devido ao seu sucesso de público mundial, a franquia produziu três *spin-offs* – *CSI Miami*, *CSI New York* e, mais recentemente, *CSI Cyber*, além de ter influenciado na criação de outras franquias que também abordam o tema da investigação criminal, como é o caso de séries como *NCSI* e *Criminal Minds*. E, embora o último episódio de *CSI – Crime Scene Investigation* tenha ido ao ar no dia 27 de setembro de 2015 nos Estados Unidos, ainda hoje é uma série que desperta o interesse do público ao redor de todo o mundo.

Dado sucesso das séries, é possível notar um crescimento do número de pesquisas que investigam séries e *sitcoms* (QUAQLIO, 2009; FROMM, 2011; ROCHA, ORENHA-OTTAIANO, 2012, 2016; FROMM; BANG, 2013; ESPERANDIO; FINATTO, 2014; ESPERANDIO, 2015). Vale também mencionar que a área investigativa criminal se mostra importante não apenas para os profissionais que nela atuam, mas também para tradutores e alunos de Tradução, bem como alunos de língua inglesa em geral. Nesse sentido, acredita-se ser relevante investigar as colocações empregadas na série *CSI – Crime Scene Investigation*, tendo em vista que os resultados desta pesquisa poderão contribuir para o aprendizado das colocações especializadas desta área, bem como auxiliar tradutores, principalmente, aprendizes, a melhor observá-las durante o processo de tradução.

Sendo assim, esta dissertação tem por objetivo extrair e analisar, do ponto de vista sintático-morfológico e léxico-semântico, as colocações especializadas presentes em um corpus constituído pelas legendas de doze temporadas da série *CSI – Crime Scene Investigation* (doravante CCSI) e em um corpus comparável compilado com o auxílio da ferramenta *Boot Cat Front End*, versão 0,71 (ZANCHETTA; BARONI; BERNARDINI, 2013), o qual recebeu a denominação de Corpus Comparável Criminal, a fim de que estas possam integrar as atividades do *Online English Collocations Workbook*. Além disso, a pesquisa objetiva, ainda, extrair e analisar as palavras-chave em ambos os corpora de pesquisa; identificar e organizar as colocações especializadas a partir de sua classificação

taxonômica de acordo com proposta de Orenha-Ottaiano (2004, 2009) fundamentada em Hausmann (1985), classificação esta que se faz necessária para a inclusão das atividades no *Online English Collocations Workbook*, uma vez que este se divide pela classificação taxonômica das colocações que o compõem; analisar de forma contrastiva as palavras-chave e as colocações obtidas em ambos os corpora de estudo; além de verificar a frequência de coocorrência destas colocações a partir do emulador de concordanciador WebCorp; e, por fim, apresentar como resultado prático a elaboração de atividades colocacionais dos tipos *Memory Game* e *Gap Fill* que foram inseridas, especificamente, na categoria *Investigation*, de acordo com a proposta do *Online English Collocations Workbook* (ORENHA-OTTAIANO, 2015).

A pesquisa justifica-se pelo fato de que, embora trabalhos que envolvam séries e *sitcoms* tenham crescido, conforme citado anteriormente, estudos que contemplem o aspecto léxico fraseológico especializado mais especificamente, as colocações especializadas, ainda são escassos (ROCHA; ORENHA-OTTAIANO, 2012, 2016; ORENHA-OTTAIANO, 2004, 2009, 2015, 2016; ROCHA 2017). Ademais, cabe mostrar a importância da área de investigação criminal, que tem se tornado atrativa para muitas pessoas que têm buscado desenvolver suas carreiras nessa área, além de ser uma área na qual profissionais de tradução podem vir a atuar, desenvolvendo trabalhos que vão além das traduções de séries e *sitcoms*. Outrossim esta área tem se tornado assunto recorrente nas conversas do dia-a-dia, devido ao fato de várias investigações e procedimentos de investigação terem se tornado conhecidos pela população, de forma que o aprendizado das colocações desta área contribui para a fluência de alunos de língua inglesa no geral, que igualmente se mostram interessados e curiosos em aprendê-las.

Neste intuito, este estudo busca conciliar os estudos colocacionais com vistas à promoção do aprendizado do aluno, com o uso de uma série televisiva, fato este que, de certa forma, promove uma aproximação entre aluno e objeto de estudo. A busca pela elaboração das atividades didáticas a partir das colocações especializadas levantadas nos corpora justifica-se, também, em razão desta pesquisa estar vinculada ao Grupo de Pesquisa “Pedagogia do Léxico, da Tradução e Linguística de Corpus”, cadastrado no Diretório de Grupos de Pesquisa do CNPq e ser 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”, sob responsabilidade da Profa. Dra. Adriane Orenha-Ottaiano, evidenciando sua relevância.

Dessa forma, a presente dissertação se organiza em cinco capítulos, além da introdução e das considerações finais. O primeiro capítulo destina-se à apresentação da série *CSI – Crime Scene Investigation* a qual compõe o corpus de estudo intitulado CCSI, que é o responsável por nortear a busca e o levantamento das colocações especializadas deste trabalho. Além disso, neste capítulo justifica-se a escolha por trabalhar com uma série como ponto de partida para o levantamento de colocações a partir dos pressupostos teóricos de Fromm (2011) e Hoffman (1998 apud FINATTO, 2015).

Já no capítulo dois, discorre-se sobre as teorias nas quais este trabalho está fundamentado. Dessa forma, o capítulo inicia-se com a Fraseologia teórica (ZULUAGA, 1980; TRISTÁ, 1988; CORPAS PASTOR, 1996; MENA; BERTRÁN, 2002; TAGNIN, 1989 [2011]), para, em seguida, serem apresentados aspectos da Fraseologia baseada em corpus (PARTINGTON, 1998; BÉRTRAN; BRETAÑA, 2008; GRANGER; PAQUOT 2008; GRIES, 2008). Após a apresentação das bases teóricas sobre a Fraseologia, passa-se a discorrer sobre as colocações (HAUSMANN, 1984, 1985; SINCLAIR, 1991; CORPAS PASTOR, 1996; HEID, 2009; TAGNIN, 1999, 2013) e unidades fraseológicas especializadas e as colocações especializadas (L'HOMME, 2000; L'HOMME; BERTRAND, 2000; BEVILACQUA, 2005; ORENHA-OTTAIANO, 2009; 2012), uma vez que estas configuram o objeto de estudo da presente pesquisa.

Em seguida, discorre-se acerca de aspectos relevantes da Linguística de Corpus (SINCLAIR, 1991; BIBER, 1998; TOGNINI-BONELLI, 2002), além de serem abordados alguns aspectos sobre os pressupostos teóricos da Abordagem Lexical (LEWIS, 1993, 2000), e a relevância do uso de corpora no ensino de colocações (ORENHA-OTTAIANO, 2012, 2013, 2015, 2016), que servem de base para a elaboração das atividades didáticas que compõem esta pesquisa,

O terceiro capítulo desta dissertação é dedicado à metodologia utilizada para o desenvolvimento de nossa pesquisa. Assim, o capítulo inicia-se com a explicação dos procedimentos utilizados para a compilação e o tratamento do CCSI por meio da ferramenta *WordSmith Tools*, versão 6.0 (SCOTT, 2012). Convém ressaltar aqui que este corpus é constituído pelas legendas de 12 temporadas da série de TV norte-americana *CSI – Crime Scene Investigation*, conhecida no Brasil como *CSI Las Vegas*, uma vez que no momento da sua compilação somente conseguimos ter acesso às legendas destas temporadas. Posteriormente, é feita a explicação sobre a compilação e o tratamento do Corpus Comparável Criminal (doravante CCC), o qual foi compilado com a ajuda da ferramenta *Boot Cat Front End*, versão 0.71 (ZANCHETTA; BARONI; BERNARDINI, 2013).

Em seguida, discorre-se sobre como se deu a extração e análise das colocações especializadas obtidas em ambos os corpora de estudo, a qual também foi realizada com o auxílio da ferramenta *WordSmith Tools*, versão 6.0 (SCOTT, 2012), bem como, sobre a utilização do emulador de concordanciador *WebCorp* como mais uma fonte de verificação da frequência de coocorrência das colocações levantadas. Este capítulo trata, ainda, sobre o *Online English Collocations Workbook*, plataforma na qual se encontram as atividades mencionadas neste trabalho, bem como apresenta os procedimentos adotados para a inclusão das colocações especializadas levantadas nestas atividades.

Por sua vez, o quarto capítulo traz a análise realizada a partir dos dados obtidos em ambos os corpora, apresentando uma análise quantitativa dos dados obtidos em ambos os corpora, de forma a evidenciar e contrastar aspectos importantes encontrados. Em seguida, passa-se à análise morfossintática das colocações obtidas em ambos os corpora, para, finalmente, desenvolver uma análise léxico-semântica das colocações especializadas que mais podem causar problemas para os alunos, considerando as dificuldades que estes podem apresentar para compreender e empregar tais colocações.

Já o quinto e último capítulo é destinado a mostrar as atividades que compõem o *Online English Collocations Workbook*, bem como a apresentar os aspectos que foram levados em conta para a inserção das colocações nas atividades da referida plataforma.

Por fim, apresentamos as considerações finais da pesquisa, avaliando o seu desenvolvimento, além de apresentar os encaminhamentos futuros para este trabalho. Logo após, são apresentadas as referências bibliográficas utilizadas para o desenvolvimento da presente pesquisa.

1 A SÉRIE *CSI – CRIME SCENE INVESTIGATION*

Conforme mencionado previamente na introdução deste trabalho, optou-se por trabalhar com a série norte-americana *CSI – Crime Scene Investigation* como ponto de partida para a busca das colocações especializadas da área criminal, o que nos leva a apresentar alguns pontos relevantes tanto sobre a série em questão, como em relação aos pressupostos teóricos que embasam esta escolha.

A série norte-americana *CSI – Crime Scene Investigation* teve seu episódio piloto lançado em outubro de 2000, nos Estados Unidos, e em abril de 2001, no Brasil, tendo se popularizado no Brasil como *CSI Las Vegas*. Após 15 temporadas, a série teve seu último episódio exibido nos Estados Unidos no dia 27 de setembro de 2015.

Figura 1 – Alguns personagens da primeira temporada da série *CSI – Crime Scene Investigation*.



Fonte: <https://maryalcantaras.wordpress.com/2011/05/07/csi-crime-scene-investigation-episodios-1%C2%AA-temporada/>

Criado por Anthony E. Zuiker, o seriado retrata o cotidiano de uma equipe de cientistas forenses que trabalham no Laboratório de Criminalista de Las Vegas (considerado o segundo maior laboratório do país, atrás apenas do FBI). No início da série, esta equipe era liderada pelo capitão James Brass, interpretado pelo ator Paul Guilfoyle, contudo, ainda na primeira temporada o entomologista Gill Grissom, vivido por William Petersen, passa a chefiar a equipe composta por Catherine Willows (Marg Helgenberger), Nicholas Stokes

(George Eads), Warrik Brown (Gary Dourdan), Gregory Sanders (Eric Szmanda), Sara Sidle (Jorja Fox) e Albert Robbins (Robert David Hall). Estes personagens constituem o núcleo principal da série por oito temporadas. A partir da oitava temporada, alguns personagens começam a deixar a série, tais como Sara Sidle (Jorja Fox) que se afasta de Las Vegas e da profissão de perita criminal após ter sido seqüestrada, e Warrik Brown (Gary Dourdan) que é assassinado no final da oitava temporada por ter descoberto corrupção por parte de outros policiais do departamento.

Assim, a partir da nona temporada, alguns novos personagens são inseridos, buscando preencher as vagas deixadas pelos antigos peritos. Merece destaque o personagem vivido por Lauren Lee Smith, Raymond Langston, que é contratado para ocupar a vaga deixada por Catherine Willows (Marg Helgenberger), a qual havia sido promovida à supervisora da equipe com a saída de Gill Grissom (William Petersen), e permanece na série até o final da 11ª temporada, quando é demitido. A partir da sua demissão, entra em cena o CSI nível 3, D. B. Russel, vivido por Ted Danson, contratado para liderar a equipe após o rebaixamento de Catherine Willows. D. B. Russel permanece como líder da equipe até o final da 15ª temporada, em 2015.

Além destes personagens que são considerados regulares na série, outros personagens considerados recorrentes também aparecem, de forma a tornar possível a apresentação de um enredo que a cada episódio retrata a investigação de um crime, repleta de detalhes que prendem a atenção dos telespectadores do início ao fim, além de conseguir abordar detalhes e dilemas da vida pessoal dos personagens que compõe o elenco regular e que se resolvem com o passar da temporada. Dessa forma, os telespectadores se vêm ansiosos pelos próximos episódios que trarão, além de um novo crime e os detalhes minuciosos da sua investigação, o desenrolar das tramas vividas por estes personagens.

Por apresentar detalhes da investigação criminal que até então eram pouco ou quase nada conhecidos pelo grande público, tais como, procedimentos de análise de impressões digitais e cenas que retratam a autópsia das vítimas, além do fato de ser exibida em diversos países, a série recebeu ao longo de suas temporadas inúmeras indicações a prêmios importantes na televisão tendo contabilizado o total de seis *Emmy Awards* (prêmio mais importante da televisão) e sete premiações no Festival de Televisão de Monte Carlo como a série dramática mais assistida no mundo. Além disso, foi responsável pelo surgimento de três *spin-offs*, a saber, *CSI Miami*, *CSI NY* e *CSI Cyber*, todos também já cancelados.

A série conta com um total de 335 episódios distribuídos ao longo de 15 temporadas. Em nossa pesquisa analisamos as legendas das 12 primeiras temporadas da série, o que gera

um total de 276 episódios. Ressaltamos, desde já, que trabalhamos com apenas 12 das 15 temporadas lançadas da série, pois no momento da compilação as legendas foi possível obter apenas as legendas referentes a estas temporadas. Em estudo futuro, espera-se ampliar o corpus composto por estas legendas, o qual se denominou CCSI (Corpus CSI), a partir da compilação e acréscimo das transcrições de áudio das três temporadas finais da série.

Como se pode perceber, a série escolhida como ponto de partida para esta pesquisa se mostra altamente conhecida pelo grande público, não apenas em nosso país, mas também no mundo todo, o que faz com que a linguagem utilizada por seus personagens seja amplamente difundida, despertando o interesse de muitos para a área da investigação criminal, que, até então, era pouco conhecida pela maioria da população.

Uma vez que a série retrata a rotina de trabalho de peritos criminais, embora seja ficcional, conta com termos e colocações usadas especificamente nesta área. Segundo Fromm (2011), desde a década de 1960 é possível perceber um aumento no uso de termos emprestados de várias áreas do conhecimento por parte das séries norte-americanas.

Fromm (2011) identificou três tipos de padrões no uso da terminologia nas séries norte-americanas:

- a. séries com terminologia totalmente ficcional: figuram neste item programas sobre o fantástico, com personagens e características que existem apenas em mundos criados a partir da imaginação. Exemplos clássicos são as séries sobre monstros, fantasmas, espíritos, mitologias, etc. [...].
- b. séries que misturam ficção e ciência: campo propício para a ficção científica, nesses programas encontramos terminologia própria de áreas reais como Física, Química, Astronáutica, Astronomia, Engenharias, Biologia, etc. No entanto, como a ficção também se faz presente, termos são criados, dentro dessas áreas, para explicar fenômenos que ainda não podem ser explicados, ou ainda fenômenos inventados. [...].
- c. séries que retratam o cotidiano de médicos, investigadores, cientistas forenses, etc.: neste caso, as terminologias usadas tentam retratar, de forma fiel, aquelas usadas pelos profissionais dessas áreas nos seus afazeres cotidianos. [...]. (FROMM, 2011, p. 2).

De acordo com a distinção apresentada pelo autor, a série escolhida como ponto de partida para a busca e o levantamento das colocações especializadas da área criminal se insere na terceira categoria, uma vez que retrata o cotidiano de cientistas forenses e os procedimentos realizados durante a investigação criminal.

Assim, entende-se que nosso corpus, composto pelas legendas da série *CSI – Crime Scene Investigation*, pode ser reconhecido por seu caráter especializado. Tal como define Hoffman (1988 apud FINATTO, 2015, p. 47), “o texto especializado é instrumento e, ao

mesmo tempo, resultado da atividade comunicativa exercida em relação a uma atividade especializada sócio-produtiva”. Neste caso, nota-se que os textos que compõem o corpus são resultantes de atividades sociais produtivas, mais especificamente trata de investigações empreendidas pelos cientistas forenses da série.

Ainda segundo Hoffman (1998), além de textos escritos e impressos, elementos de oralidade, tais como diálogos e discussões, se inserem no conceito de texto especializado (HOFFMAN, 1988 apud FINATTO, 2015). Dessa forma, as séries televisivas, as quais são compostas por elementos de oralidade, como os diálogos, e que são estudadas a partir de suas legendas, constituem-se, também, um tipo de texto especializado.

A partir do exposto, toma-se como base a relevância da série escolhida para compor o corpus de estudo e os pressupostos teóricos apresentados aqui como fatores que corroboram o uso de uma série como ponto de partida para a busca e o levantamento de colocações especializadas da área criminal, haja vista que se trata de uma série constituída por textos especializados, que consegue alcançar um número extremamente elevado de pessoas que passam a se interessar pela área e com isso, buscam conhecer e aprender os vocábulos, termos e as colocações da referida área.

Após a apresentação da série que compõe o corpus de estudo desta pesquisa e dos motivos que nos levaram a sua escolha, apresentam-se, no capítulo a seguir, os pressupostos teóricos que nortearam o desenvolvimento desta pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Tendo em vista que este trabalho permeia diferentes áreas de pesquisa, apresentamos, neste capítulo, uma revisão das teorias que servem de fundamentação para o seu desenvolvimento. Dessa maneira, trata-se de aspectos concernentes à Fraseologia e à Fraseologia Baseada em Corpus. Em seguida, são discutidos aspectos sobre as colocações da língua geral e as colocações especializadas, objeto de estudo desta pesquisa. Por fim, apresentam-se os pressupostos teóricos da Linguística de Corpus e da Abordagem Lexical, bem como suas contribuições para o desenvolvimento deste trabalho.

2.1 Fraseologia

Ainda considerada uma disciplina relativamente jovem na área da Linguística, a Fraseologia vem sendo estudada por diversos autores desde o século XVII, tendo sido registrada pela primeira vez com o estudo empreendido por Vasilevich Lomonósv (MIRONESKO, 1997 apud MONTEIRO-PLANTIN, 2014). Deste então, vários autores desenvolveram estudos fraseológicos, como Bally (1909) e Saussure (1916 [1969]), no início do século XX, assim como outros pesquisadores algumas décadas depois, tais como Zuluaga (1980), Tristá (1988), Corpas Pastor (1996), Iñesta Mena e Pamies Bertrán (2002), Tagnin (1989 [2011]), Orenha-Ottaiano (2004, 2009), entre outros.

Saussure (1916 [1969]), por exemplo, embora não aborde explicitamente o termo fraseologia, ao discorrer sobre as relações sintagmáticas, leva em consideração as expressões fixas da língua ao afirmar que “há, primeiramente, um grande número de expressões que pertencem à língua; são as frases feitas, nas quais o uso proíbe qualquer modificação, mesmo quando seja possível distinguir, pela reflexão, as partes significativas”. (SAUSSURE, 1916 [1969], p. 144). Percebe-se que Saussure já mencionava a existência das expressões fixas, as frases feitas, que, embora sejam explicadas de diversas maneiras pelos diversos autores que as estudam, são o objeto de estudo da Fraseologia.

A Fraseologia passa a ganhar *status* de ciência linguística, sendo considerada, na época, uma subdisciplina da Lexicologia, a partir do estudo das expressões fixas realizado por Bally, em 1909. Neste estudo, ele classifica as expressões fixas em três grupos, a saber: i) combinações livres; ii) agrupamentos usuais ou séries fraseológicas, e iii) unidades fraseológicas (MONTEIRO-PLANTIN, 2014), passando a ser considerado o pai da Fraseologia desde então.

Apesar da grande importância e da classificação empreendida pelo estudo promovido por Bally (1909), a Fraseologia teve suas bases teóricas e conceitos fundamentais estabelecidos, assim como passou a ter *status* de disciplina linguística independente da Lexicologia, apenas após contribuições dos estudos de linguistas soviéticos nas décadas de 1930 e 1940 (POLIVÁNOV, 1931; VINOGRADÓV, 1947), tendo Vinogradóv (1947) recebido grande destaque por seus trabalhos, além de ter sido o responsável por estabelecer que o objetivo da Fraseologia é:

o estudo das leis que condicionam a falta de liberdade das palavras e dos significados das palavras para se combinarem, e a descrição sobre esta base das combinações fixas segundo seus tipos, tanto em seu estado atual como em seu desenvolvimento histórico¹. (TELIA, 1963, apud TRISTÁ, 1988).

Desde então, muitos outros autores tem contribuído com os estudos sobre a Fraseologia, tais como Casares (1950), que apresenta uma tipologia das unidades fraseológicas que é tida como modelo para outros autores (COSERIU, 1966; TRISTÁ, 1985; ZULUAGA, 1980) que, ao desenvolver sua técnica do discurso repetido, engloba as expressões que são tradicionalmente fixas no idioma, como modismos, frases feitas, ou locuções, cujos elementos não podem ser recombinaados ou realocados, pois, assim, perderiam seu sentido (CORPAS PASTOR, 1996).

Ao apresentar informações acerca do início dos estudos da Fraseologia, busca-se evidenciar o longo período em que esta disciplina vem sendo estudada, bem como, apresentar a definição de seu objeto de estudo, haja vista que o próprio termo Fraseologia, bem como os fenômenos por ele abarcados, “não está livre de controvérsias”², conforme ressalta Corpas Pastor (1996, p. 16).

Desse modo, trazemos as definições de alguns autores sobre a Fraseologia e seu objeto de estudo. Zuluaga (1980) não traz o termo Fraseologia enquanto disciplina em seu livro, intitulado *Introducción al estudio de las expresiones fijas* (*Introdução ao estudo das expressões fixas*). Embora não mencione a disciplina, ao explicar as expressões fixas também usa o termo *unidades fraseológicas*, bem como fala da *fixação fraseológica* (ZULUAGA, 1980, p. 95), evidenciando a inserção de sua pesquisa nos estudos fraseológicos. O autor explica a escolha dessa denominação da seguinte forma:

¹ “El estudio de las leyes que condicionan la falta de libertad de las palabras y de los significados de las palabras para combinarse, y la descripción sobre esta base de las combinaciones fijas de palabras según sus tipos, tanto en su estado actual como en su desarrollo histórico”.

² “no está libre de controversia”.

As chamamos de ‘expressões fixas’ porque sua característica constitutiva é a fixação, ou também ‘unidades fraseológicas’ porque funcionam como unidades em diferentes níveis gramaticais e porque, com muitas poucas exceções perfeitamente identificáveis, estão formadas por combinações de palavras. (ZULUAGA, 1980, p. 15)³.

Dessa forma, Zuluaga (1980) já delimita o objeto de estudo de sua pesquisa e, conseqüentemente, o objeto de estudo da Fraseologia, deixando claro que trabalha com as unidades fraseológicas, e que a fixação dessas unidades ou expressões é sua característica mais marcante. No decorrer de seu livro, o autor aborda a questão da idiomaticidade dessas expressões, ressaltando que esta se constitui como outra característica importante das unidades fraseológicas. Ainda neste estudo, Zuluaga delimita e classifica dois tipos de unidades fraseológicas, as locuções e os enunciados fraseológicos.

Tristá (1988) apresenta as definições de Bally (1909) e de Vinogradov (1947) acerca da Fraseologia e conclui que, naquela ocasião, “a fraseologia constitui um ramo especial da linguística, com seus métodos e objeto de estudo”⁴ (TRISTÁ, 1988, p. 10), fundamentando seu trabalho nas definições destes autores. Além disso, apresenta três características principais que devem ser observadas para se classificar uma expressão ou combinação de palavras como unidade fraseológica. Segundo a autora, estas características são a pluriverbalidade, o sentido figurado e a estabilidade (TRISTÁ, 1988, p. 13).

Por pluriverbalidade, Tristá (1988) entende que todo fraseologismo, ou unidade fraseológica, deve ser composto por duas ou mais palavras, sendo que pelo menos uma delas deverá ser plena, e poderá estar ou não acompanhada de uma ou mais palavras auxiliares, além de ser possível encontrarmos unidades fraseológicas compostas por duas palavras plenas. Por sentido figurado, a autora entende como a presença de metáforas nas unidades fraseológicas, sendo a metáfora considerada por ela como o “fator semântico por excelência que atua na formação destas unidades”⁵ (TRISTÁ, 1988, p. 21). Já, a estabilidade, está relacionada com a noção de fixação das unidades fraseológicas defendidas pelos demais linguistas, como Zuluaga (1980), por exemplo.

Por sua vez, Corpas Pastor (1996) considera que há uma grande variedade de termos usados para denominar os fenômenos linguísticos que constituem o objeto de estudo da

³ “Las llamamos ‘expresiones fijas’ porque su rasgo constitutivo es la fijación, o, también, ‘unidades fraseológicas’ porque funcionan como unidad en diferentes niveles gramaticales y porque, con muy pocas excepciones perfectamente identificables, está formadas por combinaciones de palabras.

⁴ “[...], a fraseología constituye una rama especial de la linguística, con sus métodos y objeto de estudio.”

⁵ “[...] el factor semántico por excelencia que actúa en la formación de estas unidades.”

Fraseologia, além de haver uma variedade terminológica também para as classificações usadas nas análises desses fenômenos. Para a autora, contudo, os linguistas se mostram de acordo ao utilizarem o termo Fraseologia para se referirem à disciplina que se ocupa do “estudo das combinações de palavras” (CORPAS PASTOR, 1996, p. 16). A autora, assim, considera a Fraseologia como disciplina, e a insere como uma subdisciplina da Lexicologia, tal como Vinogradóv (1947) e os demais linguistas soviéticos de sua época.

Corpas Pastor (1996, p. 18) opta chamar as combinações de palavras estudadas em seu trabalho de *unidades fraseológicas*, justificando tal escolha pelo fato de este termo ter ganhado cada vez mais adeptos entre os linguistas e por ser mais genérico, abrangendo toda a variedade de expressões fixas que a autora aborda em sua obra. Assim, a autora define as unidades fraseológicas como

[...] unidades léxicas formadas por mais de duas palavras gráficas em seu limite inferior, cujo limite superior se situa no nível da oração composta. Estas unidades se caracterizam por sua alta frequência de uso, e de co-aparição de seus elementos integrantes; por sua institucionalização, entendida em termos de fixação e especialização semântica; por sua idiomaticidade e variação potenciais; assim como pelo grau no qual se dão todos estes aspectos nos distintos tipos. (CORPAS PASTOR, 1996, p. 20).⁶

Em contrapartida, Tagnin (2013) prefere falar em Convencionalidade e Idiomaticidade ao se referir às unidades convencionais que, embora com denominação diferente, constituem o objeto de estudo da Fraseologia. Segundo a autora, “há expressões que são convencionais por estarem intimamente ligadas a um fato social e há outras em que o que é convencional é sua forma” (TAGNIN, 2013, p. 21), ou seja, convencionou-se usar determinadas expressões em uma determinada língua, sempre em uma mesma forma, o que pode ser entendido como a noção de fixação definida por Zuluaga (1980) e Corpas Pastor (1996), por exemplo. Tagnin afirma, ainda, que “no momento em que a convenção passa para o nível do significado entramos no campo da idiomaticidade” (TAGNIN, 2013, p. 22), retomando, assim, a noção de idiomaticidade também vista em Zuluaga (1980) e Corpas Pastor (1996).

Monteiro-Plantin (2012), ao propor uma definição de Fraseologia, retoma aspectos definidos e discutidos por diferentes autores, conforme podemos observar a seguir

⁶ “[...] son unidades léxicas formadas por más de dos palabras gráficas en su límite inferior, cuyo límite superior se sitúa en el nivel de la oración compuesta. Dichas unidades se caracterizan por su alta frecuencia de uso, y de coaparición de sus elemento integrantes; por su institucionalización, entendida en términos de fijación y especialización semántica; por su idiomaticidad y variación potenciales; así como por el grado en el cual se dan todos estos aspectos en los distintos tipos”.

Trata-se de uma disciplina independente, relacionada a todos os níveis de análise linguística (do fonético ao discursivo-pragmático), cujo objetivo é o estudo das combinações de unidades léxicas, relativamente estáveis, com certo grau de idiomaticidade, formadas por duas ou mais palavras, que constituem a competência discursiva dos falantes, em língua materna, segunda ou estrangeira, utilizadas convencionalmente em contextos precisos, com objetivos específicos, ainda que, muitas vezes, de forma inconsciente. (MONTEIRO-PLANTIN, 2012, p. 33)

Ao retomar aspectos que divergiam entre os autores ou que eram considerados apenas por um ou outro autor, Monteiro-Plantin (2012) consegue trazer uma definição mais precisa e específica da Fraseologia enquanto disciplina, bem como de seu objeto de estudo.

Assim, depois de levar em consideração as características e definições apresentadas pelos autores mencionados anteriormente, consideramos a Fraseologia como a disciplina que estuda as combinações de palavras, ou expressões fixas, podendo estas combinações de palavras serem classificadas como colocações, coligações, provérbios, frases feitas, ditos, expressões idiomáticas e binômios.

Até o momento, apresentamos alguns aspectos da Fraseologia e as definições dadas a esta disciplina por alguns de seus grandes precursores. Mais recentemente, dada a interdisciplinaridade entre Linguística de Corpus e diversas áreas de pesquisa, passamos a identificar estudos que tratam da Fraseologia Baseada em Corpus, a qual é abordada na subseção seguinte.

2.2 Fraseologia Baseada em *Corpus*

Neste item, apresentamos alguns pressupostos teóricos relevantes da Fraseologia Baseada em Corpus, haja vista que esta pesquisa privilegia o trabalho conjunto entre Fraseologia e Linguística de Corpus (a qual é apresentada e discutida no subitem 2.4).

De acordo com Granger e Paquot (2008), os aspectos teóricos e tradicionais da Fraseologia, conforme mencionados até aqui, dizem respeito a uma das abordagens da Fraseologia, a Abordagem Tradicional. Segundo as autoras, há, ainda, a Abordagem Distribucional ou Baseada em Frequência, a qual é caracterizada por se valer da Linguística de Corpus e suas ferramentas como uma forma de busca e identificação das unidades fraseológicas.

Esta Abordagem Distribucional ou Baseada em Frequência, segundo Granger e Paquot (2008), teve início com o desenvolvimento do trabalho lexicográfico de Sinclair, em 1987, e ficou conhecida por utilizar a abordagem dirigida por corpus (*corpus-driven approach*), com

o objetivo de realizar a busca e a identificação das coocorrências lexicais. Além de utilizar a abordagem dirigida por corpus para a identificação das coocorrências lexicais, esta abordagem passa a considerar e analisar sequências como quadros colocacionais e coligações que, até então, eram consideradas periféricas na língua, de forma que estas passem a ser consideradas centrais para os estudos fraseológicos, pois leva em consideração a frequência com que estas são encontradas nos corpora de estudo (GRANGER; PAQUOT 2008, p. 29).

Conforme também ressalta Gries (2008), desde a pesquisa lexicográfica de Sinclair, a Fraseologia tem contado cada vez mais com as ferramentas da Linguística de Corpus, uma vez que “os procedimentos de identificação mais abrangentes para fraseologismos são certamente encontrados na Linguística de Corpus”⁷(GRIES, 2008, p. 20), o que se deve ao fato de a Linguística de Corpus (doravante LC) ser uma metodologia que privilegia a coocorrência lexical (GRIES, 2008, p. 20).

Além de Gries (2008), vários autores têm conciliado a LC e a Fraseologia em suas pesquisas, tais como Sinclair (1987; 1991), Partington (1998), Granger e Paquot (2008) Kennedy (2008), Tagnin (2013), entre outros.

Sinclair (1991) explica que seu trabalho lexicográfico desenvolvido junto ao *Projeto Cobuild* destacou-se por levar em consideração a língua em uso, por meio de corpora que permitiam a busca por unidades léxicas de forma que o contexto em que tais unidades ocorriam eram, também, considerados e analisados, método ressaltado pelo autor ao afirmar que “qualquer instância de significação da língua depende dos contextos que a rodeiam”⁸. Assim, o autor passa a considerar o contexto em que cada unidade léxica se insere, facilitando o estudo das unidades fraseológicas por meio das palavras em contexto, fornecidas pelas ferramentas resultantes do aporte metodológico da LC, contribuindo para o aumento de pesquisas na área fraseológica.

No entanto, segundo Bretaña e Bertrán (2008), em seu estudo acerca de critérios estatísticos para a extração de unidades fraseológicas a partir de corpus computadorizado, Sinclair, já na década de 1960, aplicou cálculos estatísticos advindos da LC para o estudo de coocorrência linear de palavras e, Jones e Sinclair (1974) usaram a frequência de coocorrência como um fator definidor de colocações (BÉRTRAN; BRETANÑA, 2008, p. 2). Percebe-se, assim, que os estudos de Sinclair conciliando a LC, a Lexicografia e Fraseologia tiveram

⁷ “The most comprehensive identification procedures for phraseologisms are certainly found in corpus linguistics”.

⁸ “[...]. Any instance of language depends on its surrounding context.”

início bem antes de seu *Projeto Cobuild*, o qual é considerado por Granger e Paquot (2008) o divisor de águas entre a Fraseologia Tradicional e a Fraseologia baseada em corpus.

Partington (1998) é mais um autor que realiza estudos sobre as unidades fraseológicas, de modo mais específico as colocações, por meio da Fraseologia Baseada em Corpus, sendo que, ao apresentar os diferentes tipos de definições dadas às colocações, considera a definição dada por Hoey (1991) como uma definição estatística, a qual é comumente utilizada por pesquisadores da área de LC. O autor afirma ainda que, no caso das colocações, objeto de seu estudo, “as associações habituais de uma palavra com outros itens podem, assim, ser estudadas tanto pelas chamadas concordâncias dessa palavra, como pela obtenção de listas de seus colocados mais frequentes”⁹ (PARTINGTON, 1998, p. 16).

Por mencionar as duas possibilidades de análises das colocações oferecidas pela LC, o autor dedica os capítulos dois e três de seu livro a exemplificar como as linhas de concordância auxiliam na observação e no contraste de comportamento de diferentes adjetivos, evidenciando os padrões colocacionais destes, além de serem úteis para tradutores compararem padrões de comportamento das unidades léxicas, colaborando com a escolha tradutória que farão para suas traduções (PARTINGTON, 1998, p. 27).

Kennedy (2008) desenvolveu uma pesquisa com o objetivo de verificar se as colocações constituídas por verbos lexicais são formadas arbitrariamente, ou se os colocados dos verbos estudados não são arbitrários, mas tendem a ter características comuns de forma e significado (KENNEDY, 2008, p. 22). Nesta pesquisa, o autor concilia os pressupostos da Fraseologia, mais especificamente, das colocações, com as ferramentas e alguns pressupostos próprios da LC, além de utilizar o *British National Corpus (BNC)* como corpus de sua pesquisa.

No Brasil, Tagnin (2013), ao tratar da convencionalidade da língua, faz uso das ferramentas e da metodologia da LC para o tratamento das colocações que analisa. Dessa forma, pode-se dizer que as pesquisas da autora se inserem no âmbito da Fraseologia Baseada em Corpus. Em sua obra *O jeito que a gente diz*, de 2013, a autora dedica um capítulo exclusivo à LC, enfatizando como esta contribui para as análises fraseológicas, mais especificamente, com as análises de colocações.

Assim, passamos a discorrer sobre as colocações, que recebem destaque nas pesquisas realizadas na Fraseologia Baseada em Corpus, além de serem objeto de estudo neste trabalho.

⁹ “[...] the habitual associations of a word with other items can thus be studied both by calling up concordances of that word and by obtaining lists of its most frequent collocates.”

2.3 Colocações da língua geral e colocações especializadas

Como vimos anteriormente, a Fraseologia se ocupa de diferentes unidades fraseológicas, ou fraseologismos, de forma que esta pesquisa opta por privilegiar o estudo de uma unidade fraseológica específica, as colocações.

Conforme aponta Orenha-Ottaiano (2016, p. 487), o termo *colocação*, apesar de normalmente atribuído à Firth (1957) graças às contribuições que este autor deu para a compreensão do conceito de colocação ao explicar os casos de coocorrências sintáticas, já foi mencionado na literatura anteriormente por H. E. Palmer, na década de 1950, e por Otto Jespersen, em 1917. A partir de então, diversos pesquisadores se dedicaram ao estudo das colocações, tais como Hausmann (1984, 1985), Sinclair (1991), Corpas Pastor (1996), Heid (2009) e Tagnin (1999, 2013), os quais apresentam suas próprias definições de colocação, além de caracterizá-las de formas diferentes.

Desse modo, Hausmann (1984), um dos autores que contribuiu para o estudo e a descrição das colocações, as entende como “produtos semi-acabados da língua”, considerando como aspecto mais importante destas “[...] seu estado de disponibilidade mental como um todo, não como uma criação produzida para uma tal finalidade pelo falante.” (HAUSMANN, 1984 apud ORENHA-OTTAIANO, 2012a). Para o autor, apenas reutilizamos estes “produtos semi-acabados da língua” ao nos comunicarmos, uma vez que as colocações são pré-determinadas em nosso vocabulário.

O autor trata, ainda, da descrição dos elementos que compõem a colocação, denominando-os *base* e *colocado*. Para Hausmann (1984), cada elemento possui uma função clara e hierárquica, sendo a base autônoma e responsável por determinar, enquanto o colocado é o elemento determinado pela base (HAUSMANN, 1984).

Orenha-Ottaiano (2009, p. 40), fundamentada em Heid *et all* (1991), caracteriza a base como sendo o elemento independente, com autonomia semântica, que pode ser traduzido independente de seu uso na colocação e responsável por determinar os padrões lexicais que a acompanham na colocação. Já o colocado é caracterizado como um conceito modificador, o qual pode ser semanticamente interpretável dentro da colocação, sendo traduzido de acordo com seu uso na colocação, além de ser escolhido pela base.

Sinclair (1991) propõe o princípio de livre escolha e o princípio idiomático como formas de interpretação da linguagem. O princípio de livre escolha, exemplificado pelo modelo *slot-and-filler*, é descrito pelo autor como o princípio pelo qual os textos apresentam “uma série de lacunas, as quais devem ser preenchidas por uma unidade lexical que satisfaça

suas limitações locais”¹⁰ (SINCLAIR, 1991, p. 109). Dessa forma, qualquer palavra pode ser usada para preencher estas lacunas, visto que, apenas a gramaticalidade é usada como meio de restrição. Embora apresente este princípio, o autor propõe, em seguida, o princípio idiomático, já que considera o princípio de livre escolha insuficiente para explicar a linguagem, uma vez que “virtualmente, todas as gramáticas são construídas a partir do princípio de livre escolha”¹¹ (SINCLAIR, 1991, p. 109).

Para Sinclair (1991, p. 110), o princípio idiomático é responsável pelo fato de que o “usuário de uma língua tem a sua disposição um grande número de frases semi-pré-construídas que constituem escolhas únicas, apesar de parecer serem analisadas em segmentos”¹². Segundo este princípio, usamos unidades pré-fabricadas em nossa comunicação, de forma a processarmos mais rapidamente as informações de um texto ou durante a fala, e até mesmo, a economizar esforços para tal atividade, uma vez que estas unidades pré-fabricadas são constituídas por mais de uma palavra / unidade lexical, as quais recebem o nome de colocação.

Percebemos, assim, que, para o autor, a noção de colocação está intimamente relacionada ao princípio idiomático, já que, “em algumas ocasiões, as palavras parecem ser escolhidas em pares ou grupos e estas não são necessariamente adjacentes”¹³ (SINCLAIR, 1991, p. 115).

Assim como Hausman (1984), Sinclair (1991) propõe uma nomenclatura para designar as partes que compõem uma colocação. O autor se vale do termo *node* (nódulo) para se referir à palavra que é estudada, e *collocate* (colocado) para designar a palavra que coocorre com o nódulo. Além disso, as colocações são classificadas como *upward collocations* e *downward collocations* pelo autor. As primeiras, *upward collocations*, representam o “padrão mais fraco em termos estatísticos, e as palavras tendem a ser elementos de estruturas gramaticais”¹⁴, enquanto que as *downward collocations* permitem uma “análise semântica da palavra”¹⁵ (SINCLAIR, 1991, p. 116). Dessa forma, entendemos as *upward collocations* como colocações compostas por um item gramatical como colocado, ou seja, o colocado pode ser uma preposição ou uma conjunção, como no exemplo “*back into*”, usado

¹⁰ “[...], a series of slots which have to be filled from a lexicon which satisfies local restraints.

¹¹ “Virtually all grammars are constructed on the open-choice principle.”

¹² “[...] a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments.”

¹³ “On some occasions, words appear to be chosen in pairs or groups and these are not necessarily adjacent.”

¹⁴ “[...] is the weaker pattern in statistical terms, and the words tend to be elements of grammatical frames, [...]”

¹⁵ “[...] gives us a semantic analysis of a word.”

pelo autor. Já, nas *downward collocations*, os colocados apresentam significado semântico proveniente de verbos e substantivos, como “*back home*”, também apresentado pelo autor.

Corpas Pastor ganhou destaque na área da Fraseologia e nos estudos das colocações ao publicar seu *Manual de fraseologia española* (1996), no qual se dedica ao estudo das colocações. Neste seu estudo, a autora traz informações referentes às colocações desde seus precursores, traçando um percurso histórico no qual enfatiza os trabalhos de Halliday (1966), Hausmann (1984, 1985), Cowie (1985), Sinclair (1991), entre outros, chegando a apresentar como se deu o surgimento do termo colocação nos estudos filológicos espanhóis. De acordo com a autora, a noção de colocação é “esboçada na teoria semântica de Coseriu (1966, 1977 [1986], 1978), concretamente dentro das denominadas *solidariedades léxicas*”¹⁶ (CORPAS PASTOR, 1996, 0. 63).

Além de fazer um levantamento histórico para contextualizar o estudo das colocações, a autora denomina como colocações:

[...] as combinações assim resultantes, ou seja, as unidades fraseológicas formadas por duas unidades léxicas em relação sintática, que não constituem por si mesmas, atos de fala nem enunciados; e que, devido a sua fixação na norma, apresentam restrições de combinação estabelecidas pelo uso, geralmente de base semântica: o colocado autônomo semanticamente (a base) não só determina a eleição do colocativo, como também, seleciona neste uma aceção especial, frequentemente de caráter abstrato e figurativo. ¹⁷ (CORPAS PASTOR, 1996, p. 66)

Convém ressaltar aqui que, embora a autora afirme que as colocações são compostas por duas unidades léxicas, entende-se que estas são compostas por pelo menos duas unidades léxicas, haja vista que algumas colocações são compostas por mais de duas unidades, como é o caso das colocações especializadas estendidas observadas por Orenha-Ottaiano (2009). Além disso, Corpas Pastor entende por colocação “aquela propriedade pela qual os falantes das línguas tendem a produzir certas combinações de palavras entre uma grande quantidade de combinações teoricamente possíveis”¹⁸ (CORPAS PASTOR, 1996, p. 66).

¹⁶ “[...] aparece ya esbozada en la teoría semántica de Coseriu (1966, 1978, 1986 [1977]), concretamente dentro de las denominadas *solidariedades léxicas*.”

¹⁷ “[...] a las combinaciones así resultantes, es decir, a las unidades fraseológicas formadas por dos unidades léxicas en relación sintáctica, que no constituyen, por sí mismas, actos de habla ni enunciados; y que, debido a su fijación en la norma, presentan restricciones de combinación establecidas por el uso, generalmente de base semántica: el colocado autónomo semánticamente (la base) no sólo determina la elección del colocativo, sino que, además, selecciona en éste una acepción especial, frecuentemente de carácter abstracto o figurativo.”

¹⁸ “[...] aquella propiedad de las lenguas por la que los hablantes tienden a producir ciertas combinaciones de palabras entre una gran cantidad de combinaciones posibles.”

Por sua vez, Granger (1998) em seu trabalho sobre o uso de padrões pré-fabricados na escrita de aprendizes de nível avançado de inglês como língua estrangeira, adota a definição de Van Roey (1990) para as colocações. Desse modo, Granger (1998) considera que uma colocação é “o fenômeno linguístico pelo qual um determinado item de vocabulário prefere a companhia de outro item em vez de seus ‘sinônimos’ por causa de restrições que não estão no nível da sintaxe ou significado conceitual, mas sim no de uso” (VAN ROEY, 1990, p. 46).

Já Tagnin (1999), ao abordar as colocações, aponta algumas características que devem ser observadas para a sua definição, tais como a não-idiomaticidade, a coesão, a restrição textual e coocorrência arbitrária entre seus elementos. Dessa forma, uma combinação lexical deve ter um significado composicional, apresentar uma forte ligação entre seus elementos, ocorrer em um contexto específico, e não apresentar uma razão semântica que explique seus elementos, para ser considerada uma colocação (TAGNIN, 1999).

Nesselhauf (2005) também parte para a área das combinações de palavras ao empreender um estudo sobre as colocações presentes em um corpus de aprendizes, afirmando que entende as colocações como sendo “um tipo de combinação de palavras de um certo padrão gramatical”¹⁹ (NESSELHAULF, 2005, p.25), e ressalta que, embora seu estudo privilegie as colocações verbo-nominais, esta definição pode ser aplicada a outros tipos de combinações também. Para a autora, existem duas formas principais de se entender as colocações: uma chamada de “abordagem estatisticamente orientada ou de abordagem baseada em frequência”²⁰; e outra chamada de “abordagem orientada pela significância ou de abordagem fraseológica”²¹ (NESSELHAULF, 2005, p. 12).

Segundo Nesselhauf (2005), autores como Firth (1957) e Sinclair (1991) se inserem entre os defensores da abordagem baseada em frequência, a qual entende por colocação a coocorrência de palavras em uma certa distância, e na qual é feita uma distinção entre coocorrências frequentes e coocorrências não frequentes. Já a abordagem fraseológica é defendida por autores como Cowie (1981), Mel’cuk (1998) e Hausman (1984, 1985), sendo que esta abordagem entende que as colocações são um tipo de combinação de palavras, mais comumente as que são fixas até certo grau (NESSELHAULF, 2005, p. 11-12). A partir dessa divisão, a autora se posiciona como adepta da abordagem fraseológica, após apresentar sua definição de colocação. Nesta pesquisa, nos baseamos na abordagem fraseológica, uma vez

¹⁹ “[...] a type of word combination in a certain grammatical pattern, [...]”

²⁰ “[...] ‘statistically oriented approach’ [...] or the ‘frequency-based approach’ [...]”

²¹ “[...] ‘significance oriented approach’ [...] or the ‘phraseological approach’ [...]”

que entendemos que as colocações são combinações de palavras que apresentam certo grau de fixidez.

Por sua vez, Fritzinger e Heid (2009) afirmam que o termo “colocação” tem sido usado para se referir a uma vasta gama de fenômenos, além de ser tratado com distinção pelos linguistas computacionais, e por aqueles que defendem o contexto, tal como Nesselhauf (2005) também afirma. No entanto, os autores ressaltam que, para eles, colocação é entendida de acordo com a definição proposta por Bartsch (2004), o 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”²² (BARTSCH, 2004, p.76). De acordo com os autores, esta definição baseia-se em três critérios, a saber, “co-seleção lexical (a base seleciona seu colocado), significância estatística de coocorrência e padrões sintáticos”²³ (FRITZINGER; HEID, 2009, p. 2-3).

O critério de co-seleção lexical apontado pelos autores, diz respeito ao fato de a base da colocação ser responsável por selecionar seu colocado. Dessa forma, pode-se dizer que este critério, apontado por Fritzinger e Heid (2009) relaciona-se à distinção terminológica dos termos que compõem uma colocação, a *base* e o *colocado*, feita por Hausmann (1984). Além disso, a significância estatística de coocorrência considera a frequência com que as combinações de palavras que formam determinada colocação coocorrem, ou seja, considera a estatística de frequência das colocações. Já o critério de padrão sintático apresentado por Fritzinger e Heid (2009) refere-se ao fato de as colocações obedecerem a certos padrões sintáticos da língua que pertencem ao se formarem, por exemplo, os padrões de verbo+objeto e adjetivo+substantivo estudados pelos autores.

Ao tomarmos como exemplo a colocação *a physical evidence* que ocorre 49 vezes no CCSI, 409 vezes no CCC e 277 vezes no emulador de concordanciador *WebCorp*, é possível identificar estes três critérios: a co-seleção lexical destaca-se pelo fato da base desta colocação (*evidence*) ser a responsável por determinar seu colocado (*physical*); além disso, é possível identificarmos um padrão sintático que obedece às normas da língua, adjetivo+substantivo, tal como sugerido pelos autores; o critério de estatística de coocorrência é ressaltado pela frequência com que coocorre a combinatória em análise.

Já Laufer e Waldman (2011), ao apresentaram seu trabalho relacionado ao uso das colocações na produção escrita livre de aprendizes de inglês como segunda língua, afirmam

²² “Collocations are lexically and/or pragmatically constrained recurrent co-occurrences of at least two items which are in a direct syntactic relation with each other.”

²³ “[...] lexical co-selection (a base selects its collocates), statistical significance of cooccurrence, and syntactic patterns.”

que, a partir das definições apresentadas, e das duas vertentes predominantes sobre os estudos colocacionais, já mencionadas por Nesselhauf (2005) e por Fritzinger e Heid (2009), consideram as colocações como “combinações lexicais habituais que se caracterizam pela coocorrência restrita de elementos e relativa transparência de significado”²⁴ (LAUFER; WALDMAN, 2011, P. 648). Ainda segundo as autoras, esta

coocorrência restrita distingue as colocações das combinações em que as palavras individuais são facilmente substituíveis seguindo as regras da gramática. A relativa transparência semântica das colocações, por outro lado, as distingue das expressões idiomáticas, cujo significado é muito menos transparente que o das colocações e é muitas vezes opaco porque não pode ser compreendido a partir das palavras que os compõem.²⁵ (LAUFER; WALDMAN, 2011, p. 648)

Ao observar todas as definições de colocação apresentadas aqui por diferentes autores, percebe-se que, embora haja uma variação na maneira em que esta é explicada, a noção de colocação é bem semelhante para todos os autores, sendo que a abordagem baseada em frequência e a abordagem fraseológica (NESSELHAULF, 2005) são responsáveis pelas diferenças mais substanciais na identificação de uma colocação: uma considera a ocorrência de palavras em uma dada distância e utiliza a frequência desta ocorrência como critério definidor; a outra utilização o grau de fixidez de dada combinação de palavras.

Nesta pesquisa, tal como mencionado anteriormente, adotamos a abordagem fraseológica, mais especificamente, compartilhamos dos critérios adotados por Fritzinger e Heid (2009) baseados na definição de Bartsch (2004), a qual entende por colocação uma coocorrência lexical restrita, composta por pelo menos dois itens lexicais em relação sintática entre si. A partir de nossas análises, percebemos que estas características também são observáveis nas colocações especializadas.

No que diz respeito às colocações especializadas, L’Homme e Bertrand (2000) as denominam “combinações lexicais especializadas” e as diferem das colocações da língua geral ao dizerem que as “colocações são convencionais em dada comunidade linguística; as combinações lexicais especializadas são convencionais em um dado grupo de especialistas”²⁶

²⁴ “[...] habitually occurring lexical combinations that are characterized by restricted co-occurrence of elements and relative transparency of meaning.”

²⁵ “Restricted co-occurrence distinguishes collocations from the combinations in which the individual words are easily replaceable following the rules of grammar. Relative semantic transparency of collocations, on the other hand, distinguishes them from idioms whose meaning is much less transparent than that of collocations and is very often opaque because it cannot be understood from the words that compose them.”

²⁶ “Collocations are conventional within a given linguistic community; SLCs are conventional within a group of specialists.”

(L'HOMME; BERTRAND, 2000, p. 498). Além disso, segundo Orenha-Ottaiano (2009), a principal característica que diferencia uma da outra reside no fato de, nas colocações especializadas, a base ser uma unidade terminológica, enquanto que as colocações da língua geral têm como base uma unidade lexical.

No tocante à presença de um termo como base das colocações especializadas, retomamos os estudos de Bevilacqua (2004, 2005), nos quais a autora discorre sobre as unidades fraseológicas especializadas e afirma que tais unidades fraseológicas recebem as mais diversas denominações, de acordo com cada pesquisador, as quais também recebem a denominação de colocações especializadas.

Segundo Bevilacqua (2005), as unidades fraseológicas especializadas, ou colocações especializadas como as denominamos nesta pesquisa, são:

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).

Tal como L'Homme e Bertrand (2000), Bevilacqua (2005) chama a atenção para a presença de um termo nas unidades fraseológicas especializadas. No entanto, embora L'Homme e Bertrand (2000) apontem esta característica como principal diferença entre ambos os tipos de colocações, convém ressaltar que, para as autoras, as colocações especializadas, chamadas por elas de “combinações especializadas”, não se comportam da mesma maneira que as colocações da língua geral, o que contribui para que desenvolvam um método de análise específico para as chamadas “combinações lexicais especializadas”.

Segundo Orenha-Ottaiano (2012a, p. 155), L'Homme (2000) compara aspectos de ambos os tipos de colocações a partir de cinco diretrizes: a natureza convencional das combinações; a sua composição; a composicionalidade ou não das colocações; o agrupamento de bases em uma série de unidades lexicais; e a relação semântica entre os componentes. Para isso, L'Homme (2000) se baseia em seus próprios estudos, além de se apoiar na literatura existente.

Orenha-Ottaiano (2012a) afirma que, ao concluir sua pesquisa, L'Homme (2000) defende que as colocações da língua geral e as colocações especializadas não apresentam comportamento idêntico, e ressalta que

a autora defende que as colocações da língua geral são descritas em termos de coocorrência lexical restrita, e “mesmo que em alguns casos os lexemas que compartilham características semânticas tenham colocados em comum, esta não parece ser a propriedade predominante das colocações prototípicas” (L’HOMME, 2000, p. 106). Por sua vez, 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).

Embora L’Homme (2000) acredite haver mais diferenças que semelhanças entre as colocações especializadas e as colocações da língua geral, acreditamos, assim como Orenha-Ottaiano (2012), que ambas possuem mais semelhanças do que diferenças, sendo a diferença mais significativa aquela apontada acima, sobre a base das colocações especializadas ser composta por um termo.

Contudo, nesta pesquisa, consideram-se colocações especializadas aquelas que contenham um termo em sua formação, independente deste termo ocorrer como base ou como colocado da colocação em análise, haja vista que foi possível a identificação de ambos os casos entre as colocações especializadas levantadas, e procura-se descrever as tipologias das colocações encontradas de acordo com a taxonomia proposta por Orenha-Ottaiano (2009), fundamentada em Hausmann (1985), conforme apresentado a seguir, com exemplos retirados de nosso *corpus* de estudo:

Verbais

- Verbo _{base} + Subst. _{colocado} = *to find evidence*
- Subst. _{base} + Verbo _{colocado} = *ballistics confirms*
- Verbo _{base} + Det. + Subst. _{colocado} = *to have an alibi*
- Verbo _{colocado} + Prep. + [Det.]²⁷ + Subst. _{base} = *based on the evidence*
- Verbo _{colocado} + Part. Adverbial²⁸ + Subst. _{base} = *to sift through evidence*

Nominais

- Subst. _{base} + Subst. _{colocado} = *arrest warrant*
- Subst. _{base} + Prep. + [Det.] + Subst. _{colocado} = *marks on the body*

²⁷ Neste caso pode ou não ocorrer o determinante.

²⁸ No caso dos *phrasal verbs* da língua inglesa

Adjetiva

- Adj. *colocado* + Subst. *base* = *full arrest*

Adverbial

- Verbo *base* + Adv. *colocado* = *to shoot directly*
- Adv. *colocado* + Verbo *base* = *easy kill*

Após a apresentação das definições de colocações da língua geral e das colocações especializadas, bem como, do posicionamento quanto à definição de colocação especializada adotada nesta pesquisa, passa-se a discutir sobre a Linguística de Corpus e suas contribuições para o desenvolvimento deste trabalho na subseção seguinte.

2.4 Linguística de Corpus

Desde o surgimento do primeiro corpus eletrônico, o corpus *Brown*, na década de 1960, a LC tem visto suas pesquisas serem cada dia mais ampliadas e divulgadas, graças aos avanços tecnológicos e ao desenvolvimento e aprimoramento dos computadores e de ferramentas linguísticas computacionais que auxiliam nas análises empreendidas pelos linguistas de corpus. Vale mencionar que tais ferramentas podem ser apontadas como uma das principais características da LC.

No tocante as características da LC, muitas generalizações podem ser feitas, no que tange a seu *status*. Tal como afirmam McEnery e Hardie (2012, p. 1) “é importante perceber que a linguística de corpus é um campo heterogêneo”²⁹. Para os autores, a LC “é uma área que foca em um conjunto de procedimentos, ou métodos, para estudar a língua”³⁰(MCENERY; HARDIE, 2012, p. 1), ou seja, os autores atribuem à LC *status* de metodologia, haja vista que a LC é responsável por estes procedimentos ou métodos que permitem o estudo da língua. Contudo, como bem ressaltam os autores, por se tratar de um campo heterogêneo, diferentes pesquisadores, como Sinclair (1991), Biber (1998) e Tognini-Bonelli (2002) atribuem diferentes *status* à LC, conforme apresentado a seguir.

Sinclair (1991), por exemplo, faz uma retomada de seu trabalho lexicográfico, que se tornou amplamente conhecido como o *Projeto Cobuild*, ao introduzir o leitor à LC. Tomando como exemplo este projeto, o autor chama a atenção do leitor para o fato de que, por meio da

²⁹ “[...] it is very important to realize that corpus linguistics is a heterogeneous field.”

³⁰ “[...] it is an area which focuses upon a set of procedures, or methods, for studying language [...].”

LC, vista por ele como uma metodologia que fornece ferramentas que permitem analisar grande quantidade de dados, é possível buscar evidências da língua, refletir sobre elas, identificando com maior facilidade os vocábulos que devem ou não ser inseridos em seu produto lexicográfico, de forma a evitar o risco de que um vocábulo recorrente passe despercebido pela análise manual de um linguista (SINCLAIR, 1991, p. 1-5). O autor leva seu leitor a compreender que esta se ocupa das análises de textos naturais, os quais constituem os corpora de estudo e permitem a busca de diversos vocábulos recorrentes, além de facilitar as análises quantitativas, permitindo, inclusive, a identificação daqueles vocábulos que ocorrem apenas uma vez e são tão importantes quanto os vocábulos que apresentam maior recorrência.

Biber (1998), por sua vez, considera a LC como uma abordagem caracterizada por ser empírica e analisar padrões atuais de uso em textos naturais, tendo como base de suas análises uma grande coleção de textos, conhecida como corpus, além de contar com a ajuda de computadores para suas análises, utilizando tanto técnicas automáticas e interativas, quanto técnicas qualitativas e quantitativas (BIBER, 1998, p. 4). Com isso, o autor deixa claro para o leitor que, em sua visão, a LC pode ser entendida como uma abordagem e que considera tanto a análise quantitativa quanto a qualitativa uma das principais características da LC.

Tognini-Bonelli (2002), em contrapartida, afirma que “embora a linguística de corpus pertença à esfera da linguística aplicada, ela difere de outras disciplinas parceiras sob o mesmo guarda-chuva, por poder ser vista como uma *metodologia de pré-aplicação*”³¹(TOGNINI-BONELLI, 2002, p. 1), haja vista que, diferentemente de outras metodologias que recebem um conjunto de regras pronto, a LC está em posição de definir suas próprias regras antes de aplicá-las. A autora considera, assim, a LC como “um novo empreendimento de pesquisas e abordagem filosófica para a investigação linguística”³² (TOGNINI-BONELLI, 2002, p. 1).

Diante da discussão sobre o *status* da LC e das definições expostas aqui, toma-se como base a visão de Tognini-Bonelli (2002), a qual, após as considerações feitas sobre o *status* da LC, a define como:

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, às quais se chegaram a partir de observações das instâncias reais. A observação dos fatos de

³¹ “[...] although corpus linguistics belongs to the sphere of applied linguistics, it differs from other partner disciplines under the same umbrella in that it can be seen as a *pre-application methodology*.”

³² “[...] a new research enterprise and a new philosophical approach to linguistic enquiry.”

linguagem conduz à formulação de uma hipótese para explicar esses fatos; esta, por sua vez, leva a uma generalização baseada na evidência de padrões repetidos utilizados na concordância; a última etapa é a unificação destas observações em uma formulação teórica.³³ (TOGNINI-BONELLI, 2002, p. 2).

Os dados reais autênticos mencionados por Tognini-Bonelli (2002) são os dados que compõem o corpus escolhido para cada pesquisa. Convém ressaltar aqui que, em nossa pesquisa, entendemos o corpus como sendo uma

[...] coleção de textos entendidos como 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 critérios, com um propósito específico em mente, e com o intuito de representar os ‘pedaços maiores do idioma’ selecionados de acordo com uma tipologia específica.³⁴ (TOGNINI-BONELLI, 2002, p.2)

A partir da definição de corpus dada por Tognini-Bonelli (2002), ressaltamos que, além de primar pelo uso de dados reais autênticos, é preciso levar em consideração os objetivos que se pretendem alcançar na pesquisa, no intuito de que o corpus seja condizente com estes fins. Desse modo, em nossa pesquisa, trabalhamos com dois diferentes tipos de corpora: um corpus especializado (TOGNINI-BONELLI, 2002), e um corpus comparável. Ao optarmos por trabalhar com um corpus constituído pelas legendas da série *CSI – Crime Scene Investigation*, temos, conforme concepção de Tognini-Bonelli (2002), um corpus especializado, haja vista que este corpus não contribui para a descrição da língua geral, mas se ocupa da língua especializada, de forma a permitir que olhemos mais atentamente para as idiossincrasias do uso autêntico dessa linguagem especializada (TOGNINI-BONELLI, 2002, p. 8), ou seja, este corpus nos permite buscar as colocações especializadas da área investigativa criminal, uma vez que é constituído por textos (legendas de uma série) que abordam este tema.

³³ “[...] 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.”

³⁴ “[...] 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.”

Contamos, ainda, com um corpus comparável, compilado com o auxílio da ferramenta *BootCat Front End*, versão 0.71 (ZANCHETTA; BARONI; BERNARDINI, 2013). De acordo com Tognini-Bonelli (2012)

Corpora comparáveis são corpora cujos componentes são escolhidos por serem amostras similares de suas respectivas línguas em termos de critérios externos como língua falada x língua escrita, registro, etc. No entanto, nenhum deles são traduções; nenhum alinhamento é possível, mas as correspondências podem ser estabelecidas entre as principais características linguísticas dos corpora.³⁵ (TOGNINI-BONELLI, 2012, p. 7).

Como pode se perceber pela definição apresentada pela autora, embora não seja possível traçar uma correspondência linha por linha entre os corpora que utilizamos em nossa pesquisa, o corpus comparável serve como fonte de comprovação dos dados obtidos no corpus constituído pelas legendas da série *CSI*, uma vez que é possível estabelecer correspondências linguísticas entre ambos os corpora, além do fato de que ambos têm como tema principal de seus textos a investigação criminal.

Ademais, a partir da definição de LC dada por Tognini-Bonelli (2002), é possível relacionar o uso da LC em pesquisas de diversas áreas, como Ensino e Aprendizagem de Línguas, Estudos da Tradução, 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.

No que diz respeito às pesquisas nas áreas de Ensino-Aprendizagem de línguas e de Lexicografia, por exemplo, a LC tem contribuído na compilação de dicionários especializados, como é o caso dos dicionários *Oxford Student's Dictionary of English*, *Cambridge Dictionary of American English*, *Collins Cobuild Learner's Dictionary*, *Longman Dictionary of Contemporary English*, entre outros. Não podemos deixar de mencionar, também, o desenvolvimento de materiais didáticos voltados para o ensino de língua inglesa, os quais se utilizam da LC em sua elaboração, como é o caso da coleção de livros *Touchstone* (MCCARTHY; MCCARTEN; SANDIFORD, 2005), da editora *Cambridge University Press*, que apresenta indicações e notas sobre o léxico baseadas em *corpora*, ensinando a língua inglesa a partir de amostras reais de uso.

³⁵ “Comparable corpora are corpora whose components are chosen to be similar samples of their respective languages in terms of external criteria such as spoken vs. written language, register, etc. However, none of them are translations; no alignment is possible, but correspondences can be established among the main linguistic features of the corpora.”

Além das áreas mencionadas acima, a Fraseologia também tem feito uso de ferramentas computacionais criadas para pesquisas baseadas na LC, tendo-se valido dos benefícios que estas trazem para o desenvolvimento de suas pesquisas, principalmente no que diz respeito ao estudo das colocações (FRITZINGER, HEID, 2009; LAUFER; WALDMAN, 2011; ORENHA-OTTAIANO, 2009, 2015), haja vista a facilidade de encontrá-las e observá-las por meio dos programas computacionais que permitem ao pesquisador analisar as colocações no contexto em que ocorrem como, por exemplo, no caso da ferramenta *Concord*, do programa *WordSmith Tools* (SCOTT, 2012) ou do programa *AntConc* (ANTHONY, 2012).

2.5 Abordagem Lexical

Ao propormos trabalhar com o ensino-aprendizagem das colocações, parece-nos bastante apropriado tomarmos os postulados teórico-metodológicos da Abordagem Lexical como base para nossa pesquisa, 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 Lewis (1993), ao promover essa dicotomia, tende-se a simplificar um dos mais complexos e importantes aspectos da língua, visto que, ao analisar a língua que falamos percebe-se 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; entre vocabulário e gramática; entre habilidades receptivas e produtivas; entre competência e desempenho etc. 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)³⁶

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 pensar de maneira binária, se pense 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 uso cotidiano, essa diferenciação se torna mais complicada, uma vez 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 se deixe de pensar em tal dicotomia.

Assim, Lewis passa a abordar a questão das colocações ao fazer uma analogia com um espectro, no qual

“de um lado, estão as palavras semanticamente fortes, as quais possuem um alcance colocacional limitado, são (relativamente) raras, e onde seu valor quase sempre coincide com sua significação [...]. No final do espectro estão as palavras tradicionalmente reconhecidas como gramaticais, ao invés de itens de vocabulário [...]”³⁷ (LEWIS, 1993, p. 37).

Dessa forma, o autor demonstra que a Abordagem Lexical se ocupará, além do ensino de palavras com diferentes cargas semânticas, das colocações, visto que estas são combinações lexicais que contém 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), com o propósito de ajudar os professores a incorporar o ensino das colocações em

36 “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 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.”

37 “At one end, are semantically strong words which have limited collocational range, are (relatively) rare, and where their value almost always coincides with their signification [...]. At the other end of the spectrum are words traditionally recognized as grammatical, rather than vocabulary items [...]”

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). Além de Lewis (2000), Orenha-Ottaiano (2012, 2015) aborda questões concernentes ao ensino das colocações.

Hill (2000) defende que é necessário dedicar mais tempo ao ensino de colocações desde o nível mais básico do ensino de língua inglesa, considerando que, ao focar o ensino de colocações, o professor permitirá que um aluno, com um conhecimento restrito a 2.000 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 de um 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, o autor acredita que o uso das colocações amplia a competência comunicativa dos alunos (HILL, 2000, p. 61). Para atingir esse objetivo, 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 tanto, apresentam um modelo de atividade intitulado *5-1 box*, no qual os alunos memorizam os colocados que ocorrem com cada nóculo em destaque.

Por sua vez, Orenha-Ottaiano (2012) reforça a importância do ensino explícito das colocações durante as aulas de língua inglesa, uma vez que os alunos não têm o hábito de se aterem as colocações que encontram nos textos com os quais se depararam, bem como discorre sobre a relevância do uso de corpora no ensino das colocações (ORENHA-OTTAIANO, 2012, p. 245). Em seu estudo, a autora utiliza um teste colocacional para averiguar o conhecimento de alunos dos cursos de Licenciatura em Letras e de Tradução em

uma universidade pública, de forma a apontar se as colocações devem ser ensinadas de forma implícita ou explícita. Como resultado dessa pesquisa, Orenha-Ottaiano (2012) conclui que estas devem ser ensinadas de forma explícita, como dissemos, e enfatiza que

após tomarem conhecimento dos resultados dos testes, alguns alunos se tornaram mais responsáveis pela sua própria aprendizagem e começaram a refletir sobre os possíveis colocados das palavras estudadas em sala de aula e se tornaram mais cuidadosos com a combinabilidade das palavras ao produzir seus textos escritos ou orais. Além disso, todos eles reconheceram e enfatizaram a importância da fraseologia para a aprendizagem de uma língua estrangeira, pois perceberam que a pesquisa provou que eles não têm problemas em relação a compreensão das unidades multi-palavras, mas eles têm dificuldades em produzi-las quando falam ou escrevem.³⁸ (ORENHA-OTTAINO, 2012, p. 246).

Neste sentido, espera-se que a elaboração das atividades colocacionais que fazem parte da categoria *Investigation* no *Online English Collocations Workbook*, proposto por Orenha-Ottaiano (2015), segundo relato no capítulo 5 desta investigação, contribua para o ensino-aprendizagem das colocações especializadas de forma explícita, indo de encontro ao proposto por Lewis (2000), Hill (2000) e Orenha-Ottaiano (2012).

No capítulo a seguir discorre-se sobre a metodologia empregada para o desenvolvimento deste trabalho.

³⁸ “After they were shown their test results, some students have become more responsible for their learning and started reflecting on the possible collocates for the words studied in the classroom and become more careful about the combinability of words when producing their written or oral texts. Moreover, all of them recognized and stressed the importance of phraseology to the learning of a foreign language, as they realized that the research proved they do not have problems regarding understanding the multi-word units, but they do have difficulties in producing them when speaking or writing.”

3 METODOLOGIA DA PESQUISA

Tendo em vista que o objetivo principal de nosso trabalho é a extração e análise sintático-morfológica das colocações especializadas da área investigativa criminal a partir dos dados obtidos por meio do CCSI e do CCC, este capítulo apresenta os procedimentos e as ferramentas que permitiram alcançar este objetivo. Dessa forma, o capítulo inicia-se a partir da compilação e do tratamento de ambos os corpora de estudo. Em seguida, apresentam-se as etapas que possibilitaram a análise sintático-morfológica das colocações obtidas por meio da busca empreendida nos corpora de pesquisa.

Ademais, são apresentados alguns aspectos relacionados à criação e ao design da plataforma do *Online English Collocations Workbook* pela Profa. Dra. Adriane Orenha-Ottaiano, haja vista que esta pesquisa possibilitou a inserção de mais uma área de especialidade na referida plataforma, a saber, a de *Investigation*. Logo após, apresentamos os procedimentos utilizados para a inserção das colocações nas atividades que compõem o referido *workbook online*.

3.1 Compilação e tratamento do *Corpus CSI*

Ao propor esta pesquisa de forma que tivesse como resultado final atividades didáticas que contribuíssem para o desenvolvimento dos conhecimentos fraseológicos dos alunos, mais especificamente, para o conhecimento colocacional destes, nos preocupamos em buscar uma área que, até então, não tivesse sido contemplada pelas atividades do *Online English Collocations Workbook* ao qual nossa pesquisa está atrelada e que é apresentado na seção 3.4.

Com isso, optamos por privilegiar a área de investigação criminal (*Investigation*), a partir da série norte-americana *CSI: Crime Scene Investigation*. A série, que teve sua estréia em outubro de 2000 e chegou ao fim em maio de 2015, foi escolhida para compor nosso corpus de pesquisa por ser bastante representativa da linguagem investigativa, fato este que pôde ser comprovado por meio das análises e do levantamento das colocações especializadas. Além disso, trata-se de uma série que conseguiu atingir um público-alvo amplo e variado, colaborando para despertar o interesse dos alunos pelo estudo das colocações especializadas dessa área, conforme mencionado no capítulo 1.

Assim, após a determinação do corpus de pesquisa, o qual recebeu o nome de Corpus CSI (CCSI), passamos à sua compilação. Para tal, realizamos uma busca usando a *Internet*, e conseguimos encontrar os arquivos contendo as legendas em língua inglesa das 12 primeiras

temporadas do seriado no *site* <http://www.tvsubtitles.net/>, totalizando 276 episódios transcritos e 3.061.029 palavras. Este *site* disponibiliza as legendas originais de diversas séries norte-americanas no formato *closed caption*, ou seja, todo o conteúdo de áudio da série é transcrito para as legendas. Além das legendas originais, o *site* disponibiliza as legendas em outros idiomas, tais como português, espanhol e francês, por exemplo.

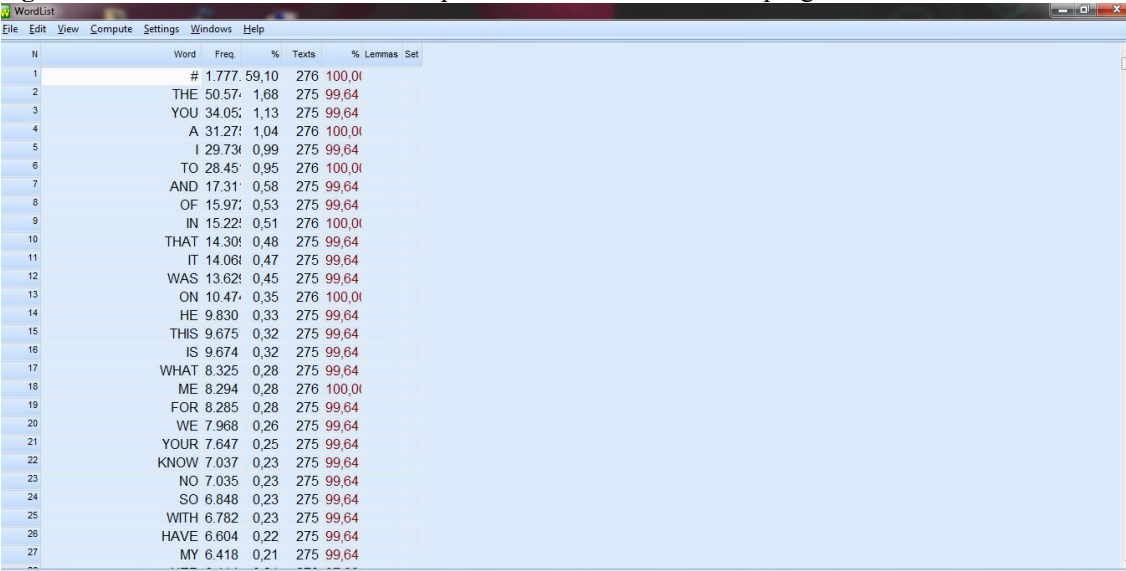
Os arquivos das legendas foram baixados no formato compactado *.rar*, sendo um arquivo por temporada, totalizando 12 (doze) arquivos neste formato. Estes arquivos contêm as legendas de todos os episódios de cada temporada, salvas individualmente no formato *.txt*, formato necessário para a utilização do programa *WordSmith Tools* (SCOTT 2012).

Embora os arquivos de cada episódio já estejam no formato necessário para análise a partir do programa *WordSmith Tools* (SCOTT 2012), foi necessário extrair estes arquivos do formato compactado *.rar* e salvá-los individualmente. Visando facilitar o trabalho, os arquivos de todos os episódios foram salvos em uma mesma pasta, identificada como CSI.txt.

Conforme mencionado, o CCSI é constituído pelas legendas originais da série, em formato *closed caption*, que apresenta todo o conteúdo de áudio de cada episódio. Ressaltamos aqui que “embora haja muita controvérsia sobre a disponibilização, por parte de legendadores alternativos (SYURI, 2011), de legendas gratuitas na *Internet*, não levamos essas discussões em consideração” (FROMM, 2011, p. 2), tendo em vista que o foco do nosso trabalho está no material linguístico fornecido por estas legendas.

Após a compilação do CCSI, passamos ao seu tratamento e análise a partir das ferramentas do programa *WordSmith Tools*, versão 6.0 (SCOTT, 2012), a saber *WordList*, *KeyWords* e *Concord*. A primeira ferramenta utilizada foi a *WordList*, ferramenta responsável pela criação de uma lista de palavras de todo o corpus, a qual contém informações sobre a quantidade de *types* e *tokens* presentes no corpus, além de fornecer o número total de palavras de corpus de pesquisa, o qual contém o total de 3.061.029 palavras. A seguir, é apresentada a imagem de uma parte desta lista de palavras, organizadas pela ordem de frequência:

Figura 2 – Tela com o resultado obtido pela ferramenta *WordList* do programa *WordSmith Tools*.



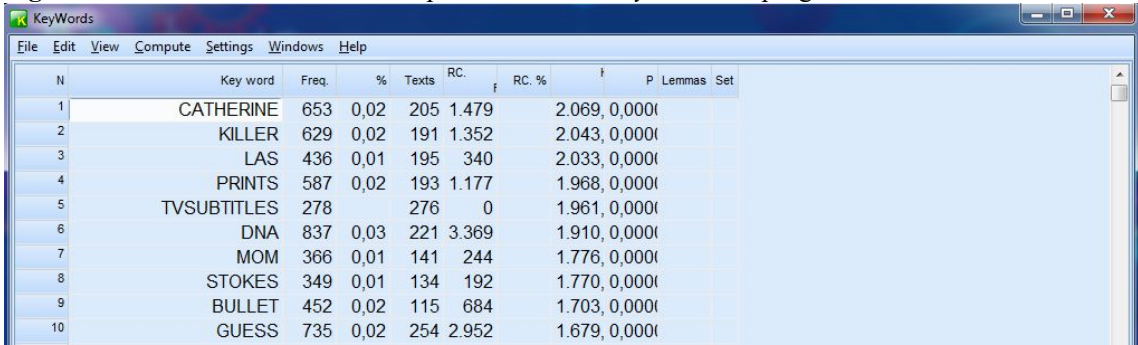
The screenshot shows the WordList window of WordSmith Tools. The window has a menu bar with 'File', 'Edit', 'View', 'Compute', 'Settings', 'Windows', and 'Help'. Below the menu bar is a table with the following columns: 'N', 'Word', 'Freq.', '%', 'Texts', '% Lemmas', and 'Set'. The table contains 27 rows of data, starting with a '#' symbol and followed by common English words. At the bottom of the window, there is a status bar showing '14,091 entries Row 1' and a tab labeled 'frequency'.

N	Word	Freq.	%	Texts	% Lemmas	Set
1	#	1.777	59,10	276	100,00	
2	THE	50.57	1,68	275	99,64	
3	YOU	34.05	1,13	275	99,64	
4	A	31.27	1,04	276	100,00	
5	I	29.73	0,99	275	99,64	
6	TO	28.45	0,95	276	100,00	
7	AND	17.31	0,58	275	99,64	
8	OF	15.97	0,53	275	99,64	
9	IN	15.22	0,51	276	100,00	
10	THAT	14.30	0,48	275	99,64	
11	IT	14.06	0,47	275	99,64	
12	WAS	13.62	0,45	275	99,64	
13	ON	10.47	0,35	276	100,00	
14	HE	9.830	0,33	275	99,64	
15	THIS	9.675	0,32	275	99,64	
16	IS	9.674	0,32	275	99,64	
17	WHAT	8.325	0,28	275	99,64	
18	ME	8.294	0,28	276	100,00	
19	FOR	8.285	0,28	275	99,64	
20	WE	7.968	0,26	275	99,64	
21	YOUR	7.647	0,25	275	99,64	
22	KNOW	7.037	0,23	275	99,64	
23	NO	7.035	0,23	275	99,64	
24	SO	6.848	0,23	275	99,64	
25	WITH	6.782	0,23	275	99,64	
26	HAVE	6.604	0,22	275	99,64	
27	MY	6.418	0,21	275	99,64	

Fonte: PrintScreen da tela do programa *WordSmith Tools*

De posse da lista de palavras, o passo seguinte foi obter a lista de palavras-chave. Essa lista é gerada por meio da ferramenta *KeyWords*, também do programa *WordSmith Tools* (SCOTT, 2012). Para gerar esta lista, é necessário utilizar a lista de palavras obtida anteriormente que deve ser contrastada com uma lista de palavras de um corpus de referência. Para esta pesquisa foi utilizada a lista de palavras do *British National Corpus (BNC)*.

Por meio da lista de palavras-chave é possível selecionar quais as palavras que serão utilizadas como núdulos de busca para as colocações especializadas. Desse modo, as palavras gramaticais são desconsideradas e são levadas em consideração apenas as palavras de conteúdo, o que permite a seleção de palavras que estão estritamente relacionadas à área da investigação criminal. A seguir, apresentamos uma imagem da tela da ferramenta *KeyWords*, na qual é possível visualizar algumas palavras de conteúdo que foram selecionadas como núdulos de busca, tais como: *killer*, *prints*, *DNA* e *bullet*:

Figura 3 – Tela com resultado obtido pela ferramenta *KeyWords* do programa *WordSmith Tools*


N	Key word	Freq.	%	Texts	RC.	f	RC. %	P	Lemmas	Set
1	CATHERINE	653	0,02	205	1.479		2.069, 0,0000			
2	KILLER	629	0,02	191	1.352		2.043, 0,0000			
3	LAS	436	0,01	195	340		2.033, 0,0000			
4	PRINTS	587	0,02	193	1.177		1.968, 0,0000			
5	TVSUBTITLES	278		276	0		1.961, 0,0000			
6	DNA	837	0,03	221	3.369		1.910, 0,0000			
7	MOM	366	0,01	141	244		1.776, 0,0000			
8	STOKES	349	0,01	134	192		1.770, 0,0000			
9	BULLET	452	0,02	115	684		1.703, 0,0000			
10	GUESS	735	0,02	254	2.952		1.679, 0,0000			

Fonte: PrintScreen da tela do programa *WordSmith Tools*

Após a obtenção da lista de palavras-chave, elaboramos uma tabela contendo as palavras-chave que foram selecionadas como núdulos de busca para as colocações especializadas, de forma a facilitar sua visualização durante a pesquisa. Abaixo, apresentamos a referida tabela, utilizada nesta pesquisa.

Tabela 1 – *KeyWords* selecionadas como núdulos de busca

Key Word	Freq.	KeyWord	Freq.	KeyWord	Freq.
<i>Abrasion</i>	50	<i>Fiber</i>	156	<i>Residue</i>	71
<i>Accelerant</i>	30	<i>Fingerprints</i>	170	<i>Robbery</i>	110
<i>Alibi</i>	84	<i>Fire</i>	206	<i>Scene</i>	849
<i>Arrest</i>	188	<i>Forensics</i>	55	<i>Serial</i>	125
<i>Attorney</i>	98	<i>Fracture</i>	57	<i>Sheriff</i>	147
<i>Autopsy</i>	129	<i>Gun</i>	903	<i>Shoot</i>	221
<i>Ballistics</i>	67	<i>Gunshot</i>	205	<i>Shooter</i>	100
<i>Blood</i>	1.830	<i>Hemorrhage</i>	18	<i>Shot</i>	789
<i>Body</i>	1.284	<i>Hemorrhaging</i>	31	<i>Skull</i>	120
<i>Bullet</i>	452	<i>Homicide</i>	194	<i>Spatter</i>	111
<i>Cell</i>	330	<i>Jail</i>	178	<i>Suicide</i>	179
<i>Contusion</i>	66	<i>Kill</i>	1.983	<i>Surveillance</i>	172
<i>Cop</i>	296	<i>Killer</i>	629	<i>Suspect</i>	610
<i>Coroner</i>	106	<i>Knife</i>	244	<i>Swear</i>	108
<i>Counselor</i>	24	<i>Lab</i>	689	<i>Tissue</i>	193
<i>Crime</i>	1.036	<i>Laceration</i>	67	<i>Trace</i>	325
<i>Dead</i>	1.391	<i>Lawyer</i>	193	<i>Trauma</i>	157
<i>Defense</i>	108	<i>License</i>	106	<i>Victim</i>	709
<i>Detective</i>	221	<i>Ligatures</i>	44	<i>Warrant</i>	332
<i>DNA</i>	837	<i>Murder</i>	894	<i>Weapon</i>	348
<i>Evidence</i>	1.112	<i>Perimortem</i>	23	<i>Wound</i>	315
<i>Eyewitness</i>	48	<i>Postmortem</i>	54	-	-
<i>Felony</i>	57	<i>Prints</i>	587	-	-

Estas palavras-chave serviram de base para a busca e o levantamento das colocações especializadas tanto no CCSI, como no CCC, o qual é apresentado e explicado na seção seguinte.

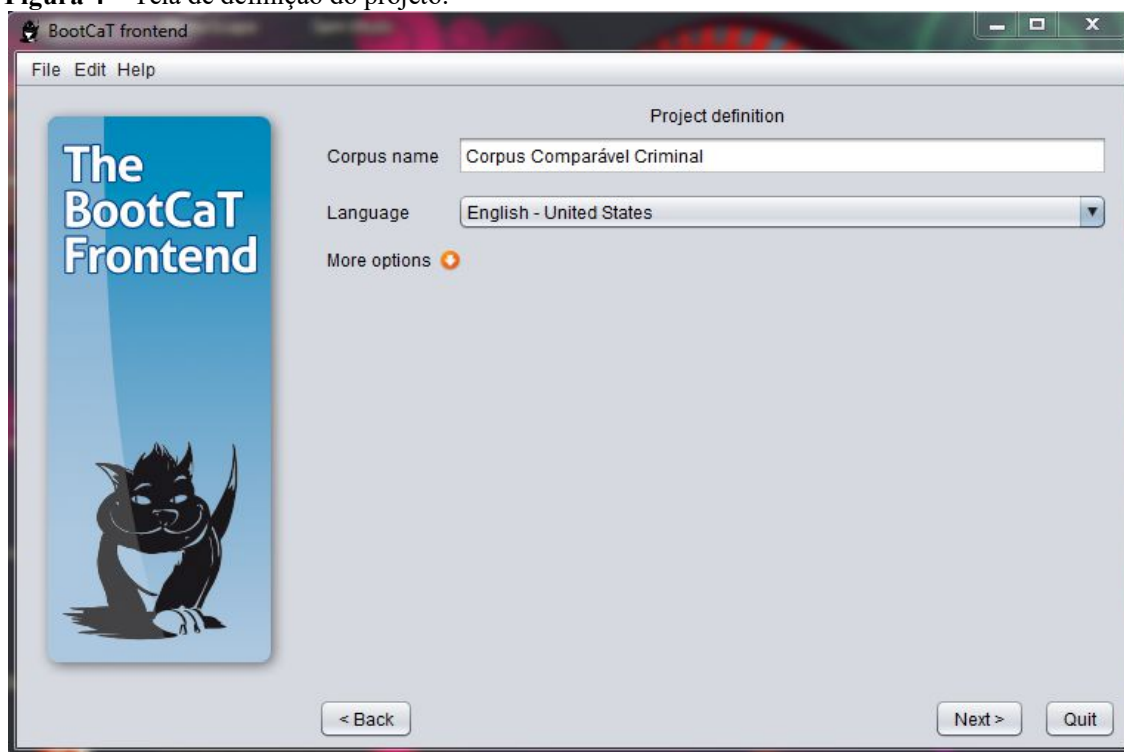
3.2 Compilação e tratamento do *corpus* comparável

Conforme mencionado, a pesquisa conta com o CCSI constituído pelas legendas de 12 temporadas da série *CSI – Crime Scene Investigation*, o qual serve de ponto de partida para a busca pelas colocações especializadas da área criminal. No entanto, visando à comprovação de uso e frequência de coocorrência destas colocações, compilamos, também, um corpus comparável, de forma a termos uma segunda fonte, além do emulador de concordanciador *WebCorp*, que comprove a frequência de uso das colocações especializadas da área investigativa criminal na língua inglesa, levantadas a partir do CCSI, além de ser possível levantar outras colocações especializadas igualmente relevantes as quais não constam no CCSI e que poderão ser utilizadas para pesquisas futuras.

Para a compilação deste corpus comparável, foi utilizada a ferramenta *Boot Cat Front End*, versão 0.71 (ZANCHETTA; BARONI; BERNARDINI, 2013). Esta ferramenta permite que o pesquisador compile um corpus via *web* a partir de sementes (*seeds*), ou palavras-chave, da língua geral ou de uma área de especialidade, em diversos *sites*.

A própria ferramenta indica os passos necessários para a compilação de um corpus. Dessa maneira, o primeiro passo é nomear o projeto e escolher a língua com a qual se deseja trabalhar. Conforme pode ser visto na imagem logo a seguir, optamos nomear este projeto de Corpus Comparável Criminal (CCC) e selecionamos a língua inglesa falada nos Estados Unidos, uma vez que esta é a língua utilizada na série *CSI – Crime Scene Investigation*.

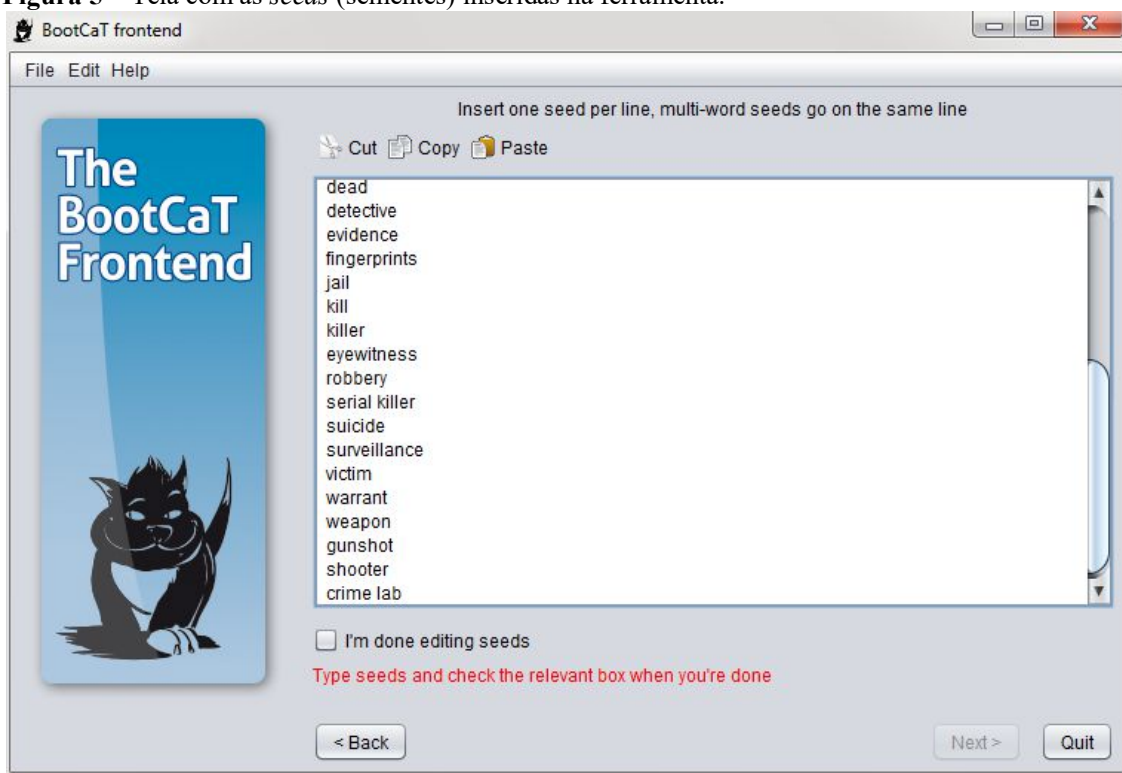
Figura 4 – Tela de definição do projeto.



Fonte: *PrintScreen* da tela do programa *BootCat*.

Em seguida, são inseridas as sementes (*seeds*), ou palavras-chave, as quais norteiam a varredura feita pela ferramenta na *web*. Para tal, foram utilizadas as palavras-chaves do CCSI como sementes, uma vez que o objetivo em relação à compilação deste corpus comparável é a comprovação das ocorrências das colocações levantadas neste corpus. A seguir, apresentamos a tela na qual as sementes são inseridas. Convém ressaltar que não há um limite máximo de sementes que possam ser inseridas na ferramenta. No entanto, quanto mais sementes inseridas, mais refinada será a busca empreendida.

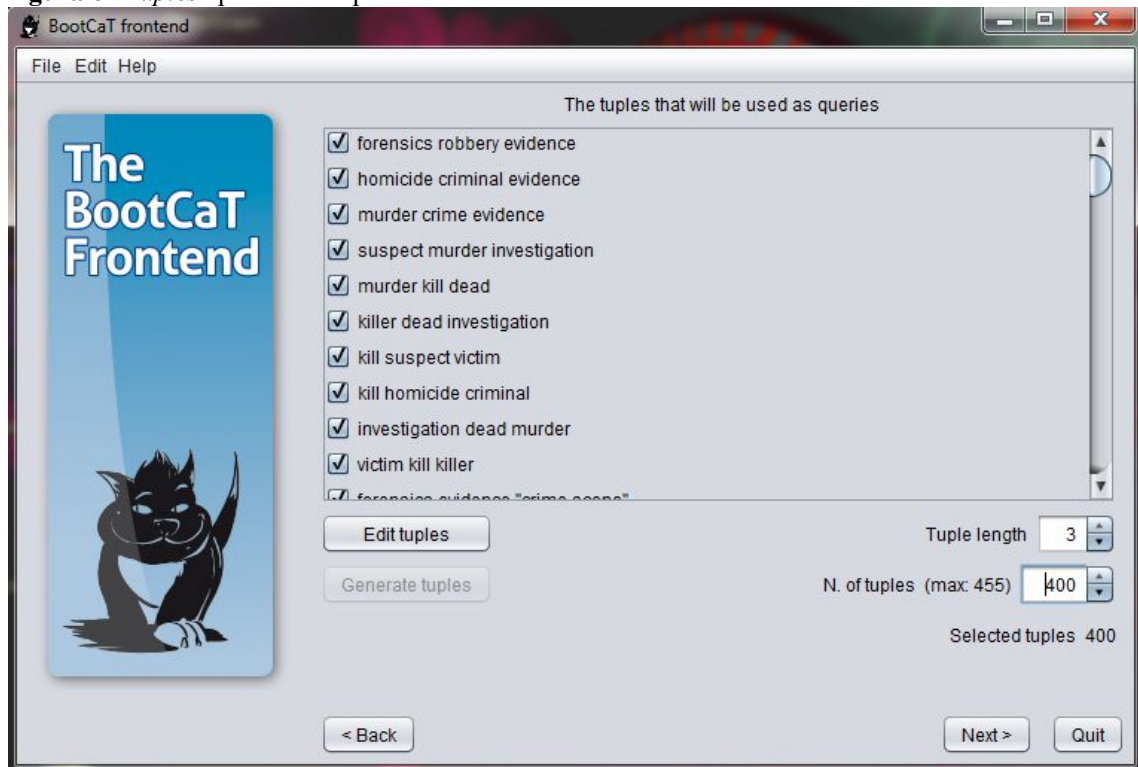
Figura 5 – Tela com as *seeds* (sementes) inseridas na ferramenta.



Fonte: *PrintScreen* da tela do programa *BootCat*.

Inseridas as sementes, são geradas as *tuples*, ou seja, combinações de, no mínimo, três palavras com as quais a ferramenta empreende a busca. Nesta etapa, o usuário pode escolher algumas configurações, tais como o tamanho e a quantidade de *tuples* que serão geradas. Para esta pesquisa, mantivemos as configurações já estabelecidas pelo programa, por julgarmos que estas atendiam às necessidades da pesquisa. Estas configurações podem ser vistas na imagem a seguir.

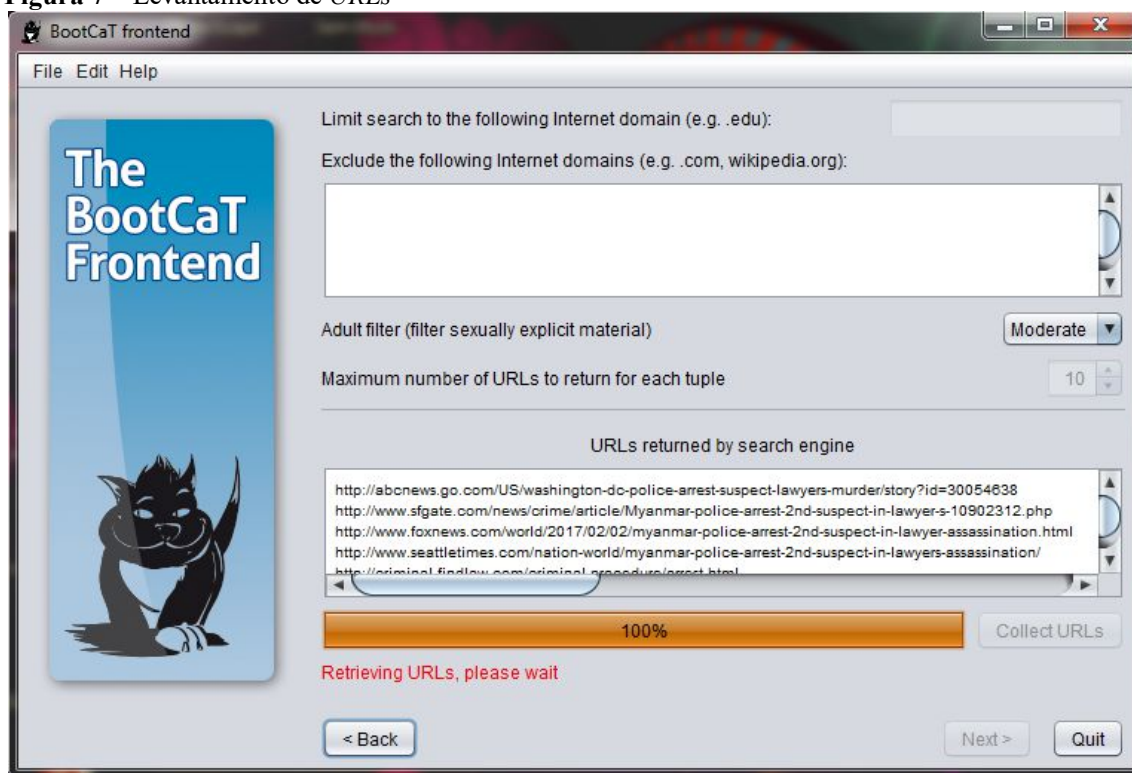
Figura 6– *Tuples* apresentadas pela ferramenta



Fonte: *PrintScreen* da tela do programa *BootCat*.

Como se pode perceber na imagem anterior, o usuário tem, também, a opção de editar as *tuples*. Concluída este passo, ao clicar em *Next*, a ferramenta começa a fazer o levantamento de *URLs* que atendam às especificações dadas a partir das sementes e das *tuples* geradas. Nesta etapa, o usuário pode verificar as *URLs* levantadas e filtrá-las, caso identifique alguma que não seja condizente com o que deseja para seu corpus, além de poder determinar a quantidade de *URLs* que será levantada para cada *tuple* apresentada.

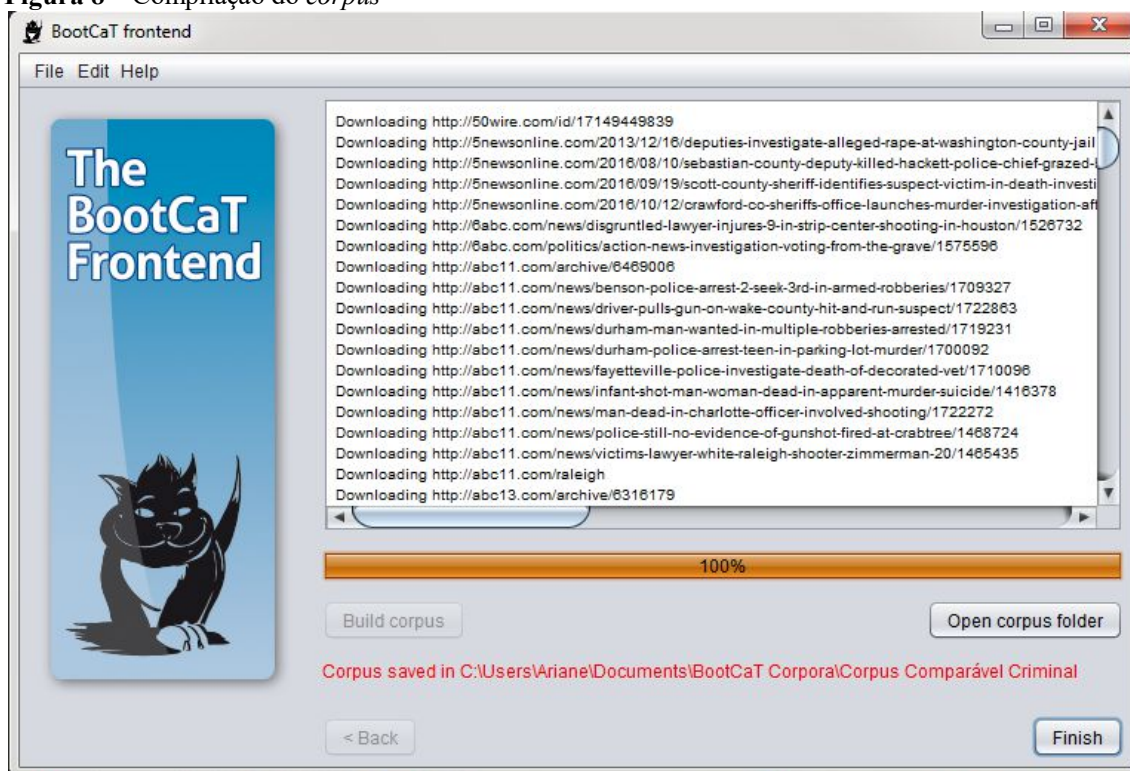
Figura 7 – Levantamento de *URLs*



Fonte: *PrintScreen* da tela do programa *BootCat*.

Ao final do levantamento das *URLs*, clica-se em *Next*, e a ferramenta começa a fazer a compilação do *corpus*. Nesta etapa, a ferramenta faz o *download* das páginas que foram listadas anteriormente.

Figura 8 – Compilação do *corpus*



Fonte: PrintScreen da tela do programa *BootCat*.

Chamamos a atenção para o fato de que a ferramenta traz o local onde o corpus foi salvo em vermelho, na própria janela, uma vez que este é salvo automaticamente no formato *.txt*, no local determinado pelo programa, o que facilita o trabalho do usuário/pesquisador.

Após finalizar a compilação deste corpus, passamos ao seu tratamento utilizando a ferramenta *WordSmith Tools*, versão 6.0 (SCOTT, 2012), da mesma forma como foi tratado o CCSI (descrita no item 3.1). A partir destes procedimentos, foi possível constatar que o CCC conta com um total de 3.886.465 palavras, sendo, assim, maior que o CCSI (3.061.029 palavras).

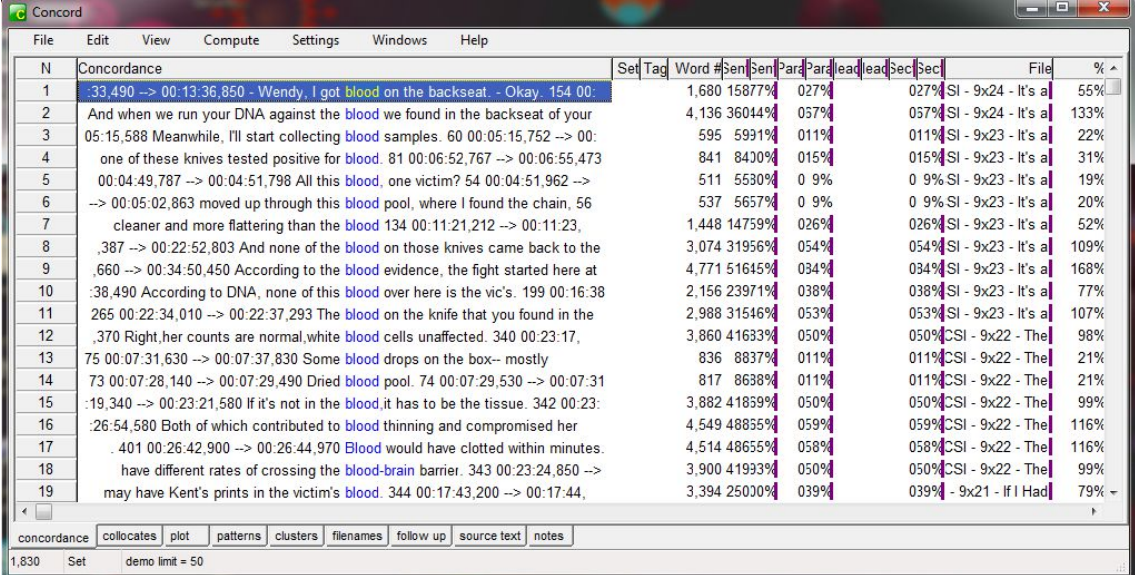
Tendo finalizado a compilação e o tratamento de ambos os corpora, passamos à extração e análise das colocações especializadas. Estes procedimentos são apresentados na seção seguinte.

3.3 Procedimento de extração e forma de análise das colocações especializadas

Para proceder ao levantamento dos candidatos a colocações especializadas, e sua posterior análise e classificação, utilizamos a ferramenta *Concord*, também do programa *WordSmith Tools*, versão 6.0 (SCOTT, 2012).

Esta ferramenta permite analisar, por meio de linhas de concordância, todas as ocorrências de cada palavra-chave em ambos os corpora de estudo. Como exemplo, tomamos a palavra-chave *blood*, a qual recebe a denominação de nódulo de busca, que ocorre 1.830 vezes no CCSI (corpus do qual partimos para o levantamento das colocações especializadas). Ao inserir esta palavra-chave como o termo a ser pesquisado, o programa retorna todas as ocorrências desta palavra, a fim de que seja possível analisar cada linha das 1.830 em que *blood* aparece, buscando pelos candidatos a colocações especializadas. A seguir, apresentamos uma imagem da ferramenta *Concord* com o nódulo de busca usado como exemplo (*blood*).

Figura 9– Tela da ferramenta *Concord* do programa *WordSmith Tools*



N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	lead	lead	Sec	Sec	File	%
1	33,490 -> 00:13:36,850 - Wendy, I got blood on the backseat. - Okay. 154 00:	1,680	15877%	027%	027%	SI - 9x24 - It's a	55%							
2	And when we run your DNA against the blood we found in the backseat of your	4,136	36044%	057%	057%	SI - 9x24 - It's a	133%							
3	05:15,588 Meanwhile, I'll start collecting blood samples. 60 00:05:15,752 -> 00:	595	5931%	011%	011%	SI - 9x23 - It's a	22%							
4	one of these knives tested positive for blood. 81 00:06:52,767 -> 00:06:55,473	841	8430%	015%	015%	SI - 9x23 - It's a	31%							
5	00:04:49,787 -> 00:04:51,798 All this blood, one victim? 54 00:04:51,962 ->	511	5530%	0 9%	0 9%	SI - 9x23 - It's a	19%							
6	-> 00:05:02,863 moved up through this blood pool, where I found the chain, 56	537	5657%	0 9%	0 9%	SI - 9x23 - It's a	20%							
7	cleaner and more flattering than the blood 134 00:11:21,212 -> 00:11:23,	1,448	14759%	026%	026%	SI - 9x23 - It's a	52%							
8	,387 -> 00:22:52,803 And none of the blood on those knives came back to the	3,074	31956%	054%	054%	SI - 9x23 - It's a	109%							
9	,660 -> 00:34:50,450 According to the blood evidence, the fight started here at	4,771	51645%	034%	034%	SI - 9x23 - It's a	168%							
10	:38,490 According to DNA, none of this blood over here is the vic's. 199 00:16:38	2,156	23971%	038%	038%	SI - 9x23 - It's a	77%							
11	265 00:22:34,010 -> 00:22:37,293 The blood on the knife that you found in the	2,988	31546%	053%	053%	SI - 9x23 - It's a	107%							
12	,370 Right, her counts are normal, white blood cells unaffected. 340 00:23:17,	3,860	41633%	050%	050%	SI - 9x22 - The	98%							
13	75 00:07:31,630 -> 00:07:37,830 Some blood drops on the box-- mostly	836	8837%	011%	011%	SI - 9x22 - The	21%							
14	73 00:07:28,140 -> 00:07:29,490 Dried blood pool. 74 00:07:29,530 -> 00:07:31	817	8638%	011%	011%	SI - 9x22 - The	21%							
15	:19,340 -> 00:23:21,580 If it's not in the blood, it has to be the tissue. 342 00:23:	3,882	41839%	050%	050%	SI - 9x22 - The	99%							
16	:26:54,580 Both of which contributed to blood thinning and compromised her	4,549	48855%	059%	059%	SI - 9x22 - The	116%							
17	,401 00:26:42,900 -> 00:26:44,970 Blood would have clotted within minutes.	4,514	48655%	058%	058%	SI - 9x22 - The	116%							
18	have different rates of crossing the blood-brain barrier. 343 00:23:24,850 ->	3,900	41993%	050%	050%	SI - 9x22 - The	99%							
19	may have Kent's prints in the victim's blood. 344 00:17:43,200 -> 00:17:44,	3,394	25030%	039%	039%	SI - 9x21 - If I Had	79%							

Fonte: PrintScreen da tela do programa *WorSmith Tools*

Conforme pode ser visto na imagem, a palavra-chave selecionada, então usada como nódulo de busca, aparece centralizada e destacada, o que facilita a análise dos elementos ao seu redor. Esta ferramenta permite uma observação detalhada de todas as palavras que aparecem ao redor da palavra-chave, o que possibilita a identificação de candidatos à colocação especializada.

Em um primeiro momento, todas as combinações de palavras extraídas da análise de cada uma das palavras-chave são consideradas candidatas à colocação, haja vista que é preciso proceder à análise de cada uma dessas combinações, para que seja possível classificá-las como uma colocação.

Assim sendo, faz-se necessário seguir alguns critérios para considerar que uma combinação seja efetivamente uma colocação. Por exemplo, observar se cada uma das combinações apresenta frequência superior a 1 (um), além de provar ser uma combinação lexical com um significado composicional, que apresente uma forte ligação entre seus elementos e que ocorra em um contexto específico (TAGNIN, 1999).

Convém ressaltar que, no tocante ao critério da frequência ser superior a 1, o leitor perceberá ao longo do texto, que algumas colocações especializadas ocorrem apenas 1 vez no CCSI ou no CCC. No entanto, estas colocações foram mantidas e apresentadas na pesquisa, devido ao fato de terem tido sua frequência de coocorrência comprovada a partir da busca feita por meio do emulador de concordanciador *WebCorp*. Dessa forma, foi possível considerar um número maior de colocações nas análises, pois não foi necessário descartar aquelas que apresentaram baixa frequência em um ou outro corpus de estudo, uma vez que sua importância no contexto em que está inserida pode ser atestada por meio da frequência obtida pelo *WebCorp*.

Após a análise de cada um dos candidatos, de acordo com os critérios apresentados anteriormente, para que pudéssemos considerar as combinações como colocações especializadas de fato, buscamos analisar a estrutura morfossintática de cada colocação, para que fosse possível fazer sua classificação taxonômica, segundo propostas de Hausmann (1985), Tagnin (1999) e Orenha-Ottaiano (2004, 2009).

A seguir, é apresentada, a título de ilustração, uma tabela contendo as informações sobre a frequência de cada colocação no CCSI, no CCC e no *WebCorp*, bem como a análise de seus elementos, segundo sua estrutura morfossintática e classificação taxonômica. Para tal, foi mantido o exemplo dado anteriormente, do nódulo de busca *blood*, e são apresentadas algumas das colocações obtidas, sendo que, a palavra-chave ocorre ora como base, ora como colocado das colocações presentes na tabela.

Tabela 2 – Exemplo das colocações obtidas a partir da palavra-chave *blood*

Nódulo Blood					
Colocação	Freq. CSI	Freq. CCC	Freq. WEBCORP	Formação	Taxonomia
<i>Blood sample</i>	9	18	288	subst. colocado + subst. base	Nominal
<i>Blood evidence</i>	29	34	203	subst. colocado + subst. base	Nominal
<i>Blood pressure</i>	14	32	2532	subst. colocado + subst. base	Nominal
<i>Blood stain</i>	10	22	399	subst. colocado + subst. base	Nominal
<i>Blood pool</i>	49	1	260	subst. colocado + subst. base	Nominal
<i>Blood trail</i>	23	7	277	subst. colocado + subst. base	Nominal
<i>Blood spatter</i>	46	46	253	subst. colocado + subst. base	Nominal
<i>Blood void</i>	1	-	76	subst. colocado + subst. base	Nominal
<i>Blood typing</i>	2	10	115	subst. colocado + subst. base	Nominal
<i>Clotted blood</i>	5	2	577	adj. colocado + subst. base	Adjetiva
<i>Blood smears</i>	4	2	235	subst. base + subst. colocado	Nominal
<i>Blood flow</i>	5	23	587	subst. colocado + subst. base	Verbal
<i>Blood vessels</i>	9	29	470	subst. colocado + subst. base	Nominal
<i>Run the blood</i>	5	-	44	verbo base + det.+subst. colocado	Verbal
<i>Draw blood</i>	1	-	251	verbo base + subst. colocado	Verbal

Como pode ser visto na tabela acima, as colocações especializadas *blood void*, *run the blood* e *draw blood* não foram encontradas no CCC, contudo, foi possível encontrar estas colocações por meio da busca realizada pelo emulador de concordanciador *WebCorp* com frequências superiores às encontradas no próprio CCSI, de forma que estas foram mantidas em nossas análises.

Tal como exemplificado nesta tabela, procedeu-se a análise de todas as colocações especializadas obtidas a partir do levantamento feito em ambos os corpora de estudo.

A seguir, é apresentado o *Online English Collocation Workbook*, plataforma na qual forma incluídas as colocações obtidas por meio da análise feita nesta pesquisa, a fim de desenvolver atividades colocacionais da área investigativa criminal.

3.4 O *Online English Collocations Workbook*

O *Online English Collocations Workbook* foi compilado pela Profa. Dra. Adriane Orenha-Ottaiano, deste Instituto, visando atender as necessidades dos aprendizes brasileiros, no que diz respeito ao ensino e à aprendizagem das colocações em língua inglesa. A criação do software foi desenvolvida pelo Grupo de Banco de Dados (GBD), sob coordenação do Prof. Dr. Carlos Roberto Valêncio, do Departamento de Computação. Até o presente momento, apenas duas fases do projeto foram desenvolvidas, compreendendo as atividades de

Memory Game (jogo da memória) e de *Gap Fill* (completar espaços). Futuramente, outras atividades serão inseridas.

Segundo Orenha-Ottaiano (2015), a compilação deste material de apoio pedagógico tem o intuito de colaborar para que os alunos consigam incorporar as colocações que compõem o *Workbook* ao seu léxico mental, de forma interativa e divertida, sendo sua elaboração justificada pela escassez de materiais que contemplem o ensino de colocações (ORENHA-OTTAIANO, 2015, p. 863).

Ainda de acordo com Orenha-Ottaiano (2015), a idealização do *Online English Collocations Workbook* se deu a partir das dificuldades que os aprendizes mostraram em relação à produção das colocações, as quais foram observadas pela própria autora ao compilar os *corpora* de aprendizes CALE (Corpus de Aprendizes de Língua Escrita) e CAT (Corpus de Aprendizes de Tradução). Desde sua criação, o *Online English Collocations Workbook* contava com cinco categorias de colocações já inseridas em ambas as atividades que fazem parte do *Workbook*: *academic language*, *business*, *general*, *medicine* e *politics*. A partir de 2015, com base nos resultados dos dados desta investigação, a categoria *Investigation* já foi inserida e as atividades estão sendo realizadas pelos alunos.

Para acessar o *Workbook* é preciso ter um *login* com usuário e senha, além de ter o acesso liberado pelo administrador. A seguir, é apresentada a tela inicial do *Workbook*, bem como a tela na qual os novos usuários fazem seu cadastro.

Figura 10 – Tela inicial do *Online English Collocations Workbook*.



← → ↻ ⓘ www.gbd.institucional.ibilce.unesp.br/workbook/index



Online English Collocations Workbook

Email address:

Password:

[Forgot your password?](#)

Log in

[Don't have an account? Please Sign Up](#)

Welcome to **Online English Collocations Workbook**, an interactive platform to learn collocations, specially designed to native Brazilian Portuguese speakers, learners of English as a foreign language.

This design of this Workbook is part of an ongoing and larger research project carried out at Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), by Ph.D. Professor Adriane Orenha-Ottaiano.

Collocations are recurrent and arbitrary combinations of words, which are lexically and/or syntactically fixed to a certain degree. They are regarded to be relevant phraseological units in the learning process as, in the process of speaking, native speakers do not simply bring separate words together, they also use "prefabricated blocks", as if they were only one word. Hence, what appears to be spontaneous is actually a stereotyped fixed and repetitive speech, and if the speaker does not have a vast repertoire of these stereotyped fixed units (collocations, for instance) at their disposable, their speech may not sound natural.

The compilation of the Online English Collocations Workbook aims that students and teachers learn collocations more effectively, so that they increase their proficiency in English and thus achieve native-like naturalness.

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

Figura 11 – Tela de cadastro de novos usuários no *Online English Collocations Workbook*.

← → ⓘ www.gbd.institucional.ibilce.unesp.br/workbook/Cadastrar ☆ ⋮

Online English Collocations Workbook

Name:

Email address:

Institution:

Password:

Confirm password:



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

Após ter sido cadastro e recebido autorização por parte do administrador do *Workbook*, o usuário é redirecionado para a tela da atividade de *Memory Game*, na qual pode escolher a quantidade de pares que deseja jogar, o nível de dificuldade, a categoria de colocações com a qual deseja jogar, bem como a taxonomia das colocações que farão parte do jogo, conforme pode ser observado na imagem seguinte:

Figura 12 – Tela inicial do *Memory Game* no *Online English Collocations Workbook*.

The screenshot shows the 'Memory Game' interface. On the left, a sidebar lists 'MEMORY GAME' and 'GAP FILL'. Under 'MEMORY GAME', there are links for 'Play a game' (highlighted with a red circle), 'Game history', and 'GAP FILL'. The main area is titled 'Memory Game' and contains settings for 'Pairs' (6, 8, 10), 'Level of difficulty' (Easy, Medium, Hard), 'Category' (Academic Language, Business, General, Medicine, Politics, Investigation), and 'Taxonomy'. A blue 'Play!' button is at the bottom. On the right, a 'Guidelines' box explains the game rules.

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

Caso o aluno deseje praticar as colocações por meio da atividade *Gap Fill*, ele deve clicar em *Play a game*, logo abaixo de *Gap Fill* e selecionar o nível de dificuldade, além da categoria e da taxonomia das colocações que farão parte da atividade, conforme mostra a imagem a seguir:

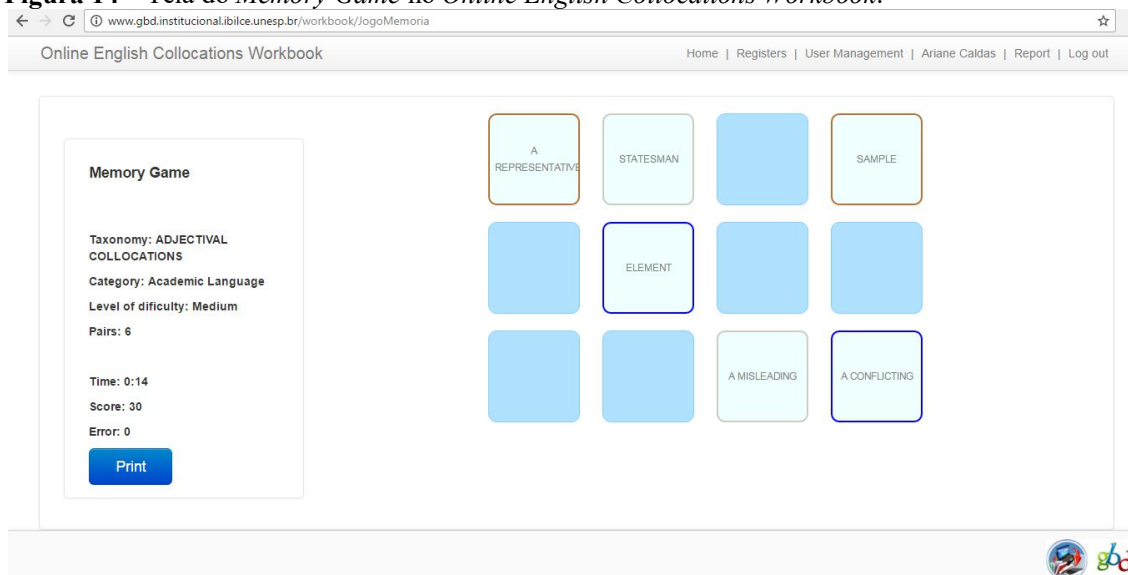
Figura 13 – Tela inicial do *Gap Fill* no *Online English Collocations Workbook*.

The screenshot shows the 'Gap Fill' interface. On the left, a sidebar lists 'MEMORY GAME' and 'GAP FILL'. Under 'GAP FILL', there are links for 'Play a game' (highlighted with a red circle), 'Game history', and 'GAP FILL'. The main area is titled 'Gap Fill' and contains settings for 'Level of difficulty' (Easy, Medium, Hard), 'Category' (Academic Language, Business, General, Medicine, Politics, Investigation), and 'Taxonomy'. A blue 'Play!' button is at the bottom. On the right, a 'Guidelines' box explains the exercise rules.

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

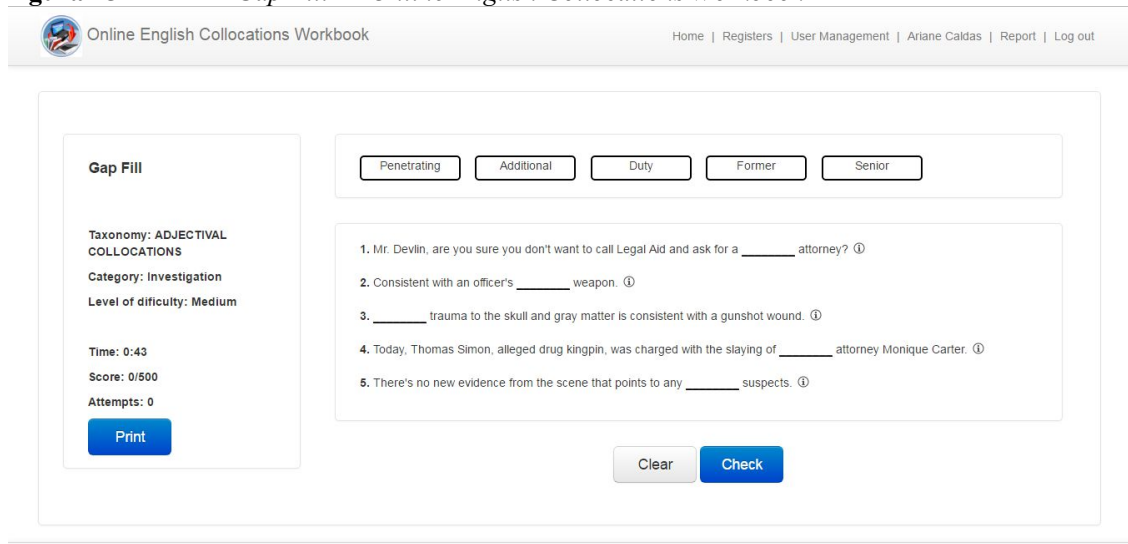
A seguir, são apresentadas as imagens contendo as telas do *Memory Game*, o qual o aluno tem 10 segundos para tentar memorizar os pares, e do *Gap Fill*, atividade na qual o aluno precisa completar as frases com as colocações corretas.

Figura 14 – 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>

Figura 15 – 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>

O aluno pode, também, verificar sua pontuação em ambas as atividades, além de poder jogá-las quantas vezes quiser.

Tendo apresentado o *Online English Collocations Workbook*, projeto maior ao qual esta pesquisa está relacionada, abordamos, a seguir, os procedimentos necessários para a elaboração das atividades que o compõem, a partir das colocações especializadas da área criminal.

3.5 Da elaboração das atividades didáticas

Uma vez que esta pesquisa tem como objetivo final a inserção das colocações especializadas encontradas nos corpora de estudos nas atividades do tipo *Memory Game* e *Gap Fill*, conforme constam no *Online English Collocations Workbook* (ORENHA-OTTAIANO, 2015), são abordados, a seguir, os passos necessários para a inserção destas atividades na referida plataforma *on-line*. Antes de cadastrar as colocações, criamos a categoria *Investigation*, na qual as atividades com as colocações especializadas identificadas nesta pesquisa foram inseridas. A imagem a seguir mostra esta categoria já inserida no *Workbook*.

Figura 16 – Tela de registro de categorias no *Online English Collocations Workbook*

The screenshot shows the 'Registers' page of the 'Online English Collocations Workbook'. On the left, a sidebar under 'MEMORY GAME' has 'Categories' selected. The main content area features a 'Category:' input field, 'Reset' and 'Save' buttons, and a table of existing categories. The table has a 'Show' dropdown set to '10' and a 'Search' field. The categories listed are: Academic Language, Business, General, Medicine, Politics, and Investigation.

Categories
Academic Language
Business
General
Medicine
Politics
Investigation

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

A partir de então, teve início o processo de inserção das colocações especializadas, obtidas por meio das análises, que farão parte das atividades do tipo *Memory Game*, processo que é feito levando-se em conta a base e o colocado de cada colocação, a sua classificação taxonômica (nominal, verbal, adjetiva ou adverbial), além do nível de dificuldade dado a cada colocação, conforme identificado previamente. Orenha-Ottaiano (2015) propõe tanto a

classificação taxonômica quanto os níveis de dificuldades, com o objetivo de conscientizar os aprendizes das estruturas das colocações (escolha da taxonomia), bem como auxiliá-los na realização da atividade ou do exercício (escolha dos níveis). Embora a classificação das colocações especializadas de acordo com níveis de dificuldade possa ser subjetiva, ela é feita a partir de critérios que consideram, principalmente, a variação e ampliação do conhecimento léxico de alunos de língua inglesa, que vão desde o nível básico até o avançado, abrangendo vocabulário que vai do mais simples ao mais sofisticado, visando auxiliar o aluno no desenvolvimento de seu conhecimento léxico e colocacional a partir do trabalho feito com as atividades propostas. A imagem a seguir mostra como é feito o cadastro das colocações para este tipo de atividade.

Figura 17 – Tela de cadastro das colocações especializadas no *Online English Collocations Workbook*.

Registers

Category: Investigation Level of Difficulty: Medium

Taxonomy: VERBAL COLLOCATIONS

Node: COVER Collocate: IN BLOOD

Reset Save

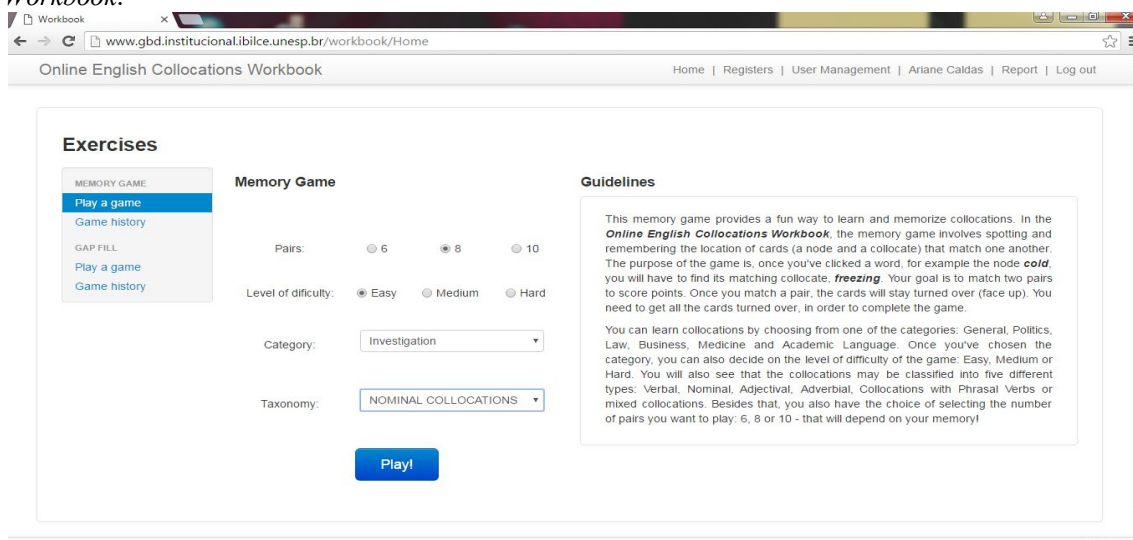
Show 10 entries Search: investigat

Category	Taxonomy	Level of difficulty	of	Node	Collocate
Investigation	ADVERBIAL COLLOCATIONS	Easy		BLOOD	DONATION
Investigation	NOMINAL COLLOCATIONS	Medium		BLOOD	DONOR
Investigation	NOMINAL COLLOCATIONS	Medium		DROP OF	BLOOD

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

Terminado esse processo, já é possível visualizar e jogar o *Memory Game*, escolhendo o nível de dificuldade desejado, bem como o tipo de colocação com o qual se gostaria de jogar, de acordo com a classificação taxonômica dada a cada colocação especializada, além da quantidade de pares com os quais se quer jogar, conforme nos mostra a imagem a seguir:

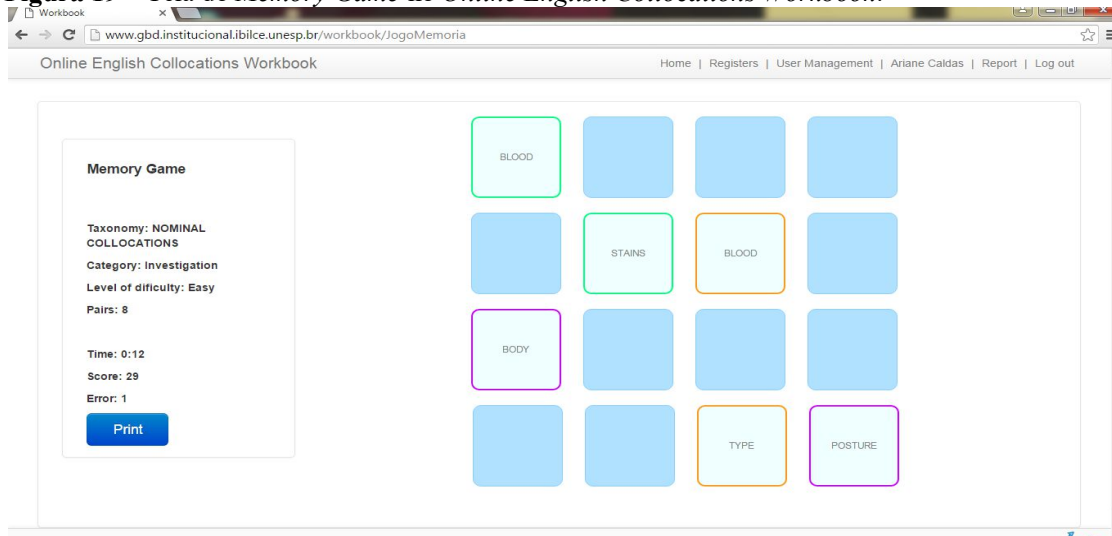
Figura 18 – Tela de seleção das colocações para o *Memory Game* no *Online English Collocations Workbook*.



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

Após a seleção das características desejadas para o jogo, é possível a memorização dos pares por um tempo determinado pelo jogador. Em seguida o jogo é iniciado e o aluno deve, além de identificar os pares corretamente, se ater ao tempo que é cronometrado por um relógio. Quanto mais rápido o aluno conseguir encontrar os pares, melhor está seu conhecimento colocacional.

Figura 19 – 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>.

No que se refere à inserção dos dados que fazem parte do *Gap Fill*, esta etapa compreende, além dos dados já mencionados para a inserção das colocações especializadas no *Memory Game*, a classificação taxonômica e a classificação de acordo com os níveis de dificuldade, a inclusão de um exemplo, retirado do CCC contextualizando a colocação especializada em questão. Os exemplos servem para auxiliar o aluno na compreensão dos significados de cada colocação especializada a partir de seu uso.

Após finalizar o processo de inserção destas colocações especializadas e seus respectivos exemplos de uso, também já é possível acessar a página para selecionar o nível de dificuldade e o tipo de colocação com o qual se deseja trabalhar, e jogar o *Gap Fill*, conforme as imagens a seguir:

Figura 20 – Tela de seleção das colocações para o *Gap Fill* no *Online English Collocations Workbook*.

The screenshot shows a web browser window with the URL www.institucional.grupogbd.com/workbook/FillInTheBlanks. The page title is "Online English Collocations Workbook". The navigation bar includes links: Home | Registers | User Management | Ariane Caldas | Report | Log out.

The main content area is titled "Exercises" and features a sidebar with two sections: "MEMORY GAME" (with links "Play a game" and "Game history") and "GAP FILL" (with links "Play a game" and "Game history"). The "GAP FILL" section is currently selected.

The "Gap Fill" section contains the following options:

- Level of difficulty: ☒ Easy ☐ Medium ☐ Hard
- Category:
- Taxonomy:
-

A "Guidelines" box on the right provides instructions: "This is a Gap Fill Exercise. Please fill in the blanks with the words from the boxes, by dragging them to the right sentence. Like the Memory Game, you can learn collocations by choosing from one of the categories: General, Politics, Law, Business, Medicine and Academic Language. Once you've chosen the category, you can also decide on the level of difficulty of the game: Easy, Medium or Hard. You will also see that the collocations may be classified into five different types: 1) Verbal; 2) Nominat; 3) Adjectival; 4) Adverbial; and 5) Collocations with Phrasal Verbs."

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

Figura 21 – Tela do *Gap Fill* no *Online English Collocations Workbook*.

Online English Collocations Workbook

Home | Registers | User Management | Ariane Caldas | Report | Log out

Gap Fill

Taxonomy: VERBAL COLLOCATIONS
Category: Investigation
Level of difficulty: Easy

Time: 0:30
Score: 0/500
Attempts: 0

Print

Fire my | Prepped her | Drop your | Received | Oxygenate her

1. The blood was unable to Oxygenate her tissue. ①
2. First time I ever had to _____ weapon. ①
3. Yeah, the coroner's assistant who _____ body collected a green, waxy substance under her fingernails. ①
4. We have visual. _____ weapon! Freeze! Freeze! Get on your knees right now! ①
5. We need to know if your company _____ tissue or bone from a Roger Lapinsky. ①

Clear | Check

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

Ao finalizar o jogo, o aluno pode verificar se suas respostas estão corretas, além de conferir seu desempenho no jogo, seja no *Memory Game* ou no *Gap Fill*, o qual fica gravado no *Workbook*, de forma a permitir que o aluno acompanhe sua evolução nas atividades, conforme as imagens abaixo:

Figura 22 – Tela da verificação das respostas do *Gap Fill*.

Online English Collocations Workbook

Home | Registers | User Management | Ariane Caldas | Report | Log out

Gap Fill

Taxonomy: VERBAL COLLOCATIONS
Category: Investigation
Level of difficulty: Easy

Time: 1:33
Score: 500/500
Attempts: 1

Print

Congratulation, you've won!

Received | Drop your | Prepped her | Fire my | Oxygenate her

1. The blood was unable to Oxygenate her tissue. ①
2. First time I ever had to Fire my weapon. ①
3. Yeah, the coroner's assistant who Prepped her body collected a green, waxy substance under her fingernails. ①
4. We have visual. Drop your weapon! Freeze! Freeze! Get on your knees right now! ①
5. We need to know if your company Received tissue or bone from a Roger Lapinsky. ①

Finish

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

Figura 23 – Tela do desempenho do aluno no *Gap Fill*.

The screenshot shows a web browser window with the URL www.institucional.grupogbd.com/workbook/BlanksHistorico. The page title is 'Online English Collocations Workbook'. The main content area is titled 'Exercises' and contains two sections: 'MEMORY GAME' and 'GAP FILL'. The 'GAP FILL' section is active, showing a table of results. The table has columns for Date, Level of difficulty, Category, Taxonomy, Score, Time, and Attempts. There are 5 entries displayed, showing scores ranging from 100 to 500 and times from 0:29 to 2:45.

Date	Level of difficulty	Category	Taxonomy	Score	Time	Attempts
20-07-2017	Easy	Investigation	VERBAL COLLOCATIONS	500	0:29	1
21-04-2017	Easy	Investigation	VERBAL COLLOCATIONS	500	1:33	1
28-03-2017	Easy	Investigation	VERBAL COLLOCATIONS	300	2:45	1
12-05-2016	Medium	Investigation	VERBAL COLLOCATIONS	500	1:39	1
12-05-2016	Medium	Investigation	VERBAL COLLOCATIONS	100	1:16	1

Showing 1 to 5 of 5 entries

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

Convém ressaltar que as mesmas colocações especializadas inseridas na atividade do tipo *Memory game* também são cadastradas na atividade do tipo *Gap Fill*. Com isso, as atividades constantes do *Online English Collocations Workbook* permitem que o aluno memorize mais facilmente as colocações estudadas, além de, por meio das atividades de *Gap Fill*, possibilitar que o aluno entenda o significado de cada colocação em seu contexto de uso, apresentado pelos exemplos extraídos dos corpora de estudo, escolhidos por serem representativos da linguagem investigativa.

No capítulo a seguir, apresentamos os dados obtidos nesta pesquisa, bem como sua análise.

4 ANÁLISE DOS DADOS

Neste capítulo, são apresentados e analisados os dados obtidos por meio dos corpora que compõem esta pesquisa, a saber, o CCSI (Corpus CSI), constituído pelas legendas de doze temporadas da série *CSI – Crime Scene Investigation*, assim como e o CCC (Corpus Comparável Criminal), compilado com o propósito de comprovar a frequência de coocorrência das colocações especializadas encontradas no CCSI.

Iniciamos o capítulo com uma análise dos dados quantitativos obtidos em ambos os corpora, passando, em seguida, para a análise morfossintática das colocações obtidas, buscando sua classificação taxonômica a partir dos pressupostos teóricos de Hausman (1984) e de Orenha-Ottaiano (2004, 2009). Posteriormente, apresentamos uma análise contrastiva entre os dados obtidos por meio do CCSI e os dados obtidos pelo CCC. A análise morfossintática é feita a partir das colocações especializadas obtidas por meio da palavra-chave *evidence*, já que está é a palavra-chave mais frequente no CCSI (1.112) e a segunda mais frequente no CCC (9.501), além de ter sido escolhida por critério de delimitação de pesquisa.

4.1 Análise quantitativa dos dados em ambos os corpora de estudo

De acordo com Biber (1998, p. 5) “o objetivo das investigações baseadas em corpus não é simplesmente reportar achados quantitativos, mas explorar a importância destes achados para a aprendizagem de padrões da língua em uso”³⁹. Embora a LC não se prenda apenas aos dados numéricos, estes revelam aspectos interessantes do corpus em estudo.

Conforme mencionado no capítulo 3, contamos com um corpus composto pelas legendas de 12 (doze) temporadas da série *CSI – Crime Scene Investigation*, o qual foi denominado Corpus *CSI* (CCSI) e com um corpus comparável, denominado Corpus Comparável Criminal (CCC), que foi compilado por meio da ferramenta *Boot Cat Front-End* versão 0.71 (BARONI; BERNARDINI; ZANCHETTA, 2013), a fim de ser mais uma fonte de comprovação da ocorrência das colocações especializadas identificadas a partir do Corpus *CSI*.

No que diz respeito ao tamanho dos Corpora, o CCSI conta com um total de 3.061.029 de palavras, enquanto o CCC apresenta um total de 3.886.465 de palavras, sendo, assim, um

³⁹ “The goal of corpus-based investigations is not simply to report quantitative findings, but to explore the importance of these findings for learning about the patterns of language use.”

pouco maior que o CCSI. Contudo, como mencionado no capítulo sobre a metodologia da pesquisa, o CCSI foi adotado como ponto de partida para a busca e o levantamento das colocações especializadas desta pesquisa. Dessa forma, foram delimitadas 67 palavras-chave a partir deste corpus, as quais também foram utilizadas para as buscas empreendidas no CCC, o que permitiu comparar os dados obtidos em ambos os corpora.

Com relação às palavras-chave selecionadas, é possível perceber que elas ocorrem com frequências bem diferentes em cada um dos corpora, tal como pode ser observado na tabela abaixo.

Tabela 3 – Frequências das palavras-chave em ambos os corpora de pesquisa

Frequência das palavras-chave em ambos os corpora de estudo								
Key Word	CCSI	CCC	KeyWord	CCSI	CCC	KeyWord	CCSI	CCC
<i>Abrasion</i>	50	54	<i>Fiber</i>	156	90	<i>Residue</i>	71	651
<i>Accelerant</i>	30	7	<i>Fingerprints</i>	170	498	<i>Robbery</i>	110	2.466
<i>Alibi</i>	84	108	<i>Fire</i>	206	1.213	<i>Scene</i>	849	5.504
<i>Arrest</i>	188	3.285	<i>Forensics</i>	55	784	<i>Serial</i>	125	1.923
<i>Attorney</i>	98	3.805	<i>Fracture</i>	57	49	<i>Sheriff</i>	147	711
<i>Autopsy</i>	129	913	<i>Gun</i>	903	5.160	<i>Shoot</i>	221	478
<i>Ballistics</i>	67	1.739	<i>Gunshot</i>	205	1.910	<i>Shooter</i>	100	1.464
<i>Blood</i>	1.830	1.471	<i>Hemorrhage</i>	18	19	<i>Shot</i>	789	3.672
<i>Body</i>	1.284	3.902	<i>Hemorrhaging</i>	31	1	<i>Skull</i>	120	163
<i>Bullet</i>	452	3.910	<i>Homicide</i>	194	2.847	<i>Spatter</i>	111	58
<i>Cell</i>	330	641	<i>Jail</i>	178	2.062	<i>Suicide</i>	179	3.305
<i>Contusion</i>	66	16	<i>Kill</i>	1.983	990	<i>Surveillance</i>	172	362
<i>Cop</i>	296	343	<i>Killer</i>	629	1.939	<i>Suspect</i>	610	3.766
<i>Coroner</i>	106	290	<i>Knife</i>	244	310	<i>Swear</i>	108	16
<i>Counselor</i>	24	32	<i>Lab</i>	689	2.328	<i>Tissue</i>	193	267
<i>Crime</i>	1.036	11.181	<i>Laceration</i>	67	17	<i>Trace</i>	325	385
<i>Dead</i>	1.391	2.096	<i>Lawyer</i>	193	1.542	<i>Trauma</i>	157	375
<i>Defense</i>	108	2.660	<i>License</i>	106	432	<i>Victim</i>	709	4.861
<i>Detective</i>	221	2.115	<i>Ligature</i>	44	13	<i>Warrant</i>	332	3.212
<i>DNA</i>	837	2.833	<i>Murder</i>	894	5.920	<i>Weapon</i>	348	2.702
<i>Evidence</i>	1.112	9.501	<i>Perimortem</i>	23	0	<i>Wound</i>	315	1.326
<i>Eyewitness</i>	48	1.570	<i>Postmortem</i>	54	44	-	-	-
<i>Felony</i>	57	1.261	<i>Prints</i>	587	279	-	-	-

De modo geral, ao comparar a frequência que as palavras-chave apresentam em ambos os corpora, fica evidente que o CCC, por ser maior, permite a identificação de uma maior ocorrência de cada palavra-chave. Diferentemente do que ocorre no CCSI que tem como palavra-chave mais frequente *evidence* com 1.112 ocorrências, a palavra-chave mais

frequente neste corpus é *crime*, que ocorre 11.181 vezes. As demais palavras-chave do CCC apresentam, também, um alto índice de frequência.

Já as palavras-chave *accelerant*, *contusion*, *hemorrhaging*, *laceration*, *postmortem*, *splatter* e *swear*, por exemplo, ocorrem em menor quantidade no CCC do que no CCSI. Acredita-se que isto se deve ao fato de o CCSI ser formado pelas legendas de uma série que, embora retrate o cotidiano de investigadores forenses, se vale de outros elementos para prender a atenção do telespectador, tais como destacar o momento da autópsia, por exemplo, contexto em que palavras como *contusion*, *hemorrhage*, *hemorrhaging*, *laceration* e *postmortem* tendem a ocorrer com maior frequência, sendo estes elementos menos enfatizados nas notícias, reportagens e relatos sobre os crimes e investigações verídicos.

Outro dado quantitativo que merece destaque é o fato de a palavra-chave *perimortem* (23 ocorrências no CCSI) não ter sido encontrada no CCC, sendo que todas as outras ocorrem em ambos os corpora. Chamamos a atenção, também, para o termo *hemorrhaging*, que ocorre apenas 2 vezes no CCC. Tal como acontece com *perimortem*, que não ocorre neste corpus, ambas as palavras foram mantidas como nós de busca por colocações especializadas tendo em vista sua frequência no CCSI (23 ocorrências para *perimortem* e 31 ocorrências para *hemorrhaging*), e a frequência com que as colocações especializadas identificadas a partir delas ocorrem no *WebCorp*.

Conforme mencionamos, uma vez que esta pesquisa é dirigida pelos dados obtidos por meio do corpus de estudo composto pelas legendas da série *CSI – Crime Scene Investigation*, a análise morfossintática apresentada a seguir é feita a partir dos dados obtidos por meio do nóculo *evidence*, haja vista que este é o nóculo de maior frequência no CCSI (1.112 ocorrências). Convém ressaltar que esta análise abrange, também, as colocações obtidas no CCC a partir deste mesmo nóculo (*evidence*), a fim de contrastar os dados obtidos em ambos os corpora.

Com relação à quantidade de colocações obtidas a partir de cada um dos corpora, foi possível identificar mais de 1.700 colocações especializadas a partir do CCSI e mais de 1.800 a partir do CCC. Além disso, cada colocação extraída apresenta frequência diversa, havendo colocações que aparecem algumas centenas de vezes e colocações que ocorrem apenas uma vez nos corpora. Faz-se apropriado dizer que, como mencionado no capítulo de metodologia da pesquisa, embora algumas colocações tenham sido encontradas apenas uma vez no CCSI, estas foram inseridas no *Online English Collocations Workbook* devido à frequência encontrada no CCC, ou, em alguns casos, à frequência obtida por meio do *WebCorp*. O mesmo se dá com algumas colocações que foram identificadas no CCC com frequência igual

a 1 e foram mantidas devido à frequência que apresentaram no CCSI e/ou no emulador de concordanciador *WebCorp*.

Ao considerarmos apenas o nóculo *evidence*, que ocorre 1.112 vezes no CCSI e 9.501 vezes no CCC, o qual é utilizado para ilustrar a análise morfossintática e léxico-semântica desenvolvida nesta pesquisa, foi possível encontrarmos 148 colocações especializadas a partir deste nóculo no CCSI, enquanto que o CCC possibilitou a identificação de 167 colocações especializadas, uma diferença pequena, se considerarmos a diferença de frequência com que *evidence* ocorre em ambos os corpora.

Dessa forma, os dados quantitativos aqui apresentados contribuíram para a comprovação de que o CCSI atende ao propósito desta pesquisa, no sentido de que, conforme será discutido neste capítulo, se trata de uma fonte expressiva do uso das colocações especializadas da área investigativa criminal, de modo a despertar o interesse destas por parte dos aprendizes de língua inglesa.

4.2 Análise morfossintática das colocações a partir do Corpus CSI

Tal como mencionado no capítulo 2, tomamos como base os pressupostos teóricos de Hausmann (1984) e Orenha-Ottaiano (2004) para empreender a análise morfossintática das colocações levantadas, a fim de organizá-las segundo sua classificação taxonômica.

Segundo já mencionamos, ao abordar a questão da formação das colocações, Hausmann (1984) afirma que estas são compostas basicamente por dois elementos, a saber, a base e o colocado, sendo que cada um desses elementos apresenta uma função bem definida e hierárquica, haja vista que a base é autônoma e determinante, e o colocado é o elemento determinado pela base (HAUSMANN, 1984 apud ORENHA-OTTAIANO, 2009, p. 39).

Assim como Orenha-Ottaiano (2009, p. 40), ao se basear em Heid *et all* (1991), entende-se, neste trabalho, que a base é um elemento independente, semanticamente autônomo, traduzível e responsável por determinar os padrões lexicais, enquanto o colocado atua como um conceito modificador, é semanticamente interpretável e traduzido a partir de seu uso dentro da colocação, além de ser determinado pela base.

Dessa forma, buscou-se analisar e descrever a estrutura morfossintática das colocações especializadas levantadas a partir do nóculo *evidence*, uma vez que este é o mais frequente no CCSI (1.112) e o segundo mais frequente no CCC (9.501), partindo dos pressupostos mencionados previamente e utilizando a tipologia proposta por Orenha-Ottaiano (2004) ao

tratar de colocações especializadas, fundamentada em Hausmann (1985), ao investigar colocações da língua geral, conforme mostramos abaixo:

Verbais

- Verbo _{base}+ Subst. _{colocado} *to find evidence*
- Subst. _{base}+ Verbo _{colocado} *evidence changes*
- Verbo _{base}+ Det. + Subst. _{colocado} *to reexamine an evidence*
- Verbo _{colocado}+ Prep. + [Det.]⁴⁰ + Subst. _{base} *based on the evidence*
- Verbo _{colocado}+ Part. Adverbial⁴¹ + Subst. _{base} *To sift through evidence*

Nominais

- Subst. _{colocado}+ Subst. _{base} *a DNA evidence*
- Subst. _{base}+ Prep. + [Det.] + Subst. _{colocado} *an evidence of the crime*

Adjetiva

- Adj. _{colocado}+ Subst. _{base} *physical evidence*

Adverbial

- Verbo _{base}+ Adv. _{colocado} *to shoot effectively*
- Adv. _{colocado}+ Verbo _{base} *tragically killed*

Convém ressaltar que esta análise em busca da classificação taxonômica das colocações especializadas se faz necessária, visando não apenas à identificação das colocações levantadas, mas, também, e principalmente, às atividades propostas no *Online English Collocations Workbook*, uma vez que estas são apresentadas de acordo com a tipologia das colocações.

A seguir, são apresentadas as colocações especializadas levantadas a partir do nódulo *evidence* no CCSI, bem como sua respectiva classificação segundo sua tipologia.

⁴⁰ Neste caso pode ou não ocorrer o determinante.

⁴¹ No caso dos *phrasal verbs* da língua inglesa.

4.2.1 Colocações especializadas nominais

De acordo com a tipologia proposta por Orenha-Ottaiano (2009), foi possível encontrar duas formações para as colocações nominais: substantivo + substantivo e substantivo + preposição + substantivo. Em nossas análises, igualmente identificamos estes dois tipos de formação entre as colocações especializadas nominais, além de termos encontrado colocações especializadas em que ora *evidence* é a base, ora é o colocado, conforme pode ser observado nos quadros a seguir. Convém ressaltar que as colocações especializadas nas quais o termo *evidence* aparece como colocado foram mantidas em nossa análise uma vez que elas foram identificadas no corpus por meio da busca e levantamento feitos a partir deste termo, de forma que, embora atue como colocado em alguns casos, é possível agrupá-las sob o termo em análise, tendo em vista sua relevância na área de especialidade em foco.

Quadro 1 – Estrutura das colocações especializadas nominais: *Evidence* como base

Colocação	Estrutura
<i>A murder evidence</i>	subst. colocado + subst. base
<i>An homicide evidence</i>	subst. colocado + subst. base
<i>A blood evidence</i>	subst. colocado + subst. base
<i>An ammo evidence</i>	subst. colocado + subst. base
<i>A case evidence</i>	subst. colocado + subst. base
<i>An evidence of assault</i>	subst. base + prep. + subst. colocado
<i>A DNA evidence</i>	subst. colocado + subst. base
<i>An evidence out of context</i>	subst. base + prep. + subst. colocado
<i>An evidence in context</i>	subst. base + prep. + subst. colocado
<i>An evidence of a struggle</i>	subst. base + prep. + det. + subst. colocado
<i>An evidence of the crime</i>	subst. base + prep. + det. + subst. colocado
<i>An interpreter of the evidence</i>	subst. colocado + prep. + det. + subst. base
<i>A hair evidence</i>	subst. colocado + subst. base
<i>A fiber evidence</i>	subst. colocado + subst. base
<i>An evidence of trauma</i>	subst. base + prep. + subst. colocado
<i>A body evidence</i>	subst. colocado + subst. base
<i>A drug evidence</i>	subst. base + subst. colocado

Quadro 2 – Estrutura das colocações especializadas nominais: *Evidence* como colocado

Colocação	Estrutura
<i>An evidence examination</i>	subst. colocado + subst. base
<i>An evidence bag</i>	subst. colocado + subst. base
<i>An evidence ledger</i>	subst. colocado + subst. base
<i>Evidence trinity</i>	subst. colocado + subst. base
<i>An evidence log</i>	subst. colocado + subst. base
<i>An evidence box</i>	subst. colocado + subst. base
<i>An evidence locker</i>	subst. colocado + subst. base
<i>An evidence room</i>	subst. colocado + subst. base
<i>An absence of evidence</i>	subst. base + prep. + subst. colocado
<i>An evidence custodian</i>	subst. colocado + subst. base
<i>An evidence marker</i>	subst. colocado + subst. base
<i>An evidence vault</i>	subst. colocado + subst. base
<i>An evidence bullet</i>	subst. colocado + subst. base
<i>An evidence Seal</i>	subst. colocado + subst. base
<i>An evidence recovery</i>	subst. colocado + subst. base
<i>A lack of evidence</i>	subst. base + prep. + subst. colocado
<i>A stitch of evidence</i>	subst. base + prep. + subst. colocado
<i>An evidence collection</i>	subst. colocado + subst. base
<i>An evidence area</i>	subst. colocado + subst. base

Tal como pode ser observado nos quadros acima, o termo *evidence* ora se apresenta como a base, em 17 colocações, conforme no quadro 1, ora como o colocado, em 20 colocações, segundo o quadro 2, haja vista que ele ora figura como elemento semanticamente autônomo e traduzível ora ocorre como um conceito modificador, tal como aponta Orenha-Ottaiano (2009).

Além disso, entre as colocações especializadas nominais apresentadas aqui, podemos notar que há uma variação da posição que o termo *evidence* assume, mesmo quando atua como base, sendo ora o primeiro elemento, ora o segundo elemento das colocações.

Nos casos em que não há uma preposição entre os substantivos que constituem a colocação, percebemos que essa variação da posição da base faz com que tenhamos colocações nominais que apresentam substantivos com função adjetiva, como é o caso de *a blood evidence* e *DNA evidence*, em que as lexias *blood* e *DNA* complementam e qualificam o tipo de evidência em questão. Esses casos ocorrem em 9 das 16 colocações especializadas nominais que tem como base o termo *evidence* (*a murder evidence*, *an homicide evidence*, *a blood evidence*, *an ammo evidence*, *a case evidence*, *a DNA evidence*, *a hair evidence*, *a fiber evidence*, e *a body evidence*). As demais colocações, nas quais *evidence* atua como base são constituídas pela formação substantivo + preposição + substantivo, sendo que, nestes casos, a função adjetiva é atribuída ao substantivo colocado, tais como *an evidence of assault*, *an evidence of a struggle* e *an evidence of the crime*, por exemplo.

Quanto aos casos em que o termo *evidence* figura como colocado e é o primeiro elemento da colocação, seguido por um segundo substantivo, percebe-se que a função adjetiva é atribuída a ele, o qual passa a qualificar o segundo substantivo. É o caso, por exemplo, das colocações *an evidence bag*, *an evidence box*, *an evidence collection*, *an evidence marker*, *an evidence custodian* e *evidence room*. Este tipo de situação ocorre em 16 das 19 colocações especializadas nominais nas quais *evidence* atua como colocado. As outras três colocações especializadas em que *evidence* atua como colocado são constituídas pela formação substantivo + preposição + substantivo: *an absence of evidence*, *a lack of evidence*, *a stitch of evidence*.

4.2.2 Colocações especializadas adjetivas

Entre as colocações especializadas extraídas a partir do nóculo de busca *evidence*, foram identificadas 34 colocações adjetivas, as quais são apresentadas no quadro a seguir, juntamente com a estrutura morfossintática de seus componentes.

Quadro 3 – Estrutura das colocações especializadas adjetivas

Colocação	Estrutura
<i>An outstanding evidence</i>	adjetivo colocado + subst. base
<i>A physical evidence</i>	adjetivo colocado + subst. base
<i>An old evidence</i>	adjetivo colocado + subst. base
<i>A new evidence</i>	adjetivo colocado + subst. base
<i>An incriminating evidence</i>	adjetivo colocado + subst. base
<i>A corresponding evidence</i>	adjetivo colocado + subst. base
<i>A transitory evidence</i>	adjetivo colocado + subst. base
<i>A hard evidence</i>	adjetivo colocado + subst. base
<i>A biological evidence</i>	adjetivo colocado + subst. base
<i>An extraneous evidence</i>	adjetivo colocado + subst. base
<i>An associated evidence</i>	adjetivo colocado + subst. base
<i>A latest evidence</i>	adjetivo colocado + subst. base
<i>An insufficient evidence</i>	adjetivo colocado + subst. base
<i>An unexplained evidence</i>	adjetivo colocado + subst. base
<i>A sufficient evidence</i>	adjetivo colocado + subst. base
<i>A recent evidence</i>	adjetivo colocado + subst. base
<i>A plastic evidence bag</i>	adjetivo colocado + subst. base
<i>A forensic evidence</i>	adjetivo colocado + subst. base
<i>A good evidence</i>	adjetivo colocado + subst. base
<i>A circumstantial evidence</i>	adjetivo colocado + subst. base
<i>A wrong evidence</i>	adjetivo colocado + subst. base
<i>An invisible evidence</i>	adjetivo colocado + subst. base
<i>An additional evidence</i>	adjetivo colocado + subst. base
<i>An unlogged evidence</i>	adjetivo colocado + subst. base
<i>A reliable evidence</i>	adjetivo colocado + subst. base
<i>An instant evidence</i>	adjetivo colocado + subst. base
<i>An empirical evidence</i>	adjetivo colocado + subst. base
<i>A disturbing evidence</i>	adjetivo colocado + subst. base
<i>A clear evidence</i>	adjetivo colocado + subst. base
<i>An equivocal evidence</i>	adjetivo colocado + subst. base
<i>A latent evidence</i>	adjetivo colocado + subst. base
<i>An important evidence</i>	adjetivo colocado + subst. base
<i>An individualized evidence</i>	adjetivo colocado + subst. base
<i>A surgical evidence</i>	adjetivo colocado + subst. base

Conforme é possível notar, as colocações especializadas adjetivas seguem a estrutura identificada e proposta por Orenha-Ottaiano (2009), havendo sempre um adjetivo que atua como colocado seguido de um substantivo base que é, por sua vez, um termo, de forma que não apresentam variação de posição do seu elemento base, a saber, o substantivo *evidence*, sendo, de certa forma, mais fácil a identificação de seus elementos.

Além disso, a julgar o número de colocações especializadas adjetivas encontradas (34) a partir de um mesmo nóculo de busca, o número de colocações deste tipo em ambos os corpora é bastante grande, haja vista que entre as colocações levantadas a partir dos demais nósculos de busca (lembrando que selecionamos um total de 67 palavras-chave que são

utilizadas como nódulos de busca) esse tipo de colocação se mostra bastante recorrente. Dessa forma, pode-se afirmar que a área de especialidade de investigação criminal é bastante rica em colocações especializadas do tipo adjetiva.

Passamos agora à análise da estrutura morfossintática das colocações especializadas verbais.

4.2.3 Colocações especializadas verbais

Nos quadros abaixo são apresentadas as colocações especializadas verbais obtidas a partir do nódulo de busca *evidence*, bem como suas respectivas estruturas morfossintáticas, separando as colocações compostas por verbos frasais, as colocações em que a base atua como sujeito e as colocações em que a base atua como objeto dos verbos colocados.

Quadro 4 – Estrutura das colocações especializadas verbais constituídas por verbos frasais

Colocação	Estrutura
<i>To sift through evidence</i>	verbo colocado + part. adverbial + subst. base
<i>To look for evidence</i>	verbo colocado + part. adverbial + subst. base
<i>To look at evidence</i>	verbo colocado + part. adverbial + subst. base
<i>To turn into evidence</i>	verbo colocado + part. adverbial + subst. base

Este quadro apresenta as colocações especializadas que possuem um verbo frasal (*phrasal verb*) em sua formação. Em nosso levantamento, identificamos quatro colocações com esta característica entre as colocações obtidas por meio do termo *evidence*. Em todos os casos apresentados, *evidence* atua como base das colocações, e tem como colocado um verbo e uma partícula adverbial, elementos que constituem os verbos frasais da língua inglesa.

Foram identificadas, também, colocações especializadas verbais nas quais a base (*evidence*) atua como sujeito dos verbos colocados. Estes casos são apresentados no quadro a seguir.

Quadro 5 – Estrutura das colocações especializadas verbais: base como sujeito

Colocação	Estrutura
<i>Evidence changes</i>	subst. base + verbo colocado
<i>The evidence suggests</i>	subst. base + verbo colocado
<i>The evidence says</i>	subst. base + verbo colocado
<i>Evidence points to</i>	subst. base + verbo colocado
<i>Evidence appears</i>	subst. base + verbo colocado
<i>Evidence collected</i>	subst. base + verbo colocado
<i>Evidence clears</i>	subst. base + verbo colocado
<i>Evidence supports</i>	subst. base + verbo colocado
<i>Evidence tells</i>	subst. base + verbo colocado
<i>Evidence persists</i>	subst. base + verbo colocado
<i>Evidence fails</i>	subst. base + verbo colocado
<i>The evidence proves</i>	subst. base + verbo colocado
<i>Evidence leads</i>	subst. base + verbo colocado
<i>Evidence shows</i>	subst. Base + verbo colocado
<i>Evidence connecting</i>	subst. base + verbo colocado
<i>The evidence indicates</i>	verbo colocado + subst. Base

De acordo com o quadro acima, o CCSI apresentou 16 colocações especializadas verbais nas quais a base atua como sujeito dos verbos colocados, predominando, assim, a formação substantivo + verbo. Convém ressaltar que, nestes casos, o verbo que acompanha a base tende sempre a ocorrer no presente, exceto para a colocação *evidence collected*, o qual ocorre no pretérito.

Em relação às demais colocações verbais encontradas, observamos que estas apresentam a base como objeto dos verbos colocados, segundo mostra o quadro a seguir:

Quadro 6 – Estrutura das colocações especializadas verbais: base como objeto

Colocação	Estrutura
<i>To find evidence</i>	verbo colocado + subst. base
<i>To reexamine an evidence</i>	verbo colocado + det. + subst. base
<i>To flush an evidence</i>	verbo colocado + det. + subst. base
<i>Based on the evidence</i>	verbo colocado + prep. + det. + subst. base
<i>To get evidence</i>	verbo colocado + subst. base
<i>To ditch an evidence</i>	verbo colocado + det. + subst. base
<i>To collect evidence</i>	verbo colocado + subst. base
<i>To follow an evidence</i>	verbo colocado + det. + subst. base
<i>To trace evidence</i>	Verbo colocado + subst. base
<i>To destroy evidence</i>	verbo colocado + subst. base
<i>To plant an evidence</i>	verbo colocado + det. + subst. base
<i>To document evidence</i>	Verbo colocado + subst. base
<i>To obliterate evidence</i>	verbo colocado + subst. base
<i>To fabricate evidence</i>	verbo colocado + subst. base
<i>To process an evidence</i>	verbo colocado + det. + subst. base
<i>To record evidence</i>	verbo colocado + subst. base
<i>To analyze an evidence</i>	verbo colocado + det. + subst. base
<i>To exclude an evidence</i>	verbo colocado + det. + subst. base
<i>To burn an evidence</i>	verbo colocado + det. + subst. base
<i>To manipulate evidence</i>	verbo colocado + subst. base
<i>To recover evidence</i>	verbo colocado + subst. base
<i>To go into evidence</i>	verbo colocado + prep. + subst. base
<i>To hide an evidence</i>	verbo colocado + det. + subst. base
<i>To chase an evidence</i>	verbo colocado + det. + subst. base
<i>To mishandle an evidence</i>	verbo colocado + det. + subst. base
<i>To compromise evidence</i>	verbo colocado + subst. base
<i>To run evidence</i>	verbo colocado + subst. base
<i>To test an evidence</i>	verbo colocado + det. + subst. base
<i>To gather an evidence</i>	verbo colocado + det. + subst. base
<i>To check an evidence</i>	verbo colocado + det. + subst. base
<i>To preserve an evidence</i>	verbo colocado + det. + subst. base
<i>To categorize evidence</i>	verbo colocado + subst. base
<i>To interpreter an evidence</i>	verbo colocado + det. + subst. base
<i>To review an evidence</i>	verbo colocado + det. + subst. base
<i>To face an evidence</i>	verbo colocado + det. + subst. base
<i>To evaluate evidence</i>	verbo colocado + subst. base
<i>To assume evidence</i>	verbo colocado + subst. base
<i>To melt evidence</i>	verbo colocado + subst. base
<i>To attack an evidence</i>	verbo colocado + det. + subst. base
<i>To disappear an evidence</i>	verbo colocado + det. + subst. base
<i>To book an evidence</i>	verbo colocado + det. + subst. base
<i>To reanalyze an evidence</i>	verbo colocado + det. + subst. Base
<i>To compel evidence</i>	verbo colocado + subst. base
<i>To steal evidence</i>	verbo colocado + subst. base
<i>To jeopardize an evidence</i>	verbo colocado + det. + subst. base
<i>To retrieve an evidence</i>	verbo colocado + det. + subst. base
<i>To tamper with evidence</i>	subst. colocado + prep. + subst. base

Conforme se pode notar, as colocações verbais em que a base atua como objeto dos verbos colocados constituem a maioria das colocações deste tipo encontradas no CCSI (47 no total) em comparação à estrutura anterior, em que *evidence* atua como sujeito dos verbos colocados (16 casos). Entre estas colocações, é possível identificar três formações diferentes, a saber: verbo + determinante + substantivo (*to reexamine an evidence, to flush an evidence*, por exemplo); verbo + preposição + substantivo (*based on the evidence, to tamper with evidence*); e verbo + substantivo (*to get evidence, to collect evidence, to document evidence*). Acredita-se que o fato de haver um maior número de colocações especializadas verbais nas quais a base é o objeto do verbo colocado se deva ao fato de termos optado por trabalhar com um substantivo que está sempre relacionado a objetos, desempenhando poucas vezes a função de sujeito em uma frase.

Ao finalizar a análise das colocações obtidas por meio do levantamento feito a partir do termo *evidence*, ressaltamos que, embora este termo tenha se mostrado bastante produtivo, ele não possibilitou encontrarmos nenhuma colocação adverbial, devido ao fato de ser um substantivo e não ocorrer com um advérbio. Dessa forma, optamos por apresentar as colocações adverbiais a partir de outros núdulos de busca, a saber, *fire, kill* e *shoot*, no intuito de demonstrar sua estrutura morfossintática, conforme pode ser verificado no subitem a seguir.

4.2.4 Colocações especializadas adverbiais

Conforme mencionado, optou-se por apresentar as colocações adverbiais obtidas a partir dos núdulos *fire, kill* e *shoot*. Estes núdulos foram escolhidos para a apresentação das colocações adverbiais por terem sido os poucos núdulos que possibilitaram o levantamento deste tipo de colocações no CCSI. Mesmo tendo sido selecionados três núdulos para a demonstração deste tipo de colocações, em toda a análise empreendida, foi possível constatar apenas 10 colocações adverbiais, como pode ser visto no quadro a seguir.

Convém ressaltar, também, que em nossas buscas não identificamos a ocorrência de colocações adverbiais formadas pela estrutura advérbio + adjetivo. Acreditamos que isto se deva ao fato de termos analisado apenas verbos (*fire, kill* e *shoot*), uma vez que o único adjetivo presente entre as palavras-chave previamente selecionadas (*dead*) não nos permitiu encontrarmos nenhuma colocação constituída por essa estrutura.

Quadro 7 – Estrutura das colocações especializadas adverbiais

Colocação	Estrutura
<i>Fire straight</i>	verbo base + adv. colocado
<i>Accidentally kill</i>	adv. colocado + verbo base
<i>Ironically kill</i>	adv. colocado + verbo base
<i>Probably kill</i>	adv. colocado + verbo base
<i>Kill efficiently</i>	verbo base + adv. colocado
<i>Easy kill</i>	adv. colocado + verbo base
<i>Intentionally kill</i>	adv. colocado + verbo base
<i>Ultimately kill</i>	adv. colocado + verbo base
<i>Successfully kill</i>	adv. colocado + verbo base
<i>Accidentally shoot</i>	adv. colocado + verbo base

Conforme é possível perceber, o nóculo *fire* apresentou apenas 1 colocação especializada adverbial, sendo que esta apresenta a estrutura verbo + advérbio (*fire straight*).

Já o nóculo *kill* é o que mais apresentou colocações deste tipo, 8 no total. Deste total, apenas a colocação *kill efficiently* é formada por verbo + advérbio; as demais colocações são compostas por advérbio + verbo (*accidentally kill*, *ironically kill*, *probably kill*, *easy kill*, *intentionally kill* e *ultimately kill*). Por sua vez, o nóculo *shoot* apresentou, também, apenas 1 colocação adverbial, sendo formada por um advérbio + verbo (*accidentally shoot*).

Dessa forma, após finalizar esta análise, verifica-se que este tipo de colocação especializada mostrou-se pouco frequente entre as colocações especializadas da área investigativa criminal obtidas a partir das palavras-chave selecionadas do CCSI – apenas 10 colocações. A seguir, é apresentada a análise morfossintática das colocações obtidas a partir do CCC

4.3 Análise morfossintática das colocações a partir do Corpus Comparável Criminal

Neste item, discorre-se sobre a análise morfossintática das colocações especializadas obtidas a partir do levantamento realizado a partir do nóculo *evidence* no CCC. Esta análise se deu a partir da mesma base teórica (HAUMANN, 1984; ORENHA-OTTAIANO, 2009), bem como da mesma metodologia empregada para a análise das colocações obtidas no CCSI, a qual foi descrita no capítulo 3.

4.3.1 As colocações especializadas nominais

Foi possível identificar um total de 43 colocações especializadas nominais compostas pelo termo *evidence* a partir do levantamento realizado no CCC, as quais são apresentadas no

quadro 8, juntamente com a análise de sua estrutura morfossintática. Tal como fizemos com as colocações obtidas no CCSI, apresentamos dois quadros, um no qual *evidence* atua como base das colocações, e outro no qual atua como colocado. Aqui, assim como ocorreu com o CCSI, *evidence* ora atua como base, ora como colocado pelo fato de terem sido agrupadas todas as colocações que contém em sua composição o termo que está sendo analisado.

Quadro 8 – Estrutura das colocações especializadas nominais: *Evidence* como base

Colocação	Estrutura
<i>Type of evidence</i>	subst. colocado + prep. + subst. base
<i>Body of evidence</i>	subst. colocado + prep. + subst. base
<i>Rules of evidence</i>	subst. colocado + prep. + subst. base
<i>Gun evidence</i>	subst. colocado + subst. base
<i>Item of evidence</i>	subst. colocado + prep. + subst. base
<i>Porn evidence</i>	subst. colocado + subst. base
<i>Evidence of a crime</i>	subst. base + prep. + det. + subst. colocado
<i>Collection of evidence</i>	subst. colocado + prep. + subst. base
<i>Preservation of evidence</i>	subst. colocado + prep. + subst. base
<i>Ballistics evidence</i>	subst. colocado + prep. + subst. base
<i>Residue evidence</i>	subst. colocado + subst. base
<i>Destruction of evidence</i>	subst. colocado + prep. + subst. base
<i>Analysis of evidence</i>	subst. colocado + prep. + subst. base
<i>Form of evidence</i>	subst. colocado + prep. + subst. base
<i>Evidence in court</i>	subst. base + prep. + subst. colocado
<i>Preponderance of evidence</i>	subst. colocado + prep. + subst. base
<i>Presence of evidence</i>	subst. colocado + prep. + subst. base
<i>Evidence from crime</i>	subst. base + prep. + subst. colocado
<i>Source of evidence</i>	subst. colocado + prep. + subst. base
<i>Identification evidence</i>	subst. colocado + subst. base
<i>Chain of evidence</i>	subst. base + prep. + subst. colocado
<i>Assault evidence</i>	subst. colocado + subst. base
<i>Handling of evidence</i>	subst. colocado + prep. + subst. colocado
<i>Fingerprint evidence</i>	subst. colocado + subst. base
<i>Firearm evidence</i>	subst. colocado + subst. base
<i>Category of evidence</i>	subst. colocado + prep. + subst. base
<i>Amount of evidence</i>	subst. colocado + prep. + subst. Base
<i>Laboratory evidence</i>	subst. colocado + subst. Base
<i>Eyewitness evidence</i>	subst. colocado + subst. Base
<i>Autopsy evidence</i>	subst. colocado + subst. Base
<i>Scientific evidence</i>	subst. colocado + subst. Base
<i>Examination of evidence</i>	subst. colocado + prep. + subst. base
<i>Property evidence</i>	subst. colocado + subst. Base
<i>Evidence of rape</i>	subst. base + prep. + subst. colocado
<i>Evidence of suicide</i>	subst. base + prep. + subst. colocado

Quadro 9 – Estrutura das colocações especializadas nominais: *Evidence* como colocado

Colocação	Estrutura
<i>Evidence analysis</i>	subst. colocado + subst. base
<i>Evidence unit</i>	subst. colocado + subst. base
<i>Evidence expert</i>	subst. colocado + subst. base
<i>Evidence code</i>	subst. colocado + subst. base
<i>Evidence photography</i>	subst. colocado + subst. base
<i>Evidence section</i>	subst. colocado + subst. base
<i>Evidence laboratory</i>	subst. colocado + subst. base
<i>Value of evidence</i>	subst. base + prep. + subst. colocado

Diferentemente do CCSI, que apresentou uma maior quantidade de colocações especializadas verbais (70 colocações no total) o CCC possui uma maior quantidade de colocações especializadas nominais - 43 no total - em contraste a 36 colocações nominais no CCSI. No entanto, tal como no CCSI, as colocações nominais encontradas a partir do nódulo *evidence* no CCC apresentam uma variação da função deste termo: ora ele ocorre como base das colocações, ora como colocado.

As colocações que tem o termo *evidence* como base ocorrem em maior quantidade – 35 colocações no total. Entre estas colocações, é possível identificar as estruturas substantivo + preposição + substantivo (*type of evidence*, *rules of evidence*, *body of evidence*, por exemplo) – 22 ocorrências desse tipo - e substantivo + substantivo (*gun evidence*, *porn evidence*, *ballistics evidence*) – 14 colocações com esta estrutura. Neste último caso, os substantivos que ocorrem como colocado de *evidence*, acabam desempenhando uma função adjetiva, tal como ocorre com as colocações encontradas no CCSI.

Por sua vez, 7 colocações, das 8 que apresentam *evidence* como colocado, têm sua estrutura formada por substantivo + substantivo (*evidence analysis*, *evidence unit*, *evidence expert*, *evidence code*, *evidence photography*, *evidence section* e *evidence laboratory*) sendo que, nestes casos, *evidence* desempenha a função adjetiva. A única colocação que foge desta estrutura é *value of evidence*, constituída por substantivo + preposição + substantivo. Contudo, ainda neste caso, *evidence* atua como colocado, embora não desempenhe uma função adjetiva.

4.3.2 As colocações especializadas adjetivas

No que diz respeito às colocações adjetivas, em todos os casos encontrados prevalece a formação adjetivo + substantivo, sendo *evidence* o substantivo base destas colocações, tal como pode ser verificado no quadro a seguir.

Quadro 10 – Estrutura das colocações especializadas adjetivas

Colocação	Estrutura
<i>Direct evidence</i>	adjetivo colocado + subst. base
<i>Digital evidence</i>	adjetivo colocado + subst. base
<i>The best evidence</i>	adjetivo colocado + subst. base
<i>Potential evidence</i>	adjetivo colocado + subst. base
<i>Convincing evidence</i>	adjetivo colocado + subst. base
<i>Available evidence</i>	adjetivo colocado + subst. base
<i>Criminal evidence</i>	adjetivo colocado + subst. base
<i>Real evidence</i>	adjetivo colocado + subst. base
<i>Strong evidence</i>	adjetivo colocado + subst. base
<i>Medical evidence</i>	adjetivo colocado + subst. base
<i>Possible evidence</i>	adjetivo colocado + subst. base

Como se pode perceber pelas informações que compõem o quadro, foram encontradas apenas 11 colocações especializadas adjetivas entre as colocações levantadas no CCC, sendo que nenhuma destas colocações pode ser encontrada no CCSI, fato que corrobora a relevância de contar com um corpus comparável para o desenvolvimento da pesquisa, uma vez que este, além de permitir a comprovação da frequência das colocações especializadas obtidas a partir do CCSI, permite que sejam consideradas nas análises colocações que também são utilizadas na área investigativa criminal e que não são usadas na série que compõe o corpus de estudo desta pesquisa.

4.3.3 As colocações especializadas verbais

Entre as colocações especializadas verbais levantadas é possível perceber a presença de colocações formadas por verbo + substantivo; substantivo + verbo; e verbo + preposição + substantivo. Além disso, identificamos apenas 10 colocações verbais no CCC, quantidade inferior à encontrada no CCSI que possibilitou a identificação de 67 colocações especializadas verbais. Destas 67 colocações verbais, 4 são constituídas por verbos frasais, um dos tipos de colocação verbal que não foi encontrado aqui. Além disso, foram identificadas 16 colocações verbais no CCSI nas quais *evidence* atua como sujeito, sendo que no caso do CCC este tipo de colocação não foi identificado. A seguir, apresentamos o quadro com as colocações verbais obtidas.

Quadro 11 – Estrutura das colocações especializadas verbais

Colocação	Estrutura
<i>To suppress evidence</i>	verbo colocado + subst. base
<i>To search for evidence</i>	verbo colocado + prep. + subst. base
<i>To give evidence</i>	verbo colocado + subst. base
<i>To admit into evidence</i>	verbo colocado + prep. + subst. base
<i>To obtain evidence</i>	verbo colocado + subst. base
<i>To present evidence</i>	verbo colocado + subst. base
<i>To produce evidence</i>	verbo colocado + subst. base
<i>To introduce evidence</i>	verbo colocado + subst. base
<i>To examine evidence</i>	verbo colocado + subst. base
<i>To provide evidence</i>	verbo colocado + subst. base

Como pode ser observado pelo quadro acima, as colocações verbais encontradas neste corpus apresentam a base *evidence* como objeto dos verbos colocados, estrutura que foi observada em 47 das colocações verbais encontradas no CCSI. Outro dado importante é que 2 entre as 10 colocações apresentadas no quadro acima apresentam uma preposição entre o verbo e seu objeto. Essa estrutura ocorreu em apenas 3 das 47 colocações verbais nas quais *evidence* atua como objeto que foram levantadas a partir do CCSI, o que mostra que essa estrutura é pouco utilizada entre as colocações verbais levantadas em ambos os corpora.

A seguir, tal como feito com as colocações adverbiais encontradas no CCSI, é apresentada a análise morfossintática das colocações especializadas adverbiais dos núdulos *fire*, *kill* e *shoot* obtidas a partir do CCC.

4.3.4 Colocações especializadas adverbiais

Para o levantamento das colocações adverbiais que ilustram esta análise, tomamos como ponto de partida os mesmos núdulos de busca, ou palavras-chave, que foram utilizados para o levantamento das colocações deste tipo no CCSI, ou seja, *fire*, *kill* e *shoot*. A seguir, são apresentadas as colocações adverbiais encontradas a partir destes núdulos, bem como sua estrutura morfossintática.

Quadro 12 – Estrutura das colocações especializadas adverbiais

Colocação	Estrutura
<i>Allegedly fire</i>	advérbio colocado + verbo base
<i>Later fire</i>	advérbio colocado + verbo base
<i>Previously fire</i>	advérbio colocado + verbo base
<i>Recently fire</i>	advérbio colocado + verbo base
<i>Reportedly fire</i>	advérbio colocado + verbo base
<i>Secretly fire</i>	advérbio colocado + verbo base
<i>Suddenly fire</i>	advérbio colocado + verbo base
<i>Accidentally kill</i>	advérbio colocado + verbo base
<i>Definitively kill</i>	advérbio colocado + verbo base
<i>Eventually kill</i>	advérbio colocado + verbo base
<i>Secretly kill</i>	advérbio colocado + verbo base
<i>Unintentionally kill</i>	advérbio colocado + verbo base
<i>Unlawfully kill</i>	advérbio colocado + verbo base
<i>Unnecessarily kill</i>	advérbio colocado + verbo base
<i>Allegedly kill</i>	advérbio colocado + verbo base
<i>Apparently kill</i>	advérbio colocado + verbo base
<i>Brutally kill</i>	advérbio colocado + verbo base
<i>Deliberately kill</i>	advérbio colocado + verbo base
<i>Intentionally kill</i>	advérbio colocado + verbo base
<i>Maliciously kill</i>	advérbio colocado + verbo base
<i>Mysteriously kill</i>	advérbio colocado + verbo base
<i>Successfully kill</i>	advérbio colocado + verbo base
<i>Tragically kill</i>	advérbio colocado + verbo base
<i>Ultimately kill</i>	advérbio colocado + verbo base
<i>Accidentally shoot</i>	advérbio colocado + verbo base
<i>Fatally shoot</i>	advérbio colocado + verbo base
<i>Secretly shoot</i>	advérbio colocado + verbo base
<i>Rarely shoot</i>	advérbio colocado + verbo base
<i>To shoot effectively</i>	verbo base + advérbio colocado

Assim como aconteceu com os demais tipos de colocações especializadas encontradas, foi possível encontrar uma maior quantidade de colocações adverbiais a partir destes três núdulos (*fire*, *kill* e *shoot*) no CCC – 27 no total. A partir do núdulo *fire*, foi possível identificar 5 colocações especializadas do tipo adverbial, prevalecendo em todas as colocações a estrutura advérbio + verbo, conforme pode se perceber pelos exemplos *allegedly fire*, *previously fire* e *recently fire*.

Já o núdulo *kill* foi o que mais retornou colocações adverbiais – 17 no total. Neste núdulo, assim como o núdulo *fire*, prevalece a estrutura advérbio + verbo (*accidentally kill*, *allegedly kill*, *brutally kill*). Por sua vez, o núdulo *shoot* apresentou 5 colocações especializadas do tipo adverbiais, sendo que uma destas 5 apresentou a estrutura verbo + advérbio (*to shoot effectively*). As outras 4 colocações apresentam a estrutura advérbio + verbo: *accidentally shoot*, *fatally shoot*, *secretly shoot* e *rarely shoot*.

Embora tenha sido possível identificar 27 colocações especializadas adverbiais no CCC, percebe-se que este tipo de colocações, ocorre muito pouco neste corpus de estudo, haja vista que, para chegarmos a estas 27 colocações, foi necessário somar as colocações obtidas a partir de 3 diferentes núdulos. Ressaltamos, porém, que assim como ocorreu com o CCSI, as colocações especializadas adverbiais do CCC foram levantadas a partir das palavras-chave previamente selecionadas e apresentadas no item 4.1, de forma que, tal como ocorreu com o CCSI, não identificamos nenhuma colocação adverbial constituída pela estrutura adjetivo + advérbio devido ao fato de termos analisado apenas verbos. Assim, pode-se dizer que este tipo de colocação não é muito comum nos corpora de estudo

Tendo realizado a análise morfossintática das colocações especializadas encontradas em ambos os corpora neste item, apresentamos, no item seguinte, uma análise contrastiva das colocações levantadas em ambos os corpora, a fim de evidenciar a importância dos corpora, assim como dos dados fornecidos pelo *WebCorp*.

4.4 Análise contrastiva das colocações especializadas nos corpora e no *WebCorp*

Neste item, é apresentada uma análise contrastiva das frequências das colocações levantadas, considerando os dados obtidos em ambos os corpora de pesquisa, bem como os dados obtidos por meio das consultas ao *WebCorp*. Conforme mencionado anteriormente, foi compilado um corpus comparável, denominado Corpus Comparável Criminal (CCC), com o intuito de ser uma fonte de comprovação de frequência de coocorrência das colocações especializadas obtidas por meio da análise do Corpus CSI (CCSI), o qual, como já mencionado, é composto pelas legendas de 12 temporadas da série *CSI – Crime Scene Investigation*.

Dessa forma, após o levantamento das colocações no CCSI, buscou-se a verificação da presença destas mesmas colocações no CCC, de forma a comprovar seu uso. Nesse processo, identificou-se uma variação de frequência de coocorrência das colocações, sendo que algumas foram identificadas apenas em um dos corpora de estudo, ou apresentaram uma baixa frequência. Este resultados nos levaram a utilizar o emulador de concordanciador *WebCorp*, a fim de comprovar a frequência de coocorrência das colocações extraídas nestes corpora.

Para melhor organizar este item, foi feita uma divisão, trabalhando com as colocações nominais, verbais e adjetivas obtidas a partir do núdulo *evidence*, em ambos os corpora, de forma separada. No que diz respeito às colocações adverbiais, partimos dos núdulos *fire*, *kill* e

shoot, apresentando as colocações encontradas a partir de sua análise, conforme já mencionado anteriormente.

Assim, iniciamos esta seção apresentando a frequência das colocações nominais, nas quais *evidence* atua como base, no CCSI, no CCC e no WebCorp.

Tabela 4 – Frequência das colocações nominais: *Evidence* como base

Colocação	Freq. CCSI	Freq. CCC	Freq. WebCorp
<i>A murder evidence</i>	1	1	92
<i>An homicide evidence</i>	1	2	58
<i>A blood evidence</i>	29	34	196
<i>An ammo evidence</i>	3	-	15
<i>A case evidence</i>	4	10	35
<i>An evidence of assault</i>	1	-	41
<i>A DNA evidence</i>	5	380	360
<i>An evidence out of context</i>	1	1	57
<i>An evidence in context</i>	1	-	86
<i>An evidence of a struggle</i>	2	9	72
<i>An evidence of the crime</i>	1	7	38
<i>An interpreter of the evidence</i>	1	-	9
<i>A hair evidence</i>	1	6	110
<i>A fiber evidence</i>	1	8	116
<i>An evidence of trauma</i>	1	-	53
<i>A body evidence</i>	1	2	119
<i>Type of evidence</i>	-	45	26
<i>Body of evidence</i>	-	60	99
<i>Rules of evidence</i>	-	55	96
<i>Gun evidence</i>	-	32	31
<i>Item of evidence</i>	-	15	101
<i>Porn evidence</i>	-	31	12
<i>Evidence of a crime</i>	-	28	24
<i>Collection of evidence</i>	-	17	25
<i>Preservation of evidence</i>	-	22	15
<i>Ballistics evidence</i>	1	89	80
<i>Residue evidence</i>	-	23	30
<i>Destruction of evidence</i>	-	23	83
<i>Analysis of evidence</i>	-	18	30
<i>Form of evidence</i>	-	20	19
<i>Evidence in court</i>	-	20	22
<i>Preponderance of evidence</i>	-	16	208
<i>Presence of evidence</i>	-	1	5
<i>Evidence from crime</i>	-	16	27
<i>Source of evidence</i>	-	14	28
<i>Identification evidence</i>	-	66	226
<i>Chain of evidence</i>	-	11	208
<i>Assault evidence</i>	-	13	27
<i>Handling of evidence</i>	-	8	19
<i>Fingerprint evidence</i>	-	36	137

<i>Firearm evidence</i>	-	15	19
<i>Category of evidence</i>	-	1	8
<i>Amount of evidence</i>	-	7	20
<i>Laboratory evidence</i>	-	12	21
<i>Eyewitness evidence</i>	-	46	74
<i>Autopsy evidence</i>	-	2	26
<i>Scientific evidence</i>	-	46	217
<i>Examination of evidence</i>	-	6	4
<i>Property evidence</i>	-	13	32
<i>Evidence of rape</i>	-	6	8
<i>Evidence of suicide</i>	-	2	13

Conforme é possível notar, a maioria das colocações encontradas no CCSI ocorre apenas uma vez neste corpus, contudo, foram mantidas na análise devido à sua relevância para a área investigativa criminal atestada pelas frequências de coocorrência identificadas no CCC e / ou no *WebCorp*. Além disso, como a busca pelas colocações foi realizada em ambos os corpora a partir da palavra-chave *evidence* algumas colocações foram identificadas apenas no CCC, tendo sido mantidas em nossa análise devido a frequência de coocorrência encontrada no *WebCorp*.

Dessa forma, destacamos as colocações *a blood evidence* que apresenta alta frequência em ambos os corpora e no *WebCorp* (29, 34 e 196, respectivamente) e *a DNA evidence*, por ser a colocação especializada que mais ocorre no CCC (380 vezes) e por ter uma alta frequência também no *WebCorp* (360), o que demonstra que se trata de uma colocação bastante utilizada no meio investigativo criminal, mesmo que não seja tão utilizada no ambiente da série *CSI – Crime Scene Investigation*. Já as colocações *identification evidence*, *chain of evidence* e *scientific evidence* não foram encontradas no CCSI, mas apresentam frequência alta no CCC e no *WebCorp*.

De modo geral, é possível perceber que o CCC possibilitou o levantamento de maior número de colocações no qual foi possível comprovar a frequência de ocorrência de várias colocações extraídas do CCSI cuja frequência era de apenas um. Além disso, há apenas cinco colocações que não foram encontradas neste corpus, mas que tiveram sua frequência de uso atestada pelo *WebCorp*, a saber, *an ammo evidence*, *an evidence of assault*, *an evidence in context*, *an interpreter of the evidence* e *an evidence of trauma*.

A tabela a seguir contém a frequência das colocações nominais nas quais *evidence* atua como colocado.

Tabela 5 – Frequência das colocações nominais: *Evidence* como colocado

Colocação	Freq. CCSI	Freq. CCC	Freq. WebCorp
<i>An evidence examination</i>	1	9	50
<i>An evidence bag</i>	5	38	289
<i>An evidence ledger</i>	2	-	14
<i>Evidence Trinity</i>	1	-	9
<i>An evidence log</i>	1	5	-
<i>An evidence box</i>	3	-	43
<i>An evidence locker</i>	1	2	-
<i>An evidence room</i>	3	13	308
<i>An absence of evidence</i>	2	3	363
<i>An evidence custodian</i>	2	-	125
<i>An evidence marker</i>	1	3	117
<i>A drug evidence</i>	2	24	219
<i>An evidence vault</i>	11	2	80
<i>An evidence bullet</i>	1	13	45
<i>An evidence Seal</i>	2	2	43
<i>An evidence recovery</i>	1	11	157
<i>A lack of evidence</i>	1	23	135
<i>A stitch of evidence</i>	1	1	62
<i>An evidence collection</i>	3	121	301
<i>An evidence area</i>	1	4	35
<i>Evidence analysis</i>	-	30	84
<i>Evidence unit</i>	-	14	95
<i>Evidence expert</i>	-	4	12
<i>Evidence code</i>	-	18	99
<i>Evidence photography</i>	-	10	47
<i>Evidence section</i>	-	14	31
<i>Evidence laboratory</i>	-	4	14
<i>Value of evidence</i>	-	4	12

Tal como ocorre com as colocações nominais apresentadas anteriormente, as colocações encontradas a partir do CCSI também apresentam menor frequência em relação às colocações encontradas no CCC. No entanto, a frequência de coocorrência das colocações especializadas aqui apresentadas é atestada pelos dados obtidos por meio da busca feita pelo WebCorp, como é o caso de *an evidence bag*, *an evidence room*, *a drug evidence*, *a lack of evidence* e *an evidence collection*, por exemplo.

Entre as 20 colocações especializadas encontradas no CCSI, apenas 4 não foram encontradas no CCC, a saber, *an evidence ledger*, *evidence trinity*, *an evidence box* e *an evidence custodian*. Contudo, embora não tenha sido possível encontrá-las no CCC, os dados obtidos por meio do WebCorp comprovam a ocorrência destas colocações: *an evidence ledger* (14 ocorrências), *an evidence trinity* (9 ocorrências), *an evidence box* (43 ocorrências) e *an evidence custodian* (125 ocorrências).

Outro dado que merece destaque são as duas colocações que não foram encontradas no *WebCorp*, *an evidence log* e *an evidence locker*. Convém dizer que estas colocações foram inseridas na análise em razão de termos utilizado como regra de corte o fato de a colocação apresentar frequência superior a 1 em pelo menos um dos corpora, ou no *WebCorp*. Dessa forma, como estas colocações ocorrem 5 e 2 vezes, respectivamente, no CCC, foram mantidas na análise, mesmo não tendo sido possível a verificação de sua ocorrência no *WebCorp*.

Já, com relação às colocações verbais nas quais *evidence* atua como sujeito, obteve-se os seguintes dados:

Tabela 6 – Frequência das colocações verbais: *Evidence* como sujeito

Colocação	Freq. CCSI	Freq. CCC	Freq. <i>WebCorp</i>
<i>Evidence changes</i>	3	-	54
<i>The evidence suggests</i>	3	26	-
<i>The evidence says</i>	14	4	26
<i>Evidence points to</i>	9	16	21
<i>Evidence appears</i>	1	5	28
<i>Evidence collected</i>	2	75	20
<i>Evidence clears</i>	1	8	79
<i>Evidence supports</i>	2	12	77
<i>Evidence tells</i>	9	3	57
<i>Evidence persists</i>	1	-	42
<i>Evidence fails</i>	1	1	50
<i>The evidence proves</i>	1	7	120
<i>Evidence leads</i>	1	12	79
<i>Evidence shows</i>	3	44	119
<i>Evidence connects</i>	2	11	111
<i>The evidence indicates</i>	1	6	81

Como se pode notar, apenas duas das colocações apresentadas na tabela acima não foram encontradas no CCC – *evidence changes* e *evidence persists*. No entanto, foram mantidas devido à sua frequência maior no *WebCorp* (respectivamente 54 e 42 ocorrências).

Já a colocação especializada *the evidence suggests*, que ocorre 3 vezes no CCSI e 26 vezes no CCC, não foi encontrada no *WebCorp*. Tal como explicitado anteriormente, ela foi mantida na análise, considerando que sua frequência em ambos os corpora foi superior a 1. Em contrapartida, as duas colocações que ocorrem com maior frequência no *WebCorp*, a saber, *the evidence proves* e *the evidence shows*, apresentam baixa frequência no CCSI (1 e 3, respectivamente) foram analisadas, devido a alta frequência com que ocorrem no *WebCorp* e, também, no CCC.

Além disso, ressaltamos que, conforme dissemos no item 4.3.3, não foi possível identificar colocações verbais nas quais *evidence* atua como sujeito que ocorrem apenas no CCC. Este corpus (CCC) possibilitou a identificação de apenas 10 colocações verbais para o nóculo *evidence*, nas quais a base atua como objeto dos verbos colocados. Dessa forma, no que diz respeito às colocações especializadas verbais nas quais *evidence* atua como sujeito dos verbos, o CCC permitiu que comprovássemos a frequência de coocorrência das colocações encontradas no CCSI, mas não nos possibilitou a identificação de nenhuma outra colocação deste tipo além das apresentadas na tabela 6.

Na tabela seguinte, apresentamos a frequência das colocações verbais que possuem *evidence* como objeto dos verbos colocados.

Tabela 7 – Frequência das colocações verbais: *Evidence* como objeto

Colocação	Freq. CCSI	Freq. CCC	Freq. WebCorp
<i>To find evidence</i>	2	11	85
<i>To reexamine an evidence</i>	2	-	47
<i>To flush an evidence</i>	1	-	43
<i>Based on the evidence</i>	3	7	52
<i>To get evidence</i>	2	3	69
<i>To ditch an evidence</i>	1	-	5
<i>To collect evidence</i>	4	57	96
<i>To follow an evidence</i>	5	-	46
<i>To trace evidence</i>	12	187	366
<i>To destroy evidence</i>	2	15	114
<i>To plant an evidence</i>	1	-	64
<i>To document evidence</i>	1	3	57
<i>To obliterate evidence</i>	1	-	8
<i>To fabricate evidence</i>	3	-	114
<i>To process an evidence</i>	3	-	41
<i>To record evidence</i>	1	-	53
<i>To analyze an evidence</i>	1	-	83
<i>To exclude an evidence</i>	1	-	38
<i>To burn an evidence</i>	1	-	57
<i>To manipulate evidence</i>	1	-	50
<i>To recover evidence</i>	1	2	49
<i>To go into evidence</i>	2	1	32
<i>To hide an evidence</i>	3	-	16
<i>To chase an evidence</i>	1	-	16
<i>To mishandle an evidence</i>	1	-	28
<i>To compromise evidence</i>	2	1	48
<i>To run evidence</i>	1	-	35
<i>To test an evidence</i>	1	-	44
<i>To gather an evidence</i>	1	-	59
<i>To check an evidence</i>	1	-	50
<i>To preserve an evidence</i>	1	-	42
<i>To categorize evidence</i>	1	-	33
<i>To interpreter an evidence</i>	1	-	32
<i>To review an evidence</i>	1	-	129
<i>To face an evidence</i>	1	-	59
<i>To evaluate evidence</i>	1	-	64
<i>To assume evidence</i>	1	-	29
<i>To melt evidence</i>	1	-	32
<i>To attack an evidence</i>	3	-	49
<i>To disappear an evidence</i>	1	-	43
<i>To book an evidence</i>	1	-	59
<i>To reanalyze an evidence</i>	1	-	8
<i>To compel evidence</i>	2	-	167
<i>To steal evidence</i>	2	-	62
<i>To jeopardize an evidence</i>	1	-	24
<i>To retrieve an evidence</i>	1	-	48

<i>To tamper with evidence</i>	2	1	236
<i>To suppress evidence</i>	-	28	45
<i>To search for evidence</i>	-	14	31
<i>To give evidence</i>	-	14	153
<i>To admit into evidence</i>	-	1	25
<i>To present evidence</i>	-	22	59
<i>To produce evidence</i>	-	12	53
<i>To introduce evidence</i>	-	10	23
<i>To examine evidence</i>	-	8	26
<i>To provide evidence</i>	-	17	53

Entre as colocações especializadas verbais em que *evidence* atua como objeto, chama a atenção o fato de grande parte das colocações encontradas no CCSI não ter sido encontrada no CCC. Como se pode observar, apenas 12 das 48 colocações levantadas a partir do CCSI, ocorrem no CCC. Contudo, a busca feita pelo *WebCorp* mostrou que todas as 48 colocações são frequentes, mesmo não tendo sido encontradas em nosso corpus comparável.

Entre todas as colocações apresentadas na tabela 9, a colocação *to trace evidence* é a que apresenta a maior frequência em cada um dos corpora e no *WebCorp*, sendo, assim, a colocação deste tipo mais utilizada para este nóculo de busca.

Já as colocações *to tamper with evidence*, *to compel evidence*, *to give evidence*, *to review evidence*, *to fabricate evidence* e *to destroy evidence* foram as colocações que se mostraram mais frequentes no *WebCorp*.

Assim, por meio dos dados apresentados nesta tabela, fica evidente a importância do uso do emulador de concordanciador *WebCorp*, uma vez que grande parte dos dados obtidos nos corpora só foram possíveis de ser comprovados graças a ele, conferindo, assim, autenticidade aos dados investigados. Além disso, é possível perceber que o CCC, mais uma vez, apresenta uma maior frequência das colocações que foram obtidas por meio de sua análise, do que a frequência das colocações que foram encontradas apenas no CCSI.

Na última tabela que apresenta os dados concernentes às colocações verbais, trazemos as colocações constituídas por verbos frasais.

Tabela 8 – Frequência das colocações verbais formadas por verbos frasais

Colocação	Freq. CCSI	Freq. CCC	Freq. <i>WebCorp</i>
<i>To sift through evidence</i>	2	-	60
<i>To look for evidence</i>	2	4	73
<i>To look at evidence</i>	2	-	59
<i>To turn into evidence</i>	1	-	48

Tal como no caso anterior, no que diz respeito às colocações verbais constituídas por verbos frasais o *WebCorp* também se mostrou de extrema importância para a verificação da frequência de coocorrência destas colocações, haja vista que estas foram encontradas com frequências baixas no CCSI e que apenas uma delas, *to look for evidence*, foi encontrada no CCC.

A seguir, são apresentados os dados concernentes às colocações adjetivas, obtidas em ambos os corpora de pesquisa e no *WebCorp*.

Tabela 9 – Frequência das colocações adjetivas

Colocação	Freq. CCSI	Freq. CCC	Freq. WebCorp
<i>An outstanding evidence</i>	1	1	52
<i>A physical evidence</i>	49	409	277
<i>An old evidence</i>	6	3	157
<i>A new evidence</i>	19	74	158
<i>An incriminating evidence</i>	3	12	188
<i>A corresponding evidence</i>	2	-	29
<i>A transitory evidence</i>	2	-	52
<i>A hard evidence</i>	5	7	-
<i>A biological evidence</i>	1	65	252
<i>An extraneous evidence</i>	1	-	45
<i>An associated evidence</i>	1	1	30
<i>A latest evidence</i>	1	-	45
<i>An insufficient evidence</i>	3	33	151
<i>An unexplained evidence</i>	2	-	50
<i>A sufficient evidence</i>	2	31	135
<i>A recent evidence</i>	1	4	103
<i>A plastic evidence bag</i>	1	-	50
<i>A forensic evidence</i>	9	250	285
<i>A good evidence</i>	4	5	131
<i>A circumstantial evidence</i>	3	130	668
<i>A wrong evidence</i>	1	-	75
<i>An invisible evidence</i>	2	-	66
<i>An additional evidence</i>	1	23	137
<i>An unlogged evidence</i>	1	-	3
<i>A reliable evidence</i>	1	17	147
<i>An instant evidence</i>	1	-	52
<i>An empirical evidence</i>	1	5	170
<i>A disturbing evidence</i>	1	1	79
<i>A clear evidence</i>	1	8	172
<i>An equivocal evidence</i>	1	-	68
<i>A latent evidence</i>	1	2	156
<i>An important evidence</i>	1	15	73
<i>An individualized evidence</i>	1	-	20
<i>A surgical evidence</i>	1	-	38
<i>Direct evidence</i>	-	60	68
<i>Digital evidence</i>	-	57	40
<i>The best evidence</i>	-	12	51
<i>Potential evidence</i>	-	34	52
<i>Convincing evidence</i>	-	11	64
<i>Available evidence</i>	-	13	61
<i>Criminal evidence</i>	-	17	50
<i>Real evidence</i>	-	21	56
<i>Strong evidence</i>	-	12	65
<i>Medical evidence</i>	-	13	77
<i>Possible evidence</i>	-	14	55

Entre as colocações adjetivas obtidas a partir dos corpora de pesquisa, a que apresenta maior frequência é *a physical evidence*, que ocorre 49 vezes no CCSI, 409 vezes no CCC e 277 no *WebCorp*. Esta colocação se destaca, também, por apresentar uma alta frequência de coocorrência no CCC se comparada a frequência obtida pelo emulador de concordanciador *WebCorp*.

Outra colocação que merece destaque devido à alta frequência apresentada, é *a circumstantial evidence*, que ocorre 668 vezes no emulador de concordanciador. Além disso, esta colocação também apresenta uma frequência expressiva no CCC (130), enquanto ocorre apenas 3 vezes no CCSI.

Como pode ser observado pela tabela, 13 das 34 colocações adjetivas encontradas no CCSI não foram encontradas no CCC. No entanto, a frequência que estas apresentaram no *WebCorp* justifica a sua inclusão na análise. Já a colocação *a hard evidence* não apresentou ocorrência no emulador de concordanciador, de forma que foi incluída na análise devido às frequências que apresentou em ambos os corpora de pesquisa, a saber, 5 no CCSI e 7 no CCC.

Por fim, são apresentadas as frequências das colocações adverbiais obtidas em ambos os corpora de estudo a partir dos núdulos *fire*, *shoot* e *kill*.

Tabela 10 – Frequência das colocações adverbiais

Frequência das colocações adverbiais			
Colocação	Freq. CCSI	Freq. CCC	Freq. <i>WebCorp</i>
<i>Fire straight</i>	1	2	2
<i>Allegedly fire</i>	-	7	55
<i>Previously fire</i>	-	1	12
<i>Recently fire</i>	-	7	20
<i>Reportedly fire</i>	-	3	28
<i>Secretly fire</i>	-	1	15
<i>Suddenly fire</i>	-	1	24
<i>Ironically kill</i>	2	-	19
<i>Probably kill</i>	7	1	93
<i>Easy kill</i>	1	-	153
<i>Ultimately kill</i>	1	3	33
<i>Accidentally kill</i>	2	1	93
<i>Definitively kill</i>	-	1	32
<i>Eventually kill</i>	-	5	78
<i>Kill efficiently</i>	1	-	25
<i>Secretly kill</i>	-	1	70
<i>Unintentionally kill</i>	-	1	44
<i>Unlawfully kill</i>	-	1	38
<i>Unnecessarily kill</i>	-	1	33
<i>Allegedly kill</i>	-	23	44
<i>Apparently kill</i>	-	2	33
<i>Brutally kill</i>	-	5	24
<i>Deliberately kill</i>	-	2	51
<i>Intentionally kill</i>	1	3	47
<i>Maliciously kill</i>	-	1	62
<i>Mysteriously kill</i>	-	1	19
<i>Successfully kill</i>	1	1	78
<i>Tragically kill</i>	-	3	28
<i>Accidentally shoot</i>	1	1	8
<i>Fatally shoot</i>	-	7	36
<i>Secretly shoot</i>	-	1	10
<i>Rarely shoot</i>	-	7	27
<i>To shoot effectively</i>	-	6	16

Conforme mostra a tabela, mesmo as colocações que não foram encontradas em um dos corpora apresentam frequência relevante no *WebCorp*, tal como ocorre com *allegedly fire*, por exemplo, que não ocorre no CCSI, tem frequência 7 no CCC e 55 no *WebCorp*. A colocação *easy kill*, que apresenta frequência de 153 no *WebCorp*, pode ser considerada a colocação especializada adverbial mais utilizada, apesar de não ter sido encontrada no CCC e de ocorrer apenas 1 vez no CCSI, haja vista o critério adotado de que a colocação deve ocorrer com frequência superior a 1 em pelo menos um dos corpora de estudo ou no emulador de concordanciador *WebCorp*.

Foi possível notar que, ao considerar as frequências obtidas para as colocações especializadas adverbiais, sem o auxílio do referido emulador de concordanciador, não seria possível considerar na análise aqui desenvolvida muitas colocações adverbiais, uma vez que 16 das 33 colocações aqui apresentadas ocorre apenas 1 vez em apenas um dos corpora de estudo, o que faria com que tivéssemos um número ainda mais reduzido de colocações especializadas deste tipo.

De forma geral, ao olhar para os dados apresentados em todas as tabelas, percebe-se que o CCC permitiu a verificação da frequência de coocorrência da maioria das colocações. Uma das razões para este resultado pode ser pelo fato de este corpus ser um pouco maior que o CCSI (3.886.465 de palavras), enquanto que o CCSI conta com 3.061.029 de palavras. No entanto, embora o CCSI tenha possibilitado encontrar uma maior quantidade de colocações especializadas da área investigativa criminal, mesmo que grande parte destas colocações ocorra uma única vez neste corpus, foi possível comprovarmos sua frequência de coocorrência por meio do CCC e pelo emulador de concordanciador *WebCorp*, fato que vem corroborar o uso do CCSI como ponto de partida para a busca e o levantamento de colocações desta área.

Para finalizar este item, apresentamos uma tabela que contém um panorama quantitativo dos dados obtidos por meio dos corpora de estudo e a partir do *WebCorp* para o nóculo de busca *evidence*, bem como para os nósulos *fire*, *kill* e *shoot*, utilizados para o levantamento das colocações especializadas adverbiais, de modo que seja possível visualizar a quantidade de cada tipologia de colocações encontrada em cada um dos corpora durante a pesquisa e no referido emulador de concordanciador.

Tabela 11 – Resumo dos dados obtidos em ambos os corpora e no WebCorp

CORPUS CSI			CORPUS CC			WEBCORP		
148 Colocações Especializadas			167 Colocações Especializadas			229 Colocações Especializadas		
36 Colocações Especializadas Nominais			70 Colocações Especializadas Nominais			77 Colocações Especializadas		
16 <i>evidence</i> como base	20 <i>evidence</i> como colocado		46 <i>evidence</i> como base	24 <i>evidence</i> como colocado		51 <i>evidence</i> como base	26 <i>evidence</i> como colocado	
67 Colocações Especializadas Verbais			35 Colocações Especializadas Verbais			75 Colocações Especializadas Verbais		
16 <i>evidence</i> como sujeito	47 <i>evidence</i> como objeto	4 <i>phrasal verbs</i>	14 <i>evidence</i> como sujeito	20 <i>evidence</i> como objeto	1 <i>phrasal verb</i>	15 <i>evidence</i> como sujeito	56 <i>evidence</i> como objeto	4 <i>phrasal verb</i>
34 Colocações Especializadas Adjetivas			32 Colocações Especializadas Adjetivas			44 Colocações Especializadas Adjetivas		
10 Colocações Especializadas Adverbiais *			30 Colocações Especializadas Adverbiais *			33 Colocações Especializadas Adverbiais *		

* As colocações adverbiais foram levantadas a partir dos núdulos *to fire*, *to kill* e *to shoot*, conforme mencionado previamente.

A tabela acima, tal como dissemos, apresenta um resumo dos dados quantitativos obtidos em ambos os corpora de estudo e no *WebCorp*. Como se pode observar, os números obtidos a partir das buscas empreendidas no emulador de concordanciador *WebCorp* são maiores do que os números obtidos em cada corpus. Isto se deve ao fato de que todas as colocações que foram levantadas em ambos os corpora terem sido buscadas no *WebCorp*, de forma que, algumas colocações encontradas em apenas um dos corpora foram encontradas no *WebCorp* enquanto outras não, além do fato de que em alguns casos não ter sido possível encontrar a colocação no emulador de concordanciador.

Com relação aos dados obtidos por meio do *WebCorp*, gostaríamos de ressaltar que este emulador de concordanciador foi utilizado para a busca e confirmação da frequência de coocorrência das colocações especializadas obtidas anteriormente a partir de ambos os corpora de estudo, não tendo sido realizadas buscas por outras colocações especializadas no

próprio *WebCorp*, o que se deve ao fato de termos analisado a área investigativa criminal a partir dos dados fornecidos por nossos corpora de pesquisa, mais especificamente, por termos tomado como ponto de partida o CCSI, tal como mencionamos anteriormente, de forma a delimitar esta investigação para que fosse viável a sua realização.

Dessa forma, encerramos a análise morfossintática empreendida neste trabalho, bem como se tem um panorama geral dos dados explicitados durante esta análise. A seguir apresentamos a análise léxico-semântica de algumas das colocações que foram apresentadas aqui, privilegiando aquelas que mais podem causar dificuldade de compreensão e produção por parte dos alunos, devido ao fato de serem compostas por um léxico mais rebuscado.

4. 5 Análise léxico semântica das colocações especializadas

Como pode ser observado pela análise morfossintática apresentada anteriormente, uma grande parte das colocações especializadas levantadas a partir do nóculo *evidence* não apresenta grandes problemas de compreensão e produção para os aprendizes. No entanto, é possível identificar algumas que se mostram mais complexas neste sentido, conforme discutiremos nesta seção. Daremos especial atenção às colocações que mais chamaram a atenção em relação aos elementos que as compõem, pelo fato de poderem apresentar certa dificuldade de compreensão e, conseqüentemente, de utilização destas por parte dos alunos.

As colocações aqui elencadas merecem atenção devido ao seu grau de dificuldade, devendo ser ensinadas individualmente, tal como é feito ao se ensinar um novo léxico, uma vez que é melhor ensiná-las aos alunos do que esperar que sejam capazes de agrupá-las sozinhos (HILL, 2000, p. 61). Partindo desta concepção, mostramos a seguir os caminhos que um aprendiz de língua inglesa percorreria para produzir ou tentar compreender o significado destas colocações, de forma a evidenciar a importância do ensino destas por parte do professor.

Para tal, contamos com os dicionários de língua inglesa *Longman Dictionary of Contemporary English* e com o *Oxford Collocations Dictionary for Students of English* como fontes de esclarecimento das unidades lexicais que compõem as colocações.

Convém ressaltar que, tendo em vista que o nóculo *evidence* não produz colocações do tipo adverbial, tal como apresentado nos itens anteriores deste capítulo, as colocações adverbiais analisadas nesta seção também foram identificadas a partir do nóculo *fire*. Além disso, os exemplos de colocações especializadas adjetivas também foram obtidos a partir de outros núculos, tais como *crime* e *weapon*, de forma a ilustrar a variedade de colocações

especializadas levantadas a partir das palavras-chave selecionados para esta pesquisa (os quais foram apresentados no item 4.1). Ademais, apresentamos alguns casos de variação morfosintática que ocorre com algumas colocações, evidenciando a diferença ou não que esta variação traz para a compreensão e uso das colocações em destaque.

4.5.1 *To mishandle an evidence*

A colocação *to mishandle an evidence* ocorre apenas 1 vez no CCSI, 5 vezes no CCC e 28 vezes no *WebCorp*. É uma colocação verbal, constituída de um verbo colocado, um determinante e um substantivo base.

Esta colocação foi selecionada por ser considerada uma combinação que poderia apresentar maior dificuldade de compreensão para os alunos, além de ser de difícil produção, caso não tivesse sido previamente ensinada, devido ao fato de ser composta por um verbo de uso um pouco mais sofisticado do que a maioria dos verbos com os quais os alunos de língua inglesa entram em contato.

O *Longman Dictionary* define *mishandle* como “lidar mal com uma situação devido à falta de habilidade ou cuidados; tratar algo grosseiramente, causando danos”⁴². Por sua vez, o *Oxford Collocations Dictionary* não apresenta o verbo em questão como um colocado para *evidence*. Tampouco traz o verbo *mishandle* como uma provável base, evidenciando a importância de sua análise e inclusão no *Online English Collocations Workbook*.

Partindo dessas definições, o aluno poderia inferir que a colocação *mishandle an evidence* se refere a “tratar uma evidência de forma inadequada”, ou a “tratar de forma errada uma evidência”. Assim, a fim de facilitar a compreensão desta colocação sem que o aluno precisasse recorrer aos dicionários e a compreendesse, sendo ainda capaz de produzi-la em seu texto, *mishandle an evidence* foi selecionada para ser incluída em uma atividade de *Gap Fill* do *Online English Collocations Workbook*, para que, de forma contextualizada, o aluno possa aprender a empregá-la e adotá-la como parte de seu discurso. A seguir, são apresentados dois exemplos de uso desta colocação.

Did Supervisor Grissom violate any procedures in mishandling the evidence?
(Exemplo extraído do CCSI)

⁴² “to deal with a situation badly, because of a lack of skill or care: [...]; to treat something roughly, often causing damage: [...]”

*According to a 2005 study by the Department of Justice, the average lab has a backlog of 401 requests for services. Plus, several state and city forensic departments have been racked by scandals involving **mishandled evidence** and outright fraud.* (Exemplo extraído do CCC)

Ao ler ambos os exemplos apresentados acima, percebe-se que a colocação pode causar maior dificuldade de produção por parte dos alunos, por ser composta por um verbo mais rebuscado, haja vista que, ao considerar o contexto em que é empregada, os alunos podem compreendê-la melhor do que quando tentam compreender seu significado de modo isolado.

Dessa forma, esta colocação especializada, além de ter sido inserida nas atividades do *Online English Collocations Workbook*, visando o desenvolvimento da habilidade dos aprendizes em produzi-la, merece destaque por parte do professor quanto ao seu entendimento por parte dos alunos, haja vista o fato de o verbo que a compõe ser mais sofisticado que os verbos das demais colocações levantadas.

4.5.2 *To jeopardize an evidence*

A colocação especializada *to jeopardize an evidence* ocorre apenas uma vez no CCSI, nenhuma vez no CCC e 34 vezes no *WebCorp*. Após análise dos contextos em que foi empregada, observamos sua importância para esta área de especialidade e, desse modo, decidimos selecioná-la, considerando também que esta colocação pode causar certa dificuldade de compreensão e produção por parte dos alunos.

Segundo o *Longman Dictionary*, *jeopardize* significa “arriscar perder ou estragar algo importante”⁴³. Já o *Oxford Collocations Dictionary* não traz essa lexia como colocado de *evidence*, nem como base de alguma colocação, novamente ressaltando sua relevância de inserção na plataforma.

Após verificarmos o significado atribuído a este verbo pelo dicionário, apresentamos dois exemplos da colocação especializada *jeopardize an evidence*, um retirado do CCSI e outro retirado das ocorrências encontradas no *WebCorp*, com o intuito de verificarmos até que ponto as definições dadas pelo dicionário ajudam ou não na compreensão do significado da colocação.

⁴³ “to risk losing or spoiling something important: [...]”.

*He was dead. And I was afraid that he being inside there could **jeopardize the evidence**.* (Exemplo extraído do CCSI).

*It should note that in order to avoid 5th amendment violations, police officers are expected to read arrestees Miranda rights immediately, failure to do so may **jeopardize evidence** in resulting court case.* (Exemplo do WebCorp).

Como os exemplos nos mostram, a colocação *jeopardize evidence* pode ser entendida como “comprometer a evidência”, ou “comprometer a prova”. E, a partir das informações fornecidas pelo *Longman Dictionary*, o aluno pode compreender esta colocação, mesmo que não chegue à tradução por nós sugerida. Contudo, o que causa maior problema é que esta colocação é composta por um verbo mais sofisticado, fato que a torna uma colocação mais sofisticada também, o que dificulta a produção da mesma por parte do aluno. Nesse sentido, justifica-se sua seleção e inclusão no *Workbook*.

4.5.3 *To tamper with evidence*

De modo similar às duas colocações anteriormente apresentadas nesta subseção, *tamper with evidence* é uma colocação verbal, composta por um verbo seguido de uma preposição e um substantivo. No que diz respeito à sua frequência, encontramos esta colocação 2 vezes no CCSI, 52 vezes no CCC e 236 vezes no *WebCorp*.

No tocante ao verbo *tamper*, este é definido pelo *Longman Dictionary* como “tocar em algo ou fazer mudanças sem permissão, de forma a causar danos deliberadamente”⁴⁴. Já o *Oxford Collocations Dictionary* novamente não traz nenhum registro do verbo *tamper*.

Nos exemplos abaixo, é possível verificarmos o contexto em que esta colocação especializada é utilizada:

*You're **tampering with evidence** of the United States government there, Mr. Stokes.* (Exemplo extraído do CCSI).

*Formal defense in the conspiracy case began on June 17, 2010, and concluded without any of the defendants testifying. Judge Lynn Leibovitz found each of the three men not guilty of charges of conspiracy, obstruction of justice and **tampering with evidence** on June 29, 2010.* (Exemplo extraído do CCC).

⁴⁴ “to touch something or make changes to it without permission, especially in order to deliberately damage it: [...]”

No caso desta colocação especializada, o aluno consegue inferir parte de seu significado a partir das definições apresentadas pelos dicionários consultados. Contudo, apenas através do contexto de uso desta é que poderá compreender o significado da colocação, sem contar com o fato de que, de posse do contexto, talvez não houvesse necessidade de recorrer a um dicionário de língua geral para tentar compreender o significado de uma colocação especializada. Dessa maneira, o aluno conseguiria inferir que *tamper with evidence* pode indicar uma ideia de “danificar a evidência”, ou “modificar a evidência”. No entanto, acreditamos que o significado mais completo desta colocação possa ser “adulterar evidência”.

4.5.4 *To flush evidence*

A colocação especializada verbal *to flush evidence* também pode causar certa dificuldade para o aluno na compreensão de seu significado, levando o aluno a recorrer a dicionários, para tentar chegar ao seu significado. No entanto, o *Oxford Collocations Dictionary* não traz junto à entrada *evidence* o verbo *to flush*, tampouco o substantivo *evidence* como um dos colocados possíveis para *to flush*.

Por sua vez, o *Longman Dictionary of Contemporary English* diz que *flush* pode significar “ruborizar, ficar vermelho; enxaguar; jorrar; fazer com que alguém deixe o local em que se esconde; ou limpar algo por meio da força da água”⁴⁵.

Contudo, ao se deparar com exemplos de uso desta colocação, o aluno perceberá que apenas com as definições apresentadas pelo dicionário para este verbo talvez não seja possível compreender a referida colocação especializada e, nesse sentido, vale reforçar a necessidade de o professor facilitador chamar a atenção à importância de se analisar a colocação como um todo, em seu contexto. Observe os exemplos abaixo, retirados do CCSI e do emulador de concordâncias *WebCorp*:

*Then where's the paraphernalia? **Flushed the evidence** before the high killed her?* (Exemplo extraído do CCSI).

*An Omaha man pleaded guilty to using counterfeit money at a Council Bluffs casino and then **flushed the evidence** once security was following him, according to court records.* (Exemplo extraído do *WebCorp*).

⁴⁵ “Become red in face, [...]; to flush a toilet; [...]; to make someone leave a place where they are hiding: [...]; to force water through a pipe in order to clean, [...].”

Esta colocação ocorre apenas 1 vez no CCSI e não foi encontrada no CCC. Por esse motivo, recorremos a um exemplo extraído das 35 concordâncias geradas pelo *WebCorp* da colocação em foco. Como pode ser visto pelos exemplos, caso o aluno tente associar algum dos significados atribuídos ao verbo *flush* pelos dicionários consultados, ele ainda poderá ter certa dificuldade de compreensão e produção da referida colocação.

Embora o aluno possa chegar a “lavar as provas/evidências”, “limpar (no sentido de destruir) as evidências”, parece-nos que esta conclusão não é completamente satisfatória. O aluno precisaria ir além, pensando em algo que encaixasse no contexto de especificidade da área, uma vez que lavar as provas não se enquadra nos contextos apresentados. Assim, partindo para um significado mais metafórico do verbo *flush*, o aluno teria que chegar à colocação “destruir as evidências/provas”. Depreender este significado mais metafórico sem a ajuda de um professor, sem que a colocação seja ensinada e trabalhada em seu contexto real de uso, poderia ser um pouco mais difícil para um aprendiz de língua inglesa, partindo apenas das definições dicionarizadas do verbo.

No entanto, ao trabalhar com as atividades contextualizadas de *Gap Fill* presentes no *Online English Collocations Dictionary*, o aluno é levado a compreender a colocação em seu contexto de uso, o que permite e facilita a sua produção, oral e escrita, e o emprego mais adequado desta colocação.

4.5.5 *To ditch evidence*

A colocação *to ditch evidence* ocorre apenas 1 vez no CCSI e nenhuma vez no CCC. Em contrapartida, foi possível verificar que esta colocação aparece 82 vezes quando buscada no *WebCorp*, de forma que se pode dizer que não se trata de uma colocação muito comum, o que nos faz olhar com mais atenção para ela.

Ao pesquisar por esta colocação no *Oxford Collocations Dictionary*, verificamos que o verbo *ditch* não ocorre como colocado do substantivo *evidence*, e que *ditch*, embora conste como uma entrada do referido dicionário, não apresenta *evidence* como um de seus possíveis colocados.

Levando em consideração que esta é uma colocação verbal, buscamos pelas definições atribuídas à *ditch*, enquanto verbo, fornecidas pelo *Longman Dictionary*, e constatamos que, segundo este dicionário, *ditch* pode ser usado para se referir a “algo que você não quer mais;

ao fim de um relacionamento romântico; a matar aula (deixar de ir para a escola quando você deveria); deixar alguém com quem se está em um lugar sem falar nada para esta pessoa; pousar um avião de forma controlada na água”⁴⁶.

Contudo, percebemos que as definições dadas pelo dicionário, embora possam dar pistas para facilitar o entendimento desta colocação por parte dos alunos, esta ainda se mostra difícil para o aluno, no que diz respeito à sua produção escrita ou oral.

A seguir são apresentados exemplos extraídos do CCSI e do *WebCorp*, respectivamente, para ilustrar o uso desta colocação.

*She goes outside, she climbs the fence, she **ditches the evidence** in the neighbor's trash, and then cuffs herself to the bed and then waits to dies?* (Exemplo extraído do CCSI).

*According to Beard, a pair of gloves were wrapped around the alleged murder weapon, and Beard held that weapon outside the window of the truck he used that night until throwing the weapon and gloves away. He says he **ditched evidence**, including the clothes he wore the night of Kelly's death, because he simply wanted to forget any memory of that night.* (Exemplo extraído do *WebCorp*).

Como se pode perceber ao ler os exemplos, as definições dadas pelo *Longman Dictionary* auxiliam na compreensão do significado da colocação e permitem que o aluno entenda a colocação *to ditch evidence* como “descarte de evidências/provas”. Porém, o significado poderá ser esclarecido a partir da leitura dos exemplos, fato que corrobora a importância de se trabalhar tal colocação por meio das atividades desenvolvidas pelo *Online English Collocations Workbook*, uma vez que este permite que o aluno entenda o uso das colocações de forma mais adequada, para que possa, em seguida, produzi-las em textos orais ou escritos.

4.5.6 *Stitch of evidence*

A colocação nominal especializada *stitch of evidence* pode trazer dificuldade tanto em relação à compreensão quanto à produção, sem contarmos com o obstáculo de uma possível tradução para o português.

⁴⁶ “to stop having something because you no longer want it: [...]; to end a romantic relationship with someone: [...]; to not go to school, a class act when you should: [...]; to leave someone you are with in a place without telling them you are going; to land an aircraft in a controlled crash into water: [...].”

O *Longman Dictionary* define *stitch* como “um pedaço curto de fio costurado a uma roupa ou a ação do fio entrando e saindo do tecido; um pedaço de fio especial usado para suturar feridas; uma dor aguda no lado do corpo; um pequeno círculo de lã formado ao redor da agulha quando se está tricotando; um estilo particular de costurar ou tricotar”⁴⁷.

A partir destas definições o aluno poderia entender *stitch of evidence* como um “ponto de evidência”. No entanto, este significado parece vago ao lermos exemplos contextualizados de uso da colocação, tais como:

Oh we got tons of motives, not a stitch of evidence. (Exemplo extraído do CCSI).

*"Every piece of evidence, every **stitch of evidence**, every whiff of evidence that goes to the finder of fact, to the jury, to the military tribunal will be reviewed by the accused, subject to confrontation, subject to cross-examination, subject to challenge," he said.* (Exemplo extraído do CCC).

Como se pode observar, “ponto de evidência” poderia ser o significado a que os alunos chegariam a partir das definições apresentadas pelos dicionários consultados. Contudo, como este significado não é satisfatório, buscou-se a tradução dada ao exemplo extraído do CCSI nas legendas em português, e obtivemos a seguinte frase:

*Temos toneladas de motivos. **Nenhuma evidência.*** (Exemplo extraído das legendas em português da série *CSI Las Vegas*).

Embora tenha sido traduzido por “nenhuma evidência”, esta tradução ainda pode não ser a melhor opção quanto ao significado da colocação especializada em questão. Parece ser uma solução de tradução para uma legenda, porém, por ser bastante simplista, não contribui para facilitar a compreensão da colocação em foco por parte do aprendiz. Por considerarmos o contexto em que é utilizada, acreditamos que esta colocação tem seu significado melhor compreendido ao entendê-la como “sem sombra de evidência” (*not a stitch of evidence*). Uma vez que *stitch* está relacionado a ponto, algo tão específico e preciso, “sem sombra de evidência” consegue dar um significado mais preciso e específico também a esta colocação.

⁴⁷ “a short piece of thread that has been sewn into a piece of cloth, or the action of the thread going into and out of the cloth; a piece of special thread which has been used to sew the edges of a wound together: [...]; a sharp pain in the side of your body, [...]; a small circle of wool that is formed around a needle when you are knitting: [...]; a particular way of sewing or knitting [...].”

4.5.7 *An evidence ledger*

An evidence ledger é uma colocação nominal, constituída por dois substantivos, sendo que o nódulo *evidence* atua como colocado, desempenhando uma função adjetiva na colocação. O *Oxford Collocations Dictionary* não traz *ledger* como uma entrada, tampouco menciona este substantivo como possível de coocorrer com *evidence*.

Por sua vez, o *Longman Dictionary* define *ledger* como “um livro no qual um negócio, banco etc registra quanto dinheiro recebe e gasta”⁴⁸. Contudo, caso o aluno queira entender a colocação partindo apenas desta definição, verá que o resultado obtido não será muito satisfatório, uma vez que uma evidência não recebe nem gasta dinheiro. Assim, para facilitar a compreensão e, conseqüentemente, a produção desta colocação por parte do aluno, é necessário considerar o contexto em que esta ocorre, como nos exemplos apresentados a seguir:

*This, uh, entry in your **evidence ledger** corresponds to the day that girl, Rio, was killed.* (Exemplo extraído do CCSI).

*View online down **evidence ledger** and evidence custody files, a copy of each related evidence custody document with all final reports to Crime Records Center where they will be kept under file number 509-10.* (Exemplo extraído do WebCorp).

Como pode ser observado a partir dos exemplos apresentados, a colocação em questão diz respeito a um “livro de provas/evidências”, de forma que, a definição dada pelo dicionário embora possa ajudar, não leva o aluno a compreender a colocação sem considerar o seu contexto de uso. Dessa forma, embora esta seja uma colocação mais simples do que as demais colocações apresentadas até aqui, esta também deve ser trabalhada com os alunos para evitar possíveis mal-entendidos.

4.5.8 *Crime Spree*

Como dissemos anteriormente, buscamos colocações adjetivas obtidas a partir de outros nódulos para empreendermos a análise léxico-semântica de forma a exemplificar a variedade de colocações obtidas em nossas análises. Além disso, podemos dizer que as colocações adjetivas obtidas a partir do nódulo *evidence* são mais fáceis de serem

⁴⁸ “A book in which a business, bank etc records how much money it receives and spends.”

compreendidas e produzidas pelos aprendizes de língua inglesa, o que contribuiu para que buscássemos uma colocação que pudesse causar certo estranhamento para os alunos.

Assim, optamos por trabalhar com a colocação *crime spree*, obtida por meio da análise do nódulo *crime*, uma vez que esta colocação ocorre 4 vezes no CCSI e 111 vezes no *WebCorp*, mostrando-se assim, bastante usual. Como pode ser perceber, nesta colocação o termo *crime* atua como colocado, fato comprovado pelo *Oxford Collocations Dictionary* que apresenta *spree* como uma de suas entradas e *crime* como um de seus adjetivos colocados.

Por sua vez, o *Longman Dictionary* define *spree* como “um curto período de tempo quando você faz muito uma atividade, especialmente gastando dinheiro ou consumindo álcool”⁴⁹. Mesmo com o auxílio da definição apresentada por este dicionário para *spree*, o aluno precisa conhecer a colocação e o contexto em que é utilizada, por isso, apresentamos os exemplos a seguir:

*Maybe our vics were the beginning of a little **crime spree**.* (Exemplo extraído do CCSI).

*An Afghan migrant evaded Austrian police twice in a 24-hour **crime spree**, assaulting numerous people before following an 18-year-old female student home and masturbating on her family's doorstep.* (Exemplo extraído do *WebCorp*).

Após ler os exemplos o aluno consegue entender o contexto em que a colocação em análise é utilizada, além de compreender melhor seu significado, o qual, “onda de crime” pode não ficar claro a partir da definição apresentada pelo *Longman Dictionary* caso o aluno pense na colocação de forma isolada, o que justifica a sua inclusão nas atividades que compõem o *Online English Collocations Workbook*.

4.5.9 *Duty weapon*

Outra colocação adjetiva escolhida para esta análise é *duty weapon*. Esta colocação, obtida por meio da análise do nódulo *weapon*, foi encontrada em ambos os corpora de estudo e no emulador de concordanciador utilizado em nossa pesquisa, ocorrendo 2 vezes no CCSI, 6 no CCC e 52 no *WebCorp*.

Ao buscarmos por esta colocação no *Oxford Collocations Dictionary* percebemos que este não apresenta *duty* como um dos possíveis colocados de *weapon*, tampouco traz esse

⁴⁹ “A short period of time when you do a lot of one activity, especially spending money or drinking alcohol”.

adjetivo como uma possível entrada. Já o *Longman Dictionary* define *duty* como “algo que você faz porque é moralmente ou legalmente correto, [...]; algo que você tem que fazer como parte de seu trabalho, [...]; uma taxa que você paga, [...]”⁵⁰.

Mesmo considerando as definições dadas para *duty* pelo referido dicionário, o aluno, provavelmente, teria dificuldade de compreender a colocação sem considerar o contexto em que é utilizada, por isso, apresentamos os exemplos a seguir:

- *Consistent with an officer's **duty weapon**?* (Exemplo extraído do CCSI).

*While this was taking place, a MINE Task Force detective was returning a **duty weapon** to its holster when an accidental discharge occurred resulting in an entry and exit wound to the lower right leg.* (Exemplo extraído do CCC).

A partir dos exemplos, o aluno consegue compreender que *duty weapon* é a “arma utilizada em serviço” pelos oficiais, o qual poderia ser inferido pelo aluno considerando a segunda definição apresentada pelo *Longman Dictionary*, mas que só fica claro ao visualizar o contexto em que é utilizada. Dessa forma, ressaltamos a importância de se ensinar as colocações de forma contextualizada, permitindo que os aprendizes compreendam e produzam as colocações com mais facilidade.

4.5.10 *Allegedly fire*

No que diz respeito às colocações do tipo adverbial que são apresentadas nesta análise, optamos por trabalhar com as que foram obtidas por meio da análise do nóculo *fire*. No caso da colocação *allegedly fire*, percebemos que o *Oxford Collocations Dictionary* não traz o advérbio *allegedly* como possível colocado para o verbo *fire*; tampouco traz o advérbio em questão como uma de suas entradas.

Por sua vez, o *Longman Dictionary* define *allegedly* como um advérbio formal “usado quando se reporta alguma coisa que a pessoa diz ser verdade, embora não tenha sido provado”⁵¹. Como pode ser observado pelos exemplos a seguir, a definição apresentada pelo *Longman Dictionary* contribui para a compreensão da colocação.

⁵⁰ “Something that you have to do because it is morally or legally right, [...]; something you have to do as part of your job, [...]; a tax you pay on something you buy, [...]”

⁵¹ “Used when reporting something that people say is true, although it has not been proved”.

*Francisco Sanchez **allegedly fired** three shots from a .40-caliber handgun at Pier 14, a tourist attraction area at the Embarcadero waterfront district.* (Exemplo extraído do CCC).

*Police charged a man with attempted murder today after shots were **allegedly fired** at two police officers. Callum Hall, 20, was charged with two counts of attempted murder, possession of a firearm with intent to endanger life and possession of a weapon in public.* (Exemplo extraído do WebCorp).

Contudo, tal como observado nos exemplos, esta colocação pode causar maior dificuldade de produção por parte dos aprendizes devido ao fato de ser utilizada em contextos considerados mais formais, fato que justifica a sua inclusão nas atividades do *Online English Collocations Workbook*.

4.5.11 *Reportedly fire*

Tal como ocorre com o advérbio apresentado na colocação anterior, também não foi possível identificar o advérbio *reportedly* como um dos colocados do verbo *fire* no *Oxford Collocations Dictionary*. Por sua vez, o *Logman Dictionary* define *reportedly* como “de acordo com o que algumas pessoas dizem”⁵². Tendo esta definição em mente, consideramos os exemplos a seguir:

*When police entered the home in pursuit, the suspect **reportedly fired** multiple gunshots at the officers.* (Exemplo extraído do CCC).

*Suffolk Police are investigating after shots were **reportedly fired** during an argument between a group of people in Cardinal Park in Ipswich this morning.* (Exemplo extraído do WebCorp).

É possível perceber, pelos exemplos apresentados, que a definição dada pelo *Longman Dictionary* contribui para a compreensão da colocação por parte dos aprendizes. No entanto, se faz importante destacar esta colocação para que os alunos trabalhem a sua produção, haja vista que ao compararem o significado atribuído ao advérbio *reportedly* com o significado atribuído ao advérbio *allegedly*, presente na colocação anterior, podem entender que ambos são utilizados da mesma forma, uma vez que fazem referencia a algo que foi reportado por outras pessoas. Dessa forma, a inserção de ambas as colocações nas atividades

⁵² “According to what some people say”.

do *Online English Collocations Workbook* visa dirimir quaisquer dúvidas quanto ao contexto em que são utilizadas.

Assim, como mencionado anteriormente, foram apresentadas nesta seção algumas das colocações especializadas que podem causar maior dificuldade de compreensão e produção por parte dos alunos. Vale mencionar, ainda, que embora em alguns casos sejam apresentados exemplos extraídos das legendas em português da série que compõem o corpus de estudo (*CSI – Crime Scene Investigation*), estes exemplos não são inseridos no *Online English Collocations Workbook*, tendo sido utilizados apenas nesta análise. Isto se deve ao fato de o referido *workbook* online ter como objetivo principal auxiliar o aluno na compreensão das colocações que o compõe a partir de seu contexto real de uso na própria língua inglesa. Em pesquisa futura, pretende-se utilizar, também, as legendas em português da série já mencionada, de forma a ampliar o estudo colocacional feito nesta pesquisa.

A seguir, são apresentados alguns casos de variação morfossintática que foram encontrados durante as análises das colocações obtidas a partir do nódulo *evidence*, os quais apresentam variação fraseológica também, como pode ser observado por meio das análises.

4.5.12 *A body evidence e body of evidence*

A colocação *a body evidence* foi encontrada uma vez no CCSI, duas vezes no CCC e 119 vezes no *WebCorp*. Já a colocação *body of evidence* não foi encontrada no CCSI e ocorre 60 vezes no CCC e 99 vezes no emulador de concordanciador *WebCorp*.

Ao olhar para ambas as colocações isoladamente, de forma descontextualizada, é possível identificarmos uma variação morfossintática, uma vez que *a body evidence* é um sintagma nominal, enquanto que *body of evidence* é um sintagma preposicionado.

No entanto, ao observar os exemplos extraídos de ambos os corpora, percebe-se que esta variação morfossintática influencia no significado da colocação. Observe os exemplos a seguir:

*I do believe that the **body evidence** corroborates the janitor's statement.*
(Exemplo extraído do CCSI)

*On the other hand, there is a considerable **body of evidence** which indicates that women, other than prostitutes, were rarely the victims of any kind of offense and were treated with the utmost respect.* (Exemplo extraído do CCC)

No primeiro exemplo, a colocação *body evidence* refere-se a uma “evidência/prova do corpo”. Já no segundo exemplo, *body of evidence* faz referência a um “corpo/conjunto de evidências/provas”. Ou seja, a partir destes exemplos é possível perceber que as colocações especializadas em destaque não apresentam apenas uma variação morfossintática, mas sim, que esta variação morfossintática implica em mudança de significado destas colocações.

A fim de comprovar a diferença de significado de ambas as colocações, apresenta-se mais um exemplo da colocação especializada nominal *body of evidence*, também extraído do CCC.

*Nothing is more compelling than an eyewitness who says in court, "that's him" and points at the defendant. But a growing **body of evidence** now shows the unreliability of eyewitness testimony -- and the horrendously wrong jury verdicts that eyewitnesses produce.* (Exemplo extraído do CCC).

Neste exemplo, tal como no primeiro exemplo dado para esta colocação, fica evidente que *body of evidence* tem um significado bem diferente de *a body evidence*. Aqui, novamente, a colocação em destaque refere-se a um “conjunto de evidências/provas”.

Dessa forma, embora ambas as colocações sejam consideradas de fácil compreensão e produção por parte dos alunos quando analisadas isoladamente, é possível perceber que a inserção destas nas atividades que compõem o *Online English Collocations Workbook* se faz justificável, devido à diferença de significados atribuídos a cada uma. Ademais, o que a princípio parecia ser apenas uma variação morfossintática, que nos levaria a crer que ambas as colocações fossem parassinônimas, se mostra, na verdade, uma variação de significado bem mais forte, uma vez que estas colocações são utilizadas em contextos completamente diversos, apesar de serem formadas pelos mesmos componentes.

4.5.13 *Examination of evidence X Evidence examination*

Diferentemente do caso apresentado anteriormente, em que a variação morfossintática das colocações reflete em uma diferença de significados, ao comparar os exemplos das colocações *examination of evidence* e *evidence examination* é possível perceber que, neste caso, há apenas uma variação morfossintática, sendo uma colocação um sintagma preposicionado e a outra um sintagma nominal, como pode ser observado nos exemplos a seguir.

*The participants are assumed to have basic knowledge of weapons and **examination of evidence**.* (Exemplo extraído do CCC)

*The crime lab is staffed with four Crime Scene Investigators (CSIs). Photography and photographic processing, latent fingerprint processing, crime scene investigation and processing, autopsy evidence collection, crime scene reconstruction, confirmatory drug testing, and **evidence examination** are just some of the duties carried out by the CSIs.* (Exemplo extraído do CCC)

Ao ler ambos os exemplos, constata-se que ambas as colocações fazem referência a um “exame das evidência/provas”. Neste caso, o que as diferenciam são as frequências que cada uma apresentou em ambos os corpora e no emulador de concordanciador *WebCorp*. A colocação *examination of evidence* não foi encontrada no CCSI, mas ocorre 6 vezes no CCC e 4 vezes no *WebCorp*; já a colocação *evidence examination* ocorre uma vez no CCSI, 9 vezes no CCC e 50 vezes no *WebCorp*, mostrando ser uma colocação mais usual do que a anterior.

A partir do momento que se percebe a variação morfossintática, verifica-se se esta pode ou não resultar em uma diferença de significados e, consequentemente, de usos das colocações. Como neste caso isto não ocorre, a colocação que será incluída nas atividades da referida plataforma *on-line* será aquela que apresenta uma maior frequência de ocorrência, no caso, *evidence examination*. Dessa forma, embora ambas as colocações sejam de fácil compreensão, o aluno entrará em contato com aquela mais frequente, o que contribuirá para aprimorar seu conhecimento colocacional.

Assim, finalizamos a análise léxico-semântica, que teve o intuito de discutir sobre as dificuldades que os alunos podem encontrar de compreensão e produção de algumas das colocações especializadas levantadas durante a pesquisa, bem como de apresentar dois casos de variação morfossintática com diferentes implicações, visando, também, a compreensão e utilização destas colocações por parte dos alunos. Dessa forma, o próximo capítulo aborda os critérios considerados para a seleção das colocações que compõem as atividades do *Online English Collocations Workbook*, bem como a ilustração da elaboração de algumas das atividades que o compõem.

5 AS ATIVIDADES DO *ONLINE ENGLISH COLLOCATIONS WORKBOOK* – CATEGORIA *INVESTIGATION*

Neste capítulo, são apresentados os critérios utilizados para a seleção das colocações especializadas que compõem as atividades de *Memory Game* e *Gap Fill*, integrantes do *Online English Collocations Workbook* na categoria *Investigation*, bem como alguns exemplos destas.

Após o levantamento e a análise das colocações especializadas encontradas, buscou-se classificar cada uma delas, de acordo com três níveis de dificuldades, fácil, médio e difícil, segundo os níveis sugeridos pelo *Online English Collocations Workbook*. Embora esta seja uma classificação subjetiva, foram adotados alguns critérios necessários que levam em consideração a variação e a ampliação do conhecimento léxico dos alunos de língua, além de considerar o conhecimento e a experiência da pesquisadora e serem revisados pela criadora da referida plataforma *on-line* que também orienta esta pesquisa. Vale ressaltar que os níveis atribuídos às colocações estão sendo testados pelos alunos de um curso de Bacharelado em Tradução e de um curso de Licenciatura em Letras, o que nos possibilita uma maior adequação quando ao grau de dificuldade atribuído às colocações especializadas que compõem as atividades. Além disso, esta classificação é necessária para facilitar a inclusão das colocações especializadas no *workbook online*, uma vez que, segundo sua proposta as colocações são divididas de acordo com os níveis de dificuldade citados acima, além de obedecerem à classificação taxonômica dada a partir da análise da sua formação.

Ressaltamos que as atividades são divididas nestes três níveis de dificuldade como uma forma de possibilitar que alunos, com diferentes níveis de conhecimento da língua inglesa (desde o básico até ao avançado), possam praticar, haja vista que, tal como afirma Hill (2000), a colocação “não é um bônus adicional ao qual passamos a focar quando os alunos se tornaram suficientemente avançados. As colocações devem desempenhar um papel importante em nosso ensino desde a primeira aula”⁵³ (HILL, 2000, p. 60).

Outro critério adotado para a inserção das colocações diz respeito à quantidade de colocações que são inseridas a partir de cada nódulo. Como dissemos no capítulo anterior, alguns nódulos geraram mais de 100 colocações especializadas, como no caso do nódulo *evidence*, que gerou 148 colocações, das quais 90 foram inseridas no *Workbook*. Outro nódulo que pode ser utilizado para exemplificar a diferença entre colocações obtidas e colocações

⁵³ “[...] is not an added bonus which we pay attention to once students have become sufficiently advanced. Collocations should play an important part in our teaching from lesson one”.

incluídas nas atividades é o nódulo *body*, que resultou em 126 colocações especializadas encontradas a partir do *Corpus CSI* e teve apenas 71 colocações inseridas nas atividades do *workbook* distribuídas entre as taxonomias já apresentadas e os três níveis de dificuldade mencionados. Esta seleção de colocações especializadas inseridas se justifica pelo fato de que, caso inseríssemos todas as colocações levantadas destes nódulos que resultaram em mais de 100 colocações, as atividades poderiam se tornar de difícil resolução devido ao fato de ocorrer mais de um caso em que se usaria uma mesma base no mesmo exercício, o que poderia prejudicar o aprendizado do aluno.

A seguir apresentamos as colocações especializadas do tipo nominal que foram incluídas nas atividades do *Online English Collocations Workbook*.

Quadro 13 – Colocações nominais do nódulo *Evidence* incluídas no *Online English Collocations Workbook*.

<i>A murder evidence</i>	<i>Ballistics evidence</i>
<i>An homicide evidence</i>	<i>Analysis of evidence</i>
<i>A blood evidence</i>	<i>Autopsy evidence</i>
<i>An ammo evidence</i>	<i>Scientific evidence</i>
<i>An explanation of the evidence</i>	<i>An evidence examination</i>
<i>A case evidence</i>	<i>An evidence bag</i>
<i>A DNA evidence</i>	<i>An evidence box</i>
<i>An evidence of a struggle</i>	<i>An evidence room</i>
<i>An evidence of the crime</i>	<i>Fingerprint evidence</i>
<i>An evidence of trauma</i>	<i>An absence of evidence</i>
<i>A body evidence</i>	<i>A drug evidence</i>
<i>A Gun evidence</i>	<i>An evidence vault</i>
<i>An item of evidence</i>	<i>A lack of evidence</i>
<i>Preservation of evidence</i>	<i>A stitch of evidence</i>

Como se pode notar no quadro acima, foram incluídas 28 colocações nominais em ambas as atividades do *Online English Collocations Workbook*, a saber *Memory Game* e *Gap Fill*. Tal como dissemos anteriormente, estas colocações são divididas por níveis ao serem inseridas no referido *workbook*, os quais serão apresentados mais a frente. A seguir, apresentamos as colocações verbais que foram inseridas nas atividades da plataforma online.

Quadro 14 – Colocações verbais do nóculo *Evidence* incluídas no *Online English Collocations Workbook*.

<i>Evidence appears</i>	<i>To run evidence</i>
<i>Evidence supports</i>	<i>To test an evidence</i>
<i>Evidence tells</i>	<i>To check an evidence</i>
<i>Evidence fails</i>	<i>To compel evidence</i>
<i>Evidence proves</i>	<i>To jeopardize an evidence</i>
<i>Evidence shows</i>	<i>To tamper with evidence</i>
<i>Evidence connects</i>	<i>To search for evidence</i>
<i>Evidence indicates</i>	<i>To sift through evidence</i>
<i>To find evidence</i>	<i>To look for evidence</i>
<i>To reexamine an evidence</i>	<i>To look at evidence</i>
<i>To flush an evidence</i>	<i>To turn into evidence</i>
<i>Based on the evidence</i>	<i>The evidence suggests</i>
<i>To get evidence</i>	<i>Evidence points to</i>
<i>To ditch an evidence</i>	<i>To process an evidence</i>
<i>To collect evidence</i>	<i>To analyze evidence</i>
<i>To follow an evidence</i>	<i>To burn an evidence</i>
<i>To trace evidence</i>	<i>To manipulate evidence</i>
<i>To destroy evidence</i>	<i>To recover evidence</i>
<i>To plant an evidence</i>	<i>To hide an evidence</i>
<i>To document evidence</i>	<i>To mishandle an evidence</i>
<i>To fabricate evidence</i>	<i>To obliterate evidence</i>
<i>The evidence says</i>	

A partir de todas as colocações especializadas obtidas por meio da busca e levantando empreendidos em ambos os corpora, foram selecionadas 43 colocações verbais para serem incluídas nas atividades do *Online English Collocations Workbook*. O fato de termos selecionado uma quantidade maior de colocações verbais para comporem as atividades do referido *workbook online* se deve ao fato de termos encontrado uma quantidade maior de colocações deste tipo. Dessa forma, no caso do nóculo *evidence*, prevalecem as colocações verbais sobre os demais tipos, fato que não ocorre, necessariamente, com os demais nós analisados e que tiveram suas colocações inseridas nas atividades. A quantidade de colocações verbais varia muito de um nóculo para outro. A seguir, apresentamos as colocações adjetivas que foram selecionadas para comporem as atividades da plataforma online.

Quadro 15 – Colocações adjetivas do nódulo *Evidence* incluídas no *Online English Collocations Workbook*.

<i>A physical evidence</i>	<i>A circumstantial evidence</i>
<i>An old evidence</i>	<i>An additional evidence</i>
<i>A new evidence</i>	<i>A reliable evidence</i>
<i>An incriminating evidence</i>	<i>An empirical evidence</i>
<i>A recent evidence</i>	<i>A clear evidence</i>
<i>A biological evidence</i>	<i>A latent evidence</i>
<i>An insufficient evidence</i>	<i>An important evidence</i>
<i>A sufficient evidence</i>	<i>A wrong evidence</i>
<i>A forensic evidence</i>	<i>A corresponding evidence</i>
<i>A good evidence</i>	

No que diz respeito às colocações adjetivas, foram selecionadas um total de 19, a partir das que são mais recorrentes entre todas as colocações levantadas. O critério de recorrência foi utilizado para a seleção de todas as colocações no caso dos nódulos que possibilitaram um maior número de colocações, o que permite que o aprendiz entre em contato com as colocações que são mais utilizadas na área de investigação criminal.

Já com relação às colocações adverbiais, foram incluídas nas atividades todas as colocações deste tipo apresentadas e analisadas no capítulo anterior, as quais são também apresentadas na lista a seguir, haja vista que foram analisadas colocações de três diferentes nódulos, e que estas, por ocorrerem em menor número em nossos corpora de pesquisa, se mostram relevantes para comporem as atividades.

Quadro 16 – Colocações adverbiais inseridas no *Online English Collocations Workbook*.

<i>Fire straight</i>	<i>Definitively kill</i>	<i>Maliciously kill</i>
<i>Allegedly fire</i>	<i>Eventually kill</i>	<i>Mysteriously kill</i>
<i>Previously fire</i>	<i>Kill efficiently</i>	<i>Successfully kill</i>
<i>Recently fire</i>	<i>Secretly kill</i>	<i>Tragically kill</i>
<i>Reportedly fire</i>	<i>Unintentionally kill</i>	<i>Accidentally shoot</i>
<i>Secretly fire</i>	<i>Unlawfully kill</i>	<i>Fatally shoot</i>
<i>Suddenly fire</i>	<i>Unnecessarily kill</i>	<i>Secretly shoot</i>
<i>Ironically kill</i>	<i>Allegedly kill</i>	<i>Rarely shoot</i>
<i>Probably kill</i>	<i>Apparently kill</i>	<i>To shoot effectively</i>
<i>Easy kill</i>	<i>Brutally kill</i>	-
<i>Ultimately kill</i>	<i>Deliberately kill</i>	-
<i>Accidentally kill</i>	<i>Intentionally kill</i>	-

Assim, ao incluirmos apenas uma parte das colocações de nódulos que apresentaram muitas colocações, como é o caso de *evidence*, e ao incluir todas as colocações do tipo

adverbial encontradas em ambos os corpora, possibilita-se que o aluno interaja com uma maior quantidade de colocações especializadas da área investigativa criminal e permite-se que as atividades sejam desenvolvidas sem causar dúvidas nos alunos, uma vez que a seleção das colocações é feita aleatoriamente a partir do tipo e do nível de dificuldade selecionados pelo aluno, o qual é apresentado nos quadros a seguir:

Quadro 17 – Colocações nominais classificadas de acordo com seu nível de dificuldade.

Fácil	Médio	Difícil
<i>A murder evidence</i>	<i>An explanation of the evidence</i>	<i>An ammo evidence</i>
<i>an homicide evidence</i>	<i>A case evidence</i>	<i>An evidence of a struggle</i>
<i>A blood evidence</i>	<i>An evidence of the crime</i>	<i>A lack of evidence</i>
<i>A DND evidence</i>	<i>An evidence of trauma</i>	<i>A stitch of evidence</i>
<i>A body evidence</i>	<i>An item of evidence</i>	
<i>A gun evidence</i>	<i>Preservation of evidence</i>	
<i>Ballistics evidence</i>	<i>Scientific evidence</i>	
<i>Analysis of evidence</i>	<i>An evidence examination</i>	
<i>Autopsy evidence</i>	<i>An absence of evidence</i>	
<i>An evidence bag</i>	<i>A fingerprint evidence</i>	
<i>An evidence box</i>	<i>An evidence vault</i>	
<i>An evidence room</i>		
<i>A drug evidence</i>		

Como se pode notar pelo quadro acima, 13 das 28 colocações nominais obtidas por meio do nóculo *evidence* foram classificadas como sendo de nível fácil, 11 de nível médio e apenas 4 de nível difícil. Essa divisão, embora não resulte em uma quantidade igual de colocações para cada nível de dificuldade, permite que as atividades sejam apresentadas ao aluno sem muitas repetições, uma vez que ao selecionar a taxonomia e o nível desejado, o aluno recebe as colocações escolhidas de forma aleatória pela plataforma, de acordo com sua seleção, que contém colocações de diferentes nósculos.

No quadro abaixo, apresentamos a classificação atribuída às colocações especializadas verbais obtidas por meio do nóculo *evidence*.

Quadro 18 – Colocações verbais classificadas de acordo com seu nível de dificuldade.

Fácil	Médio	Difícil
<i>Evidence points to</i>	<i>Evidence suggests</i>	<i>To flush evidence</i>
<i>Evidence appears</i>	<i>Evidence tells</i>	<i>To ditch an evidence</i>
<i>Evidence supports</i>	<i>Evidence proves</i>	<i>The evidence says</i>
<i>Evidence fails</i>	<i>Evidence shows</i>	<i>To plant an evidence</i>
<i>Evidence indicates</i>	<i>Evidence connects</i>	<i>To obliterate evidence</i>
<i>To find evidence</i>	<i>To get evidence</i>	<i>To process an evidence</i>
<i>To reexamine an evidence</i>	<i>To follow an evidence</i>	<i>To mishandle an evidence</i>
<i>Based on the evidence</i>	<i>To trace evidence</i>	<i>To run evidence</i>
<i>To collect evidence</i>	<i>To document evidence</i>	<i>To jeopardize an evidence</i>
<i>To destroy evidence</i>	<i>To fabricate evidence</i>	<i>To tamper with evidence</i>
<i>To analyze evidence</i>	<i>To hide an evidence</i>	<i>To sift through evidence</i>
<i>To burn an evidence</i>	<i>To compel evidence</i>	
<i>To manipulate evidence</i>	<i>To search for evidence</i>	
<i>To recover evidence</i>	<i>To look for evidence</i>	
<i>To test an evidence</i>	<i>To look at evidence</i>	
<i>To check an evidence</i>	<i>To turn into evidence</i>	

Dentre as 43 colocações especializadas verbais levantadas é possível perceber um equilíbrio maior entre as que foram classificadas como fáceis e as que receberam classificação de nível médio, uma vez que há exatamente 16 colocações classificadas segundo cada um desses níveis. Isso se deve ao fato de termos conseguido um número maior de colocações verbais durante o levantamento e, também, por estas colocações terem apresentado maior frequência de coocorrência. As outras 11 colocações verbais selecionadas foram incluídas nas atividades de nível difícil. Como se pode notar, este nóculo forneceu uma grande quantidade de colocações verbais que foram bem distribuídas entre os três níveis de dificuldade das atividades de *Memory Game* e *Gap Fill* que compõem o *Online English Collocations Workbook*. Apresentamos, a seguir, a classificação atribuída às colocações adjetivas.

Quadro 19 – Colocações adjetivas classificadas de acordo com seu nível de dificuldade.

Fácil	Médio	Difícil
<i>An old evidence</i>	<i>An empirical evidence</i>	<i>An incriminating evidence</i>
<i>A new evidence</i>	<i>A reliable evidence</i>	<i>A forensic evidence</i>
<i>A recent evidence</i>	<i>A latent evidence</i>	<i>A circumstantial evidence</i>
<i>A biological evidence</i>	<i>A corresponding evidence</i>	
<i>An insufficient evidence</i>		
<i>A sufficient evidence</i>		
<i>A good evidence</i>		
<i>An additional evidence</i>		
<i>A clear evidence</i>		
<i>An important evidence</i>		
<i>A wrong evidence</i>		
<i>A physical evidence</i>		

Conforme mostra o quadro acima, entre as colocações especializadas adjetivas selecionadas para comporem as atividades do *workbook online*, a maioria delas, 12, foram classificadas como pertencentes ao nível fácil, enquanto 4 ficaram no nível médio e apenas 3 no nível difícil. De forma geral, as colocações adjetivas podem ser consideradas mais fáceis por parte dos alunos, de forma que seu grau de dificuldade maior seja atribuído àquelas colocações compostas por lexias mais complexas ou utilizadas em situações mais específicas.

Passamos a seguir, a apresentar a divisão de acordo com os níveis de dificuldade das colocações adverbiais obtidas por meio dos nódulos *fire*, *kill* e *shoot*.

Quadro 20 – Colocações adverbiais classificadas de acordo com seu nível de dificuldade.

Fácil	Médio	Difícil
<i>Recently fire</i>	<i>Fire straight</i>	<i>Allegedly fire</i>
<i>Secretly fire</i>	<i>Previously fire</i>	<i>Reportedly fire</i>
<i>Suddenly fire</i>	<i>Ironically kill</i>	<i>Ultimately kill</i>
<i>Probably kill</i>	<i>Eventually kill</i>	<i>Kill efficiently</i>
<i>Easy kill</i>	<i>Unintentionally kill</i>	<i>Unlawfully kill</i>
<i>Accidentally kill</i>	<i>Deliberately kill</i>	<i>Allegedly kill</i>
<i>Definitively kill</i>	<i>Maliciously kill</i>	<i>To shoot effectively</i>
<i>Secretly kill</i>	<i>Mysteriously kill</i>	
<i>Unnecessarily kill</i>	<i>Fatally shoot</i>	
<i>Apparently kill</i>	<i>Rarely shoot</i>	
<i>Brutally kill</i>	<i>Successfully kill</i>	
<i>Intentionally kill</i>		
<i>Tragically kill</i>		
<i>Accidentally shoot</i>		
<i>Secretly shoot</i>		

Entre as 33 colocações adverbiais levantadas a partir dos núdulos *fire*, *kill* e *shoot*, 15 foram classificadas como sendo de nível fácil, 11 de nível médio e 7 de nível difícil. Como o quadro mostra, a divisão feita permite que o aluno interaja com colocações dos três núdulos mencionados em todos os níveis de dificuldade. Mais uma vez, como ocorreu com os demais tipos de colocação especializada apresentados neste capítulo, as colocações de nível fácil ocorrem em maior número que as demais, porém, neste caso de uma forma mais equilibrada quando comparado ao número de colocações de nível médio.

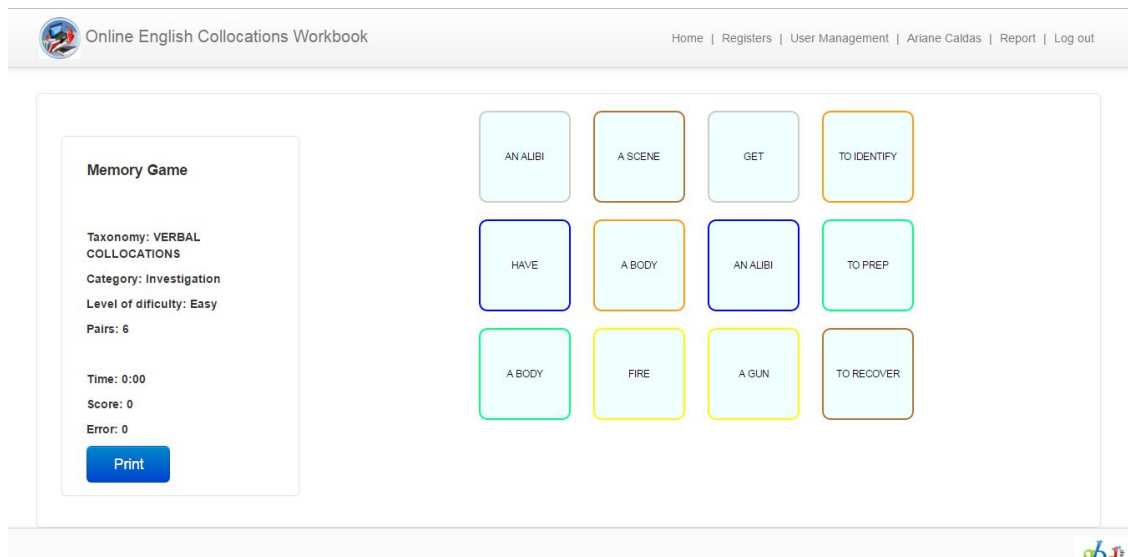
Dessa forma, após a inclusão das colocações de acordo com a taxonomia e o nível de dificuldade, tal como apresentamos, as atividades propostas no *Workbook*, ao colocar em evidência as colocações especializadas, visam promover um aprendizado autônomo por parte dos alunos, de forma a conscientizá-los da importância do conhecimento e do uso das colocações para se tornar fluente na língua inglesa.

A seguir, é apresentado um exemplo da atividade de *Memory Game*, destacando as colocações do tipo verbais que receberão a classificação de nível fácil. Convém dizer que, ao iniciar esta atividade, além de escolher a área de especialidade com a qual deseja trabalhar, selecionar o tipo das colocações de acordo com a sua classificação taxonômica e selecionar o nível de dificuldade destas, o aluno opta por jogar com 6, 8 ou 10 pares de colocações.

A imagem a seguir apresenta a tela de memorização dos pares das colocações, sendo que, neste caso, selecionamos o nível fácil e 6 pares de colocações verbais, uma vez que, ao

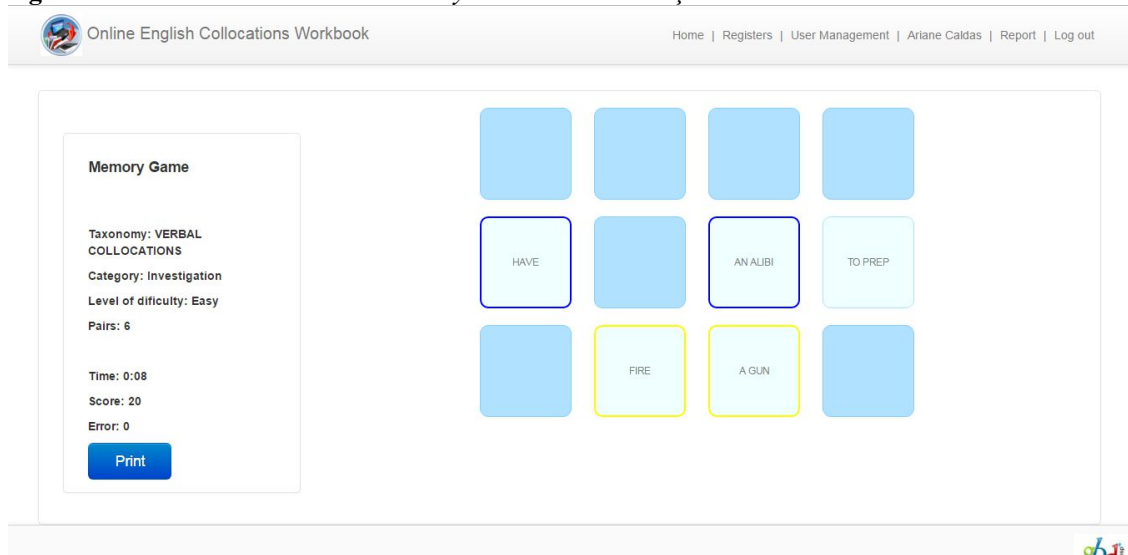
iniciar a atividade, o aluno tem 10 segundos para memorizar os pares das colocações apresentadas, sendo que estes são compostos pela base e o colocado de cada colocação.

Figura 24 – Tela de memorização dos pares das colocações especializadas



Após memorizar os pares que são identificados por meio do uso de uma mesma cor, o aluno inicia a atividade, buscando encontrar a base e o colocado de cada colocação.

Figura 25 – Tela da atividade de *Memory Game* com colocações verbais de nível fácil



Como pode se perceber pelas imagens acima, o aluno tem o tempo cronometrado, à esquerda da tela, bem como sua pontuação e seus erros também apresentados durante a execução da atividade. Este tipo de atividade permite, de forma lúdica e divertida, que o aluno memorize a formação das colocações, de forma a poder acessá-las e utilizá-las ao se comunicar com maior agilidade. Ao terminar esta atividade, o aluno, caso queira, pode

escrevê-las em um caderno, sem consultá-las novamente, de forma a verificar se realmente conseguiu memorizar as colocações trabalhadas na atividade. Este exercício feito após a realização das atividades é uma forma de manter ativo o conhecimento apreendido, além de se constituir em um fonte de consulta futura sobre as colocações.

No que diz respeito às atividades do tipo *Gap Fill*, independentemente do nível de dificuldade e do tipo de colocação especializada (verbal, adjetiva, adverbial ou nominal), a atividade sempre contará com 5 sentenças. Assim, o aluno precisa identificar o colocado correto para cada base. Antes de iniciar a atividade, o aluno deve selecionar as características desta, como nível de dificuldade, área de especialidade das colocações e a tipologia destas.

Após selecionar as características, o aluno clica em *play*, e é redirecionado para a seguinte tela, na qual tem aleatoriamente as frases enumeradas, e os colocados dispostos acima das frases.

Figura 26 – Tela da atividade *Gap Fill* com colocações especializadas adjetivas de nível fácil

The screenshot displays the 'Gap Fill' activity interface. On the left sidebar, the title 'Gap Fill' is shown. Below it, the taxonomy is 'ADJECTIVAL COLLOCATIONS', the category is 'Investigation', and the level of difficulty is 'Easy'. It also shows a timer at '0:04', a score of '0/500', and 'Attempts: 0'. A 'Print' button is located at the bottom of the sidebar. The main area features five word options in buttons: 'decapitated', 'Unidentified', 'Facial', 'Obvious', and 'Fresh'. Below these are five numbered sentences with blanks for the words. At the bottom of the main area are 'Clear' and 'Check' buttons.

Online English Collocations Workbook

Home | Registers | User Management | Ariane Caldas | Report | Log out

Gap Fill

Taxonomy: ADJECTIVAL COLLOCATIONS
Category: Investigation
Level of difficulty: Easy

Time: 0:04
Score: 0/500
Attempts: 0

Print

decapitated Unidentified Facial Obvious Fresh

1. Nick's going to work with me on that _____ body at Summercliff. ①

2. Obvious impact to the face. Severe _____ contusions. ①

3. Look, it's a _____ wound to the back of the head. ①

4. And with T.C.'s history, he becomes the _____ suspect. ①

5. If it's on the train tracks near a _____ body with the victim's blood on it, it's usually a crime. ①

Clear Check

O aluno, então, insere os colocados nos espaços corretos e, em seguida, poderá verificar suas respostas, como na imagem a seguir:

Figura 28 – Tela de correção da atividade *Gap Fill* com colocações especializadas adjetivas de nível fácil

Online English Collocations Workbook

Home | Registers | User Management | Ariane Caldas | Report | Log out

Gap Fill

Taxonomy: ADJECTIVAL COLLOCATIONS
Category: Investigation
Level of difficulty: Easy

Time: 1:42
Score: 200/500
Attempts: 1

Print

Unidentified Facial Fresh Obvious decapitated

1. Nick's going to work with me on that Unidentified body at Summercliff. ①
2. Obvious impact to the face. Severe Facial contusions. ①
3. Look, it's a Fresh wound to the back of the head. ①
4. And with T.C.'s history, he becomes the Obvious suspect. ①
5. If it's on the train tracks near a decapitated body with the victim's blood on it, it's usually a crime. ①

Finish

Nesta etapa, o aluno tem a opção de corrigir as sentenças em que inseriu os colocados inadequados, ou pode optar por verificar as respostas. Como dissemos, sempre são apresentadas 5 sentenças para que o aluno possa inserir os colocados de maneira correta.

A partir das colocações trabalhadas na atividade de *Gap Fill*, os alunos podem produzir novas frases, novas sentenças, diferentes das que são utilizadas na atividade, a fim de verificar se conseguiram compreender o contexto de uso das colocações, além de aprimorar e praticar o uso das colocações trabalhadas pela atividade.

Assim, por meio destes dois tipos de atividades o aluno desenvolve seu conhecimento sobre colocações de forma lúdica e divertida, facilitando seu aprendizado. Ademais, estas atividades compõem a primeira e a segunda fase de implantação do *Online English Collocations Workbook* e, futuramente, novas atividades serão adicionadas, permitindo que o aluno continue aprimorando seus conhecimentos colocacionais.

Gostaríamos de reforçar, também, o fato de que neste capítulo, tal como ocorreu durante o desenvolvimento desta dissertação, enfatizamos as colocações que foram obtidas a partir do nódulo *evidence* e as colocações adverbiais obtidas a partir dos nódulos *fire*, *kill* e *shoot*. Contudo, tal como mencionado anteriormente, e como pode ser observado nas imagens das telas das atividades de *Memory Game* e *Gap Fill*, já foram inseridas no *Online English Collocations Workbook* colocações de outros nódulos, o que permite ao aluno a realização das

atividades contando com uma variedade de colocações especializadas da área de investigação criminal.

Neste sentido, as atividades propostas pelo referido *workbook* online permitem, também, que se desenvolva o ensino-aprendizagem das colocações de forma explícita, tal como sugere Orenha-Ottaiano (2012) e de acordo com as propostas de Lewis (2000) e Hill (2000), possibilitando que o aluno entre em contato com este tipo de unidades fraseológicas de forma a torná-lo consciente de sua presença em seu dia-a-dia independente da área em que atue, permitindo-lhe desenvolver sua competência fraseológica ampliando seu conhecimento sobre a língua inglesa.

CONSIDERAÇÕES FINAIS

O trabalho que aqui se encerra teve como principal objetivo extrair e analisar, do ponto de vista morfossintático e léxico-semântico, as colocações especializadas presentes no Corpus *CSI* e no Corpus Comparável Criminal, a fim de que estas pudessem integrar as atividades do *Online English Collocations Workbook*, na categoria *Investigation*. Para tal, buscou-se, durante o desenvolvimento desta pesquisa, trabalhar a partir das bases teórico-metodológicas da Linguística de Corpus e da Fraseologia, visando à promoção da extração e análise de colocações especializadas da área criminal, a partir de um corpus constituído pelas legendas de 12 (doze) temporadas da série de TV *CSI - Crime Scene Investigation*, e de um Corpus Comparável Criminal.

Esta dissertação inicia-se com a apresentação da série que serviu de base para a compilação do corpus de estudo, a saber, a série de TV norte-americana *CSI – Crime Scene Investigation*. Neste primeiro capítulo justifica-se, ainda, a escolha dessa série para compor o corpus de estudo fundamentada a partir dos pressupostos de Fromm (2011) e Hoffman (1998 apud FINATTO, 2015).

Já, no segundo capítulo é feita a revisão das teorias que servem de base para o desenvolvimento deste trabalho, começando pela Fraseologia teórica (ZULUAGA, 1980; TRISTÁ, 1988; CORPAS PASTOR, 1996; MENA; BERTRÁN, 2002; TAGNIN, 1989 [2011]), passando por aspectos da Fraseologia baseada em corpus (PARTINGTON, 1998; BÉRTRAN; BRETANA, 2008; GRANGER E PAQUOT 2008; GRIES, 2008), para enfatizar, em seguida as colocações da língua geral e especializadas, uma vez que estas constituem o objeto de estudo deste trabalho. Para tal, contamos com autores como L’Homme (2000); L’Homme e Bertrand (2000); Bevilacqua (2005) e Orenha-Ottaiano (2009, 2012).

Este capítulo é encerrado com a apresentação de alguns aspectos da Linguística de Corpus (SINCLAIR, 1991; BIBER, 1998; TOGNINI-BONELLI, 2002) e da Abordagem Lexical (LEWIS, 1993, 2000) e trabalho de Orenha-Ottaiano (2012, 2015) uma vez que nortearam a compilação das atividades didáticas que são o resultado final desta pesquisa.

No terceiro capítulo, discorre-se sobre a metodologia utilizada para a compilação e o tratamento de ambos os corpora utilizados nesta pesquisa, a saber, o Corpus *CSI*, constituído pelas legendas de 12 (doze) temporadas da série de TV norte-americana *CSI – Crime Scene Investigation*, e o Corpus Comparável Criminal, o qual foi compilado com o auxílio da ferramenta *Boot Cat Front End* versão 0.71 (ZANCHETTA; BARONI; BERNARDINI,

2013). Em seguida, explicamos como se deu a extração e análise das colocações identificadas nos corpora, por meio da ferramenta *WordSmith Tools*, versão 6.0 (SCOTT, 2012).

Ainda no terceiro capítulo é feita uma breve apresentação do *Online English Collocations Workbook*, plataforma criada pela Prof^a. Dr^a Adriane Orenha-Ottaiano (2015), na qual as colocações extraídas e analisadas nesta pesquisa foram incluídas para comporem as atividades da categoria *Investigation*. Além disso, discorre-se sobre como se deu o processo de inclusão destas colocações na referida plataforma.

Por sua vez, o quarto capítulo traz a análise dos dados obtidos durante a pesquisa. Em um primeiro momento, são apresentados os dados quantitativos de ambos os corpora, destacando-se informações importantes, tais como, as palavras-chave selecionadas, bem como suas frequências em cada um dos corpora. Além disso, por meio destes dados quantitativos, é justificada a escolha da palavra-chave *evidence* para exemplificar as análises sintático-morfológica e léxico-semântica realizadas com base nas colocações especializadas encontradas. Com relação à análise sintático-morfológica, percebeu-se que, em alguns casos o termo em análise, também chamado de nódulo ou palavra-chave, ocorreu ora como base, ora como colocado das colocações levantadas. Ressalta-se que as colocações que apresentaram o termo em análise como colocado foram mantidas e analisadas por julgar que, independente da posição que o termo ocupa na colocação, seja base ou colocado, a sua presença caracteriza uma colocação especializada, e esta variação na sua posição não interfere na compreensão das colocações.

Ainda com relação à análise morfossintática empreendida, foi possível perceber, também, que a tipologia de colocações que se mostra mais frequente é a nominal, que apresentou um total de 36 colocações obtidas a partir do CCSI e 70 obtidas a partir do CCC e a verbal, que apresentou um total de 67 colocações no CCSI e 35 no CCC. Já as colocações especializadas adjetivas ocorrem em menor número, 34 colocações no CCSI e 32 no CCC, fato que nos leva a crer que, mesmo ocorrendo em menor número que as colocações do tipo nominal e verbal, estas são recorrentes nos corpora de estudo. No entanto, o dado que mais chama à atenção é que as colocações do tipo adverbial praticamente não são encontradas entre as colocações levantadas.

Para que fosse possível apresentar a análise de colocações adverbiais, foi necessário recorrer às colocações deste tipo obtidas por meio da busca feita a partir de três diferentes nódulos, a saber, *fire*, *kill* e *shoot*. Estes nódulos foram os únicos que possibilitaram a identificação de colocações especializadas adverbiais e, mesmo somando as colocações obtidas a partir dos três, foram encontradas apenas 33 colocações adverbiais em ambos os

corpora. Convém ressaltar que esta busca foi empreendida a partir das 67 palavras-chave selecionadas e que, por isso, não foi possível identificarmos uma maior quantidade de colocações adverbiais a partir de outros verbos, tampouco foi possível a identificação deste tipo de colocação formada pela estrutura advérbio + adjetivo, uma vez que o único adjetivo presente entre as palavras-chave selecionada, *dead*, não nos permitiu encontrar nenhuma colocação deste tipo.

Dessa forma, ao se considerar que o Corpus CSI (CCSI) possibilitou o levantamento de mais de 1.700 colocações e que o Corpus Comparável Criminal (CCC) permitiu o levantamento de mais de 1.800 colocações, as 33 colocações especializadas adverbiais encontradas em ambos é um número baixo, fato que nos leva a crer que este tipo de colocações não é comum entre as colocações especializadas da área investigativa criminal.

Ainda neste quarto capítulo é apresentada a análise léxico-semântica de algumas colocações que podem causar maior dificuldade de compreensão e produção por parte dos aprendizes, em razão de serem constituídas por um vocabulário mais rebuscado e sofisticado. Além destas colocações, são apresentados dois casos de variação morfossintática de duas colocações, sendo que em um dos casos a variação entre um sintagma preposicionado e um sintagma nominal não interfere no significado e uso das colocações, como em *examination of evidence* e *evidence examination*; enquanto que no outro, percebe-se que há uma mudança de significado e uso ao optar pelo sintagma nominal ou pelo sintagma preposicionado, como em *a body evidence* e *body of evidence*.

Assim, chega-se ao quinto e último capítulo deste trabalho, que foi elaborado para tratar de algumas das atividades que compõem o *Online English Collocations Workbook* na categoria *Investigation*, além de conter aspectos que foram levados em consideração para a inclusão das colocações nas atividades no referido *Workbook*, principalmente no que diz respeito à classificação taxonômica destas e à divisão das colocações especializadas em níveis de dificuldade, fácil, médio e difícil, de forma a permitir que o aluno desenvolva seu conhecimento colocacional.

Desse modo, ao concluir este trabalho, espera-se ressaltar a importância da Linguística de *Corpus* para a Fraseologia, e a relevância das duas áreas para o desenvolvimento de pesquisas na área dos estudos sobre colocações especializadas, principalmente, no que tange à elaboração de materiais didáticos para o ensino de colocações, fornecendo, assim, subsídios teórico-práticos para a área da Pedagogia do Léxico Baseada em *Corpus*. Ademais, a partir da inserção de mais uma categoria (*Investigation*) no *Online English Collocations Workbook*, espera-se contribuir para o desenvolvimento do ensino-aprendizagem das colocações

especializadas da área investigativa criminal para os tradutores aprendizes, tradutores profissionais, bem como professores e alunos de inglês como língua estrangeira.

Em pesquisa futura, pretende-se utilizar as colocações obtidas nesta pesquisa em novas atividades que deverão ser desenvolvidas e incorporadas no *Online English Collocations Workbook*, conforme proposta do Projeto de Pesquisa “A compilação de materiais didáticos e glossários especializados baseados em corpora e sua contribuição para uma Pedagogia do Léxico e da Tradução”, além de utilizá-las para a compilação de um glossário de colocações especializadas da área criminal, de forma que estas colocações possam ser consultadas não apenas por alunos de língua inglesa, mas também, por profissionais da área que necessitem compreendê-las e utilizá-las em seu dia-a-dia.

REFERÊNCIAS

ANTHONY, L. **AntConc 3.4.4w**. 2014.

BALLY, C. **Précis de stylistique**. Genève: Eggiman, 1905.

_____. **Traité de stylistique française**. Paris: Klincksieck, 1909. v. 2

BANG, M.; FROMM, G. Terminologia em série: House M. D. **Entreletras**, v. 4, n. 2, p. 114-133, 2013.

BARROS, L. A. **Curso básico de Terminologia**. São Paulo: EDUSP, 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.

BEVILACQUA, C. R. Fraseologia: perspectiva da língua comum e da língua especializada. **Revista Língua e Literatura**, v. 6 e 7, n. 10/11, p. 73-86, 2004/2005.

_____. Unidades fraseológicas especializadas: estado da questão em relação a sua definição, denominação e critérios de seleção. **TradTerm**, v.11, p. 237-253, 2005.

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

BRETAÑA, J. M. P.; BERTRÁN, A. P. Combined statistical and grammatical criteria for the retrieval of phraseological units in an electronic corpus. In: GRANGER, S.; MEUNIER, F. **Phraseology: an interdisciplinary perspective**. Amsterdam / Philadelphia: John Benjamins, 2008.

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

CABRÉ, M. T. **La terminología: representación y comunicación**. Barcelona, Institut Universitari de Lingüística Aplicada. 1999.

_____. **La terminología: teoría, metodología, aplicaciones**. Barcelona, Antártida. 1993.

CASARES, J. **Introducción a la lexicografía moderna**. Madrid: Consejo Superior de Investigaciones Científicas, 1950.

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

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

CORPAS PASTOR, G. **Manual de fraseología española**. Madrid: Editorial Gredos, 1996.

CONZETT, J. Integrating collocation into a reading & writing course. In: LEWIS, M. **Teaching Collocation: further developments in the lexical approach**. Hove: Language Teaching Publications. 2000.

COSERIU, E. Structure lexicale et enseignement du vocabulaire. Actes du Premier Colloque International de Linguistique Appliquée, Nancy, 26-31 octobre, 1964. Faculté des Lettres et des Sciences Humaines de l'Université de Nancy. **Annales de l'Est**, Memorie, v. 31, p. 175-217, 1966.

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

ESPERANDIO, I. B. **Legendas de seriados de tema sobrenatural: uma abordagem terminológica para tradutores**. 2015. 229 f. Dissertação (Mestrado em Teorias Linguísticas do Léxico) – Universidade Federal do Rio Grande do Sul, Porto Alegre. 2015.

_____.; FINATTO, M. J. B. A definição terminológica na legendagem de seriados. **Cadernos de Letras**, n. 22, p. 17-38, 2014.

FINATTO, M. J. B.; ZILIO, L. **Textos e termos por Lothar Hoffman**. Porto Alegre: Palotti, 2015.

FIRTH, J. R. **Papers in linguistics 1934-1951**. Londres, Oxford University Press, 1957

FRITZINGER, F.; HEID, U. Automatic grouping of morphologically related collocations. **Proceedings of the Corpus Linguistics 2009 Conference**, Liverpool/UK, 2009.

FROMM, G. Ficção, tradução, terminografia e linguística de corpus: confluências. **Anais do SILEL**. Uberlândia: EDUFU, v. 2, v 2. 2011.

GRANGER, S. Prefabricated patterns in advanced EFL writing: collocations and formulae. In: Cowie A. P. (ed.). **Phraseology: theory, analysis and applications**, Oxford: Oxford University Press, 1998, p. 145-160

_____.; MEUNIER, F. **Phraseology: an interdisciplinary perspective**. Amsterdam / Philadelphia: John Benjamins, 2008.

_____.; PAQUOT, M. Disentangling the phraseological web. In: GRANGER, S.; MEUNIER, F. **Phraseology: an interdisciplinary perspective**. Amsterdam / Philadelphia: John Benjamins, 2008.

GRIES, S. T. Phraseology and linguistic theory: a brief survey. In: GRANGER, S.; MEUNIER, F. **Phraseology: an interdisciplinary perspective**. Amsterdam / Philadelphia: John Benjamins, 2008.

HALLIDAY, M. A. K. et al. Lexis as a linguistic level. In: _____. et al. **In memory of JR Firth**, p. 162, 1966, v. 148

HAUSMANN, F. J. Kollokationen im deutschen wörterbuch: ein beitrag zur theorie des lexikographischen biespiels. In: BERGENHOLTZ, H; MUGDAN, J. (eds). **Lexikographie und Grammatik**. Tübingen: Niemeyer, 1985.

_____. Wortschatzlernen ist kollokationslernen: zum lheren und lernen französischer wortverbindungen. **Praxis des neusprachlichen Unterrichts**, 31, p. 305-406. 1984.

HEID, U.; MARTIN, W.; POSCH, I. An overview of approaches towards the description of collocations. In: **Feasibility of standards for collocational description of lexical items**. In: **Eurotra 7 – Report**, Stuttgart/Amsterdam, 1991.

HILL, J.; LEWIS, M.; LEWIS, M. Classroom strategies, activities and exercises. In: **Teaching Collocation: Further Developments in the Lexical Approach**. Hove: Language Teaching Publications, 2000.

HOEY, M. **Patterns of lexis in text: describing English language**. Oxford: Oxford University Press, 1991.

HOFFMANN, L. Die rolle der fachsprachen seit der mitte des 20. Jahrhunderts. In: BESCH, W. et al (Orgs) **Sprachgeschichte**. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. 2 ed. Berlin: Walter de Gruyter, 2000. v. 2.

IÑESTA MENA, E. V.; PAMIES BERTRÁN, A. **Fraseología y metáfora: aspectos tipológicos y cognitivos**. Granada: Granada Linguistica, 2002.

KENNEY, G. Phraseology and language pedagogy: semantic preference associated with English verbs in the British National Corpus. In: GRANGER, S.; MEUNIER, F. **Phraseology in Foreign Language Learning and Teaching**. Amsterdam: John Benjamins. 2008.

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**. Ann Arbor 61:2, 2011, pp. 647-672.

LEWIS, M. **Teaching Collocation: further developments in the lexical approach**. Hove: Language Teaching Publications. 2000.

_____. **The Lexical Approach: the state of ELT and a way forward**. London: Language Teaching Publications. 1993.

LEWIS, M. There is nothing as practical as a good theory. In: **Teaching Collocation: further developments in the lexical approach**. Hove: Language Teaching Publications, 2000.

L'HOMME, M. C. Understanding specialized lexical combinations. In: **Terminology**, Amsterdam, v. 6, n. 1, 2000, p. 89-110.

L'HOMME, M. C.; BERTRAND, C. Specialized lexical combinations: should they be described as collocations or in terms of selectional restrictions? In: **Proceedings**. Ninth Euralex International Congress, p. 497-506. 2000

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

MCENERY, T.; HARDIE, A. **Corpus linguistics: method, theory and practice**. Cambridge University Press. 2012

MIRONESKO, E. **Fraseología rusa: teoría y práctica**. Barcelona: Librería Universitaria, 1997.

MONTEIRO-PLANTIN, R. S. **Fraseologia: era uma vez um Patinho Feio no ensino de língua materna**. Fortaleza: Imprensa Universitária, 2014.

NAVARRO, S. **Glossário bilíngue de colocações da hotelaria: um modelo à luz da linguística de corpus**. 2011. 249 f. Dissertação de Mestrado – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo, 2011.

NESSELHAULF, N. **Collocations in a learner corpus**. Amsterdam: John Benjamins Publishing Company, 2005.

ORENHA-OTTAINO, A. **A compilação de um glossário bilíngüe de colocações, na área de jornalismo de negócios, baseado em corpus comparável**. 2004. 246 f. Dissertação (Mestrado em Estudos Linguísticos e Literários) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2004.

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

_____. Collocations and the design of teaching materials for second language learners. In: **TALC10 Proceedings of the 10th International Conference on Teaching and Language Corpora**, Varsóvia, Polônia. Warsaw: Institute of Applied Linguistics, 2013a, p. 93-103.

_____. 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**, Belo Horizonte, v. 23, n. 3, p. 833-881, 2015.

_____. 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, p. 241-251, 2012.

_____. **Online English Collocations Workbook**. Disponível em: <http://www.gbd.institucional.ibilce.unesp.br/workbook/index>. Acesso em: 30 ago. 2017.

_____. 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**. Campinas: Editora Pontes, v.2, p.147- 163, 2012a.

ORENHA-OTTAIANO, A. The compilation of an online corpus-based bilingual collocations dictionary. In: CORPAS PASTOR, G. (Org). **Computerized and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives**. Genebra: Editions Tradulex, 2016, v. 1, p. 486-493.

_____. **Unidades fraseológicas especializadas: colocações e colocações estendidas em contratos sociais e estatutos sociais traduzidos no modo juramentado e não-juramentado**. 2009. 282 f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociência, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2009.

_____.; ROCHA, J. M. P. Colocações especializadas na área médica extraídas a partir do corpus House M.D. **Cadernos do IL**, Porto Alegre, n.44, p. 295-318, 2012.

_____. Ensino de inglês como LE e contribuições pedagógicas de um glossário bilíngue de colocações. **Signótica**, Goiânia, v. 27, n. 2, p. 485-510, 2015.

OXFORD. **Oxford collocations dictionary for students of English**. Oxford: Oxford, 2012.

_____. **Oxford student's dictionary of English**. Oxford: Oxford, 2013.

PARTINGTON, A. **Patterns and meanings: using corpora for English language research and teaching**. Philadelphia: John Benjamins Publishing Company, 1998.

POLIVÁNOV, E. D. **Travaux du Cercle Linguistique de Prague**, 1931.

QUAQLIO, P. **Television dialogue: the sitcom Friends vs. natural conversation**. Amsterdam / Philadelphia: John Benjamins, 2009

ROCHA, J. M. P. **Fraseologia jurídico-comercial e proposta de um glossário de colocações especializadas trilingue baseado em corpus**. 2017. 292 f. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto. 2017.

SAUSSURE, F. **Curso de linguística geral**. São Paulo: Cultrix, 1969.

SAYURI, J. Legendários. **Superinteressante**. São Paulo, 2011.

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

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

_____. **Looking up: an account of the COBUILD project in lexical computing**. London: Collins Cobuild, 1987.

SUMMERS, D. **Longman Dictionary of Contemporary English**. Essex: Pearson Education Limited, 2003.

TAGNIN, S. E. O. Collectiong data for a bilingual dictionary of verbal collocations: from scraps of paper to corpora research. In: LEWANDOSKA-TOMASZCZYK, B.; MELIA, P. J. (eds). *Palc' 99: Pratical Applications in Language Corpora. Papers from the International Conference at the University of Lodz*, 15-18 April 1999. Frankfurt and Main: Peter Lang GmbH, p. 399-407, 1999.

_____. **Expressões idiomáticas e convencionais**. São Paulo: Ática, 1989.

_____. **O jeito que a gente diz**: combinações consagradas em inglês e português. Barueri: DISAL, 2013.

TELIA, V. N. Acerca de los tipos y modos de formación de los fraseologismos. In: _____. **Problemas de la formación de los fraseologismos en ruso**. Moscú, 1963.

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

TRISTÁ, A. M. **Fraseología y contexto**. Cuba: Editorial de Ciencias Sociales La Habana, 1988.

_____.; CARNEADO MORE, Z. V. **Estudios de fraseologia**. La Habana: Academia de Ciencias de Cuba, 1985.

VAN ROEY J. **French-English contrastive lexicology**: an introduction. Louvain-la-Neuve : Peeters.1990.

VINOGRADOV, V. V. Ob osnovnuikh tipakh frazeologicheskikh edinit v russkom yazuike. In: SHAKHMATOV, A. A. **Sbornik statey i materialov**. Moscow: Nauka, 1947, p. 339-64.

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**, v. 45, n. 2, p. 291 – 312, 2011.

WebCorp. Disponível em: <http://www.webcorp.org.uk/live/>. Acesso em: 20 abr. 2016.

WOOLARD, G. Collocation – encouraging learner independence. In: **Teaching Collocation**: further developments in the lexical approach. Hove: Language Teaching Publications, 2000.

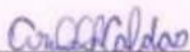
ZANCHETA, E.; BARONI, M.; BERNARDINI, S. **Boot Cat Front End**, versão 0.71. 2013

ZULUAGA, A. **Introducción al estudio de las expresiones fijas**. Frankfurt: Peter Lang, 1980.

TERMO DE REPRODUÇÃO XEROGRÁFICA

Autorizo a reprodução xerográfica do presente Trabalho de Conclusão, na íntegra ou em partes, para fins de pesquisa.

São José do Rio Preto, 06 / 09 / 2014



Assinatura do autor