



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

Emanoel Henrique Alves

**Uma investigação da tradução de colocações a partir de um *corpus* paralelo
de textos jornalísticos de cunho político**

São José do Rio Preto

2021

Emanoel Henrique Alves

Uma investigação da tradução de colocações a partir de um *corpus* paralelo de textos jornalísticos de cunho político

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.

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

São José do Rio Preto

2021

A474i Alves, Emanuel Henrique
Uma investigação da tradução de colocações a partir de um corpus paralelo de textos jornalísticos de cunho político. / Emanuel Henrique Alves. -- São José do Rio Preto, 2021
155 f. : tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientadora: Adriane Orenha-Ottaiano

1. Pedagogia do léxico e da tradução. 2. Corpus de aprendizes. 3. Fraseologia. 4. Colocações. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Emanoel Henrique Alves

**Uma investigação da tradução de colocações a partir de um *Corpus* paralelo
de textos jornalísticos de cunho político**

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”, Câmpus de São José do Rio Preto.

Comissão Examinadora

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

Prof^a. Dr^a. Maria Cristina Parreira da Silva
UNESP – Campus de São José do Rio Preto

Prof. Dr. Odair Luiz Nadin da Silva
UNESP – Campus de Araraquara

São José do Rio Preto

12 de novembro de 2020

A Deus, aos meus pais, José e Alda, pelo apoio nessa jornada. À família Araújo, em especial ao meu avô, Alcides. À família Alves, as minhas tias, Maria, Josefina, Lourdes às minhas primas, Valcenice e Mariana, D. Carol Martins e família. Aos meus avós, Jeronima, Maria e Manoel (*in memoriam*).

AGRADECIMENTOS

A minha orientadora, Prof^ª. Dr^ª. Adriane Orenha-Ottaiano, por me aceitar como orientando, por sua maestria em me auxiliar nesse processo de pesquisa e por acreditar nesta proposta de trabalho e em minha capacidade de desenvolvê-la.

Minha eterna gratidão à coordenadora, Prof^ª. Dr^ª. Patrícia Viana Belan, e à Prof^ª. Dr^ª. Leila Maria Gumushiam Felipini por, gentilmente, aceitarem supervisionar a aplicação do projeto no Centro Universitário Sagrado Coração e pelo carinho em me ajudar a expandir os meus horizontes com a realização desta investigação.

À Prof^ª. Dr^ª. Ketilin Mayra Pedro pela emissão da carta de anuência e aos alunos de Letras-Tradutor (2018 e 2019) pelo tempo dedicado a realizarem as traduções dos textos com tanto empenho.

Também agradeço a Prof^ª. Dr^ª. Stella Esther Ortweiler Tagnin pelo tempo dedicado à leitura do texto e pelas sugestões dadas durante a apresentação na modalidade painel, no VIII Seminário de Estudos Linguísticos da UNESP, em Araraquara. De igual modo, minha gratidão à Prof^ª. Dr^ª. Angela Maria Tenório Zucchi por suas indicações como debatedora no XI Seminário de Estudos Linguísticos da UNESP.

Agradeço ainda as professoras Dr^ª. Maria Cristina Parreira e Dr^ª. Paula Tavares Pinto, por aceitarem participar da banca de qualificação e conduzi-la com tanta clareza.

Também agradeço as minhas amigas, irmãs de coração e companheiras de mestrado, Carina Beltramini e Giseli Aparecida Cecílio, pelas conversas, aprendizados e momentos de pesquisas juntos.

Aos orientandos da minha orientadora, em especial às colegas, Prof^ª. Me. Elaine Cristina Oliveira e Prof^ª. Me. Ariane Donizete Delgado Ribeiro Caldas, pelo auxílio prestado, e ao Prof. Me. Jean Michel Pimentel Rocha, por compartilhar do mesmo tipo de projeto.

Agradeço também aos professores, Dr^ª. Marize Dall'Aglio Hattner, Dr. Eduardo Penhavel, Dr^ª. Paula Tavares Pinto, Dr^ª. Maria Emília Pereira Chanut e Dr^ª. Ana Frankenberg-Garcia, pelas disciplinas ministradas. Ao Prof. Dr. Eduardo Batista pelo minicurso ofertado e à Prof^ª. Dr^ª. Diva Cardoso de Camargo pela bela apresentação sobre a sua trajetória de pesquisa.

À Prof^ª. Dr^ª. Maria Cristina Parreira da Silva e ao Prof. Dr. Odair Luiz Nadin da Silva por aceitarem cordialmente compor como titulares a banca de defesa deste trabalho, bem como às suplentes, Prof^ª. Dr^ª. Regiane Aparecida Santos Zacarias e Prof^ª. Dr^ª. Paula Tavares Pinto.

À minha tia Lídia Alves de Araújo, José Araújo e Paulo Alves de Araújo. Agradeço ainda, de coração, aos meus amigos, Eduardo Pescinini Ruli, Lucas de Rocco Moraes e Mariana de Aragão Bologna, pela compreensão e amizade desenvolvida em todas as etapas.

Por fim, agradeço, ainda, às minhas amigas Ilze Baci, Carla Renata Rodrigues e Sônia Alves Pachoal pelo incentivo e apoio durante a trajetória de mestrado.

“O leitor há de se lembrar de que mencionamos que certas palavras parecem combninar-se de forma natural, não havendo, via de regra, explicação” (TAGNIN, 2013, p. 63).

RESUMO

Esta pesquisa buscou verificar as traduções das colocações empregadas no *Corpus* de Aprendizes de Tradução 2 (CAT 2), por meio de um modelo descritivo de análise estatística de dados. O CAT 2 é uma extensão do projeto CAT 1 e ambos fazem parte do projeto guarda-chuva “Compilação de Materiais Didáticos e Dicionários Especializados de Colocações Baseados em *Corpora* (COMADEC)”, iniciado em 2012, na Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Campus de São José do Rio Preto. O estudo se pautou nas teorias da Linguística de *Corpus* (MCENERY; WILSON, 1996; MCENERY; HARDIE, 2012; SINCLAIR, 2004; TOGNINI-BONELLI, 2001) e dos Estudos Fraseológicos (ORENHA-OTTAIANO, 2004, 2015; MARTELLI, 2007; TAGNIN, 2013), com foco nas colocações que corroboram com reflexões para as questões pedagógicas do léxico e da tradução baseada em *corpus*. Os participantes da investigação são alunos do curso de Bacharelado em Letras com Habilitação para Tradutor, do Centro Universitário Sagrado Coração (UNISAGRADO), de Bauru. Os partícipes deveriam apresentar níveis de proficiência em língua inglesa entre B2 e C2. Para isso, o pesquisador ofereceu um teste escrito (ALLAN, 2004) a fim de classificá-los entre os referidos níveis. O CAT 2 contém dois *subcorpora*, sendo um de 3 mil palavras, com seis textos originais em Português brasileiro, composto por seis textos jornalísticos com extensão de 500 a 800 palavras e, outro, com seus respectivos textos traduzidos para o inglês, com 95.221 mil palavras em língua inglesa. Para avaliar as colocações no CAT 2 (*subcorpus* de TOs e *subcorpus* de TTs), a metodologia contou com as ferramentas *Keywords*, *Word Sketch* e *Concordance* do gerenciador de *corpus* *Sketch Engine* (KILGARRIFF *et al.*, 2014) e o *corpus* de referência *English Web 2013* (SUCHOMEL; POMIKÁLEK, 2012; JAKUBÍČEK *et al.*, 2013) alocado nessa mesma plataforma e que serviu para comprovar se os exemplos de colocações levantadas no CAT 2 para a discussão de dados eram empregadas ou não em língua inglesa, com vistas à apontar ainda possíveis dificuldades tradutórias. Depois de observadas as colocações, propomos modelos de atividades baseados nos problemas tradutórios encontrados por meio da *Online English Collocations Platform* e, em seguida, outra atividade com base no *English Web 2013* e na ferramenta *SKELL*. A análise de dados ocorreu a partir de seis exemplos de colocações extraídas do CAT 2, com o intuito de descrever as escolhas tradutórias utilizadas pelos aprendizes com relação às unidades fraseológicas selecionadas. Como resultado, o presente estudo comprovou que ao buscarem correspondentes para colocação “fazer a segurança” em inglês, os estudantes empregaram os equivalentes *do/make security*, sendo que não encontramos ocorrências combinadas entre essas partes de léxico no *corpus* de referência. Pelo contrário, o *English Web 2013* apresentou somente as palavras como colocados de *security*. Assim, observou-se a viabilidade de incentivar mais trabalhos com base nesse modelo de *corpus*, uma vez que dificuldades são comuns nesse grupo, muitas vezes, devido à ingenuidade do tradutor aprendiz. Portanto, ficou ainda comprovado que o modelo de exploração linguística empregado pode contribuir para a melhoria da competência fraseológica do aluno de tradução, servindo de indício para a elaboração de atividades e materiais pedagógicos com base em *corpus* para fins tradutórios.

Palavras-chave: Pedagogia do Léxico e da Tradução. *Corpus* de Aprendizes. Fraseologia. Colocações. Português Brasileiro. Língua Inglesa.

ABSTRACT

This research aimed to verify the translations of the collocations used in the Translation Learner Corpus 2 (CAT 2), using a descriptive model of statistical data analysis. CAT 2 is a continuation of the CAT 1 project, and both are part of the project called *The Compilation of Corpus-based Didactic Materials and Specialized Dictionaries of Collocations* (COMADEC), started in 2012, at São Paulo State University “Júlio de Mesquita Filho” (UNESP), in São José do Rio Preto. The study was based on the theories of Corpus Linguistics (MCENERY; WILSON, 1996; MCENERY; HARDIE, 2012; SINCLAIR, 2004; TOGNINI-BONELLI, 2001) and Phraseological Studies (ORENHA-OTTAIANO, 2004, 2015; MARTELLI, 2007; TAGNIN, 2013), focusing on collocations, which corroborate reflections on the pedagogical issues of the lexicon and corpus-based translation. The research participants are students of B.A. in Translation, from the University Center Sagrado Coração (UNISAGRADO), in Bauru. Participants should have levels of proficiency in English between B2 and C2. For this purpose, the researcher offered a written test (ALLAN, 2004), in order to classify them among those levels. CAT 2 contains two subcorpora, one of them with 3,000 words, with six original texts in Brazilian Portuguese, consisting of six journalistic texts ranging from 500 to 800 words, and the other one with their respective texts translated into English, with 95,221 words in English. To assess CAT 2 collocations (TOs subcorpus and TTS subcorpus), the methodology used the *Keywords*, Word Sketch and Concordance tools of the *Sketch Engine* corpus manager (KILGARRIFF *et al.*, 2014), and the corpus of reference *English Web 2013* (SUCHOMEL; POMIKÁLEK, 2012; JAKUBÍČEK *et al.*, 2013) allocated on the same platform and served to prove whether the examples of collocations found in CAT 2 for the discussion of data were used or not in English, in order to point out possible translation difficulties. After observing the collocations, we propose activity models based on the translation problems found through the *Online English Collocations Platform* and then another activity based on the *English Web 2013* and the *SKELL* tool. The data analysis occurred from six examples of collocations extracted from CAT 2 to describe the learners’ translation choices in relation to the selected phraseological units. As a result, the present study proved that when searching for correspondents for “fazer a segurança” in English, students used the equivalents do/make security, and we did not find any combined occurrences between these parts of the lexicon in the reference corpus. On the contrary, *English Web 2013* presented only the words as collocated of security. Thus, the feasibility of encouraging more work based on this corpus model was observed since collocational difficulties are common in this group, often due to the translation learner’s naivety. Therefore, it was also proven that the linguistic exploration model employed could contribute to the improvement of the phraseological competence of the translation learner, serving as an indication for the elaboration of educational activities and materials based on corpus for translation purposes.

Keywords: Lexicon and Translation Pedagogy. Learner Corpus. Phraseology. Collocations. Brazilian Portuguese. English Language.

LISTA DE FIGURAS

Figura 1 – Convencionalidade, Fraseologismos e Colocações.....	33
Figura 2 – Níveis de convencionalidade.....	34
Figura 3 – Imagem de apresentação de trabalho	66
Figura 4 – <i>Common European Framework of Reference</i> (CEFR)	70
Figura 5 – Guia de usuário do <i>Sketch Engine</i>	71
Figura 6 – <i>Dashboard do Sketch Engine</i>	73
Figura 7 – Etapa de organização dos textos	75
Figura 8 – Etapa de criação do <i>corpus</i> de estudo	75
Figura 9 – Etapa de adicionar textos no <i>corpus</i>	76
Figura 10 – Etapa de compilação estatística dos textos no <i>corpus</i>	76
Figura 11 – Primeiras palavras-chave do CAT 2	78
Figura 12 – Tópico download da ferramenta <i>Keywords</i>	78
Figura 13 – Relações gramaticais com a palavra <i>security</i> na <i>Word Sketch</i>	83
Figura 14 – Item avançado da ferramenta <i>Concordance</i>	84
Figura 15 – Linhas de concordância de <i>international</i>	84
Figura 16 – <i>Collocations</i> na ferramenta <i>Concordance</i>	84
Figura 17 – <i>Logdice</i> no item <i>Collocations</i> da ferramenta <i>Concordance</i>	85
Figura 18 – Imagem de apresentação da plataforma	86
Figura 19 – <i>Login</i> da plataforma	87
Figura 20 – Inserção da atividade <i>Gap Fill</i>	87
Figura 21 – Inserções da atividade <i>Memory Game</i>	88
Figura 22 – Tela inicial da plataforma <i>SKELL</i>	89
Figura 23 – Tela inicial da <i>Online English Collocations Platform</i>	120
Figura 24 – Tela inicial do exercício <i>Memory Game</i>	121
Figura 25 – Tela instrucional do exercício <i>Memory Game</i>	121
Figura 26 – Tela inicial de início exercício <i>Memory Game</i>	122
Figura 27 – Tela inicial de início exercício <i>Memory Game</i>	122
Figura 28 – Tela dos <i>cards</i> embaralhados – Exercício <i>Memory Game</i>	123
Figura 29 – Resultado do exercício <i>Memory Game</i>	123
Figura 30 – Tela inicial do exercício <i>Gap Fill</i>	124
Figura 31 – Tela inicial de início exercício <i>Gap Fill</i>	125
Figura 32 – Tela dos exercícios de <i>Gap Fill</i>	125
Figura 33 – Tela inicial de início exercício <i>Gap Fill</i>	126
Figura 34 – Resultado do exercício <i>Gap Fill</i>	126
Figura 35 – Coocorrências com a palavra <i>Immigrants</i> no <i>Word Sketch</i>	128
Figura 36 – Linhas de concordância da colocação <i>immigrant visa</i>	128
Figura 37 – Verbos com “segurança”	129
Figura 38 – Sugestão de colocado para a base “ <i>security</i> ”	130
Figura 39 – Linhas de concordância a palavra-chave <i>border</i>	131
Figura 40 – Relações gramaticais com a palavra-chave <i>border</i>	132
Figura 41 – Relações gramaticais com a palavra-chave <i>border</i>	132

LISTA DE TABELAS

Tabela 1 – Resultados do <i>Placement Test</i> com os alunos do 3º ano	90
Tabela 2 – Resultados do <i>Placement Test</i> com os alunos do 2º ano	91
Tabela 3 – Frequência geral de contagem do <i>Subcorpus</i> de TOs e TTs	91
Tabela 4 – Formas verbais, traduções, frequências e <i>LogDice</i> da colocação “arrumar confusão”	94
Tabela 5 – Traduções, exemplos e frequências da colocação “protagonismo internacional” ..	97
Tabela 6 – Taxonomia, traduções, frequência e <i>LogDice</i> da colocação “protagonismo internacional” ..	98
Tabela 7 – Sugestões, frequência e <i>LogDice</i> da colocação “traçar a estratégia” no <i>English Web 2013</i>	101
Tabela 8 – Sugestões, frequência e <i>LogDice</i> das colocações no <i>English Web 2013</i>	108
Tabela 9 – Sugestões, frequência e <i>LogDice</i> das colocações no <i>English Web 2013</i>	111
Tabela 10 – Sugestões, frequência e <i>LogDice</i> das colocações no <i>English Web 2013</i>	114
Tabela 11 – Taxonomia, sugestões, frequências e <i>LogDice</i> das colocações no <i>English Web 2013</i>	116

LISTA DE QUADROS

Quadro 1 – Diferentes denominações dos estudos fraseológicos.....	39
Quadro 2 – Características das UFs 1.....	40
Quadro 3 – Características das UFs 2.....	40
Quadro 4 – Tipologias das UFs.....	41
Quadro 5 – Estrutura fraseológica de Pavel (1993).....	44
Quadro 6 – Estrutura fraseológica de Blais (1993).....	44
Quadro 7 – Caracterização dos agentes participantes da pesquisa.....	68
Quadro 8 – Grade curricular do curso de Letras-Tradutor.....	69
Quadro 9 – Títulos dos textos originais.....	74
Quadro 10 – Extração das <i>keywords</i> do CAT 2 pelo <i>Microsoft Excel</i>	79
Quadro 11 – Quadro resumo da tipologia do <i>corpus</i> de estudo.....	80
Quadro 12 – Colocações nos TOs e suas traduções no CAT 2.....	92
Quadro 13 – Colocação “arrumar confusão” (TO).....	92
Quadro 14 – Colocação “arrumar confusão”: traduções, frequência e exemplos em inglês no CAT 2.....	92
Quadro 15 – Linhas de concordância extraídas do <i>English Web 2013</i>	94
Quadro 16 – Quadro resumo da colocação “arrumar confusão”.....	96
Quadro 17 – Colocação “protagonismo internacional” (TO).....	96
Quadro 18 – Linhas de concordância extraídas do <i>English Web 2013</i>	98
Quadro 19 – Quadro resumo da colocação “protagonismo internacional”.....	100
Quadro 20 – Colocação “traçar a estratégia” (TO).....	100
Quadro 21 – Traduções, exemplos e frequências da colocação “protagonismo internacional”.....	101
Quadro 22 – Linhas de concordância extraídas do <i>English Web 2013</i>	102
Quadro 23 – Linhas de concordância extraídas do <i>English Web 2013</i>	103
Quadro 24 – Quadro resumo da colocação “traçar a estratégia”.....	104
Quadro 25 – Quadro resumo da colocação “traçar a estratégia”.....	104
Quadro 26 – Colocação “fechamento de fronteiras” (TO).....	104
Quadro 27 – Traduções, exemplos e frequências da colocação “fechamento de fronteiras”.....	105
Quadro 28 – Sugestões, frequência e <i>LogDice</i> da colocação “traçar a estratégia” no <i>English Web 2013</i>	105
Quadro 29 – Linhas de concordância extraídas do <i>English Web 2013</i>	106
Quadro 30 – Resumo da colocação “fechamento de fronteiras”.....	107
Quadro 31 – Colocação “amparar imigrantes” (TO).....	107
Quadro 32 – Traduções, exemplos e frequências da colocação “amparar imigrantes”.....	107
Quadro 33 – Linhas de concordância extraídas do <i>English Web 2013</i>	108
Quadro 34 – Resumo da colocação “amparar imigrantes”.....	109
Quadro 35 – Colocação “Projetar a vida” (TO).....	110
Quadro 36 – Traduções, exemplos e frequências da colocação “projetar a vida”.....	110
Quadro 37 – Linhas de concordância extraídas do <i>English Web 2013</i>	111
Quadro 38 – Resumo da colocação “projetar a vida”.....	112
Quadro 39 – Colocação “sucessivos atritos” (TO).....	113

Quadro 40 – Traduções, exemplos e frequências da colocação “sucessivos atritos”	113
Quadro 41 – Linhas de concordância extraídas do <i>English Web 2013</i>	114
Quadro 42 – Resumo da colocação “sucessivos atritos”	115
Quadro 43 – Colocação “fazer segurança” (TO).....	115
Quadro 44 – Traduções, exemplos e frequências da colocação “fazer segurança”	116
Quadro 45 – Linhas de concordância extraídas do <i>English Web 2013</i>	117
Quadro 46 – Resumo da Colocação “fazer segurança”	117
Quadro 47 – Quatro etapas de discussões das colocações.....	118
Quadro 48 – Resumo das UFs identificadas a partir da palavra <i>immigrant</i>	129
Quadro 49 – Contexto para tradução	129

LISTA DE ORGANOGRAMAS

Organograma 1 – Algumas ferramentas do gerenciador de <i>Corpus Sketch Engine</i>	77
Organograma 2 – Estrutura do <i>corpus</i> de estudo	80
Organograma 3 – Processo geral de análise	81

LISTA DE ABREVIATURAS

CAT 1	<i>Corpus</i> de Aprendizes de Tradução 1
CAT 2	<i>Corpus</i> de Aprendizes de Tradução 2
CEFER	<i>Common European Framework of Reference</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COMADEC	Compilação de Materiais Didáticos e Dicionários Especializados de Colocações Baseados em <i>Corpora</i>
DDL	<i>Data Driven Learning</i>
F.	Frequência
FRASCORP	Fraseologia e Colocações Baseado em <i>Corpora</i>
LC	Linguística de <i>Corpus</i>
LE	Língua Estrangeira
TCLE	Termo de Consentimento Livre e Esclarecido
TOs	Textos originais
TTs	Textos Traduzidos
UF	Unidade Fraseológica
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
UNISAGRADO	Centro Universitário Sagrado Coração
UT	Unidade Terminológica

SUMÁRIO

1 INTRODUÇÃO	17
2. FUNDAMENTAÇÃO TEÓRICA.....	21
2.1 A Linguística de <i>Corpus</i>	21
2.2 A Linguística de <i>Corpus</i> e suas conexões com a Tradução	24
2.3 <i>Corpus</i> de aprendizes	26
2.4 Léxico e Tradução: um olhar pedagógico baseado em <i>corpus</i>	29
2.5 A convencionalidade	33
2.5.1 A <i>Fraseologia</i> e algumas contribuições.....	35
2.5.2 As diferentes tipologias dos fraseologismos	38
2.5.3 A <i>Fraseologia</i> na língua geral e na língua especializada.....	42
2.6 A interface entre a <i>Fraseologia</i> e a Linguística de <i>Corpus</i> na contemporaneidade	46
2.7 As colocações.....	500
2.7.1 A <i>taxonomia das colocações</i>	53
2.7.2 As <i>colocações de língua geral e de língua especializada</i>	55
2.8 Os erros colocacionais em <i>corpus</i> de aprendizes.....	59
3. METODOLOGIA.....	65
3.1 O projeto <i>Corpus</i> de Aprendizes de Tradução	65
3.2 Processo de coleta de dados e os documentos legais do Comitê de Ética	67
3.2.1 Aspectos relevantes sobre a instituição locus da pesquisa e o curso	68
3.2.2 Os procedimentos e os critérios de seleção dos participantes.....	69
3.3 As ferramentas e os <i>corpora</i> empregados na pesquisa	71
3.3.1 A ferramenta de análise lexical de <i>corpus</i> <i>Sketch Engine</i>	71
3.3.2 Procedimento de compilação do <i>corpus</i> e análise de dados.....	74
3.3.3 Levantamento e análise dos dados no <i>Corpus de Aprendizes de Tradução 2 (CAT 2)</i>	77
3.3.4 Estrutura e etiquetagem do <i>corpus</i> do estudo	79
3.3.5 O <i>corpus</i> de referência utilizado como base na pesquisa.....	81
3.3.6 <i>LogDice</i> : procedimento estatístico de análise das colocações.....	82
3.3.7 Estratégia de atividade: O <i>corpus</i> de referência <i>English Web 2013</i> , ferramenta <i>SKELL</i> e a <i>Online English Collocations Platform</i> empregados para fins pedagógicos.....	85

3.3.8	<i>Descrição da Online English Collocations Platform e suas atividades com jogos</i>	86
3.3.9	<i>Breve descrição de uso da ferramenta SKELL</i>	88
4.	ANÁLISE E DISCUSSÃO DE DADOS	90
4.1	Os resultados do teste de proficiência	90
4.1.1	<i>Informações do CAT 2 e colocações selecionadas para análise e discussão dos dados</i>	91
4.2	Descrições quantitativa, qualitativa e estatística das colocações da pesquisa.....	92
5.	PROPOSTA DE ATIVIDADES A PARTIR DA ONLINE ENGLISH COLLOCATIONS PLATFORM E EXERCÍCIOS COM O USO DO ENGLISH WEB 2013 E DA FERRAMENTA SKELL	119
5.1	As atividades colocacionais com base no CAT 2 inseridas na <i>Online English Collocations Platform</i>	120
5.2	Proposta de atividade baseada no <i>corpus</i> de referência <i>English Web 2013</i>	127
5.3	Proposta de atividade baseada na ferramenta <i>SKELL</i>	130
6	CONSIDERAÇÕES FINAIS	134
	REFERÊNCIAS	137
	ANEXOS A – Carta de anuência do Centro Universitário do Sagrado Coração.....	152
	ANEXO B – Termo de Consentimento Livre e Esclarecido.....	153
	ANEXO C – Termo de Consentimento Livre e Esclarecido: dados do projeto de pesquisa.....	154
	ANEXO D – Parecer de aprovação.....	155

1 INTRODUÇÃO

O desenvolvimento de pesquisas com interesse na exploração científica por meio do *corpus* de aprendizes tradutores vem se tornando cada vez mais necessário, sobretudo quanto à investigação das unidades fraseológicas, tais como expressões, provérbios, colocações, entre outros.

Os fraseologismos, “combinações formadas por pelo menos duas palavras gráficas” (AMADEU-SABINO, 2011, p. 389), são transmitidos de geração em geração. A colocação, por sua vez, fenômeno linguístico escolhido a ser analisado neste estudo, é um dos grandes gargalos, de acordo com as ciências do léxico, uma vez que o emprego adequado dessa unidade fraseológica é sempre um desafio para a tradução, principalmente para aprendizes dessa área, os quais precisam lidar ainda com outros aspectos linguísticos no texto (ORENHA-OTTAIANO, 2012a, 2012b, no prelo; TAGNIN, 2013).

As análises científicas, em relação ao uso das colocações, por exemplo, vêm crescendo e beneficiando estudantes dessa área (ALVES; ORENHA-OTTAIANO, 2020; MURAKAMI, 2016; ORENHA-OTTAIANO, 2012a, 2012b, 2015, no prelo; TAGNIN, 2013, dentre outros). Porém, há um longo caminho a ser trilhado. O exercício de encontrar uma tradução para um termo de uma língua para outra é uma atividade que requer muita investigação por parte do tradutor, pois não se trata apenas de elencar um item lexical isolado, mas de dois ou mais que se combinam, muitas vezes, restritamente, cujo significado deve ser entendido pela combinação como um todo.

Dessa forma, a escolha de um conceito exato e adequado para um determinado contexto exige cuidados, principalmente, por causa dos aspectos convencionais da língua (TAGNIN, 2013) e da forte carga cultural apresentada por uma determinada fraseologia. Um exemplo disso é a colocação *have an accident*, em língua inglesa, cuja tradução culturalmente consagrada pelo uso em português, quando envolve um sujeito prejudicado, é “sofrer um acidente”. Neste caso, o verbo *have*, se isolado de *accident* perde o sentido colocacional, uma vez que, nesse caso, não deve ser traduzido como “ter”, mas empregado com o sentido de “sofrer”.

Sendo assim, nos referimos à colocação como uma combinação produzida por falantes reais em um contexto autêntico, que pode apresentar problemas de tradução, principalmente se o tradutor for aprendiz. Por esta razão, torna-se viável refletir sobre as dificuldades que os alunos brasileiros apresentam em relação ao emprego dos termos no idioma inglês à luz da pesquisa científica, enquanto parte da prática tradutória.

Fillmore (1979) alega que, quando um aprendiz desconhece as convenções de uma língua, este, por sua vez, torna-se um “falante ingênuo”, isto é, aquele que acaba realizando combinações literais e inexistentes na língua-alvo. Tagnin (2002) explora esse conceito de Fillmore (1979), aplicando-o à tradução e designando o “tradutor ingênuo” como aquele que desconhece as colocações na língua-alvo e que perde a naturalidade do texto traduzido, por se ater ao texto fonte.

Além do mais, mesmo que a internet, em sua rapidez, tenha se tornado a fonte de informação mais acessada na busca de palavras em outro idioma, essa ferramenta não é suficiente para identificar uma colocação com exatidão. Na tradução de textos para o inglês, um estudante que consultar fontes *on-line*, de forma involuntária, pode, muitas vezes, cometer equívocos na tentativa de combinar palavras intuitivamente. Esse fato pode ocorrer pela falta de conhecimento dos aspectos de convencionalidade, isto é, do contexto de uso da combinação a ser empregada.

Sendo assim, o objetivo principal desta pesquisa é investigar as traduções das colocações empregadas no *Corpus* de Aprendizes de Tradução 2 (CAT 2), por meio de um modelo descritivo de análise estatística de dados para fins pedagógicos. Como objetivos específicos, elencamos os seguintes:

1. Compilar um *corpus* de aprendizes e levantar as colocações;
2. Analisar as colocações de língua geral e especializada presentes nos textos originais CAT 2, com vistas ao levantamento de problemas ou dificuldades no emprego desta fraseologia;
3. Criar atividades a partir do CAT 2, utilizando os modelos de *Gap Fill* e *Memory Game* existentes na *Online English Collocations Platform*, conforme análise e discussão de dados;
4. Propor atividades colocacionais com base nas dificuldades discutidas por meio do *corpus* de referência e ferramentas de *corpus on-line*.

Para cumprir com os objetivos elencados, encontramos possibilidades de atingi-los por meio de um *corpus* de estudantes, pois julgamos ainda ser um modelo que pudesse contemplar esta proposta de trabalho, além de carecer de mais contribuições, tendo como sujeito de pesquisa o tradutor aprendiz, ainda em processo de formação na graduação.

Por conseguinte, esta investigação científica tem como base o CAT – *Corpus* de Aprendizes de Tradução (ORENHA-OTTAIANO, 2012b, 2015, no prelo), cuja metodologia e parâmetros foram delineados no âmbito do projeto guarda-chuva de “Compilação de

Materiais Didáticos e Dicionários Especializados de Colocações baseados em *corpora*” (COMADEC), iniciado em 2012 e ainda em vigência, sob a coordenação da Prof^a. Dr^a. Adriane Orenha-Ottaiano, da Universidade Estadual Paulista (UNESP). Este *corpus*, intitulado *Corpus* de Aprendizes de Tradução 1 (CAT 1), com aproximadamente mais de 100.000 palavras, é composto de traduções de textos jornalísticos, nas direções português brasileiro-inglês, cujos originais foram extraídos de jornais brasileiros *on-line*, sendo as traduções realizadas pelos discentes dos cursos de Bacharelado em Letras com Habilitação para Tradutor e Licenciatura em Letras, da referida universidade.

A viabilidade da expansão desse estudo nos permitiu a criação de um segundo, o CAT 2, atendendo aos mesmos parâmetros e critérios de coleta e metodologia do primeiro (ORENHA-OTTAIANO, 2012a, 2012b, no prelo). Para a sua formação, em forma de parceria, convidamos alunos do curso de Letras – Tradutor do Centro Universitário Sagrado Coração (UNISAGRADO), de Bauru (SP).

Por esse motivo, pautamos este trabalho nos estudos realizados no âmbito da linha de pesquisa “Pedagogia do Léxico e da Tradução baseada em *corpora*”, que prevê que as dificuldades encontradas em *corpus* de aprendizes dão a oportunidade de “elaborar atividades que atendam às suas necessidades, quanto à elaboração de materiais de ensino e de tradução baseados em *corpora*, os quais atestam o uso real e corrente dos idiomas tratados” (ORENHA-OTTAIANO; PINTO, 2018, p. 11).

Desse modo, a relevância social desta pesquisa se dá ao tentar observar as dificuldades dos acadêmicos tradutores com o uso das colocações e, por meio do *corpus*, contribuir com sugestões pedagógicas, visto que há poucas atividades de exploração na área da Linguística de *Corpus* (doravante LC) nesta área no Brasil.

A LC (BAKER, 1993; MCENERY; HARDIE, 2012) vêm oferecendo subsídios para a verificação de dados reais e a construção de *corpora* computadorizados com base no sistema probabilístico (HALLIDAY, 1991). É fato que o uso do computador na referida área tornou peça fundamental para o estudo do léxico, principalmente, quando nos referimos ao uso da colocação na língua. Tal postura possibilitou a análise da fraseologia por meio da compilação de textos em larga escala, facilitando a investigação dos aspectos convencionais do uso, especialmente quando empregadas milhares de vezes em um mesmo contexto.

A presente dissertação está organizada em quatro capítulos. No Capítulo 1, a fundamentação teórica aborda a LC, seguida dos Estudos da Tradução baseada em *corpus*, abrange reflexões sobre a Pedagogia da Tradução. Ainda são tratadas as questões da convencionalidade do léxico, as contribuições da Fraseologia e suas tipologias, retomando

características dessa área com base na língua geral e de especialidade, bem como a interface entre a Fraseologia e LC na contemporaneidade. No campo do objeto de estudo, a saber, as colocações (ORENHA-OTTAIANO, 2004, 2012a, 2012b, 2015, no prelo; MARTELLI, 2007; TAGNIN, 2013; HAUSMANN, 1985), exploramos questões concernentes a seus domínios, a taxonomia, as tipologias e as implicações pedagógicas, como também as expressões na Língua Geral e de Língua Especializada e, ainda, exemplos de pesquisas que discutem sobre dificuldades na tradução dessa fraseologia.

No Capítulo 2, a metodologia envolveu uma breve explanação do projeto CAT, detalhes sobre o processo de coleta de dados, os parâmetros para sua organização, os documentos legais do Comitê de Ética, bem como os aspectos relevantes sobre a instituição, o curso, os procedimentos e os critérios de seleção dos participantes. Em seguida, apresentamos as ferramentas e os *corpora* utilizados, assim como os detalhes sobre a plataforma *Sketch Engine*, o procedimento de análise de dados e o *corpus* empregado como aporte na investigação.

Por último, discute-se o item estatístico *LogDice*, que serviu como parâmetro de medida estatística para comprovar os termos situados. Neste mesmo item, descrevemos as formas percorridas para a elaboração das estratégias da atividade a partir do CAT 2, a fim de propormos modelos de atividades e exercícios com base na *Online English Collocations Platform*, no *corpus* de referência *English Web 2013* e na ferramenta *Sketch Engine for Language Learning (SKELL)*.

A discussão e análise de dados, no Capítulo 3, permitiu-nos verificar os resultados do teste de proficiência aplicados aos alunos participantes, dados estatísticos do *corpus* e, também, discorrer sobre o comportamento léxico-fraseológico das combinações selecionadas no CAT 2, para, então, apontarmos as contribuições por meio do *English Web 2013*.

No Capítulo 4, desenvolvemos três propostas de atividades, baseadas na discussão de dados do CAT 2, por meio da *Online English Collocations Platform*, do *corpus* de referência empregado na pesquisa e da *SKELL*, apresentando possíveis sugestões de aplicabilidade no contexto educacional da tradução.

Por fim, apresentamos as considerações finais, com reflexões do que foi explorado neste trabalho, possíveis contribuições futuras, referências adotadas no decorrer do texto e anexos com os documentos legais empregados.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresentamos as questões teóricas que conduziram esta investigação. Por essa razão, propomos uma explanação sobre os aspectos da Linguística de *Corpus* e a função do *corpus*, bem como a relação dessa área com a Tradução e Ensino de LE. Também tratamos de questões relacionadas ao léxico e suas características convencionais, baseadas em contribuições da Fraseologia e suas implicações na língua geral e na de especialidade, com apontamento para os diferentes tipos de fraseologismos, e discorrendo ainda sobre a interface entre os estudos fraseológicos e a Linguística de *Corpus* na contemporaneidade. Ademais, elencamos alguns pontos sobre as colocações, sua funcionalidade nas línguas geral e especializada, a taxonomia que sustenta este trabalho, além de dissertar sobre pesquisas que abordaram erros colocacionais em inglês como LE.

2.1 A Linguística de *Corpus*

A Linguística de *Corpus* (LC) vem se tornando uma área de grande interesse, principalmente pela exploração da linguagem produzida por falantes reais, mediante suas variedades linguísticas (MCENERY; WILSON, 1996; MCENERY; HARDIE, 2012; MEYER, 2004).

Meyer (2004) entende que os linguistas adeptos das teorias e descrições da LC podem encontrar recursos úteis para a investigação linguística. Ao pesquisar uma língua, como o inglês, por exemplo, o autor acredita que o percurso de análise deve ser com base em dados reais, e não artificiais. O grande auge da LC, segundo Meyer (2004), ocorreu no início dos anos de 1960 com o surgimento do *Corpus Brown*, criado em língua inglesa, cujos pioneiros de sua elaboração foram W. Nelson Francis e Henry Kučera. Nessa época, a Linguística era dominada pela gramática gerativa e qualquer outro tipo de abordagem gramatical que surgisse era vista como inaceitável.

Com base em Francis (1992), Meyer (2004) acrescenta que, quando foi criado, o *Corpus Brown* foi considerado como uma prática inútil e imprudente, pois a análise intuitiva, como já propunha a gramática gerativa, era a única fonte legítima para se explorar um idioma. Contudo, Meyer (2004) ressaltou o fato de haver muitos linguistas apegados à crença da gramática gerativa. Porém, afirmou que, com o tempo, a adesão à LC aos estudos descritivos e teóricos de linguagem tornou-se mais frequente na área Linguística.

Com o avanço da tecnologia, a LC sofreu grande impacto da Linguística Computacional. Segundo McEnery e Hardie (2012), os métodos com base em *corpus* se tornaram essenciais em muitas áreas da Linguística Computacional, por meio das inovações – como a questão probabilística (HALLIDAY, 1991), que se mostrou de grande utilidade para os linguistas de *corpus*, sendo considerado um aspecto relevante na obtenção de dados empíricos, contrapondo o foco na análise intuitiva da linguagem, como proposta pela Linguística Gerativa (CHOMSKY, 1995).

Depois dessa prévia introdutória, passamos a tratar especificamente da LC neste trabalho. McEnery e Hardie (2012, p. 1, tradução nossa), definem a LC como “[...] uma base apropriada a fim de estudar um conjunto específico sobre questões de pesquisa”¹. A palavra *corpus*, principal recurso nessa área, pode ser definida como um conjunto de textos em larga escala, que são coletados para análise linguística de forma computadorizada (MCENERY; HARDIE, 2012). Esse processo é chamado de compilação, cujos resultados são geralmente apresentados com base em estatísticas calculadas pelos programas de análise lexical.

Dessa forma, Sinclair (1991, p.13) ainda aponta que o fato de se ter bem claros os objetivos para a compilação de um *corpus*, auxilia na seleção adequada do objeto linguístico de um idioma que se pretende estudar (fraseologias- colocações, expressões idiomáticas etc.), pois nem sempre é possível analisar todos os fenômenos linguísticos em um único *corpus*, uma vez que, ao criá-lo, também é preciso delimitar o assunto a ser explorado.

McEnery e Hardie (2012) salientam que o *corpus* serve como um suporte importante, nem tanto por ser uma coletânea de textos, mas por causa das propriedades que ele adquire quando é bem estruturado e cuidadosamente construído. Esse olhar mais atento para o *corpus* é de suma importância, pois é preciso estabelecer objetivos bem definidos ao criá-lo. Seria basicamente a simples, porém árdua, tarefa de traçar alguns questionamentos, tais como: *por quê; para quê?; para quem?* um *corpus* deve ser criado. Por consequência, entendemos que o autor chama de “notável”, ou seja, significativo, pelo fato de que um *corpus* adequado deve conter dados necessários de forma que atenda às metas de sua análise.

Para comprovar a relevância científica no uso do *corpus* pela LC, em um estudo sobre a delimitação do *corpus*, Bowker (2000) identificou que o trabalho com os *corpora* eletrônicos pode gerar uma base de dados mais confiável, pois anteriormente, antes do uso do computador nessa área, a compilação manual era exaustiva. Sobre esse tópico da delimitação, a autora descreve uma situação em que iniciou uma de suas pesquisas com a construção de

¹ Do original: “[...] an appropriate basis on which to study a specific set of research questions” (MCENERY; HARDIE, 2012, p. 1).

corpus voltado para alunos de tradução, incluindo textos de mais de uma temática referentes ao campo da computação. Porém, descobriu no decorrer da experiência a necessidade de delimitar os temas dos textos que foram inseridos no *corpus*, certa de que a produção e a qualidade da investigação seriam mais relevantes.

Por essa razão, recorreremos a Sinclair (2004), quando mostra a necessidade de se criar um *corpus* com o fim de estudar a linguagem, sendo que o conteúdo selecionado deve representar a língua inserida nele. Dessa forma, o autor tratou de questões sobre as políticas de amostragem do *corpus*. O estudioso considera três aspectos importantes na composição do *corpus*: a amostragem, a representatividade e o balanceamento, os quais, segundo ele, são inevitáveis no *design* do *corpus*.

Quanto à amostragem, Sinclair (2004) a conceitua basicamente sob os critérios de indicação do idioma ou variedade linguística, critérios de seleção, e de natureza e dimensões das amostras. Pensando no *corpus* de aprendizes, por exemplo, este pode ser constituído de um idioma específico, ou variedades linguísticas em *subcorpus*, adotar critérios na seleção e no nivelamento dos participantes, o que de certa forma revelará a natureza das amostras e, atrelado ao objetivo, a dimensão do *corpus* por ser definida.

A noção de representatividade tem como base o *design* e a composição do *corpus*. Sinclair (2004) afirma que a documentação do *design* e da composição envolve as formas de organização do conteúdo e a justificativa dos critérios de escolha.. Tagnin (2015, p. 29) ressalta que “[...] um *corpus* deve ser representativo daquilo que se pretende investigar, contendo uma amostragem significativa do universo a ser estudado”.

Já o balanceamento, assim como a representatividade, Sinclair (2004) afirma que esses aspectos podem auxiliar na elaboração do *design* do *corpus*, bem como na seleção de seus componentes. Tagnin (2015, p. 357) exemplifica que esse balanceamento trata-se de um “[...] processo pelo qual se garante que dois *corpora* sejam construídos de maneira similar quanto a origem, gênero, extensão, período de produção dos textos, ou quaisquer outros critérios que sejam relevantes para a pesquisa a que se destinam”.

De modo geral, o uso do *corpus* na LC e suas funcionalidades abarcam uma gama de possibilidades de estudos linguísticos com vistas a investigações científicas em diversos campos. A priori, ela pode ser útil enquanto ciência do léxico na criação de *corpus* (*corpora*), que também se caracteriza por um viés interdisciplinar de exploração linguística (MEUNIER; GRANGER, 2008), mas também contempla áreas como a Fraseologia (CORPAS PASTOR, 1996; BEVILACQUA, 1996; GRANGER; PAQUOT, 2008; ORENHA-OTTAIANO, 2004, 2009, 2012a, 2012b, 2015, no prelo; TAGNIN, 2013), a Terminologia (CABRÉ, 1993, 1999,

2009; BARROS, 2004; BOURIGAULT; JACQUEMIN; L'HOMME, 2001), a Terminografia (BARROS, 2004; FUERTES-OLIVERA; TARP, 2014), a Lexicologia (KORNEVA, 2016; CORPAS PASTOR, 1996), a Lexicografia (CASARES, 1950) e os Estudos da Tradução (BERNARDINI, 2004; TAGNIN, 2002; ORENHA-OTTAIANO, 2004, 2009), utilizando-se tanto da língua geral como a de especialidade, em diversas esferas do conhecimento e seus mais diferentes gêneros textuais e discursivos.

1.2 A Linguística de *Corpus* e suas conexões com a Tradução

Com base em Oliveira (2015) e Laviosa *et al.* (2017), a relação entre Tradução e LC teve início a partir dos estudos de Baker (1993), que percebeu a necessidade de “uma disciplina que a tenha como objeto de estudo: os Estudos da Tradução” (OLIVEIRA, 2015, p.26), passando a instigar pesquisas voltadas para a análise descritiva da tradução com base em *corpus*. Outrora, Berber Sardinha (2002, p. 15) já previra que “[...] os estudos tradutológicos como área acadêmica de pesquisa, quanto à prática tradutória, têm muito a ganhar com um contato maior com a Linguística de *Corpus*”. Laviosa *et al.* (2017, p. 27, tradução nossa) dividiram essa fusão de áreas em três momentos:

- O primeiro *the dawn* (“o amanhecer”), e único que carrega um pseudônimo, ocorreu entre os anos de 1993 a 1996, quando Mona Baker introduziu a ideia de usar o *corpus* como metodologia para estudos descritivos da tradução;
- O segundo momento se deu com a publicação da tese de doutorado de Laviosa-Braithwaite (1996) até o final dos anos de 1990;
- E, em terceiro, houve um crescimento desse tipo de pesquisa a partir dos anos 2000, em que se prezou pelo uso com base na ideia linguística e análise crítica do discurso, tendo como marco a terminologia *Corpus-based critical translation studies*.

Ao abordar essa perspectiva da Tradução, Paiva (2009) aponta que o foco de pesquisas nessa área passou de uma visão meramente prescritiva para uma exploração mais descritiva da língua. Laviosa (2017, p. 2) salienta que “a sinergia entre Estudos Descritivos da Tradução e Linguística de *Corpus* serviu de estímulo à criação de uma variedade de recursos de *corpus*, ao desenvolvimento de uma metodologia de pesquisa descritiva”. Essa perspectiva descritiva da tradução pode ser posta em prática quando “o acesso a *corpora* computadorizado deve

permitir-nos explorar uma série de outras questões teóricas difíceis de lidar com base em estudos de pequena escala” (BAKER, 1993, p. 247, tradução nossa)².

Quanto ao *corpus* nos estudos da tradução, Baker (1993, p. 237) destaca que:

[...] distanciar-se dos textos de origem e da noção de equivalência é fundamental para preparar o terreno de trabalho com *corpus* porque permite que a disciplina deixe de lado sua obsessão de longa data baseada na ideia de estudar isoladamente as instâncias individuais (uma tradução comparada a um texto fonte de cada vez) e cria um requisito que pode encontrar cumprimento no trabalho de *corpus*, ou seja, o estudo de um grande número de textos do mesmo tipo. É precisamente nesse ponto que o trabalho de *corpus* entra em cena (BAKER, 1993, p. 237, tradução nossa)³.

Orenha-Ottaiano (2012b, p. 1) corrobora com essa nova visão acerca da interface LC e Estudos da Tradução, afirmando que o linguista de *corpus* utiliza metodologias que permitem investigações de maneira mais ampla e menos intuitiva, como aquelas baseadas nas teorias de Noam Chomsky. Baker (1993, p. 241) afirma que, para a Linguística, as coletâneas de *corpora* informatizados, seguidas do desenvolvimento de uma metodologia de pesquisa, fornecem meios de superar nossas limitações humanas e minimizar nossa dependência da intuição. Como veremos mais a frente, o uso da LC na tradução diminui a incidência de erros combinatórios, principalmente as combinações fixas, que são compreendidas dentro de suas convenções de uso em uma língua.

Do ponto de vista linguístico, Laviosa (2017) também entende que os Estudos de Tradução com base em *corpora*, que tem a noção da língua em uso como norte, podem ser usados como recursos para recuperar e examinar equivalentes lexicais, terminológicos, fraseológicos, sintáticos e estilísticos, bem como garantir a qualidade de uma tradução. Devido a esses proveitos, a autora apresenta ainda que os *corpora* passaram a fazer parte da grade curricular de programas de pós-graduação a fim de treinar tradutores e satisfazer as necessidades atuais da indústria linguística globalizada e tecnológica.

Por causa dessa demanda, os *corpora* eletrônicos podem auxiliar os tradutores na verificação desses e outros fenômenos linguísticos, considerando as combinações formadas pelo léxico mental (ORENHA-OTTAIANO, 2015) do usuário de uma língua – incapaz de armazenar todas as combinações existentes. Por essa razão, uma vez que o tradutor

² Do original: “[...] access to computerised corpora should enable us to explore a number of other theoretical issues which are difficult to deal with on the basis of small-scale studies.

³ Do original: “The move away from source texts and equivalence is instrumental in preparing the ground for corpus work because it enables the discipline to shed its longstanding obsession with the idea of studying individual instances in isolation (one translation compared to one source text at a time) and creates a requirement which can find fulfilment in corpus work, namely the study of large numbers of texts of the same type. This is precisely where corpus work comes into its own” (BAKER, 1993, p. 237).

constantemente lida com alto volume de informação, em meio ao avanço tecnológico, a parceria entre cientistas da computação e linguistas de *corpus*, que resulta na melhoria de ferramentas de análise de *corpus*, tem dado ao tradutor mais segurança no uso da língua.

Sendo assim, o “[...] tradutor conta hoje com bem mais recursos” (SOBRAL, 2008, p. 14), e é nesse cenário que a LC ganha espaço, por fornecer ferramentas de análise computadorizada a esse profissional, que são bem mais úteis em primeira instância do que a criação de listas de glossários de forma manual, ou mesmo alimentar programas de memória de tradução intuitivamente ou sem parâmetro de análise estatística da frequência do léxico empregados na língua, como oferecido pela LC.

Por sua vez, Sobral (2008) afirma que, por muito tempo, os aspectos teóricos da tradução ficavam a cargo de acadêmicos alienados, sem apresentar nenhuma relação com a prática dos tradutores. O autor ainda relata que com o passar dos anos, a visão sobre a tradução se profissionalizou, causando mais associação entre teoria e prática. No âmbito acadêmico, com base em Bernardini (2002), Oliveira (2015, p. 27) questiona sobre a formação do tradutor, “se devemos treinar ou capacitar tradutores”. Para o autor, o treino refere-se ao ato de oferecer fórmulas para certas situações, enquanto a capacitação visa habilidades como a adaptação e o desenvolvimento da capacidade do tradutor.

Portanto, a experiência que um tradutor pode adquirir em seu primeiro contato com a área de Tradução desde a sua graduação e ao longo dos anos em sua atuação profissional, traz consigo uma série de oportunidades que conduzem ao desenvolvimento de competências e habilidades requeridas a esse profissional.

2.3 Corpus de aprendizes

Dentre os vários percursos de averiguação linguística a partir da LC, destacamos os possíveis modelos de *corpora* existentes, os quais podem nortear a ótica da investigação linguística para examinar a linguagem real por meio da frequência. Os possíveis modelos de *corpus* são geralmente denominados como paralelos, comparáveis, bilíngues, multilíngues geral e/ou especializado, “Faça você mesmo” (*DIY – Do it yourself corpora*) e de aprendizes. Sendo assim, adotado como premissa básica nesta pesquisa, aponta-se a relevância de analisar um *corpus* de aprendizes, que pode servir até mesmo de base pedagógica para apoiar o estudante em sua trajetória acadêmica.

Nesse viés pedagógico, Gavioli (2005, p. 1) aponta que a pesquisa baseada em *corpus* tem contribuído para a produção de materiais, porém, sem muito êxito em sua aplicação em

sala de aula ainda. Apesar disso, a autora questiona a escassez de ferramentas para fins didáticos e o baixo interesse de professores e alunos, que desconhecem as vantagens dos *corpora* nas atividades escolares. Para comprovar essa hipótese, Granger (2002), afirma que uma pesquisa baseada na análise da linguagem produzida por estudantes é um ponto fundamental a ser considerado no ensino de um segundo idioma, com vistas para dois objetivos. O primeiro deles é o de corroborar para a melhor compreensão do processo de aquisição dessa segunda língua e daquilo que a influencia. Já o segundo, é tido como uma fonte de dados úteis, a fim de que os profissionais possam elaborar instrumentos de ensino e aprendizagem com foco nas dificuldades comprovadas por meio do *corpus* de aprendizes.

Granger (2002) ressalta que o interesse nesse tipo de abordagem se deu com os novos recursos oferecidos por esse modelo de *corpus* que começou a ser explorado de forma computadorizada, com sua aparição no final dos anos de 1980, enquanto parte da LC. Todavia, a autora acrescenta que apenas recentemente o *corpus* de aprendizes passou a ser adotado por teóricos e falantes de uma segunda língua. Isso se deve aos investimentos em ferramentas de análise computadorizada na LC: *WordSmith Tools* (SCOTT, 2012); *AntConc* (ANTHONY, 2014); *Sketch Engine* (KILGARRIFF *et al.*, 2014), dentre outros gerenciadores de *corpus* que proporcionam uma visão mais abrangente dos textos, permitindo inferências no comportamento linguístico do aprendiz, bem como a criação de materiais para fins pedagógicos.

Alguns *corpora* delas foram destacados por Orenha-Ottaiano (2015):

- o *International Corpus of Learner English*, orientado por Granger (1993);
- o *Longman Learner's Corpus*, supervisionado por Summers (1999);
- o *Cambridge Learner's Corpus*, coordenado por Nicholls (1999);
- o *Multilingual Learner Corpus*, sob a coordenação da Prof^a. Stella E. O. Tagnin, da Universidade de São Paulo (USP);

Para Tagnin (2013), um *corpus* de aprendizes deve ser composto por redações não corrigidas, com textos de diversos níveis de aprendizado. Segundo a autora, ao se comparar um *corpus* de aprendizes não nativos com outros *corpora* de aprendizes nativos, geralmente podemos identificar os problemas no uso de uma LE.

No que diz respeito à exploração de *corpora* de aprendizes de tradução, Orenha-Ottaiano (2015), por sua vez, realizou um trabalho parecido, ao compilar dois *corpora* com

traduções de alunos do português para o inglês, sendo um com textos de aprendizes do curso de Tradução, *Corpus* de Aprendizes de Tradução (CAT), e outro com textos de estudantes do curso de licenciatura em Letras, *Corpus* de Aprendizes de Língua Estrangeira (CALE). Embora diferente da sugestão de Tagnin (2013), o estudo mostrou a viabilidade de uma análise contrastiva entre o *corpus* de aprendizes com outro de referência, mas de língua geral.

Independentemente se objetivo de pesquisa seja referente ao ensino ou à tradução, esse tipo de análise contrastiva por meio do *corpus* de aprendizes conduz a observações mais coerentes que comprovam, na maioria das vezes, o surgimento de problemas de emprego de dados linguísticos, no que tange a frequência de uso ou não desses dados. Os erros mais comuns, conforme cita Tagnin (2013), ocorrem quando os alunos “subusam”, “sobreusam” ou evitam itens na escrita da redação, assim como a ausência de combinações convencionais da língua-alvo.

Sobre os dois primeiros exemplos, Orenha-Ottaiano e Paiva (2012) denominam os termos “sobreuso” como aqueles que apresentam padrões de maior frequência em relação ao seu uso por falantes nativos, enquanto que “subuso” refere-se a itens de menor frequência, os quais são notados, segundo elas, com textos traduzidos para uma segunda língua que são “comparados à produção de falantes nativos de língua inglesa” (ORENHA-OTTAIANO; PAIVA 2012, p. 190).

Um *corpus* paralelo de aprendizes de tradução oferece subsídios mais detalhados ao docente com relação ao comportamento linguístico entre as escolhas lexicais do texto-fonte e o traduzido. Nesse panorama, o conceito de paralelo fundamenta-se pela prática de realizar inferências nessa ponte tradutória. Para Granger (2002), esse modelo de *corpus* conduz à identificação de possíveis dificuldades tradutórias, na comparação de dados linguísticos entre uma língua e outra, como também pode ser usado como guia para elaboração de atividades com base em problemas tradutórios.

Um exemplo disso pode ser observado em obras com foco nas colocações, pois por meio desse tipo de análise, de maneira contrastiva, os professores podem, por exemplo, trabalhar “com as colocações em sala de aula de modo mais eficiente e auxiliar seus alunos a empregá-las de modo mais preciso e produtivo” (ORENHA-OTTAIANO, 2015, p. 833). Orenha-Ottaiano (2012a, 2012b, 2015, no prelo) também traz reflexões interessantes sobre a viabilidade do trabalho com *corpus* de aprendizes. A autora comenta o enfrentamento de possíveis problemas no uso dos padrões combinatórios do léxico, ao se tentar utilizar uma determinada colocação, por exemplo, na escrita de aprendizes, levantados em projetos de mestrado (ALVES; ORENHA-OTTAIANO, 2020) e doutorado (ROCHA; VIANA;

ORENHA-OTTAIANO, 2020; ROCHA, 2021) em programas linguísticos de pós-graduação e em pesquisas por meio de programas de iniciação científica (BRESSAN; ORENHA-OTTAIANO, 2013; BRITO, 2012; NAKANISHI, 2013).

Além desses recursos e modelos de *corpus* disponíveis, a LC conta também com propostas teóricas como a Abordagem Movida por Dados (*Data Driven Learning – DDL*) (JOHNS, 1991; BOULTON, 2009, 2015), que visa conduzir o estudante a “descobrir a língua” por meio do *corpus* apresentado em aula, momento em o docente tem a tarefa de colaborar para que essa prática de aprendizagem ocorra. Por outro lado, uma segunda vertente, ou seja, a Abordagem Lexical (*Lexical Approach*) (LEWIS, 2000) objetiva tornar o usuário do idioma capaz de produzir e de compreender combinações fixas a partir do seu contexto de uso.

De modo geral, essas pesquisas nos conduzem a validar as vantagens que um estudo com foco em *corpus* de aprendizes pode vir a promover para a esfera científica, resultando em melhorias de materiais didáticos e dicionários para fins pedagógicos. Assim, podemos concluir que a relação entre LC e Ensino de Língua Estrangeira tem um vasto caminho a ser trilhado e, se feito, pode conduzir o aluno a uma tradução mais consistente, de forma que ao produzir padrões combinatórios em outro idioma, atenda aos aspectos convencionais de uma língua-alvo.

Depois de destacar a relevância do *corpus* de aprendizes no ensino de língua estrangeira para tradutores, passamos a pensar sobre algumas teorias que embasam esse tipo de trabalho e que caminhos percorridos podem fundamentar esse conceito a partir da pedagogia do léxico e da tradução baseada em *corpus*.

2.4 Léxico e Tradução: um olhar pedagógico baseado em *corpus*

Do ponto de vista do ensino, Orenha-Ottaiano e Pinto (2018) entendem o léxico e a tradução da seguinte maneira: a *Pedagogia da Tradução Baseada em Corpora* estaria relacionada ao desenvolvimento de habilidades tradutórias; já a *Pedagogia do Léxico Baseada em Corpora* mantém um foco parecido, porém, mais amplo ao propor o aprendizado do léxico que os falantes nativos ou fluentes empregam com frequência, por meio de pesquisas e materiais didáticos, cujo foco é que se alcance a competência linguístico-fraseológica.

O léxico, na Linguística, ocupa lugar de destaque, pois é a unidade básica e estudada em ciências como morfologia, fonologia, fonética, gramática, por exemplo, mas é na Lexicologia, segundo Korneva (2016), que o léxico é principalmente explorado como unidade

de palavras combinadas em uma dada língua. Essa relação com mais de uma palavra, de acordo com a autora, pode ocorrer de três formas: etimológica (palavras com a mesma raiz), temática (palavras com o mesmo tema), gramatical (palavras que são partes determinadas na oração) e lexical (palavras que compartilham o significado).

Berber-Sardinha (2009), com base em Halliday *et al.* (2004), aborda as implicações de um item lexical, como uma unidade “léxico-gramática”, a qual forma o vocabulário, isto é, o léxico propriamente dito. Esse léxico, conforme o autor, não se refere à grafia ou à pronúncia das palavras, mas a “unidade de sentido, que resulta de escolhas do falante ou escritor”. Por consequência, em qualquer língua é comum nos depararmos com unidades lexicais formadas por palavras combinadas, que juntas compõem um sentido único e não podem ser explicadas apenas pela gramática.

Pelo contrário, a gramática sozinha, em seu conjunto de regras, não consegue explicar fenômenos linguísticos de forma combinada, como as colocações. Orenha-Ottaiano (2004, p. 1) afirma que “a gramática sem o potencial combinatório, sintagmático e paradigmático do léxico, desempenha poucas funções no sistema”. Sendo assim, considerar a utilização do léxico em uma esfera convencional, na maioria das vezes, pensado pela sua frequência de uso e na disseminação de uma colocação de uma geração para outra, a torna também uma unidade combinatória que só pode ser explicada por meio do contexto de uso.

Se pensarmos em uma conexão entre o léxico e a tradução baseada em *corpus*, corroboramos para o que também salienta Orenha-Ottaiano e Pinto (2018, p. 131), sobre a “Pedagogia do Léxico e da Tradução” vinculada aos estudos aplicados:

[...] uma vez que se dedica às questões linguísticas na formação do tradutor, assim como as questões pedagógicas que envolvem o léxico, ou seja, o ensino-aprendizagem do vocabulário de uma língua estrangeira. Acreditamos que o uso de ferramentas computacionais que permitem as buscas do léxico bilíngue em coletâneas de textos disponibilizados no computador, os chamados *corpora* eletrônicos, permitirá que os aprendizes de uma língua estrangeira, tradutores e professores de línguas estrangeiras possam realizar inúmeras pesquisas.

Em relação ao uso e abordagem baseada em frequência, Aston (1999) também levantou a hipótese de que os recursos dos *corpora* servem como ferramentas significativas para aprendizes, em suas potencialidades e limites, pelo fato de auxiliarem significativamente as práticas pedagógicas. Por esse viés, as possibilidades de investigação que a LC pode oferecer durante a trajetória docente é a capacidade de criar *corpora* próprios, conforme os seus objetivos de trabalho.

Por sua vez, Frankenberg-Garcia (2012) discorre sobre formas de integrar o uso prático dos *corpora* ao ensino diário de línguas, e afirma que essa utilização precisa ser mais disseminada, pois segundo a autora, mesmo aqueles que conhecem essa prática não a empregam em seus ensinamentos. Ela aponta algumas pesquisas sobre atividades baseadas em *corpora* que podem ser encontradas nos estudos de Tribble e Jones (1997), Anderson e Corbett (2009), Bennet (2010), Reppen (2010), mas chama a atenção sobre aplicá-las à realidade da sala de aula, devido a pouca integração de exercícios desse tipo nesse contexto específico. Por essa razão, Frankenberg-Garcia (2012) defende a inclusão de *corpora* na sala de aula, pois considera uma maneira de expor os alunos à linguagem autêntica, acreditando ainda que os docentes podem compilar seus próprios *corpora* ao invés de utilizar de materiais prontos.

Segundo Frankenberg-Garcia (2012), alguns exemplos podem ser observados em contribuições científicas de alguns autores como: Krieger (2003) que provou o uso da palavra *any* utilizada em frases afirmativas e não exclusivamente negativas e interrogativas; por seu turno, Barlow (2000), que observou a viabilidade de analisar correspondentes em L1 e L2, por exemplo, pela comparação das palavras em inglês *head* e *chef* em francês; já O'Keeffe, McCarthy e Carter (2007) apresentou como os *corpora* podem auxiliar alunos na compreensão de certos padrões combinatórios restritos, como as colocações, identificando, por exemplo, que o verbo *strike*, comumente é usado precedido do substantivo *bargain*.

Outro exemplo dado pela autora, com base em Reppen (2010), contempla amplamente as questões metodológicas da abordagem pelo uso de *corpora* para fins de aprendizagem:

propôs uma atividade em sala de aula, em que os alunos realizaram um pequeno *corpus* de artigos jornalísticos, a fim de que pudessem extrair uma lista de palavras e conseqüentemente procurarem palavras difíceis, que são frequentes no *corpus*, antes da leitura dos artigos, seguida e outra atividade, em que os alunos deveriam observar as diferenças de registro em listas de frequência separadas de textos acadêmicos e discursos transcritos (FRANKENBERG-GARCIA, 2012, p. 3 tradução nossa)⁴.

De certa forma, apoiadas pela LC, notamos claramente que essas sugestões podem auxiliar pedagogicamente o ensino de LE em diversas áreas que requerem o aprendizado de uma segunda língua, principalmente para tradutores aprendizes (FRANKENBERG-GARCIA, 2012, p. 3).

⁴ Do original: “[...] proposes a classroom activity where learners make a small corpus of newspaper articles and then extract a word list in order to look up difficult words that are frequent in the corpus before they start reading the articles, and another one where learners are required to notice register differences in separate frequency lists of academic texts and transcribed speech”(FRANKENBERG-GARCIA, 2012, p. 3).

No âmbito estrutural, por exemplo, Oliveira (2015) questiona a falta de espaço para discutir o ensino de tradução com base em *corpus* nos cursos de graduação, ao sugerir um método de ensino de tradução para tradutores aprendizes, com base na LC. Na tentativa de estudar a escrita e examinar o processo de construção do texto de chegada, o pesquisador utilizou-se de ferramentas de análise de *corpus* e planilhas do *Microsoft Excel* de forma didática para auxiliar tradutores em formação.

Dessa maneira, notamos que uma abordagem baseada em *corpus* surge mais como uma oportunidade de capacitação do que de treinamento durante o processo tradutório, que vai além da esfera meramente tecnicista: **i)** criar um *corpus* e compilá-lo; **ii)** observar os dados de maneira estatística, que provam a frequência e uso das palavras (lista de palavras, palavras-chave e análise por meio de linhas de concordância); **iii)** elaboração de glossários especializados; **iv)** criação e melhoria de materiais didáticos, dentre outros. Desse modo, o uso do *corpus* tem mais a ver com as questões de formação e desenvolvimento do que apenas com técnica, pela amplitude de habilidades promovidas por essa metodologia, sem contar o uso de *corpora on-line* disponíveis para pesquisa.

É fato que aprender a manusear e, até mesmo, criar um *corpus* próprio, exige um conjunto de técnicas, a princípio. Porém, o que temos percebido é um grande interesse da pesquisa científica que tem explorado os estudos da tradução com base em *corpus* (ALVES; ORENHA-OTTAIANO, 2020; BAKER, 1993; BERNARDINI, 2004; CAMARGO, 2005, 2007; ORENHA-OTTAIANO, 2012a, 2012b, 2015; TAGNIN, 2013). Essa disseminação das investigações em tradução fundamentada em *corpus* remete ao que apontamos anteriormente sobre a formação do tradutor visando o desenvolvimento de sua competência tradutória quanto ao emprego do léxico.

Outrossim, Berber Sardinha (2002,) já apontava sobre a existência da unanimidade entre tradutores e linguistas de *corpus*, quanto ao uso de *corpus* eletrônico na tradução, alertando sobre o aumento do uso desse recurso nessa e em outras áreas de estudos linguísticos (MCENERY; WILSON, 1993). Entretanto, o autor já alertara para uma lenta aceitação por parte de pesquisadores (BERBER SARDINHA, 2002).

No Brasil, alguns trabalhos sobre o uso de *corpus* nos estudos da tradução são: Berber Sardinha (2002), Paiva (2009), Viana e Tagnin (2015), Orenha-Ottaiano (2009, 2012a, 2012b, 2015, no prelo), entre outros. Basicamente, são pesquisas sobre questões voltadas para a compilação e manuseio de ferramentas de análise de *corpus* e aplicadas ao seu uso na língua geral ou na especializada. Entretanto, na maior parte dos casos, as investigações com *corpus*

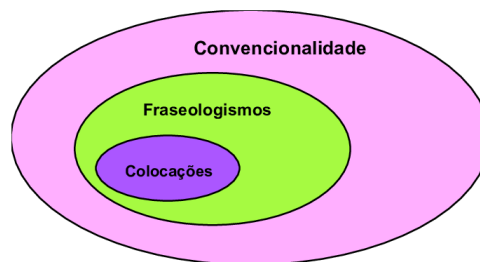
influenciam direta ou indiretamente o ensino e aprendizagem do léxico na tradução, muitas vezes, disseminados em pesquisas amparadas por projetos de programas de pós-graduação.

Uma vez que temos um arcabouço teórico sobre a manipulação de dados por meio dos estudos com *corpus*, isso fundamenta a aplicabilidade desse recurso linguístico voltado para alunos de Tradução de um modo mais prático e didático, a fim de possibilitar situações de ampliação do repertório lexical durante as práticas, de modo a refletir acerca do emprego adequado do léxico, principalmente as unidades fraseológicas (UFs).

2.5 A convencionalidade

A figura abaixo, proposta por Orenha-Ottaiano (2004), ilustra de modo resumido a relação conceitual entre convencionalidade e fraseologismos para entendimento da funcionalidade da colocação:

Figura 1 – Convencionalidade, Fraseologismos e Colocações



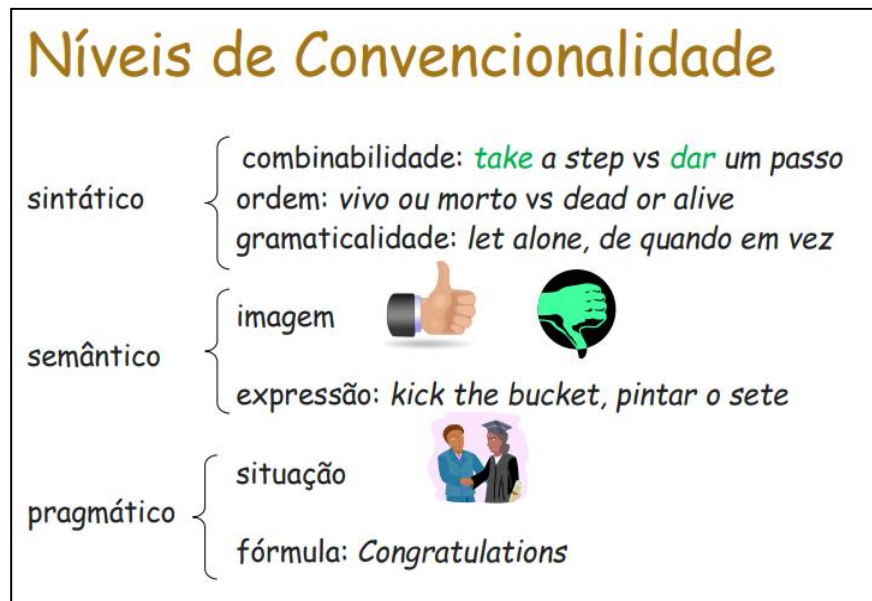
Fonte: Orenha-Ottaiano (2015, p. 849).

De acordo com Orenha-Ottaiano (2004), a convencionalidade ocupa maior proporção nesta figura, pois está relacionada aos costumes consagrados na sociedade. Esses hábitos guiam o uso do léxico por meio das palavras combinadas, produzidas por falantes reais, seja no uso da língua em comunidades específicas ou de modo geral, por exemplo, dependendo da variação linguística do local. A exemplo disso, a partir de Biderman (2001), Azevedo e Silva (2017, p.39) apontam que “o nível lexical, situado entre o morfológico e o sintático, corresponde à somatória de toda a experiência acumulada de uma sociedade e do acervo de sua cultura através das idades”. Desse modo, a ideia de convencionalidade está diretamente relacionada aos aspectos da língua por causa do uso lexical. De acordo com Tagnin (2013), a noção de convenção pode ser explicada apenas pela língua nos níveis social (*quando* dizer algo) e linguístico (*como* dizer). Dois exemplos desse elo, segundo a autora, encontram-se na

expressão “feliz natal”, uma convenção social que exprime comemoração, e a expressão “mundos e fundos”, sem opção de inversão, além de estarem unidas por uma conjunção.

Observamos na figura abaixo três níveis de convencionalidade em um modelo indicado por Tagnin (2017):

Figura 2 – Níveis de convencionalidade



Fonte: captura da tela de apresentação de trabalho (TAGNIN, 2017).

Percebe-se que em todos os níveis as combinações estão relacionadas aos aspectos convencionais, e “[...] aos usos e costumes sociais já preestabelecidos e consagrados pela comunidade” (ORENHA-OTTAIANO, 2015, p. 845). É nesse ponto que o aprendiz geralmente se confunde ao empregar combinações lexicais fixas na língua, seja por desconhecer as combinações no idioma de chegada, por falta de atenção na hora de buscar uma combinação equivalente ou porque esteja preocupado apenas em identificar a tradução de uma palavra, de modo isolado.

Matuda (2011) apresenta como os tradutores poderiam pensar na convencionalidade durante o processo tradutório:

Os tradutores, assim como os intérpretes, os redatores, os jornalistas e os documentalistas são usuários indiretos das terminologias, pois a eles interessa o uso adequado dos termos **fraseologias** e expressões idiomáticas para que o texto esteja de acordo com as normas de convencionalidade que regem a produção do tipo e do gênero textual em questão na cultura em que é produzido (MATUDA, 2011, p. 81, grifo nosso).

No caso da fraseologia, a autora entende o emprego desse tipo de combinação lexical de maneira convencional, respeitando as variáveis ligadas ao emissor, receptor, situação e objetivo da comunicação. Dessa forma, o texto a ser traduzido não é um emaranhado de palavras desconexas.

Pelo contrário, apresenta uma estrutura lexical prevista na língua-fonte e na língua-alvo. Sendo assim, o tradutor, frente ao desafio de buscar uma combinação lexical em outro idioma, deve levar em conta, segundo Matuda (2011, p. 81), “o ato tradutório tendo como ponto de partida o ato de comunicação”, aspecto fundamental, a nosso ver, para que a convencionalidade no uso do léxico seja priorizada nesse processo.

Dessa forma, embasado por um *background* empírico e observacional da descrição da linguagem, o tradutor, consciente da convencionalidade da língua, passa a desenvolver um texto mais fluente, ao se deparar, com o desafio de traduzir uma determinada combinação fixa, como a colocação (ORENHA-OTTAIANO, 2015) e outras fraseologias existentes.

Orenha-Ottaiano (2012b) ainda destaca a inovação metodológica ofertada pela LC para a área da tradução, com o intuito de descobrir as UFs e seus padrões combinatórios (HEYLEN; MAXWELL, 1994), como também a repetição de combinações na língua por suas características semânticas. A autora ressalta que, do ponto de vista teórico, a LC pode auxiliar na delimitação, explicação e definição de fraseologismos como, por exemplo, as colocações.

Frente a esses levantamentos, logo notamos um vasto caminho a ser trilhado, no que tange ao uso desse tipo de combinação a partir da tradução de aprendizes, o que auxilia na criação de materiais didáticos para esse público.

2.5.1 A Fraseologia e algumas contribuições

De acordo com Cowie (1998), o campo de atuação da Fraseologia ocorreu tanto na antiga União Soviética como também em países da Europa Ocidental. De acordo com o autor, o aumento das atividades acadêmicas nos últimos anos, na Europa Ocidental e nos EUA, despertou o interesse dos linguistas aplicados ocidentais e em outros países da Europa Oriental.

Cowie (1998) ainda pontua que, na Europa Ocidental, as evidências das investigações nessa área se deram pelo reconhecimento da Fraseologia como disciplina acadêmica dentro da Linguística. O uso do termo *phraseological* (fraseológico), segundo ele, estava presente não somente em pesquisas consolidadas e difundidas, mas também em publicações de vários

dicionários especializados sob uma ou outra perspectiva teórica (COWIE; MACKIN, 1975; SINCLAIR; MOON, 1989, 1995).

Ademais, segundo Cowie (1998), os estudos sobre a Fraseologia resultaram em colaborações teórico-práticas por meio de livros didáticos sobre semântica lexical (CRUSE, 1986), investigações sobre Lexicologia, campo que se dedica às unidades combinatórias (CARTER, 1987) e a exploração do vocabulário no ensino de línguas (CARTER; MCCARTHY, 1988).

Não obstante, Cowie (1998, p. 2) ressalta ainda o surgimento de outras obras, como o livro oficial de Gläser (1986a), as monografias disponíveis dedicadas a idiomas para fins específicos (FLEISCHER, 1982; GRÉCIANO, 1983; CORPAS PASTOR, 1996), processos interlínguas (HOWARTH, 1996) e outras variedades funcionais (MÜLLER, 1993).

Entretanto, esse mesmo autor afirma ainda que há falta de um levantamento exato das várias correntes teóricas sobre a Fraseologia em língua inglesa, com informações de estudos atuais da época, nem mesmo de disciplinas correlatas (análise computacional, aprendizado de idiomas, estilística e lexicografia), áreas que estavam contribuindo fortemente para os estudos fraseológicos (COWIE, 1998).

Como afirma Cowie (1998), cientificamente, o foco da Fraseologia deixou de envolver pesquisas puramente descritivas e passou a demonstrar mais interesse na análise de UFs, combinações de palavras e expressões frasais. O autor atribui esse fato à “crescente conscientização sobre a prevalência de combinações memorizadas na língua escrita e falada, e a um reconhecimento mais amplo que desempenham na aquisição da L1 e L2, assim como na produção da língua na vida adulta”, conforme aponta Cowie (1998, p. 1, tradução nossa)⁵.

Cowie (1998) aponta a presença das questões linguísticas na Fraseologia e comenta sobre o surgimento da ideia de proficiência nativa em um idioma, que depende de um armazenamento de unidades pré-fabricadas, isto é, UFs que apresentam certo grau de complexidade e estabilidade internas na língua. O autor acrescenta também o fato da Fraseologia ter sido reconhecida como disciplina acadêmica na Linguística.

A apropriação a partir do termo “fraseológico”, enquanto adjetivo, segundo Cowie (1998), tornou-se evidente tanto em pesquisas fortemente disseminadas, quanto no surgimento de uma gama de dicionários especializados, e pelo maior interesse nos livros didáticos sobre semântica lexical, na Lexicologia e no vocabulário voltado para o ensino de línguas.

⁵ Do original: “An increasing awareness of the prevalence of ready-made memorized combinations in written and spoken language and a wider recognition of the crucial part they play in first- and second-language acquisition and adult language production” (COWIE, 1998, p. 1).

Outro aspecto relevante considerado por Cowie (1998) é que a Fraseologia auxilia na análise de questões “linguísticas-culturais”. Dessa forma, o pesquisador refere-se a um ponto de vista de investigação fundamentada em um “paradigma antropocêntrico mais amplo”, que tem como eixo central a ideia de que toda língua tem como foco a reflexão e extensão da visão de mundo compartilhada pela comunidade linguística. Nessa questão, Cowie (1998) faz menção às bases estruturais de uma cultura nas expressões de uso, isto é, convencionais, as quais são transmitidas por gerações por intermédio de normas linguísticas e culturais de uso. Segundo o autor, a Fraseologia se destaca pelo fato de servir como um mecanismo relevante que contribui para a formação e a manutenção da identidade cultural.

Entretanto, a Fraseologia passou a ser explorada no campo da Lexicologia, uma ciência linguística que a destacou como subdisciplina (CORPAS PASTOR, 1996), que se dedica ao estudo de combinações de palavras estáveis na língua, as quais se assemelham às palavras independentes de maneira que, quando combinadas, apresentam um único significado, assim como uma palavra isolada.

Monteiro-Plantin (2017, p. 33) entende a Fraseologia como “uma disciplina independente, relacionada a todos os níveis de análise linguística (do fonético ao discursivo-pragmático)”. Trata-se de uma disciplina com foco na capacidade humana de produzir combinações involuntariamente:

o objetivo é o estudo das combinações de **unidades léxicas**, relativamente estáveis, com certo grau de idiomacidade, 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 2017, p. 33, grifo nosso).

Isso demonstra a necessidade de se intensificar as pesquisas sobre os estudos fraseológicos, uma vez que permitem contribuições com base no léxico, explorando padrões combinatórios, cuja produção sempre ocorre no uso da língua em sociedade humana. Sendo assim, é possível ainda que tais pesquisas se apresentem pautadas em um cunho mais funcionalista, mostrando a relevância de uma visão linguística mais aplicada ao uso.

A fim de resumir de modo geral os aspectos da Fraseologia, relacionamos algumas informações acordadas por Orenha-Ottaiano (2004):

- A Fraseologia é uma área de pesquisa situada dentro da Linguística, voltada à investigação de combinações lexicais recorrentes;

- Os estudos sobre Fraseologia têm ganhado projeção mais recentemente;
- Linguistas como Saussure (1969 [1916]), Pottier (1978), entre outros, já discorriam a respeito de combinações estáveis na língua, talvez não de forma tão explícita, porém mencionavam existir combinações não-livres na língua;
- Na década de 1940, a Fraseologia passa a ser considerada um ramo da ciência linguística, destacando-se o linguista russo V. V. Vinogradov (1947), que lançou os conceitos fundamentais da Fraseologia;
- O termo Fraseologia é polissêmico;
- Seu primeiro significado é muito semelhante a um dos significados de Terminologia. Refere-se a um conjunto de combinações fraseológicas de palavras de uma dada língua ou de termos de uma língua de especialidade;
- Essas combinações são recorrentes, mais ou menos estabilizadas/cristalizadas e abrangem as colocações, as coligações, as expressões idiomáticas, os binômios, os provérbios, dentre outros;
- Algumas expressões fraseológicas são decodificáveis, são transparentes, como a maioria das colocações e dos binômios, por exemplo.

Portanto, neste tópico tentamos apresentar brevemente alguns apontamentos teóricos da Fraseologia, enquanto área e disciplina, bem como suas contribuições científicas dentro na Linguística. No próximo item tratamos mais especificamente das tipologias adotadas pela Fraseologia e os aspectos que a definem.

2.5.2 As diferentes tipologias dos fraseologismos

Considerando o interesse nos estudos fraseológicos, encontramos na literatura diferentes autores que apresentam denominações distintas ao termo *unidade fraseológica*, como afirma Bevilacqua (1996, p. 10): locução fraseológica (BALLY, 1951); locução (CASARES, 1950); expressão fixa (ZULUÁGA, 1975); fraseologismo (TRISTÁ, 1988); combinação fixa de lexemas (ETTINGER, 1982); lexias complexa e textual (POTTIER, 1978); sintagma fixo, expressão idiomática, unidade complexa (FIALA, 1987); entidade fraseológica, fraseologismo (GOUADEC, 1994); frasema (GRÉCIANO, 1993), entre outros.

Orenha-Ottaiano (2004, p. 15), por sua vez, também se preocupou em apontar diversos pesquisadores e suas respectivas denominações fraseológicas, cujos termos, mesmo que não totalmente equivalentes, segundo a autora, receberam diferentes nomenclaturas: unidades

multipalavras (COWIE, 1998); expressões convencionais (*conventional expressions*), unidades (lexicais) frasais (*phrasal (lexical) units*) (PAWLEY, 2001); expressões multipalavras (LEWIS, 2000); estruturas formulaicas lexicalizadas (*lexicalized sentence stems*) (PAWLEY; SYDER, 1983); agrupamentos linguísticos (AZEVEDO, 1986); blocos pré-fabricados ou simplesmente *prefabs* (BOLINGER, 1976); unidades fraseológicas (VINOGRADOV, 1947; AMOSOVA, 1963; CARNEADO; TRISTÁ, 1983, TERMINASOVA, 1992); unidades lexicais *multivocabulares* (*multiword lexical units*) (ZGUSTA, 1967); expressões feitas (*ready-made utterances*) (LYONS, 1968); *frasemas* (MELCHUK, 1992); lexemas frasais, unidades frasais, unidades cognitivas (*cognitive units*) (KNOWLES, 1990); unidades lexicais multipalavras (*multi-word lexical units*), pacotes lexicais (*lexical bundles*) (BIBER, 1995); “*tijolinhos*” (*building blocks*) (PAWLEY; SYDER, 1983); “*pedaços*” (*chunks*) (NATION, 2001), *polipalavras* (*polywords*); “*tijolinhos fixos e por vezes fossilizados*” (KJELLMER, 1987); *lexias* textuais (POTTIER, 1978).

No entanto, Corpas Pastor (1996, p. 16) já havia mencionado que o termo Fraseologia sofreria controvérsias como outros fenômenos linguísticos, pois acreditava-se que os linguistas não chegavam a um acordo sobre um termo geral com relação à classificação que seria empregada em suas análises. A estudiosa pontua algumas terminologias mais genéricas, porém variadas, em sua obra *Manual de Fraseología Española*, utilizadas por pesquisadores dessa área na literatura espanhola. Percebemos que tais denominações foram elencadas pela autora e estão dispostas em três grupos, tendo as seguintes palavras como eixo: pluriverbal ou multipalavras, expressão fixa e fraseologia (CORPAS PASTOR, 1996, p. 16):

Quadro 1 – Diferentes denominações dos estudos fraseológicos

Palavra núcleo	Algumas referências e denominações
Pluriverbal ou Multipalavras	“expressão pluriverbal” (<i>Expresión pluriverbal</i>) (CASARES, 1992 [1950]); “unidade pluriverbal lexicalizada e habitual” (<i>unidad pluriverbal lexicalizada y habitualizada</i>) (HAENSCH <i>et al.</i> , 1982; CORPAS PASTOR, 1995) ou “unidade pluriverbal” (<i>unidad léxica pluriverbal</i>) na obra de Hernández (1989); “unidades de multipalavras” (<i>multiword unit</i>) (COWIE, 1985); dentre outros.
Fixa	“expressão fixa” (<i>expresión fija</i>) (ZULUAGA, 1980; GARCÍA-PAGE SÁNCHEZ, 1990; MARTÍNEZ MARÍN, 1991); “expressão fixa” (<i>fixed expression</i>) (ALEXANDER, 1978; MOON, 1992); dentre outros.
Fraseologismos	“unidade fraseológica ou fraseologismo” (<i>unidad fraseológica o fraseologismo</i>) (ZULUAGA, 1980); dentre outros.

Fonte: adaptado pelo autor com base em Corpas Pastor (1996, p. 16, tradução nossa).

Mesmo assim, Corpas Pastor (1996) entende que a fraseologia, assim como outros fenômenos lexicais, é passível de controvérsia devido à falta de consenso entre os linguistas

da área, como observamos nas denominações citadas anteriormente. De todas as denominações indicadas, a autora ressalta o uso da terminologia *Expresion Fija* (expressão fixa), pelo fato de enfatizar a ideia de algo fixo, afirmando que apenas em casos excepcionais os graus de fixação ocorrem (CORPAS PASTOR, 1996, p. 16).

Segundo Corpas Pastor (1996), mesmo os termos unidade “pluriverbal lexicalizada e habitual” apresentam uma natureza multiverbal e características mais explícitas, a saber: i) uso de duas palavras ortográficas; ii) apresentam certo grau de lexicalização; iii) se caracterizam pela alta frequência de cooperação na língua. Entretanto, a pesquisadora destaca ainda que a terminologia *unidade fraseológica*, por sua característica genérica, passou a ser mais usada por autores espanhóis, na Europa Ocidental, na antiga União Soviética e países do Leste como aponta Amadeu-Sabino (2011, p. 387) com base em Corpas Pastor (1996, p. 19). Desse modo, passamos a adotar esta última opção no decorrer desta pesquisa.

Outro ponto que nos chama a atenção foram algumas características que expõem com mais precisão sua representação na língua, como podemos observar a seguir:

Quadro 2 – Características das UFs 1

As unidades fraseológicas diferenciam-se de outras lexias	
•	Por serem expressões formadas por <i>várias palavras</i> (embora discordemos parcialmente dessa afirmação, por acreditarmos que devam ser consideradas as combinações formadas por, <i>pelo menos, duas palavras</i>);
•	Por estarem institucionalizadas (ou convencionalizadas) pela comunidade falante;
•	Por serem estáveis em diversos graus (que englobam tanto a <i>estabilidade formal</i> – o que significa que algumas unidades fraseológicas podem ter um grau menor de estabilidade, ao passo que outras podem ser altamente estáveis ou até mesmo fixas; quanto a <i>estabilidade semântica</i> – que reflete nas peculiaridades de sentido que apresentam);
•	Por apresentarem certas particularidades sintáticas ou semânticas;
•	Pela possibilidade de variação de seus elementos integrantes, seja como variantes lexicalizadas na língua ou como modificações ocasionais no contexto.

Fonte: Amadeu-Sabino (2011, p.387, grifos da autora).

Além disso, Amadeu-Sabino (2011) com base em Corpas Pastor (1996) apresenta algumas características específicas (frequência, institucionalização, estabilidade, idiomaticidade, variação, gradação) das UFs e suas respectivas definições que introduzem as peculiaridades desse fenômeno linguístico que vai além do simples conceito básico de unidades formadas por duas palavras ou mais. Vejamos o quadro a seguir:

Quadro 3 – Características das UFs 2

Caraterísticas	Definição
Frequência	Tanto de uso dessas combinatórias, quanto de coocorrência de seus elementos constitutivos;

Institucionalização	Sinônimo de convencionalização pela comunidade falante, graças, principalmente, à frequência de ocorrência e a sua fixidez formal e semântica;
Estabilidade	No sentido de <i>fixidez formal</i> e de <i>fixidez de conteúdo</i> (ou peculiaridades semânticas), aspectos estes estritamente relacionados, dado que a <i>fixidez formal</i> leva à <i>mudança semântica</i> ;
Idiomaticidade	Propriedade semântica de certas unidades fraseológicas em que o significado global não corresponde à somatória do sentido de cada um de seus elementos constituintes;
Variação	Uma característica presente em algumas UFs que possuem variantes (variações léxicas), entendidas como formas alternativas parcialmente idênticas em sua estrutura e componentes e que não apresentam diferenças de sentido;
Gradação	Que se baseia nos critérios de <i>grau de restrição colocacional</i> (ausência de restrição, restrição parcial e restrição total); de <i>fixidez sintático estrutural</i> (regular, regular com restrições e irregular) e de <i>opacidade semântica ou idiomaticidade</i> (semanticamente transparentes, metafóricas, semitransparentes e opacas).

Fonte: Amadeu-Sabino (2011, p. 388, grifos da autora).

Esses aspectos mais pontuais das UFs corroboram com Bevilacqua (2005) que também apresenta outras denominações aos tipos de fenômenos linguísticos estudados à luz da Fraseologia:

Para alguns autores, a fraseologia limita-se às expressões idiomáticas próprias de uma língua; outros consideram que ela inclui os provérbios, os ditados, as locuções e as lexias compostas. Há ainda quem considere que tais unidades possuam estruturas extremamente variáveis, podendo incluir palavras, grupos de palavras, locuções, expressões, orações, segmentos de frases, frases, conjunto de frases e assim por diante (BEVILACQUA, 2005, p. 74).

Por conseguinte, Tagnin (2013) afirma que a Fraseologia se ocupa do estudo de qualquer UF fixa em uma língua, como as colocações, as expressões convencionais e idiomáticas, os binômios, os provérbios, a prosódia semântica, dentre outros.

Monteiro-Plantin (2017), por sua vez, categoriza como Fraseologia as parêmiias ou sentenças proverbiais, expressões idiomáticas, colocações, pragmatemas, estereótipos e clichês, bordões, *slogans* etc. Portanto, para entendermos um pouco mais da funcionalidade das UFs na língua, indicamos algumas tipologias tratadas por Orenha-Ottaiano (2004, p. 20-21):

Quadro 4 – Tipologias das UFs

Fraseologismos	Características	Exemplos
Expressões idiomáticas	- Estruturas sintagmáticas complexas; única unidade lexical; apresentam um sentido particular dentro de uma realidade específica.	<i>Let the cat out of the bag</i> (deixar escapar um segredo); geralmente sem intenção de fazê-lo.
Colocações	- Combinações recorrentes; arbitrárias e convencionais; lexicalmente e/ou sintaticamente fixas	<i>run a business</i> (<i>dirigir um negócio</i>); <i>a</i>

	até certo grau; têm alcance colocacional mais ou menos restrito (ORENHA-OTTAIANO, 2017).	<i>booming business (um negócio próspero)</i> , etc.
Marcadores conversacionais	- Expressões que indicam a contribuição e intenção do falante dentro de uma conversação (apud KELLER, 1979).	<i>To my view ... (Na minha opinião...)</i>
Fórmulas de rotina	- Expressões de situações sociais específicas; altamente previsíveis em determinados discursos; fórmulas situacionais obrigatórias.	<i>Bom apetite (Português)</i> <i>Enjoy yourself (Inglês)</i>
Verbos frasais (<i>phrasal verbs</i> ; <i>two-word verb</i> ; <i>three-word verbs</i>)	- Classe que não existe em português; formados por um verbo seguido de uma partícula adverbial; *Os <i>three-word verbs</i> (verbo + partícula adverbial + preposição – indica a regência do <i>phrasal verb</i>).	<i>take off</i> (= decolar) * <i>put up with</i> (= tolerar)
Provérbios	- Frase curta; de origem popular; frequentemente com ritmo e rima; rica em imagens; sintetiza um conceito a respeito da realidade ou uma regra social ou moral.	Ex.: <i>The early bird catches the worm (Deus ajuda quem cedo madruga)</i> .
Frases feitas	- Frases feitas diferem das fórmulas de rotina porque, se não as utilizarmos em um dado contexto, não implicarão a ruptura dos padrões comportamentais convencionais.	Ex: <i>It's none of your business! (Não é da sua conta!)</i> <i>Mind your own business! (Cuide da sua vida!)</i>
Binômios	- Formados por duas palavras da mesma categoria gramatical e ligadas por uma conjunção ou preposição (apud MALKIEL, 1959).	Ex: <i>profit and loss</i> (lucros e perdas) <i>live and learn</i> (vivendo e aprendendo).

Fonte: Orenha-Ottaiano (2004, p. 20-21).

A nossa intenção com esse recorte de pesquisas foi a de proporcionar uma visão geral dos diferentes tipos de fraseologismos, considerando o conceito e a funcionalidade dessa disciplina, a fim de colaborar também com questionamentos deste estudo. Por exemplo, a questão da baixa exploração das UFs empiricamente na graduação, em especial das colocações, à luz da Linguística de *Corpus*.

2.5.3 A Fraseologia na língua geral e na língua especializada

Monteiro-Plantin (2017, p. 1) afirma que a “Fraseologia é também o nome da disciplina que tem por objeto de estudo os fraseologismos ou as UFs de uma língua, ao tratarmos da língua comum, em oposição à língua de especialidade, objeto a cargo da Terminologia”.

Para Saussure (1988, p. 148-149), a Fraseologia, enquanto unidade combinatória, era intitulada como um “conjunto de agrupamentos”, já indicando os aspectos do léxico combinado ao afirmar que “não falamos por signos isolados, mas por grupos de signos, por massas organizadas que são elas próprias signos”. Em Saussure (1988, p. 149) também

encontramos que a língua é o resultado de comparações associativas ou sintagmáticas, pois é um conjunto de relações usuais e constitui o funcionamento da mesma. No caso das associações sintagmáticas, segundo o autor, quase todas as unidades da língua dependem das palavras ao redor da cadeia falada e das partes sucessivas compostas por ela.

Bevilacqua (2016) apresenta alguns conceitos relacionados à fraseologia da língua comum, tão importantes quanto Ferdinand Saussure, sobre o qual ela também discorre em seu trabalho. Aproveitamos alguns deles para demonstrar a visão fraseológica tanto da língua comum (BALLY, 1951; CASARES, 1950; POTTIER, 1978; FIALA, 1987), como também da língua de especialidade (PAVEL, 1993; BLAIS, 1993)

Bally (1951), na visão de Bevilacqua (1996, p. 23) elenca as UFs por agrupamento, afirmando que “a assimilação dos fatos da língua ocorre, principalmente, pelas associações e agrupamentos”. Segundo o autor, esses agrupamentos “podem ser passageiros, mas também, pela repetição, passam a ter um caráter usual e a formar unidades indissolúveis”.

Para exploração do campo fraseológico, Bevilacqua (1996, p. 27) retoma a ideia das lexias como uma unidade lexical memorizada, a partir de Pottier (1978), com base em dois dos quatro grupos apresentados pelo autor:

- *complexa*: constitui-se em uma sequência em vias de lexicalização e possui graus diferenciados. Ex: *um complexo industrial, tomar medidas, sinal vermelho*;
- *textual*: consideradas como lexias complexas que alcançam o nível de enunciado ou de um texto. Como exemplos desse grupo estão os *hinos, preces, charadas, provérbios* etc.

De acordo com Bevilacqua (1996), Pottier (1978) apud BEVILACQUA, 1996), as lexias podem ser memorizadas e se classificam por graus diferentes: *rígidas* (apresenta uma sequência invariável. Ex.: *procurado pela justiça*), ou *variável* (uma parte se mantém estável e a outra não. Ex.: *tudo leva a crer/supor/pensar que*).

Nessa perspectiva, Bevilacqua (1996, p. 29) resume as UFs na visão de Saussure (1988), Bally (1951) e Pottier (1978) apresentadas pelas seguintes características: **a)** pluriverbalidade; **b)** estabilidade sintático-semântica; **c)** inseparabilidade de seus elementos; **d)** não comutabilidade; e **e)** frequência. Assim, os três autores “mencionam ainda a possibilidade de uma UF ser formada por uma parte fixa e outra variável, podendo ter, portanto, diferentes graus de fixação” (BEVILACQUA, 2016, p. 29).

Em contrapartida à fraseologia de língua comum, temos, então, a fraseologia de língua de especialidade que se difere da primeira, pois se trata de uma linguagem que é “utilizada para a comunicação cotidiana entre os falantes de determinada comunidade linguística” (BEVILACQUA, 2016, p. 12). A autora ainda aponta que o critério de pluriverbalidade presente na fraseologia de língua geral também está presente na de especialidade, pelo fato de conter mais de uma palavra em seu agrupamento linguístico. A diferença está no fato de que, a fraseologia de língua de especialidade pode ser compreendida por uma palavra “núcleo”, intitulada como “Unidade terminológica” (UT) que, segundo ela:

Significa dizer que a **UT**, constitui-se segundo as regras morfológicas e sintáticas de uma língua, além de referir-se a um conceito que, no caso, é **específico de um domínio**, ou seja, ela assume um novo significado ao passar a fazer parte de determinada área e, portanto, de **uma terminologia**” (BEVILACQUA, 1996, p. 36, grifo nosso).

Ao definir a função lexical da fraseologia de língua de especialidade, com base em Pavel (1993) Bevilacqua (1996, p. 37) denomina que as “UTs são consideradas como núcleos de co-ocorrentes usuais ou privilegiados nos textos de uma especialidade”. Vejamos a representação estrutural a seguir:

Quadro 5 – Estrutura fraseológica de Pavel (1993)

<p>UT + adj (agregado bidimensional) V + UT (absorver um agregado) UT + V (agregado + adotar uma configuração)</p>

Fonte: Pavel (1993) com base em Bevilacqua (1996, p. 37).

Em seguida, acrescenta a ideia proposta por Blais (1993), Bevilacqua (1996, p. 37) se apodera do termo “fraseologismo”, ao invés de UF, e o define “como a combinação de elementos linguísticos, própria a um domínio, em que um deles é um termo núcleo, relacionada semântica e sintaticamente e para os quais existe uma restrição paradigmática”. Os exemplos estruturais apresentados pela autora são:

Quadro 6 – Estrutura fraseológica de Blais (1993)

<p>termo núcleo + adj (ruído surdo) n + termo núcleo (golpe surdo) n+ prep + art + termo núcleo (vibração das rodas) verbo + art + termo núcleo (apertar o freio)</p>
--

Fonte: Blais (1999) com base em Bevilacqua, (1996, p. 37).

Partindo dessas definições discutidas até aqui, observamos que a fraseologia especializada possui uma forte ligação com embasamentos terminológicos, partindo da averiguação de padrões combinatórios em textos de áreas específicas. Colson (2008, p. 192, tradução nossa) ressalta a relevância de “[...] investigar a diversidade da fraseologia, concentrando-se em casos específicos da linguagem”⁶, que comprova o surgimento de pesquisas de interesse nesse tipo de linguagem.

A tradução torna-se, nesse contexto, uma área privilegiada, por concentrar análises de fraseologia de língua de especialidade, devido ao seu aspecto comparativo. Rebecchi e Silva (2017, p.102) afirmam que as “fraseologias especializadas também são essenciais para a escrita e tradução de textos”⁷. Colson (2008) considera, ainda, que todo tipo de comparação entre padrões combinatórios entre línguas é considerado como fraseologia contrastiva, mas criticou muitos estudos tradicionais meramente descritivos e comparativos, sem a preocupação com as abordagens teóricas. Sendo assim, notamos que a pesquisa acaba por validar toda atividade de comparação entre uma língua e outra, resultando em um produto específico, como os glossários e dicionários especializados (BOURIGAULT; JACQUEMIN; L'HOMME, 2001). Dessa forma, além de pensar no produto final em uma área de especialidade, os parâmetros pragmáticos devem ser também considerados (SALAMANCA MARTÍNEZ; SUAREZ DE LA TORRE, 2017).

De igual modo, Colson (2008) aponta contribuições que levantam questões teóricas relacionadas à Fraseologia, que destacam problemas de classificação e terminologia, bem como a interação da Fraseologia com o contexto (BURGER, 1998; COWIE, 1998; GLÄSER, 1984, 1985).

Segundo Colson (2008), na atual conjuntura, o avanço da fraseologia especializada ocorre com base nos interesses na criação de dicionários especializados (MOON, 2008), principalmente nas questões relacionadas ao processamento de dados das línguas naturais de forma computadorizada (TSCHICHOLD, 2008), especialmente com o surgimento de dicionários eletrônicos de aprendizes (HEID, 2008), conforme aponta o autor. Esse fato se deve pelo aparecimento de pesquisas voltadas para as questões interdisciplinares da Fraseologia que, de certa forma, vem contribuindo para a produção de materiais em áreas de especialidade.

⁶ Do original: “[...] to investigate the diversity of phraseology by concentrating on specific cases across languages” (COLSON, 2008, p. 192).

⁷ Do original: “Specialized phraseologies are also essential for writing and translating texts” (REBECHCHI; SILVA, 2017, p. 102).

Portanto, notamos que os estudos sobre a Fraseologia podem contribuir para as pesquisas que exploram tanto a língua geral quanto áreas específicas e que, de certa forma, apresenta um forte caráter interdisciplinar, principalmente por meio da área de especialidade, tornando-se também uma disciplina de grande interesse entre os linguistas de *corpus* (BEVILACQUA, 1996; MONTEIRO-PLANTIN, 2017; ORENHA-OTTAIANO, 2004, 2009; TAGNIN, 1999, 2013, dentre outros).

2.6 A interface entre a Fraseologia e a Linguística de *Corpus* na contemporaneidade

O avanço nas investigações sobre as unidades lexicais combinatórias ainda carrega uma grande carga conceitual dos primórdios da Fraseologia, desde seu início com os estudos de V.V. Vinogradov a partir dos anos de 1940. Notamos, em seção anterior, que mesmo os teóricos mais antigos continuam servindo de base para recentes investigações que tem a Fraseologia como foco. De acordo com Burger *et al.* (2007), a pesquisa com essa disciplina no cenário internacional passou a se desenvolver com mais velocidade a partir dos anos de 1970. Desde esse interesse contínuo pela Fraseologia, entre os anos de 1970 e 1980, a área ganha força por meio de conferências internacionais e publicações científicas.

No final do século XX e início do século XXI, esse fato pode ser claramente observado, não somente na expansão dos interesses pela área, mas principalmente nas questões interdisciplinares, contando com contribuições em áreas especializadas e culminando também no incentivo desses eventos, sendo eles: o *International Symposia on Phraseology* (1994, 1996) e o *International EUROPHRAS Congresses* (desde 1992), segundo Cowie (1998, p. 1). Há ainda outros como o *Yearbook of Phraseology* (2017, 2018) e as quatro versões do Congresso Brasileiro de Fraseologia (2011, 2013, 2016, 2018), o qual vem apresentando contribuições da Fraseologia na língua portuguesa, ou em contraste com outros idiomas.

Meunier e Granger (2008) alegam que, de modo interdisciplinar, a Fraseologia vem sendo explorada em todos os campos da linguagem, ocupando um amplo espaço de destaque em diversas áreas, desde o processamento das línguas naturais até o ensino de línguas estrangeiras. Portanto, passou a ser recentemente reconhecida como uma disciplina, tida como uma subárea da Lexicologia, dedicada ao estudo das combinações de palavras, como já vimos em seção anterior.

A pesquisadora Corpas Pastor, em entrevista recente (CORPAS PASTOR; ORTIZ ALVAREZ, 2017) sobre o assunto, afirmou que o grande desenvolvimento da Fraseologia,

intitulada como “contemporânea”, está atrelado à influência da escola russa, observada, principalmente, com base nos estudos alemães e cubanos. Ela alega ainda que, no século XXI, ocorreu um desencadeamento da Fraseologia nas línguas alemã, espanhola, italiana e francesa.

As autoras também consideram a interface interdisciplinar da Fraseologia certas de que, atualmente, a disciplina atingiu certo grau de “maturidade” por meio da exploração de novos horizontes, contribuindo para pesquisas, tais como: “os estudos contrastivos e de tradução; a gramática de construções; os estudos cognitivos e psicolinguísticos; e as aplicações da Linguística de *Corpus* e do Processamento de Linguagem Natural” (CORPAS PASTOR; ORTIZ ALVAREZ, 2017, p. 263). Dessa forma, o que temos observado, até então, é a maneira como se molda essa Fraseologia contemporânea, não porque tem se preocupado com textos atuais, mas sim tem se apresentado cada vez mais com uma forte ligação com as áreas afins que citamos acima.

Igualmente, a Fraseologia atingiu também um nível de maturidade, em razão da aceitação e apropriação de seus conceitos por parte de linguistas e, principalmente, pelos estudiosos da LC. É nesse cenário, pós-moderno e híbrido, que a LC ganha espaço com os estudos fraseológicos, pois, segundo Corpas Pastor e Ortiz Alvarez (2017, p. 263), essa área veio para modificar os estudos com base na Linguística geral, provocando uma mudança paradigmática em todas as esferas da pesquisa científica. As autoras pontuam que, “as unidades fraseológicas, terminologia adotada por esta pesquisa, representam o uso vivo da linguagem, e, portanto, requerem um conjunto de dados representativos que só os *corpora* podem proporcionar” (CORPAS PASTOR; ORTIZ ALVAREZ, 2017, p. 263).

Elas apostam, ainda, em um futuro baseado no uso dos *corpora* de grandes proporções, mas pontuam a necessidade da sensibilidade à variação linguística. É nesse ponto, que cabe ao pesquisador a responsabilidade no cuidado com os objetivos ao compilar um determinado *corpus*, seguindo critérios adequados, como já discorremos em seção anterior.

De modo geral, pensando em Fraseologia, enquanto área ou disciplina, como quer que seja, esta tem possibilitado pesquisas em diversos assuntos, veemente fortalecida por uma base interdisciplinar, como descreve os pesquisadores Meunier e Granger (2008) na obra *Phraseology, an interdisciplinary perspective*, organizado após um congresso de Fraseologia em outubro de 2005, em Louvain-la-Neuve, na Bélgica, que reúne vários artigos, dispostos em uma série de capítulos.

A obra ainda traz outras vozes de pesquisa em relação ao uso das UFs à luz da LC. Na seção 1, *Phraseology and linguistic theory: A brief survey*, Gries (2008) discute o papel que a

Fraseologia ocupa, tanto em diferentes teorias com base linguística (linguística generativa, linguística cognitiva), bem como no paradigma metodológico desempenhado da LC. Já o tópico 3, *Phraseology across languages and cultures*, Colson (2008) abordou assuntos relacionados à forma com que a Fraseologia translinguística possui ligação direta com os estudos da tradução, auxiliando ativamente em práticas da tradução por meio da LC multilíngue, a Fraseologia contrastiva e o processamento de linguagem natural, por uma perspectiva perspicaz. Para ele, as investigações multilíngues a partir de *corpora* podem oferecer uma análise estatística das várias categorias de combinações específicas definidas, como também uma metodologia confiável.

Por meio desse panorama, fica evidente que “a LC tem se mostrado uma metodologia confiável para o estudo da fraseologia nas últimas décadas” (NAVARRO, 2019, p. 188, tradução nossa)⁸. Nos últimos 30 anos houve uma sequência de pesquisadores (GRANGER; PAQUOT, 2008), dentre outros, conforme aponta Carrió-Pastor (2019), que passaram a focar, por exemplo, em estudos voltados às colocações dentro da Fraseologia.

Entretanto, ainda existe um ponto de atenção, pois nesse avanço da Fraseologia, sobre a necessidade de encontrar “um consenso com relação à classificação das unidades fraseológicas, ao enfoque pragmático dessas unidades, à introdução delas nas aulas de LM e LE para poder desenvolver a competência fraseológica dos usuários da língua” (CORPAS PASTOR; ORTIZ ALVAREZ, 2017, p. 268), seria possível compreender com mais precisão a relevância social dessa área nas questões de ensino.

Porém, ainda não seria um caminho muito sólido pensar em desenvolver a competência fraseológica dos usuários de uma segunda língua ou LE, sem antes refletirmos em sua produção. Qualquer competência envolve um conjunto de habilidades específicas para cada caso e, se desenvolvidas de forma adequada, podem contribuir para um uso mais consciente da língua, principalmente durante a formação de um estudante em suas práticas tradutórias na vida acadêmica. As competências requeridas para um tradutor em formação diferem daquelas que aplicam apenas ao ensino de línguas como segunda língua ou LE. O tradutor tem uma responsabilidade bem maior em relação ao texto produzido.

Paquot e Granger (2012) afirmam que os estudos fraseológicos em L2 passaram a produzir descobertas relevantes, pois apresentam resultados significativos quando realizados de forma comparada, ou seja, pela comparação entre um *corpus* com traduções de alunos e outro de referência em língua nativa. As autoras distinguem ainda dois aspectos da

⁸ Do original: “In the last decades, corpus linguistics has proved itself to be a reliable method for the study of phraseology” (NAVARRO, 2019, p. 188).

Fraseologia, com base em *corpus*, que são essenciais para pesquisas sobre UFs, a saber: a “co-ocorrência” entre palavras e a “frequência” entre elas, as quais devem ser analisadas de forma separada, simplesmente pelo fato de exigirem do usuário da língua, competências distintas, o que resulta em dois tipos de desafios, principalmente para alunos de um segundo idioma.

Para elas, a coocorrência, por sua vez, é responsável pelo agrupamento de dois itens lexicais, que podem ser adjacentes, mas geralmente não apresentam essa característica. Paquot e Granger, (2012) afirmam que esse aspecto foi intitulado por Firth (1957) como *mutual expectancy* (expectativa mútua), quando ocorre uma preferência de uma palavra a ser combinada com outra; por exemplo, no caso das colocações, estas são definidas por combinações lexicamente restritas, devido à limitação inserida em uma construção gramatical específica (verbo-objeto, advérbio-adjetivo, adjetivo-substantivo, etc.).

Contudo, a nova abordagem apresentada por Sinclair (1987 apud PAQUOT; GRANGER, 2012, p. 70), traz o termo colocação como uma significação de palavras que coocorrem estatisticamente. As autoras fazem referência aos métodos com base em *Mutual Information* (informação mútua), que no caso das colocações, por exemplo, permitem extraí-las não por critérios linguísticos, mas por meio de métodos estatísticos e passam a considerar tanto as colocações arbitrárias como as livres.

No caso da recorrência, por exemplo, conforme descreve Paquot e Granger (2012), trata-se da repetição de palavras agrupadas, isto é, duas ou mais palavras que são frequentes. Esse fato pode ser observado por meio do *corpus*, que permite visualizar a frequência de certos padrões combinatórios existentes na língua em larga escala. Esses padrões podem ser chamados de *lexical bundles* (BIBER *et al.*, 1999 apud PAQUOT; GRANGER, 2012, p. 16), os quais podem apresentar-se “gramaticalmente completos” (*that’s a good idea, on the other hand*), ou incompleto (*in the case of the, it should be noted that*). As autoras acrescentam que esses *lexical bundles*, intitulados como *conventionalized building blocks* (blocos de construção convencionalizados), aparecem na língua como fórmulas de rotina durante a produção da linguagem (ALTENBERG, 1998, p. 122 apud PAQUOT; GRANGER, 2012, p. 16).

Temos observado nos estudos acerca dos pressupostos teóricos da LC que eles vêm contribuindo veementemente para as pesquisas da Fraseologia, por meio da identificação e observação de dados linguísticos como esses. Por consequência, o avanço das tecnologias digitais na área linguística aumentou o interesse por parte de muitos pesquisadores na investigação das UFs.

As autoras Meunier e Granger (2008, p. 19) consideram a LC como a responsável pelo desenvolvimento dos estudos sobre as UFs, oferecendo métodos automatizados de extração e análise de dados linguísticos aos pesquisadores da área, principalmente por abranger um número estatístico bem maior dessas unidades. Evidenciamos esse fato devido ao hibridismo dos interesses nos estudos sobre *corpus*, como observamos, por exemplo, no grande projeto *Multilingual Student Translation Corpus* (MUST) que abarca vários *corpora* em um só, em interface com pesquisas de várias línguas, especialmente em um período altamente globalizado.

A vantagem desse tipo de abordagem e o estreitamento entre as áreas, Linguística e a Computação, oferecem subsídios em programas de observação de dados linguísticos de alta tecnologia e veemente mais assertivo em relação à análise de informações com base em *corpus*. Esses programas (vide seção sobre Linguística de *Corpus*) têm sido construídos com foco em avanços nos estudos das UFs. Uma das ferramentas mais eficazes atualmente é o gerenciador de *corpus Sketch Engine* (KILGARRIFF *et al.*, 2014), uma plataforma *on-line*, gratuita para teste (um mês apenas), que oferece uma capacidade de analisar fenômenos linguísticos, proporcionando resultados que diminuem, e muito, o tempo de busca por um padrão combinatório.

Desse modo, percebe-se que os estudos linguísticos sobre a Fraseologia com base em *corpus*, enquanto uma disciplina que atravessa muitos aspectos da linguagem, vem crescendo no campo da pragmática, da fala e cognitiva dos estudos de linguagem (LUQUE NADAL, 2012). O uso da Fraseologia ainda é um dos problemas enfrentados, vista como um “grande obstáculo” por aprendizes de inglês como L2, como bem destaca Gilquin, Granger e Paquot, (2007, p. 329), ao observar um *corpus* de estudantes de inglês. Portanto, nesse cenário, ainda é preciso que ocorra cada vez mais o desenvolvimento de pesquisas que testem o valor das UFs no âmbito de formação acadêmica de tradutores, atreladas à teoria e metodologia da LC, levantadas no contexto da investigação científica, com vistas à prática em sala de aula.

2.7 As colocações

A ideia de colocação surgiu no século XX com os estudos de Firth (1957), pioneiro na exploração linguística sobre esse fenômeno, “embora já tivesse sido utilizado em estudos

linguísticos” (MARTELLI, 2017, p. 13, tradução nossa)⁹. Anos depois, os chamados *Neo-Firthians* (seguidores de Firth) deram prosseguimento em seus estudos sobre esse termo, com destaque para o professor universitário e pesquisador John McHardy Sinclair, o qual teve maior participação nessa trajetória, seguido dos pesquisadores Michael Hoey, Susan Hunston, Bill Louw, Michael Stubbs, Wolfgang Teubert e Elena Tognini-Bonelli, alunos de John McHardy Sinclair na universidade de Birmingham, na Inglaterra (MCENERY; HARDIE, 2012, p. 122).

Mais tarde, com os avanços tecnológicos por meio da LC, principalmente por meio do processo de identificação probalístico (HALLIDAY, 1991), isto é, estatístico, os estudos sobre o uso das colocações se intensificaram devido à programas de gerenciamento de *corpus* (*Wordsmith tools*, *AntConc*, *Sketch Engine*, dentre outros), os quais possuem ferramentas de análise de dados linguísticos, como propõe a Linguística de *Corpus*. Observamos essa funcionalidade em Granger (2002, p. 21, tradução nossa), pois explica que “os dados do *corpus* fornecem muito mais objetividade para a escolha de vocabulário, com maior foco nas combinações de palavras de todos os tipos”¹⁰, sendo as colocações um exemplo.

O termo colocação pode ser basicamente entendido como combinações de duas ou mais unidades de palavras, que geralmente “andam juntas” (TAGNIN, 2013). Para Shitu (2015) a colocação também significa duas palavras que se “posicionam juntas”. Segundo a autora, a origem de *collocation* está em: *col* (junto), *loc* (colocar); *ate* (verbo sufixo) e *ion* substantivo sufixo (NATION, 1990). Já se percebe de imediato, uma intenção de significado presente na própria descrição das duas primeiras partes da palavra *collocation* em inglês, facilmente observadas em *col* e *loc*.

Dessa forma, esse agrupamento de palavras para a formação de um único significado não ocorre apenas por uma palavra isolada, mas pelas características na associação entre duas ou mais palavras, vista como uma “abstração do nível sintagmático” (FIRTH, 1968 apud McENERY e HARDIE, 2012, p. 123). Tal abstração pode ser notada na colocação *village green* (TAGNIN, 2013), que pelo fato de ser tão restrita, possui uma carga semântica imagética. Um aprendiz que não esteja familiarizado com tal combinação poderia não compreender, logo de imediato, que se trata da colocação “praça pública” em português.

Percebe-se que as colocações são um tipo de unidade semi-pré-fabricada, comprovando, assim, o conceito de arbitrariedade (NESSELHAUF, 2005). Um exemplo

⁹ Do original: “[...] although it had already been used in linguistic studies” (MARTELLI, 2007, p. 13).

¹⁰ Do original: “Corpus data has provided a much more objective basis for vocabulary selection, has led to greater, attention to word combinations of all types” (GRANGER, 2002, p. 21).

disso, com base em Cowie (1981, 1994), ocorre com a colocação *perform a task*, em que o verbo *perform* não pode ser substituído por *make* ou *do* (MARTELLI, 2007, p. 17). Esse processo de arbitrariedade na junção dessas palavras envolve ainda a convencionalidade da colocação e sua frequência na língua (ORENHA-OTTAIANO, 2004, 2009; TAGNIN, 2002, 2013).

O que notamos, então, é que uma vez que a colocação é uma combinação restrita, o desconhecimento desse fenômeno pode comprometer seu uso de forma adequada em um texto traduzido para outro idioma. Quando um aluno não aprende a enxergar essas restrições do léxico combinado, como as colocações, causará certo estranhamento no leitor fluente da língua-alvo.

Então, no que tange o uso aleatório da colocação na aprendizagem de uma língua, a autora cita Fillmore (1979), que apresenta em seus estudos o conceito “falante ingênuo”, aquele que tem dificuldades ao utilizar um determinado fraseologismo em outro idioma, por exemplo, a colocação. Na tradução, essa ingenuidade pode ocorrer no momento da leitura e interpretação do texto original até o uso de uma colocação correspondente na tradução para outra língua (TAGNIN, 2002, p. 193), na tentativa de utilizar combinações arbitrárias presentes em um idioma de forma livre e aleatória.

De acordo com Tagnin (2002, p. 192), no caso das colocações, a ingenuidade acontece quando este não se “sabe que há determinadas combinações fixas ou semi-fixas conhecidas como colocações”, conforme os exemplos a seguir apontados pela autora: substantivo + substantivo (*credit card*), substantivo + adjetivo (*nursing home*, *silent movie*), substantivo, como sujeito + verbo (*a river flows*) ou verbo + substantivo como objeto (*pay a visit*), verbo + advérbio (*pay dearly*), e adjetivo + advérbio (*deeply hurt*).

Por isso, nesta pesquisa, nos baseamos no conceito de colocações de Orenha-Ottaiano (no prelo):

trata-se de combinações recorrentes, arbitrárias, convencionalizadas, abrangentes, lexicalmente e/ou sintaticamente fixas até certo grau e que podem ter um alcance colocacional mais ou menos restrito. São, ainda, combinações específicas de um idioma e de uma cultura e, como tal, a colocabilidade de seus elementos pode variar significativamente de um idioma para outro, sendo constituído, portanto, por sua própria rede de colocações.

Um exemplo de restrição no uso da colocação ocorre quando dizemos *amigo secreto* e não *amigo escondido*, ou seja, a troca de uma combinação por outra pode causar

estranhamento para o nativo em português. A seguir, tratamos da relação entre as colocações de língua geral e as colocações da língua especializada.

2.7.1 A taxonomia das colocações

De modo geral, Hausmann (1985) e Orenha-Ottaiano (2004, 2009, 2015) apresentam a taxonomia das colocações em quatro classes gramaticais (verbais, nominais, adjetivas e adverbiais), conforme descrito abaixo:

VERBAIS – com quatro formas básicas:

- Verbo **colocado** + Substantivo **base**: *acquire shares*;¹¹
- Substantivo **base** + Verbo **colocado**: *investments dropped*;
- Verbo **colocado** + Preposição + Substantivo **base**: *dispose of shares*;
- Verbo **colocado** + Partícula Adverbial + Substantivo **base**: *set up a business*;
- Verbo **colocado** + Adjetivo **base**: *grow strong*.

NOMINAIS – com duas formas básicas:

- Substantivo **base** + Substantivo **colocado**: *share subscription*;
- Substantivo **colocado** + Preposição + Substantivo **base**: *holder of shares*; *an attack of flu*.

ADJETIVAS – com uma forma:

- Adjetivo **colocado** + Substantivo **base**: *bearer shares*; *lifethreatening illness*.

ADVERBIAIS – com três formas básicas:

- Advérbio **colocado** + Adjetivo **base**: *fully eligible*;
- Verbo **base** + Advérbio **colocado**: *drop dramatically much*;
- Advérbio **colocado** + Verbo **base**: *fully paid*; *duly appointed*.

Para identificar uma colocação, Tagnin (1999) afirma que é preciso observar as seguintes características:

- Recorrências;
- Não idiomaticidade, ou seja, seu significado tem que ser composicional;
- Coesão, é necessário que haja uma ligação muito forte entre seus elementos, muito mais forte do que se esperaria de uma combinação qualquer;
 - Restrição contextual, deve haver uma probabilidade de que ocorram dentro de um contexto específico;
 - Coocorrência arbitrária entre seus elementos, ou seja, não há razão semântica que explique tal ocorrência.

¹¹ Os exemplos foram extraídos do *corpus* de negócios investigado em Orenha-Ottaiano (2004).

Além dessas características, uma colocação é formada por dois elementos (determinante e determinado). Orenha-Ottaiano (2015), com base em Hausmann (1985), exemplifica que “aquele que determina é chamado de base, elemento autônomo, enquanto o outro, determinado, atua como colocado, e que somente pode ser interpretado semanticamente pela colocação”. Um exemplo oferecido pela autora é a combinação *issue shares*, em que *issue* (base) determina a função da palavra *shares* (colocado). Na tentativa de esclarecer as funções linguísticas, Orenha-Ottaiano (2015) resume a concepção de base e colocado da seguinte forma:

Base:

- um elemento independente;
- semanticamente autônoma;
- traduzível, independentemente de seu uso na colocação; e
- determina padrões lexicais que podem combinar com ela.

Quanto ao colocado:

- funciona como um conceito modificador;
- é semanticamente interpretável somente dentro da colocação;
- sua tradução depende do uso na colocação;
- é escolhido por uma dada base para formar uma colocação (HEID; MARTIN; POSCH, 1991).

Segundo Hausmann (1985 apud TAGNIN, 2002, p. 195), a base na colocação possui uma carga semântica maior, determinando de que tipo de colocação se trata. Por exemplo, segundo a autora, na combinação de verbo (base) + um substantivo (colocado), refere-se à uma colocação verbal, e na colocação de adjetivo + substantivo, uma colocação adjetiva. Por exemplo, nas colocações a partir do nódulo *political*, enquanto adjetivo base (por exemplo, *political institutions*, *political beliefs*, *political leader*, *political system*, *political party*). A autora afirma que as colocações nominais e adjetivas são mais frequentes no que denomina “inventário fraseológico”.

Tagnin (2002, p. 194) afirma que as colocações são “palavras que co-ocorrem em frequência maior do que se tratasse de uma combinação aleatória”. Dessa forma, o levantamento das colocações com base na frequência do *corpus* permite a identificação de dados mais concretos no uso desse fenômeno linguístico. Analisar as colocações baseado nos padrões de sua coocorrência no *corpus*, segundo McEnery e Hardie (2012), é um processo de descarte da busca por colocações de modo intuitivo, por exemplo, quando se recorre a um falante nativo para comprovar se uma colocação é usada ou não na língua. Para o autor, muitos grupos definem as colocações como “uma ampla variedade de co-ocorrências de diferentes padrões que podem ser extraídos de um *corpus*” (MCENERY; HARDIE, 2012, p. 123, tradução nossa)¹².

A investigação de colocações por meio do *corpus* pode garantir ao pesquisador e ao estudante de tradução mais bagagem teórica e um domínio mais seguro no uso e exploração do léxico. O aspecto relevante é que não é possível ter um falante nativo à disposição o tempo todo, e este também não daria conta de todas as fraseologias existentes em sua própria língua, o que acabaria sendo uma consulta que se pautaria na intuição de qualquer forma.

2.7.2 As colocações de língua geral e de língua especializada

Como abordamos anteriormente, no tópico sobre Fraseologia de língua geral e de especialidade, consideramos viável dedicar esta seção às colocações da língua geral em contraste com as colocações de língua especializada. Em linhas gerais, tanto as colocações de língua geral quanto as de especializada se definem por apresentar uma base estrutural, porém, o que as diferencia é o fato das colocações de língua especializada serem unidades lexicais caracterizadas pelo contexto da língua geral, pois possui a mesma forma, mas sentidos diferentes, enquanto que as colocações de língua especializada se identificam por conter um termo ou uma unidade terminológica (UT) específica como base na combinação.

De modo geral, para Lorente *et al.* (2017), a colocação tornou-se o grande objeto de interesse entre os linguistas. Muitas reflexões da colocação na língua são referentes à UF em diversas abordagens em que esta é utilizada. A autora ainda afirma que a colocação é o assunto principal no campo da aquisição de primeira e segunda línguas, para tradutores e lexicográficos, revisão e correção de textos, como também para os teóricos e estudiosos do léxico (LORENTE *et al.*, 2017) que lidam com o tratamento das colocações de língua geral e

¹² Do original: “[...] a wide range of different co-occurrence patterns that may be extracted from a corpus” (HARDIE; MCENERY, 2012, p. 123).

das colocações de língua especializada separadamente, ou ambas concomitantemente, dependendo do texto e do gênero em que são veiculadas.

Como temos percebido no decorrer da leitura para este trabalho, a gama de pesquisas sobre as UFs é grande, principalmente quanto ao uso das colocações. Lorente *et al.* (2017) ressaltam o alto interesse nas colocações gerais, comumente encontradas em todos os gêneros. Hoje, com o amparo da LC, a identificação de colocações de modo geral ficou mais acessível. Martelli (2007, p. 16, tradução nossa) comenta sobre as várias definições do termo colocação e afirma que, na maioria das vezes, “[...] é usado para se referir a algum tipo de relação estabelecida entre itens lexicais ao longo do eixo sintagmático”¹³. Porém, a autora chama a atenção para as colocações baseadas em duas abordagens *frequency-based approach* e *the phraseological approach*. Isso se deu devido ao grande interesse dos linguistas de *corpus* que, conforme apresentamos neste trabalho, passaram a se dedicar veemente para uma observação da colocação enquanto fenômeno linguístico.

Sendo assim, o ganho que os dicionários bilíngues de língua geral obtiverem a partir dos estudos empíricos propostos pela LC, um destaque para milhares de entradas de verbetes que tratam das colocações de língua geral, tornando-se, dessa forma, materiais de uso mais confiável. Corpas Pastor (2015) listou alguns exemplos de colocações, as quais foram extraídas de dicionários bilíngues, como o *Collins Spanish-English Dictionary*, o *Larousse Spanish-English Dictionary* e o *Oxford Spanish-English Dictionary*. Por meio desse material ela mostra que a colocação *pay homage* (“prestar homenagem”) tem seu equivalente em espanhol como *rendir homenaje*. Se observarmos atentamente, essas colocações podem aparecer tanto na escrita quanto na fala, sendo utilizadas de modo geral.

Por outro lado, em virtude do aspecto interdisciplinar que a Fraseologia alcançou nos últimos anos, as colocações de língua especializada passaram a ser um dos focos de atenção entre os linguistas, principalmente os de *corpus*, assumindo um perfil terminológico. Orenha-Ottaino (2009) classifica as colocações de língua especializada como um subconjunto da língua geral e explica que os tradutores precisam adquirir conhecimento sobre terminologia, simplesmente pelo motivo das colocações de língua especializada “estarem inseridas em uma comunidade linguística e, desse modo, serem parte do seletivo e criativo inventário léxico-semântico” (ORENHA-OTTAINO, 2009, p. 44). Segundo a autora, ambas as colocações apresentam mais semelhanças do que diferenças em questões de aspecto, porém, julgou

¹³ Do original: “[...] it is used to refer to some sort of relation established between lexical items along the syntagmatic axis” (MARTELLI, 2007, p. 16).

necessário que fossem realizadas novas pesquisas para investigar as diferenças entre as colocações de língua geral e colocações de língua especializada.

Ademais, com base nos estudos de L’Homme (2000), Orenha-Ottaiano (2009) elencou cinco aspectos que mostram as semelhanças e diferenças entre as colocações de língua geral e de língua especializada, os quais trouxemos um breve comentário apenas com as ideias centrais:

- **natureza convencional das colocações** (refere-se ao fato das colocações pertencerem a um comunidade linguística específica);
- **composição das colocações** (composta sob a estrutura de Hausmann, “base” e “colocado” (1984, 1985);
- **composicionalidade ou não composicionalidade das combinações** (as colocações de língua geral apresentam uma “base” composicional, sendo o colocado não composicional, enquanto que nas colocações de língua especializada, a composicionalidade não se mostra um quesito relevante, pois, neste caso, observa-se uma base estável, mas seus coocorrentes, mesmo sendo unidades terminológicas, geralmente transmitem outro significado);
- **o agrupamento de palavras-chave (bases) em uma série de unidades lexicais** (refere-se aos aspectos combinatórios ou não entre base e colocado, podendo apresentar, às vezes, uma relação semântica entre as partes da colocação);
- **generalização das relações semânticas entre os componentes** (as colocação de língua geral apresentam em termos de coocorrência léxica restrita – “[...] mesmo que em alguns casos os lexemas que compartilham características semânticas tenham colocados em comum, não é um fator predominante [...]” (ORENHA-OTTAIANO, 2009, p. 49). Já a colocação especializada é descrita em termos de coocorrência lexical livre – porque a “não-composicionalidade” não é um critério que predomina na identificação de uma combinação lexical especializada; e também pelo fato “de ser altamente produtiva, a definição dos grupos de termos semanticamente relacionados está associada a coocorrentes” (ORENHA-OTTAIANO, 2009, p. 50).

Sobre as colocações especializadas, Lorente *et al.* (2017) abordaram questões pela ótica do discurso especializado e explicam tratar-se de “[...] um conjunto de textos orais e escritos, em um tópico, produzido por especialistas nesse campo específico de conhecimento especializado em ambientes comunicativos profissionais e/ou acadêmicos”¹⁴ (LORENTE *et al.*, 2017, p. 202, tradução nossa). Os autores, considerando a presença de UTs nas colocações especializadas em relação ao nível lexical, atribui características em espanhol vs. inglês comuns em textos especializados, dos quais destacamos alguns: a) combinações livres (*escribir un artículo/write an article*; b) colocações gerais: *rigurosamente cierto/certainly*

¹⁴ Do original: “[...] a set of oral and written texts, on a specific topic, which are produced by experts in that particular field of specialized knowledge in professional and/or academic communicative settings” (LORENTE *et al.*, 2017, p. 202).

true; c) colocações especializadas (*dictar [una] sentencia/pass judgment*; e d) terminologias de multipalavras: *alcohol metílico/methyl alcohol* (LORENTE *et al.* 2017, p. 202-203).

Diante dos aspectos citados acima, a autora ainda considera “o léxico especializado como um léxico componente da linguagem, e não como um sistema separado e distinto” (LORENTE *et al.*, 2017, p. 203). Nesse ponto, Lorente *et al.* (2017, p. 208), ao referenciar L’Homme (2000), afirmam que:

O acesso às colocações – especialmente se de alguma forma a explicação de seu significado é fornecido [em dicionários, glossários] – pode ajudar a resolver problemas relacionados à produção de unidades lexicais a serem usadas corretamente na combinação de um termo específico ou outra unidade lexical (L’HOMME, 2009, p. 237, tradução nossa)¹⁵.

Embora as colocações de língua especializada possam ser estudadas separadamente, sabemos que é no texto que estas adquirem seu devido valor de uso convencional. Outro ponto abordado é com relação às questões convencionais e restrições combinatórias dessas colocações em recursos terminológicos, materiais didáticos para aprendizes e mediadores da linguagem, assim como qualquer disciplina específica. Esse fato é bem evidente em cursos de tradução, os quais geralmente apresentam em suas grades disciplinas específicas nas áreas de jornalismo, jurídico-comercial, científico-técnico, dentre outros.

Nessas disciplinas, os alunos são submetidos a textos com uma alta carga de especialização e, assim, o desenvolvimento da competência para traduzir textos de áreas específicas é cada vez mais necessário e o conhecimento de colocações de língua especializada também. O termo competência ainda é um assunto pouco explorado nas colocações de língua especializada por parte dos pesquisadores da LC. Aqueles que o fazem, limitam-se apenas ao campo da Fraseologia baseada no ensino por meio do *corpus*, porém, há poucas pesquisas pensadas a partir da competência, visando a relação de nível de proficiência com os resultados do *corpus*.

Se remetermos essas contribuições pela ótica da tradução, logo notamos que tanto para as colocações de língua geral quanto para as de língua especializada, é necessário ter cuidado ao utilizá-las em um texto traduzido. Contudo, cabe ao tradutor, que na maioria das vezes se depara com colocações especializadas, a cautela e a atenção na escolha desse tipo de colocação, pois o fato de produzir um texto que será utilizado publicamente ou por um terceiro, exige um olhar mais atento aos padrões combinatórios. Dessa forma, cabe ainda a

¹⁵ Do original: “The access to collocations – especially if some form of explanation of their meaning is provided – can help solve problems related to the production of correct lexical units to be used in combination with a specific term or another lexical unit” (L’HOMME, 2009, p. 237).

esse profissional o conhecimento de materiais disponíveis hoje, como os *corpora on-line*, dicionários monolíngues e especializados.

Lorente *et al.* (2017) sustentam a ideia de que as colocações são uma peça fundamental em qualquer idioma e, por isso, tanto as colocações gerais e as especializadas devem ser exploradas considerando os autores que atuam nessas áreas (BENSON, 1995; MOON, 2008; L'HOMME, 2009 apud LORENTE, 2017, p. 2).

Um dos motivos que leva os estudiosos a pensar dessa forma é devido ao avanço da globalização no final do século XX, resultando no aumento do conhecimento especializado. Acrescentam ainda que, a prática da lexicografia se concentrou na preparação de dicionários especializados ou terminológicos. Por outro lado, a *internet* vem sendo parte fundamental nessa trajetória. Além da possibilidade de consultar essas fontes de modo *on-line* ou comprando essas obras em formato digital, o acesso aos *corpora* disponíveis facilita ainda mais o trabalho na investigação de colocações especializadas. Plataformas como o *Sketch Engine* (KILGARRIFF *et al.*, 2014) e o projeto COMET (TAGNIN, 2010) permitem buscas de palavras-chave que conduzem a uma escrita mais segura, principalmente ao tradutor.

Sabemos que não seria possível esgotar esse assunto aqui, mas tentamos explicar alguns pontos que requerem do tradutor a proximidade com as colocações, seja no âmbito da língua geral ou de especialidade. A seguir, tratamos da taxonomia das colocações e de algumas características em sua composição.

2.8 Os erros colocacionais em *corpus* de aprendizes

Um dos obstáculos sobre problemas encontrados no comportamento lexical das colocações na tradução ocorre quando o sujeito dessa prática ainda é aluno, conforme Şen Bartan (2019, p. 74, tradução nossa) demonstra que “os aprendizes têm problemas para usar as colocações e cometem erros colocacionais”¹⁶.

Em muitos casos, contribuições como essa são voltadas para o ensino de LE de modo geral, contudo, apresentando uma forte preocupação com o emprego das colocações. Desse modo, em ambas as áreas, ensino de LE e tradução, essas pesquisas podem auxiliar na formação de aprendizes de tradução, uma vez que são realizadas no âmbito acadêmico e estão embasadas pelos pressupostos teóricos da LC. Independente do público de coleta de dados desta investigação, restringindo somente a alunos de tradução, julgou-se conveniente apontar

¹⁶ Do original: “[...] Learners have problems producing collocations and make collocation errors”.

algumas pesquisas que tratam sobre problemas no uso das colocações, identificados em textos traduzidos ou escritos em inglês como LE por estudantes não nativos.

Ao tentar usar uma colocação, seja de língua geral ou especializada, os aprendizes, sem a orientação de um *corpus* como parâmetro de pesquisa, podem apresentar problemas na identificação de uma colocação correspondente na língua de chegada. Por essa razão, os tradutores em processo de formação podem apresentar equívocos no uso de uma colocação por desconhecerem-as ou até mesmo dificuldade de leitura e interpretação do texto. Mesmo que pareça um termo antigo, a colocação ainda é um campo da Fraseologia que vem sendo de grande interesse para as ciências do léxico, principalmente por causa da dificuldade na produção de colocações em inglês como língua estrangeira.

Quando um aprendiz de LE utiliza uma colocação inadequadamente em uma determinada língua, como o inglês, por exemplo, certa estranheza será percebida por alguém já avançado em tal idioma, principalmente para quem é fluente ou nativo na mesma. É muito comum em uma roda de conversa com pessoas falando de algum assunto em inglês, por exemplo, ver um ou mais participantes perguntando a tradução de uma palavra isolada em inglês. Em alguns casos, a curiosidade pode ser pela tradução de uma colocação, e por desconhecer as questões fraseológicas, o discurso fica vago simplesmente por desconhecerem-as, mesmo que informalmente.

Essas questões se agravam quando inseridas no contexto da prática tradutória do texto escrito. Sendo assim, a preocupação com vocabulário isolado e o desconhecimento das colocações, enquanto unidades combinatórias arbitrárias, pode comprometer o avanço na fluência da língua-alvo, bem como na leitura do texto.

Farrokh (2012, p. 55) considera as colocações um dos maiores problemas no aprendizado de línguas estrangeiras, pelo fato de ser um fenômeno linguístico arbitrário, atuando como um obstáculo aparentemente “intransponível” na aquisição da fluência nativa. A autora acrescenta que o inglês é um idioma repleto de colocações, que atuam na língua por meio da frequência de palavras, isto é, palavras-chave, as quais coocorrem com mais frequência em relação ao esperado, como observamos nas pesquisas com a LC.

Em sua introdução, ela afirma claramente que no uso colocacional um falante não nativo usaria em inglês “*Can I help you?*”, enquanto um nativo comumente diria “*Can I give you a hand?*”. Então, observamos que mesmo o primeiro caso seja naturalmente aceito, no segundo, a colocação *give a hand* (“dar uma mão”) em português, além de soar com mais naturalidade, demonstra um domínio maior do inglês por parte do falante. A autora ainda aponta outros dois exemplos dos aspectos de restrição da colocação, sendo eles: o fato de

dizemos *last year* e não *last hour*, como também no uso da preposição *by* (de) nas combinações *by car* (de carro), *by train* (de trem), uma vez que o mesmo já não ocorre em *on foot* (a pé).

Phoocharoensil (2011), por sua vez, examinou erros de colocações na escrita de alunos tailandeses, estudantes de inglês como língua estrangeira, os quais foram divididos em dois grupos com proficiências diferentes. Concluiu que os discentes, independentemente de sua proficiência em LE, apresentaram dificuldades na tentativa de transferir as colocações da L1 para a L2.

O autor aponta os seguintes erros mais comuns nessa transferência da seguinte maneira: a) adição (ao combinar *leave from home*, mesmo que *leave* seja um verbo transitivo, não é seguido de *from*); b) omissão de preposições (o uso de *listen music*, sem o uso de *to* entre as palavras); escolha incorreta de palavras (*play computer*, não muito comum); c) redundância colocacional (*place dwelling*, como lugar de habitação, moradia). Ademais, o pesquisador apontou problemas com o uso de sinonímia (*peaceable home*, como casa de paz; *authentically believe*, acreditar autenticamente); e generalização excessiva a partir do uso do advérbio *very* (*The room that I very like is the dining room; I very love to stay at home*), os quais, segundo ele, incentivou ainda mais a presença de equívocos colocacionais no em L2.

Shitu (2015) também decidiu investigar erros de colocação, porém, em redações de estudantes de nível avançado da Faculdade Federal de Educação de Kano, na Nigéria. Foram coletadas cerca de 900 redações em um total de 300 participantes. Os textos foram transcritos durante três palestras, com temas e tempos de apresentação semelhantes em todas elas, porém ocorridas em diferentes ocasiões.

Dos equívocos identificados, foram separados apenas um por vez, mesmo que houvesse repetição no decorrer do texto e, posteriormente, o número de ocorrências foi convertido em porcentagem para obtenção dos dados. Como resultado, identificaram-se erros colocacionais nos textos dos alunos, confirmando suspeitas de interferências da língua materna, excesso de generalização e falta de conhecimento sobre colocação.

Para corroborar, Shitu (2015, p. 327) exemplificou alguns problemas colocacionais. Os exemplos apresentaram problemas de uso incorreto (palavra tachada) em uma das palavras *versus* palavra correta (palavra sublinhada) (~~*doing*~~ *inspection* vs. *conduct* *inspection*; ~~*bad peers*~~ vs. ~~*bad*~~ *friends*; *Height* ~~*quest*~~ vs. *High* *demand*), bem como erros em toda a estrutura (*a very bad idea about* vs. *lessons*; *negative impression*). Logo, o autor ressalta que isso se deve ao fato de haver diferenças entre a L1 e L2, fazendo com que o estudante de inglês como

segundo idioma recorra primeiramente ao seu idioma nativo antes de traduzir uma colocação para L2 (SHITU, 2015, p. 3275).

Şen Bartan (2019) se dedicou a analisar erros de colocação em um *corpus* escrito e composto por traduções de textos literários na direção tradutória turco – inglês, a fim de descrever e explicar os erros identificados. O *corpus* foi compilado por estudantes do Departamento de Tradução e Interpretação em Língua Inglesa, por meio do qual se detectaram mais erros no uso de colocações verbais do tipo, “verbo + substantivo”, e depois adjetivas, “adjetivo + substantivo”.

Em Murakami (2016) também encontramos uma análise parecida com a anterior, porém com públicos e objetivos diferentes. O autor analisou o uso das colocações verbais por meio de um *subcorpus* (EF- Cambridge Open Language Database) com redações de alunos brasileiros. Os estudantes selecionados eram de nível avançado e, mesmo assim, não empregaram esse tipo de colocação adequadamente, por causa do uso de “verbo inapropriado”.

O autor elencou cinco tipos de desvios colocacionais, os quais foram extraídos a partir do *corpus* de estudo e analisadas com base em cinco categorias que medem a escala de aceitabilidade linguística da colocação: “Aceitável”; “Questionável”; “Inaceitável”; “Possivelmente aceitável”; “Possivelmente inaceitável” (MURAKAMI, 2016, p. 69).

Dentre as categorias citadas acima apontamos somente as de aspecto inaceitável, para exemplificar diretamente erros colocacionais. São elas: a) desvio I - verbo inadequado (~~porposes~~/has a gourgous view); b) substantivo inadequado (have ~~place~~/a chance_{Am,Br}); c) desvio III - verbo com substantivo adequado com estrutura inadequada (get ~~to a~~/a loan); d) desvio IV - verbo e substantivo inadequados (~~make some huge house taske~~/ do the housework); e) desvio V - verbo como substantivo (~~gave him a chase~~/chased him; f) desvio VI - substantivo como verbo (~~my legs pain~~/ feel pain in my leg/my legs hurt); g) desvio VII - inclassificável (can we do for give our caring about this kindly of people); neste último exemplo, observamos o verbo *do* seguido de *can* e acompanhado de *give*, sendo que a intencionalidade seria dizer *what can we do to provide those kind of people with care* (MURAKAMI, 2016).

Entretanto, um dos problemas comuns, como já mencionamos nesta seção, é sobre o apego ao texto original. Caso o tradutor não se atente para isso, mesmo em processo de formação, pode não escrever com tanta naturalidade na língua-alvo, como salienta Tagnin (2002, p. 193):

por estranho que pareça, mesmo como falante nativo da língua alvo, o tradutor pode ter problemas no nível da produção para conseguir soluções naturais, caso se atenha tanto ao texto de partida a ponto de não perceber que, entre formas igualmente gramaticais, uma delas é de uso mais corrente.

Notamos que na citação acima o tradutor é considerado um “tradutor ingênuo” da língua alvo, isto é, alguém que, a princípio, apresenta certo domínio relevante sobre as combinações lexicais em relação a outros falantes, pois, este, por sua vez, também pode encontrar dificuldades na escolha exata de uma colocação para um contexto específico.

O uso de um *corpus* composto diretamente com textos de aprendizes oferece um respaldo com mais fundamento ao invés de utilizar por um longo período de tempo um *corpus* nativo com alunos. Por isso, quando tratamos sobre o estudo das colocações, principalmente hoje, baseado nas teorias de abordagem por frequência, é preciso também considerar o conteúdo do *corpus*, atrelado ao seu objetivo, de modo que represente a linguagem na qual os textos são selecionados (SINCLAIR, 2004).

Na produção de um texto em outro idioma, entram em cena os aspectos linguísticos, tais como a Fraseologia, na difícil tarefa de traduzir combinações pré-fabricadas (TAGNIN, 2002, p. 193), obedecendo às questões de uso. Esse uso deixa de ocorrer quando o tradutor, “[...] não pode compreender muitas fórmulas discursivas por não conhecer as convenções sociais que determinam seu uso na língua alvo” (TAGNIN, 2002, p. 193).

Retomamos a questão em que a autora pontua sobre os problemas do tradutor com a “habilidade de compreensão”, ou seja, caso a compreensão do texto-fonte seja falha, possíveis inadequações no uso de combinações arbitrárias, como as colocações, por exemplo, ocorrerão, afetando a comunicação e a fluência da leitura do texto traduzido.

Assim, mesmo com objetivos e públicos diferentes, as pesquisas acima nos proporcionaram um panorama de como muitos pesquisadores, na preocupação em analisar a escrita de um aprendiz de inglês não nativo, tem buscado indicar caminhos de observação do uso colocacional, a partir dos problemas encontrados com base em *corpus*.

E mesmo com estudantes de alto nível de proficiência, observamos que esses equívocos ainda ocorrem com frequência. A ideia principal ao apresentar essas investigações, nos mostra a relevância de se auxiliar, cada vez mais, os alunos de tradução em formação acadêmica, quanto às particularidades das colocações na língua-alvo e da necessidade de buscar por correspondentes adequados ao tentar traduzi-las.

Com certeza, seria impossível abordar todas as pesquisas relacionadas ao uso das colocações a partir da escrita de aprendizes. Desse modo, no próximo capítulo tratamos das

questões metodológicas do *corpus* de estudo desta investigação.

3. METODOLOGIA

Neste capítulo, iniciamos com os pontos sobre a origem do projeto, os aspectos relevantes à instituição e ao curso, assim como os critérios adotados para a seleção dos participantes, bem como a descrição do processo metodológico e as ferramentas empregadas na pesquisa.

3.1 O projeto *Corpus de Aprendizes de Tradução*

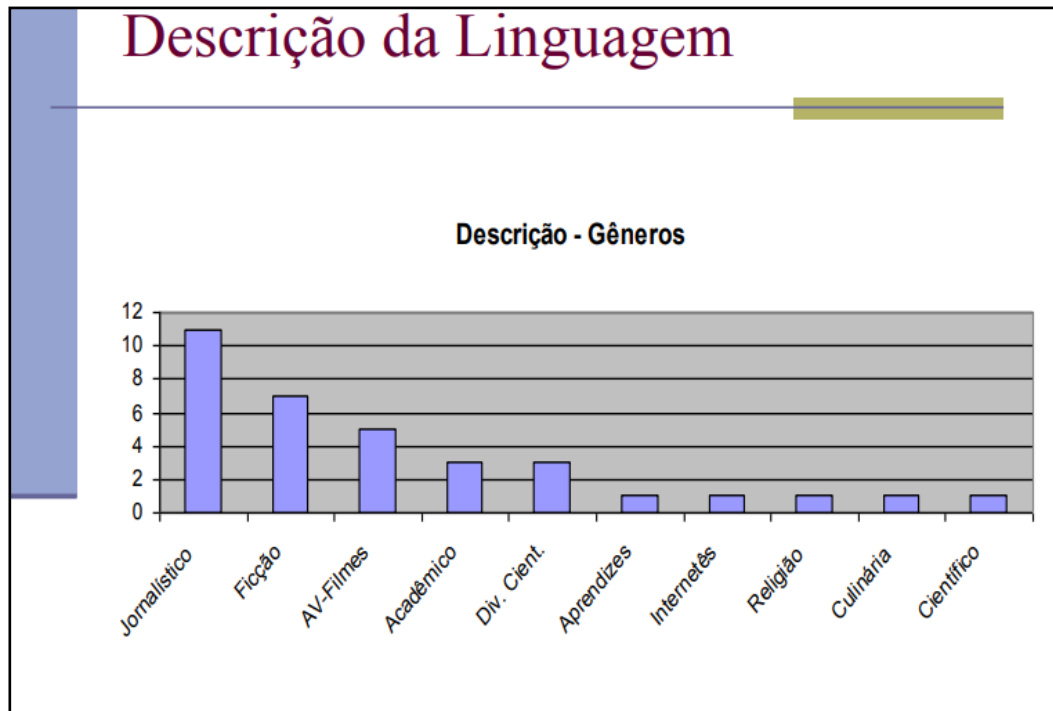
Esta pesquisa introduz o CAT 2, parte do projeto guarda-chuva COMADEC, iniciado em 2012, sob coordenação da Prof^{ra}. Dr^a. Adriane Orenha-Ottaiano, da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Campus de São José do Rio Preto. Este trabalho tem dado início a diferentes análises, desde subprojetos de iniciação científica na graduação, até mesmo estudos desenvolvidos por estudantes de mestrado e doutorado, em nível de pós-graduação.

O primeiro subprojeto teve início com a compilação do CAT 1, que conta com observações focadas na investigação da tradução de aprendizes universitários. O objetivo de criar um *corpus* desse tipo foi analisar as traduções produzidas por acadêmicos brasileiros, realizadas a partir de textos extraídos de jornais e revistas em língua portuguesa (de 500 a 800 palavras), tendo como sujeitos alunos em formação no curso de Letras, com bacharelado em Tradução.

A escolha por trabalhar com materiais jornalísticos se deu por se tratar de uma esfera de atuação humana, cujos gêneros escritos é um dos objetos utilizados para a construção do *corpus* desta área como proposto pelo *British National Corpus* (BNC). Tal atuação tem o foco na descrição da linguagem, como notamos na Figura a seguir, conforme nos mostra Tagnin (2011).

Ao explorar estudos que relacionam o Jornalismo e a Linguística de *Corpus* (LC), Murad (2014) com base em Roberts (2002), aborda as tipologias do texto traduzido, que têm como base o texto original, sendo a língua geral ou especializada. No âmbito das colocações, objeto linguístico que é foco desta pesquisa com o CAT, nota-se que estes dois domínios da linguagem têm sido altamente identificados.

Figura 3 – Imagem de apresentação de trabalho



Fonte: captura de tela de apresentação de trabalho (TAGNIN, 2011).

De acordo com Orenha-Ottaiano (no prelo), o processo tradutório seguiu as seguintes etapas:

- As traduções deveriam ser realizadas em duas horas ininterruptas;
- As traduções não deveriam ser feitas com auxílio de outros colegas;
- Autorização para utilização somente de ferramentas que já empregavam no laboratório (memórias de tradução, dicionários, excluindo-se instrumentos de tradução automática, tais como o *Google Tradutor* etc.);
- A entrega das traduções deveria seguir os prazos estipulados pela equipe do projeto.

A metodologia do CAT 1, segundo Orenha-Ottaiano (2012b, 2015), prevê como ponto de partida para análise dos Textos Traduzidos (TTs), o levantamento de palavras-chave. Em decorrência disso, concluiu, em suas análises, que os sujeitos de sua pesquisa (no caso, os aprendizes) apresentaram dificuldades de compreensão quanto às questões convencionais desta combinação e seu grau de fixidez na língua de chegada.

3.2 Processo de coleta de dados e os documentos legais do Comitê de Ética

Diante da verificação destes dados colocacionais com base no *corpus* de aprendizes de tradução, pensamos em ampliar este estudo para alunos de um Curso de Tradução de uma instituição de Ensino Superior privada. Sendo assim, com o consentimento da orientadora deste trabalho, entramos em contato com a Prof^ª. Dr^ª. Patrícia Viana Belan, responsável pela coordenação do curso de Letras-Tradutor do Centro Universitário Sagrado Coração (UNISAGRADO), localizado na cidade de Bauru, interior de São Paulo, para que, em forma de parceria, a instituição fosse o *locus* para a coleta de dados que formariam o CAT 2.

Depois de algumas reuniões entre o pesquisador e a coordenação para os devidos alinhamentos, a proposta foi aceita pela universidade e o trabalho deu início seguindo alguns critérios burocráticos:

- Apresentação do projeto à coordenação do curso de Letras-Tradutor e, conseqüentemente, à coordenação geral do Centro de Ciências Humanas do Centro Universitário Sagrado Coração (UNISAGRADO), sob a responsabilidade da Prof^ª. Dr^ª. Ketilin Mayra Pedro;
- Carta de anuência concedendo a liberação para aplicá-lo após a aprovação;
- Submissão de documentação necessária (carta de anuência, rascunho do projeto, Termo de Consentimento Livre e Esclarecido – TCLE) ao *Comitê de Ética* da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), campus de São José do Rio Preto;
- Aprovação do Comitê de Ética;
- Início da coleta dos dados.

A presente pesquisa de mestrado compreende um estudo que envolveu a participação de seres humanos para sua produção. Por este motivo, conforme os requisitos legais exigidos para este tipo de atividade científica, este estudo foi submetido ao *Comitê de Ética* da UNESP e teve sua aprovação no dia 09 de setembro de 2018, data que permitiu o início da coleta de dados na universidade *locus* de análise. Todo o processo tradutório ocorreu durante o Estágio Obrigatório ou em aulas de Prática de Tradução, sob a supervisão da Prof^ª. Dr^ª. Leila Maria Gumushian Felipini, responsável pelas disciplinas. Vale ressaltar que as professoras envolvidas no processo de reunião dos materiais de estudo foram grandes incentivadoras para

a realização deste trabalho.

3.2.1 Aspectos relevantes sobre a instituição locus da pesquisa e o curso

A coleta de dados foi integralmente realizada na UNISAGRADO. O curso de Letras-Tradutor teve início no ano de 1987 e, de acordo com as informações do *site*, encontra-se entre os mais antigos do Brasil. A fim de caracterizar brevemente a Instituição de Ensino, bem como os estudantes aos quais atende, fornecemos abaixo uma breve caracterização do grupo em estudo:

Quadro 7 – Caracterização dos agentes participantes da pesquisa

Informações	Instituição de Ensino Superior
Tipo de Instituição	Particular
Modalidade de ensino	Presencial
Período de estudos	Noturno
Duração do curso	Três anos
Nível de proficiência	Intermediário a proficiente

Fonte: elaborado pelo autor.

Também julgamos relevante apresentar os objetivos do curso e as premissas que direcionam para a formação integral do aluno enquanto profissional em Tradução. Vejamos, a seguir, o objetivo geral e os objetivos específicos, indicados no *site* da universidade (UNISAGRADO, 2020):

Objetivo Geral:

Formar bacharéis capacitados para a tradução de textos nas Línguas Portuguesa e Inglesa e para o desenvolvimento de pesquisas na sua área de atuação.

Objetivos específicos:

- Formar profissionais para o uso das línguas estudadas nos seus principais níveis: fonológico, morfológico, sintático e semântico;
- Habilitar profissionais que identifiquem nas literaturas estudadas os aspectos mais relevantes da expressão cultural e artística dos povos que elas representam;
- Preparar profissionais para o uso das línguas conforme os padrões de norma culta e, também, conforme aspectos sociolinguísticos e outras variantes, dependendo do contexto;
- Formar profissionais para a tradução e versão de textos, ou quaisquer outros materiais que lhe sejam confiados, em língua portuguesa ou inglesa nas mais diversas áreas, identificando as principais características e especificidades que as envolvem;
- Preparar profissionais criativos e versáteis, facilitando seu futuro desempenho, e tornando-os aptos a adequar-se às constantes mudanças no campo social.

Desse modo, observando a viabilidade da aplicação do subprojeto CAT em uma universidade particular, tivemos o cuidado de escolher também uma instituição privada em que o curso de Tradução atendesse aos critérios do macroprojeto, de que os sujeitos da pesquisa fossem impreterivelmente alunos de graduação, além da observação clara do desenvolvimento dos aspectos linguísticos, culturais, dentre outros. Portanto, excluímos todo e qualquer público que esteja em processo de formação que não seja de nível acadêmico.

Por meio dos objetivos propostos, a grade do curso também oferece uma diversidade de disciplinas, as quais prezam pela formação integral do estudante no âmbito profissional, elencadas no quadro a seguir em três grandes áreas:

Quadro 8 – Grade curricular do curso de Letras-Tradutor

1º semestre	2º semestre
Compreensão e Produção Oral em Língua Inglesa I Comunicação e Expressão Introdução à Linguística I Língua Portuguesa: Morfossintaxe Métodos e Técnicas da Pesquisa Teoria da Tradução I Análise do Discurso Compreensão e Produção Escrita em Língua Inglesa	Compreensão e Produção Oral em Língua Inglesa II Ferramentas de Tradução Introdução à Linguística II Língua Portuguesa: Gramática Contextualizada Teoria da Tradução II
3º semestre	4º semestre
Análise do Discurso Antropologia Ética e Cultura Religiosa Fonética e Fonologia da Língua Inglesa Prática da Tradução I: Jornalística/Literária Sociolinguística Teoria da Literatura	Compreensão e Produção Escrita em Língua Inglesa Cultura Inglesa e Norte-Americana Introdução às Literaturas de Língua Inglesa Introdução às Literaturas de Língua Portuguesa Língua Portuguesa: Filologia Prática da Tradução II: Científico-Técnica
5º Semestre	6º Semestre
Cultura Brasileira Estágio I Estágio: Orientação e Supervisão I Métodos e Técnicas da Pesquisa em Tradução Morfologia da Língua Inglesa I Prática da Tradução III: Audiovisual Sociologia da Responsabilidade Social	Atividades Acadêmico-Científico-Culturais Estágio II Estágio: Orientação e Supervisão II Morfologia da Língua Inglesa II Prática da Tradução IV: Jurídico-Comercial Trabalho de Conclusão de Curso Versão

Fonte: elaborado pelo autor com base na grade curricular do curso de Letras-Tradutor (UNISAGRADO)¹⁷.

3.2.2 Os procedimentos e os critérios de seleção dos participantes

Em 2018, a partir das primeiras reuniões de alinhamento para o início da pesquisa de campo e coleta de dados, a coordenação informou que, naquele ano, a graduação de Letras-Tradutor tinha apenas duas turmas disponíveis para a aplicação: segundo e terceiro ano. Um

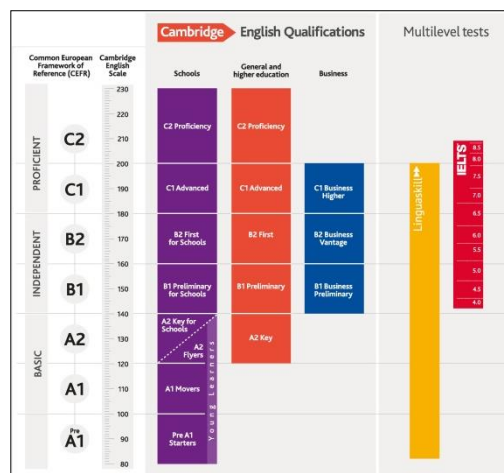
¹⁷ UNISAGRADO. Grade Curricular do curso de Letras-Tradutor. Centro Universitário Sagrado Coração, Bauru (SP), 2020. Disponível em: <https://unisagrado.edu.br/graduacao/letras-tradutor>. Acesso em: 11 abr. 2020.

dado importante diz respeito à grade curricular apresentada anteriormente. Notamos que os participantes já haviam cursado ou estavam cursando a disciplina de “Prática da Tradução I: Jornalística”, fato que reforça a investigação de possíveis erros de tradução realizados pelos aprendizes.

Pensando que poderíamos não alcançar um número significativo de palavras para a formação do *corpus*, solicitamos à coordenação que a Carta de Anuência oferecida em nome da instituição abrangesse alunos do curso de modo geral. Dessa forma, poderíamos aplicar o projeto em qualquer turma disponível, porém, utilizando apenas os dados dos alunos que alcançassem níveis de proficiência, no mínimo, intermediário (B2), segundo a classificação dos níveis de idiomas.

Para identificar os níveis dos sujeitos da pesquisa, um *Placement test* (ALLAN, 2004) foi aplicado aos estudantes, a fim de comprovarmos o nível de proficiência no idioma em questão. O teste, composto por 100 questões, foi usado para averiguar a habilidade escrita dos acadêmicos, com foco na gramática aplicada ao uso e suas características lexicais. Foram selecionados alunos com proficiência B2 a C2. A imagem a seguir apresenta a estrutura do *Common European Framework of Reference (CEFR)*¹⁸, cuja escala de proficiência serviu de norte para classificação dos estudantes.

Figura 4 – Common European Framework of Reference (CEFR)



Fonte: captura da tela dos níveis de proficiência (CEFR, 2020).

¹⁸ *Common European Framework of Reference (CEFR)*. Cambridge Assessment English, 2020. Disponível em: <https://www.cambridgeenglish.org/exams-and-tests/cefr/>. Acesso em: 11 abr. 2020.

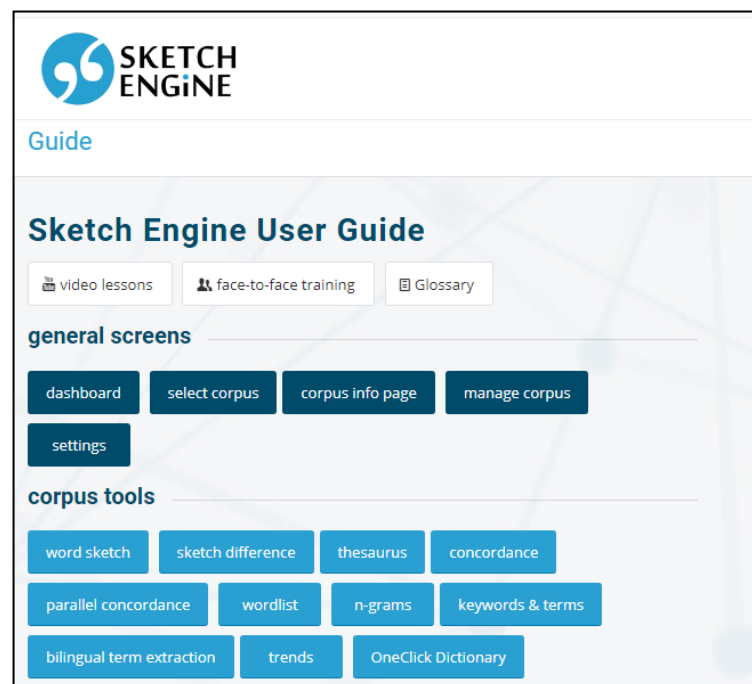
3.3 As ferramentas e os *corpora* empregados na pesquisa

A partir de agora, discorreremos sobre a estrutura do *corpus* estudado. Em seguida, serão apresentados os usos e as funções da ferramenta *Sketch Engine*, bem como uma explanação das características do *corpus* de referência *English Web 2013*, utilizado para comparar as colocações observadas no CAT 2. Por fim, tratamos da plataforma *Online English Collocations Platform*, adotada para a inserção das atividades após a análise das colocações levantadas.

3.3.1 A ferramenta de análise lexical de corpus *Sketch Engine*

A plataforma *on-line Sketch Engine* é um gerenciador de *corpus* e *software* de análise de texto, que serviu como base para compilar as traduções do CAT 2. É um gerenciador contemporâneo e com interfaces de trabalho de fácil manuseio, com uma ampla gama de *corpora* disponíveis, contando também com ferramentas de busca e análise observacional de dados linguísticos.

Figura 5 – Guia de usuário do *Sketch Engine*



Fonte: captura de tela do *Sketch Engine*.

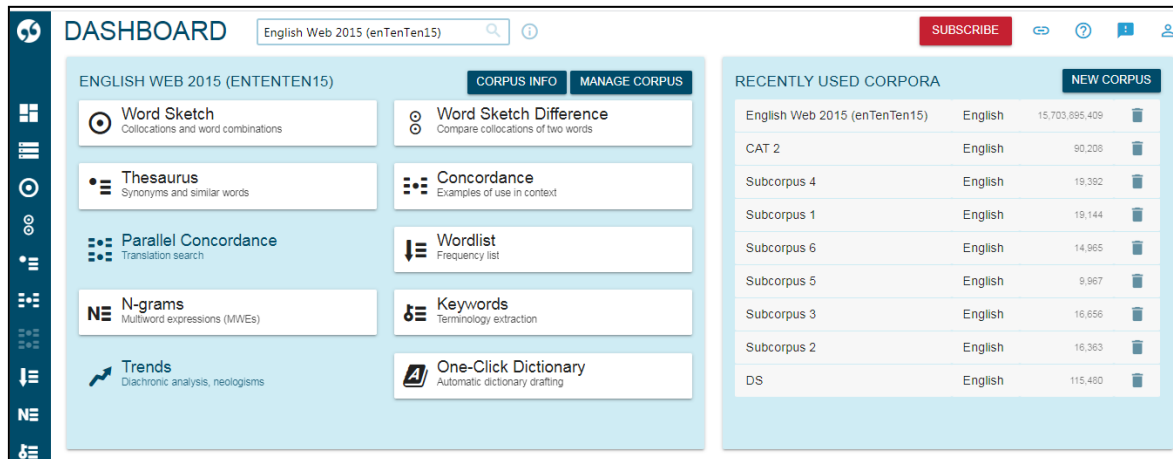
Na figura acima, observamos que, de acordo com o guia geral da plataforma, a

visualização se dá pelos ícones: *video lessons* (vídeo aulas), *face-to-face training* (treinamentos) e *glossary* (glossário). Em seguida, na tela geral (*General Screens*), as funções iniciais do *Dashboard* (painel controle), área inicial das funções de manuseio do *corpus*, é feita por meio dos ícones: *select* (seleção), *info page* (informações) e *Manage* (gerenciar), assim como a função de *Settings* (definições). Na terceira parte, no tópico *Corpus Tools* (Ferramentas do *Corpus*), resumidamente, podemos verificar as definições dessas ferramentas do seguinte modo:

- *Word sketches* – uma página que apresenta o comportamento gramatical e colocacional de uma palavra;
- *Word sketch difference* – comparação e contraste dos termos que formam uma expressão;
- *Distributional Thesaurus* – buscador automatizado de itens com significados semelhantes ou os que aparecem em contextos similares;
- *Concordance search* – buscador de exemplos de uma forma de palavra, lema, frase, *tag* ou uma estrutura mais complexa;
- *Collocation search* – buscador de terminologias candidatas à colocação que apresentando as coocorrências mais frequentes a partir do vocábulo de pesquisa;
- *Word lists* – identificador da frequência de elementos que podem ser filtrados por critérios, como categorias específicas (substantivo, verbo, adjetivo, advérbio), por exemplo;
- N-gramas – apresenta listas com a repetição de expressões com várias palavras;
- *Terminology / Keyword extraction (both monolingual and bilingual)* – buscador automático de palavras-chave e termos com várias palavras de textos (com base na contagem de frequências e critérios linguísticos).

Alguns desses itens descritos acima são disponibilizados de acordo com o modelo utilizado pelo usuário. Em nosso caso, como usuário acadêmico, tivemos acesso apenas aos itens da Figura a seguir, suficientes para esta pesquisa.

Figura 6 – Dashboard do Sketch Engine



Fonte: captura de tela do *Sketch Engine*.

A vantagem desse gerenciador está em sua capacidade de analisar os dados linguísticos, uma vez que conta com uma interface altamente dinâmica e de fácil acesso e manuseio, como ocorre com os aplicativos, por exemplo. Entretanto, mesmo com toda capacidade que este *software* oferece, Machado e Bevilacqua (2018) consideram os seguintes pontos positivos e negativos de sua aplicabilidade:

Pontos positivos¹⁹:

- lematiza automaticamente o *corpus* inserido. Com isso, todas as variações morfológicas de uma palavra são recuperadas, uma vez que são reconhecidas sob a forma de lema;
- recupera, a partir de uma única busca do NT, todos os colocados (em função da lematização) que serão candidatos a conformar UFE eventivas;
- reconhece unidades fraseológicas que apresentam outros itens inseridos entre o NE e NT, isto é, até aquelas mesmo com *span*.

Pontos negativos:

- é um *software* pago, cuja licença aumenta de valor conforme o tamanho do trabalho. Portanto, quanto mais dados para analisar, mais cara é a assinatura;
- exige conexão à rede de internet para realizar as buscas;
- por lematizar o *corpus*, recupera o ruído nas buscas por função sintática, seja por

¹⁹ Siglas terminológicas. NT: núcleo terminológico; UFE: Unidades Fraseológicas Especializadas Eventivas; NE: Núcleo Eventivo; *Span*: “[...] distância que há entre a base da colocação [...]” (MACHADO; BEVILACQUA, 2018).

recuperar diferentes classes gramaticais ou por confundir em alguns casos a lematização.

Por outro lado, não seria possível acreditar que a máquina, por si só, seja capaz de apontar tudo que estamos buscando linguisticamente. Por isso, além de utilizarmos esse recurso em nossa metodologia, entendemos que as análises qualitativas com bases teóricas consolidadas norteiam o pensamento crítico e reflexivo no estudo dos fenômenos linguísticos, como as unidades fraseológicas, em especial as colocações estudadas.

De modo geral, o *Sketch Engine* apresenta uma interface mais atual, contendo o *corpus* de referência e outros que são passíveis de serem alimentados, bem como espaço para inserção de novos *corpora*. Além do suporte computacional, a plataforma também conta, impreterivelmente, com a atuação de linguistas da área da LC. Com a soma interdisciplinar entre as análises estatística e linguística, o *Sketch Engine* vem se tornando a plataforma mais utilizada no meio, devido à sua praticidade analítica.

3.3.2 Procedimento de compilação do corpus e análise de dados

Foram selecionados seis TOs para a coleta de dados, os quais apresentam os seguintes títulos:

Quadro 9 – Títulos dos textos originais

Título dos textos originais
Texto 1 - O Brasil no mundo
Texto 2 - Trump e Putin são os czares modernos?
Texto 3 - Extrema direita alemã amplia atos contra imigrantes
Texto 4 - Plano de deportação de Trump deve jogar 300 mil imigrantes na ilegalidade
Texto 5 - Nova legislação migratória visa fim dos guetos na Dinamarca
Texto 6 - Retórica virulenta aproxima Trump de países do Leste Europeu

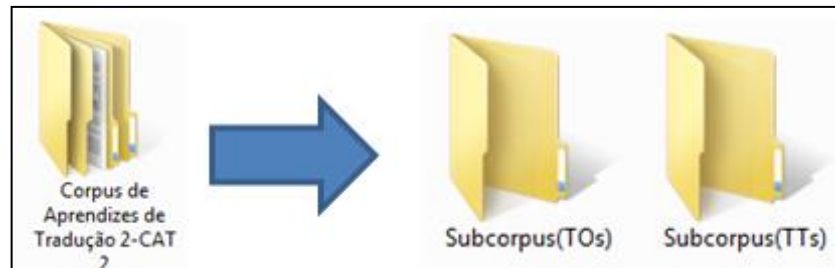
Fonte: elaborado pelo autor.

Esses textos foram aplicados durante a coleta de dados na universidade *locus* da pesquisa e enviados à professora responsável pela aplicação dos textos e coletas de suas respectivas traduções em duas etapas: traduções dos textos 1, 2 e 3 (80 traduções), realizadas pelos alunos do 3º ano em 2018; traduções dos textos 4, 5 e 6 (68 traduções), realizadas pelos alunos do 2º ano no decorrer do ano de 2019. O processo resultou em total de 148 traduções.

Depois desse momento, iniciou-se o processo de compilação do CAT 2 no *Sketch Engine*. Primeiramente, reunimos o *subcorpus* de TOs separadamente e, em seguida, o de TTs por pastas. Essa separação ocorreu devido ao levantamento de palavras-chave que se deu em

dois momentos. Não foi preciso renomear os documentos com os nomes dos participantes, uma vez que mesmo com as capturas de tela foi possível manter as identidades preservadas.

Figura 7 – Etapa de organização dos textos



Fonte: agrupamento do *subcorpus* de TOs e TTs (elaborada pelo autor).

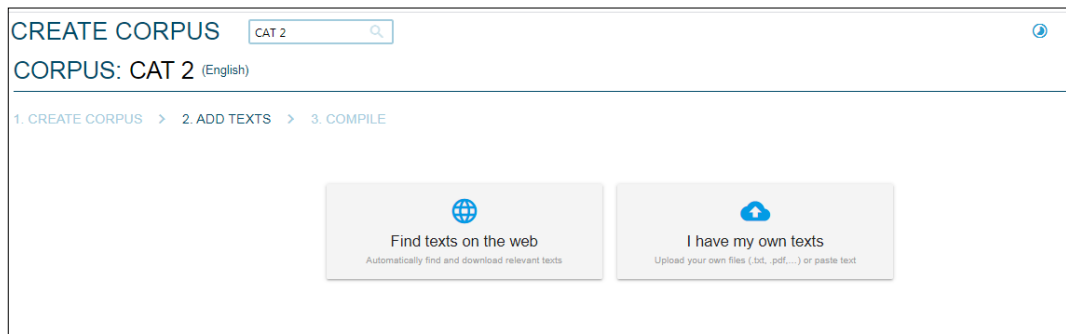
A partir dessa separação, iniciamos a compilação de ambos os *subcorpora* separadamente. Essa parte contou com três etapas, sendo a primeira delas a criação do *corpus*, conforme mostra a imagem a seguir:

Figura 8 – Etapa de criação do *corpus* de estudo

Fonte: captura de tela do *Sketch Engine*.

A Figura acima mostra as possibilidades de escolher o tipo de *corpus*, bem como indicar a língua que os textos podem ser compilados. Depois, passamos para a segunda interface, que permitiu realizar o *upload* dos textos no gerenciador:

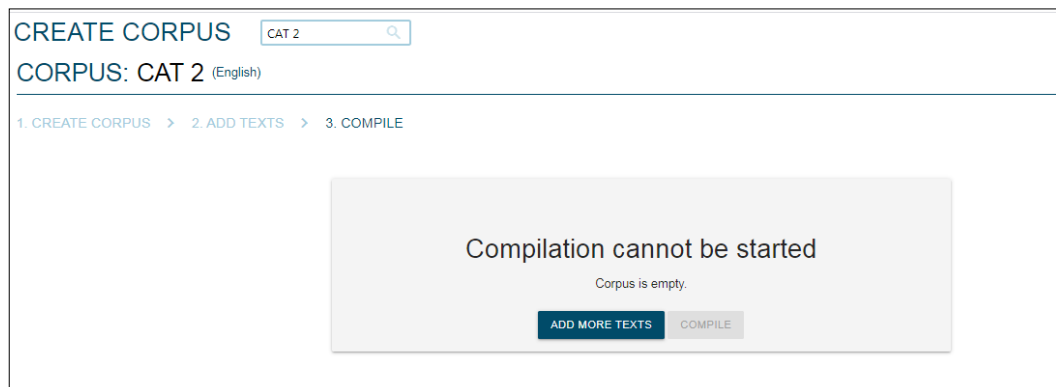
Figura 9 – Etapa de adicionar textos no *corpus*



Fonte: captura de tela do *Sketch Engine*.

Nesta etapa foi possível adicionar os textos, optando pela segunda opção, *I have my own texts*, à direita, uma vez que se trata, como já esclarecemos, de um *corpus* de estudo com textos próprios. Por último, a terceira fase permite a compilação dos textos, momento em que o gerenciador realiza a leitura estatística no *corpus*. Dessa forma, é possível utilizar as ferramentas apresentadas no subcapítulo sobre o *Sketch Engine*.

Figura 10 – Etapa de compilação estatística dos textos no *corpus*



Fonte: captura de tela do *Sketch Engine*.

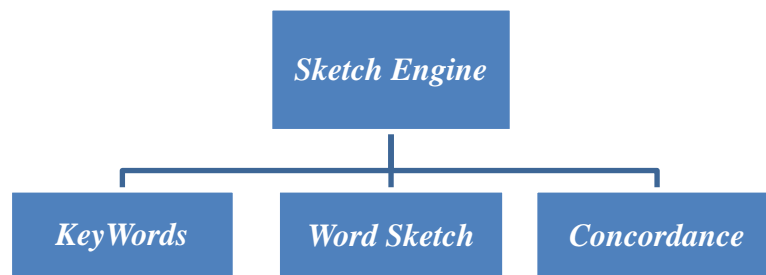
Após esse processo, iniciou-se a análise das colocações, a partir do levantamento das palavras-chave dos TTs, utilizando o *Sketch Engine*, que ocorre com o uso da ferramenta *Keywords*, inserida na tela inicial *Dashboard*. Feito isso, é possível identificar os problemas obtidos por meio da comparação entre TTs e seus respectivos TOs.

Em seguida, alguns problemas de tradução foram selecionados, como amostragem para a análise desta pesquisa, a fim de comprovar se as opções tradutórias levantadas eram ou não frequentes no *corpus* de referência *English Web 2013*. Para realizar esse processo, empregou-se o *Sketch Engine* e outros recursos como dicionários.

3.3.3 Levantamento e análise dos dados no Corpus de Aprendizes de Tradução 2 (CAT 2)

Para investigar as colocações desta pesquisa, decidimos utilizar apenas algumas dessas ferramentas, tais como: *Keywords*, *Word Sketch* e *Concordance*, pois julgamos que seriam suficientes como mostra o Organograma 1:

Organograma 1 – Algumas ferramentas do gerenciador de *Corpus Sketch Engine*



Fonte: elaborado pelo autor.

Dadas as nomenclaturas acima, seguimos com suas respectivas definições e funções abaixo:

- *Word Sketch*: identifica a categorização das colocações de palavra e suas relações gramaticais;
- *Keywords*: são os conceitos chaves de maior frequência no *corpus*;
- *Concordance*: linhas do *corpus* acompanhadas de algum contexto à esquerda e outro à direita, o qual lista o termo de busca recorrente de modo centralizado;

Certos de que esses itens seriam suficientes para esta análise, iniciamos o processo pelo levantamento das palavras-chave do *subcorpus* de TOs, que nos conduziu a encontrar suas respectivas traduções no *subcorpus* de TTs, utilizando a ferramenta *Keywords* do *Sketch Engine*.

Primeiramente, realizamos a busca por meio da frequência gerada a partir da opção *keywords*, no *subcorpus* de TOs e, em seguida, no *subcorpus* de TTs, a fim de extrairmos as colocações nos textos originais. Como o processo de levantamento de palavras-chave de ambos os *subcorpora* foi o mesmo, passamos a verificar como esse processo ocorreu apenas a partir do *subcorpus* de TTs, conforme pode ser observado na Figura a seguir.

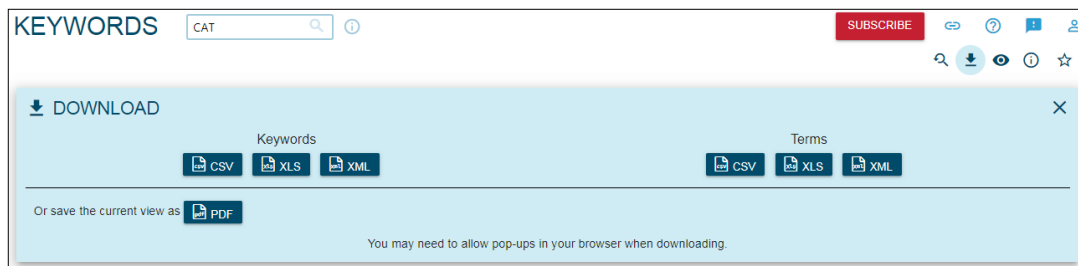
Figura 11 – Primeiras palavras-chave do CAT 2

Word	Word	Word	Word	Word
1 Montefiore ...	11 Brexit ...	21 Rubens ...	31 Honduran ...	41 Franzese ...
2 Trump ...	12 xenophobic ...	22 lulopetista ...	32 Temer ...	42 Lidia ...
3 TPS ...	13 Linke ...	23 Romanov ...	33 FIESP ...	43 Fórum ...
4 federative ...	14 Unibes ...	24 Románov ...	34 Lafer ...	44 Sebag ...
5 Wagenknecht ...	15 Gurovitz ...	25 federalism ...	35 FGV ...	45 tyrant ...
6 Chemnitz ...	16 Lafontaine ...	26 illiberal ...	36 Salvador ...	46 Celso ...
7 Putin ...	17 Thale ...	27 czar ...	37 Ziulkoski ...	47 Fraga ...
8 Barbosa ...	18 Irice ...	28 HYPERLINK ...	38 Bolsonaro ...	48 Weisskircher ...
9 AfD ...	19 Kerwin ...	29 Salvadorans ...	39 Goldenstein ...	49 Manès ...

Fonte: captura da tela *Keywords* do *Sketch Engine*.

Em seguida, foi feito o *download* em forma de planilha, gerada diretamente em formato *Excel*, o que permitiu a seleção mais cautelosa das palavras candidatas à colocação.

Figura 12 – Tópico download da ferramenta *Keywords*



Fonte: captura de tela do *Sketch Engine*.

Por meio da tabela, listamos as palavras em ordem de frequência, simplesmente pelo fato de estarmos trabalhando com as mais recorrentes. Dessa forma, também notamos que esses conceitos no CAT 2 apresentam claramente a temática do *corpus*, ou seja, podem trazer termos comumente usados na área da política, podendo evidenciar sua representatividade como base de dados para este estudo. A seguir, apresentamos o levantamento de 135 exemplos das palavras-chave encontradas no *corpus*:

Quadro 10 – Extração das keywords do CAT 2 pelo Microsoft Excel

Lista das palavras-chave CAT 2				
<i>country</i>	<i>refugee</i>	<i>Disaster</i>	<i>reelect</i>	<i>grandmother</i>
<i>immigrant</i>	<i>defense</i>	<i>Ghetto</i>	<i>critic</i>	<i>informal</i>
<i>president</i>	<i>integration</i>	<i>Central</i>	<i>flexibility</i>	<i>irresponsible</i>
<i>immigration</i>	<i>incorporate</i>	<i>Concentration</i>	<i>friendship</i>	<i>occupant</i>
<i>leader</i>	<i>renew</i>	<i>Incidente</i>	<i>globalize</i>	<i>positioning</i>
<i>movement</i>	<i>expulsion</i>	<i>Despise</i>	<i>populism</i>	<i>promulgation</i>
<i>criticize</i>	<i>nationalist</i>	<i>Studies</i>	<i>accountability</i>	<i>tropical</i>
<i>election</i>	<i>visa</i>	<i>Criticismo</i>	<i>boyish</i>	<i>charisma</i>
<i>citizen</i>	<i>politics</i>	<i>Illegally</i>	<i>confusion</i>	<i>closure</i>
<i>Party</i>	<i>demonstration</i>	<i>Inhabitant</i>	<i>demolish</i>	<i>credibility</i>
<i>Institute</i>	<i>federalism</i>	<i>Legislation</i>	<i>diplomat</i>	<i>diplomacy</i>
<i>foreigner</i>	<i>Prime</i>	<i>Pact</i>	<i>intimidate</i>	<i>discriminatory</i>
<i>Washington</i>	<i>dismantle</i>	<i>Thesis</i>	<i>mandate</i>	<i>dissident</i>
<i>France</i>	<i>defend</i>	<i>Warn</i>	<i>punctuality</i>	<i>evident</i>
<i>radical</i>	<i>leftist</i>	<i>Opposition</i>	<i>revoke</i>	<i>fascinating</i>
<i>decade</i>	<i>liberal</i>	<i>Summit</i>	<i>showman</i>	<i>funeral</i>
<i>protest</i>	<i>crush</i>	<i>Democratic</i>	<i>solidarity</i>	<i>hypothesis</i>
<i>xenophobic</i>	<i>empire</i>	<i>Iliberal</i>	<i>authorization</i>	<i>irresponsibility</i>
<i>speech</i>	<i>clash</i>	<i>Trade</i>	<i>autocrat</i>	<i>Latino</i>
<i>White</i>	<i>dynasty</i>	<i>Adaptation</i>	<i>closed</i>	<i>manipulate</i>
<i>eastern</i>	<i>isolate</i>	<i>Rebuild</i>	<i>columnist</i>	<i>Ministry</i>
<i>tyrant</i>	<i>activist</i>	<i>Parallel</i>	<i>imperative</i>	<i>Republic</i>
<i>historian</i>	<i>deportation</i>	<i>Permission</i>	<i>implication</i>	<i>rupture</i>
<i>legally</i>	<i>external</i>	<i>Adapt</i>	<i>voting</i>	<i>stake</i>
<i>economist</i>	<i>voter</i>	<i>Emphasize</i>	<i>ambassador</i>	<i>tragedy</i>
<i>border</i>	<i>democracy</i>	<i>Permanence</i>	<i>biography</i>	<i>unite</i>
<i>Congress</i>	<i>assimilation</i>	<i>Confront</i>	<i>devastating</i>	<i>advisory</i>

Fonte: elaborado pelo autor com base no *Sketch Engine*.

Dentre elas, foram selecionadas apenas as de conteúdo (*function words*), ou seja, àquelas que serviriam como uma possível fonte para a identificação das estruturas morfológicas: verbal, nominal, adjetiva e adverbial, como previsto pela taxonomia (HAUSMANN, 1985; ORENHA-OTTAIANO, 2009, 2015).

Em seguida, utilizamos os referidos termos para verificar os resultados colocacionais com aporte nos dados estatísticos encontrados por meio da *Word Sketch*, cuja verificação também nos possibilitou encontrar linhas de concordância mais específicas ao contexto da colocação.

3.3.4 Estrutura e etiquetagem do corpus do estudo

Em nosso objeto de estudo, elencamos um quadro resumo com algumas tipologias que comprova a representatividade do *corpus* e sua estrutura interna, uma vez que essa última

pode e deve estar atrelada aos objetivos de pesquisa de sua compilação, evitando, assim, que o conceito de *corpus* restrinja a um conjunto de textos em larga escala. Vejamos o quadro abaixo:

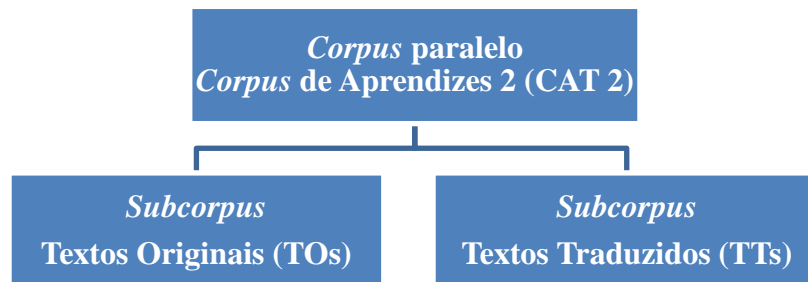
Quadro 11 – Quadro resumo da tipologia do *corpus* de estudo

Modo	Escrito
Tempo	Sincrônico
Seleção	Amostragem
Conteúdo	Monolíngue
Autoria (<i>corpora</i>)	Aprendiz
Disposição interna	Paralelo
Finalidade	Pesquisa
Modo	Escrito

Fonte: elaborado pelo autor.

Dentre as tipologias acima, destacamos a “disposição interna”, apresentada de modo paralelo que, de maneira geral, se trata de uma análise entre os textos originais (TOs) e textos traduzidos (TTs). Observamos o Organograma, a seguir:

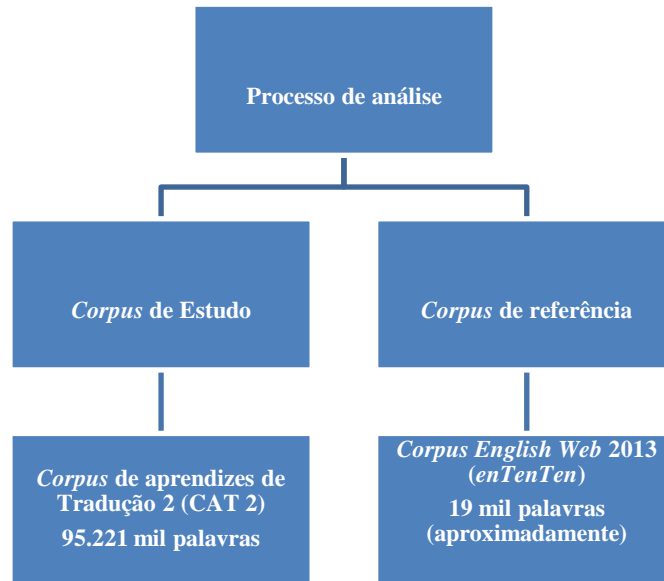
Organograma 2 – Estrutura do *corpus* de estudo



Fonte: elaborado pelo autor.

Depois de definida a estrutura paralela do CAT 2, optamos pela escolha do *English Web 2013*, que está inserido no *Sketch Engine*, por meio do qual foi possível comparar as dificuldades de tradução, bem como observar a frequência de coocorrência ou não da colocação do CAT 2. Vejamos o Organograma:

Organograma 3 – Processo geral de análise



Fonte: elaborado pelo autor.

Depois de definida essa estrutura final que apresenta uma visão geral da estrutura do CAT 2 na pesquisa, descrevemos com mais detalhes o *Corpus* de Referência empregado para comprovação e comparação de uso das colocações extraídas no CAT 2 que foram discutidas na etapa de análise.

3.3.5 O corpus de referência utilizado como base na pesquisa

Para esta investigação, decidimos por utilizar a ferramenta *English Web 2013*, com aproximadamente 19 bilhões de palavras, que compõe a família *Em Ten Ten*, de corpora comparáveis, disponível em mais de 30 idiomas, chegando a mais de 10 bilhões de termos por cada trabalho como referência. Trata-se de textos criteriosamente coletados da internet, como prevê os objetivos da elaboração de *corpus*.

O *English Web 2013* escolhido como *corpus* referência para esta análise é construído por meio de tecnologia especializada, que coleta somente conteúdos linguísticos relevantes. Está inserido no gerenciador de corpora on-line *Sketch Engine*, no qual realizamos a compilação do nosso estudo.

De acordo com Jakubíček *et al.* (2013), a busca ocorre por meio de uma ferramenta de rastreamento (*Spirderling tools*), que foi criada especificamente para busca de dados linguísticos. Por meio do *jusText* é possível realizar o *download* dos textos, permitindo a remoção de “conteúdos irrelevantes”, tais como *links* de navegação, anúncios, cabeçalhos,

rodapés etc. Outro ponto é sua estrutura de *Metadata*, um item capaz de levantar informações relevantes, em relação ao ano de publicação, autores, domínios, tipos (falado ou escrito) e registro formal e informal.

Também, observamos em uma descrição disponível no site do *Sketch Engine*, que o *English Web 2013* e as ferramentas *Keywords*, *Wordlist*, *Concordance* e *Word Sketch* são uma das indicadas para verificação de dados linguísticos por meio deste *corpus* de referência, como já prevemos anteriormente em nossa metodologia de análise do gerenciador em questão.

3.3.6 *LogDice: procedimento estatístico de análise das colocações*

Agora, para detectar se as combinações levantadas tratavam de colocações ou não, utilizamos a ferramenta *Word Sketch* para verificar os padrões e relações gramaticais entre as partes combinadas. Dentro desse mecanismo, nos apoiamos no item estatístico *LogDice*. Esta é uma medida estatística que corrobora com a linguística do ponto de vista computacional, quanto à observação de listas de candidatos à colocação (RYCHLÝ, 2008), um item ainda não muito explorado em pesquisas que envolvem assuntos sobre aprendizagem de línguas (GABLASOVA; BREZINA; MCENERY, 2017).

Segundo as informações do gerenciador de *corpus*, trata-se, especificamente, de um de um tipo de *score* que indica a força da colocação na língua, por meio da amostragem. Basicamente, quanto mais alto o resultado de uma dada colocação, maior é a probabilidade combinatória das partes envolvidas. Contudo, mesmo uma contagem mais baixa, não é algo negativo, pois, isto significa que as terminologias também se combinam com muitas outras, ocorrendo, assim, certo tipo de variação entre as palavras combinadas.

Do ponto de vista prático, o apontamento do *LogDice* é relevante por se tratar de “medidas de associações estatísticas que estabelecem a força de associação entre o nóculo e seus colocados” (BERBER SARDINHA; ACUNZO; FERREIRA, 2016, p. 179-180). Os referidos autores explicam que o *score* do *LogDice* independe do tamanho do *corpus*, pois opera segundo as frequências relativas entre as partes da colocação “nóculo” + “colocado”.

No que tange a frequência e o *LogDice* das colocações analisadas no CAT 2, em comparação com o *corpus* de referência, consideramos o mínimo de ≥ 3 , como nota de corte (FRANKENBERG-GARCIA, 2018), para ponderar qualquer sugestão empregada pelos alunos adequadamente e/ou indicar novas colocações, sendo 14 o valor máximo (RYCHLÝ, 2008). Todos esses itens podem ser verificados a seguir, conforme a marcação em vermelho.

Figura 13 – Relações gramaticais com a palavra *security* na *Word Sketch*

The screenshot shows the Word Sketch interface for the word "security" in the "English Web 2013 (enTenTen13)" corpus. The search results are categorized into three panels:

- verbs with "security" as object:**

ensure	19,410	8.24	...
back	5,250	7.67	...
compromise	4,233	7.67	...
threaten	4,525	7.5	...
enhance	8,383	7.4	...
guarantee	5,732	7.38	...
improve	17,186	7.18	...
strengthen	3,578	6.84	...
trade	2,503	6.71	...
provide	37,058	6.63	...
maintain	8,012	6.6	...
- verbs with "security" as subject:**

guard	2,043	7.7	...
measure	3,024	7.69	...
force	2,018	6.96	...
lend	734	6.02	...
benefit	881	5.84	...
monitor	731	5.84	...
scan	611	5.8	...
feature	2,476	5.68	...
- "security" and/or ...:**

safety	42,200	9.91	...
privacy	19,589	9.48	...
medicare	15,597	9.26	...
peace	20,222	9.07	...
stability	11,928	8.68	...
protection	8,922	7.97	...
security	11,056	7.92	...
reliability	5,912	7.72	...
comfort	6,792	7.64	...

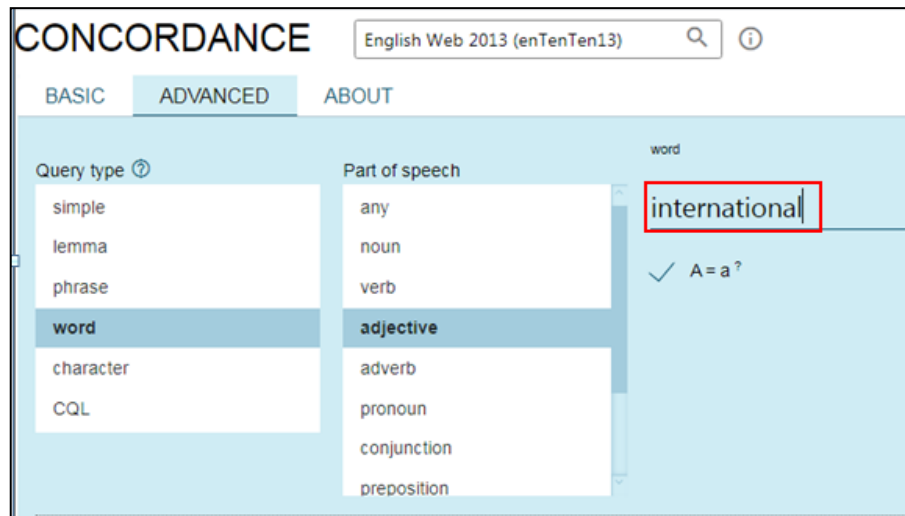
Fonte: captura de tela do *Sketch Engine*.

Além dessa, exploramos o item de busca avançada da ferramenta *Concordance*, em alguns momentos durante a análise, a fim de também comprovar a validade das colocações empregadas, sendo que esta apresenta o item *LogDice*.

No caso de colocações adjetivas, por exemplo, essa parte da ferramenta *Concordance* no *Sketch Engine* pode ser muito útil, principalmente pelo fato de que o colocado desse tipo de colocação coocorre sempre anteposto à palavra de base; diferente das outras formas colocacionais empregadas na taxonomia desta pesquisa, que permitem um espaçamento entre o colocado e a base da combinação.

A partir da imagem a seguir, observamos que o percurso para identificar colocações ocorre por meio tópico *query type*, clicando no subtópico *word*, seguido do tópico *part of speech*, onde, como exemplo, utilizamos a palavra *international* para descrever esse trajeto.

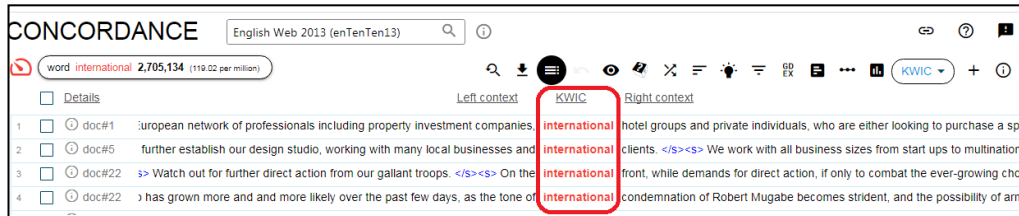
Figura 14 – Item avançado da ferramenta *Concordance*



Fonte: captura de tela do *Sketch Engine*.

Comumente, esta ação gerou as linhas de concordância com a palavra pesquisada, como notamos no exemplo:

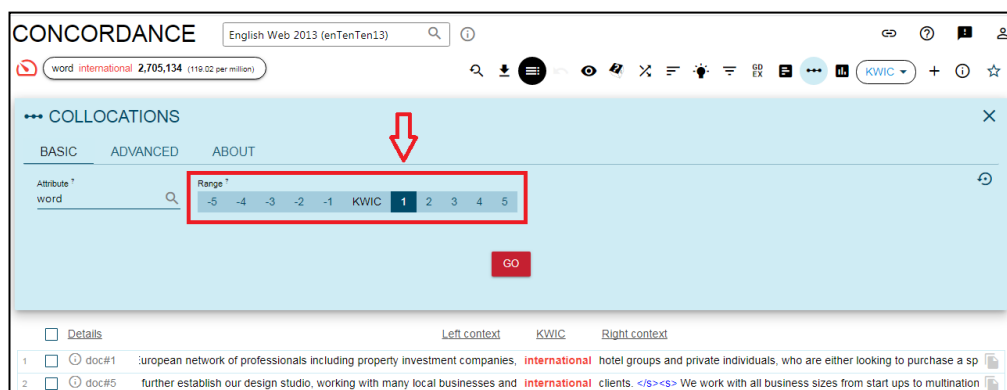
Figura 15 – Linhas de concordância de *international*



Fonte: captura de tela do *Sketch Engine*.

Como resultado, foi gerada uma nova tela com as linhas de concordância da palavra *international*.

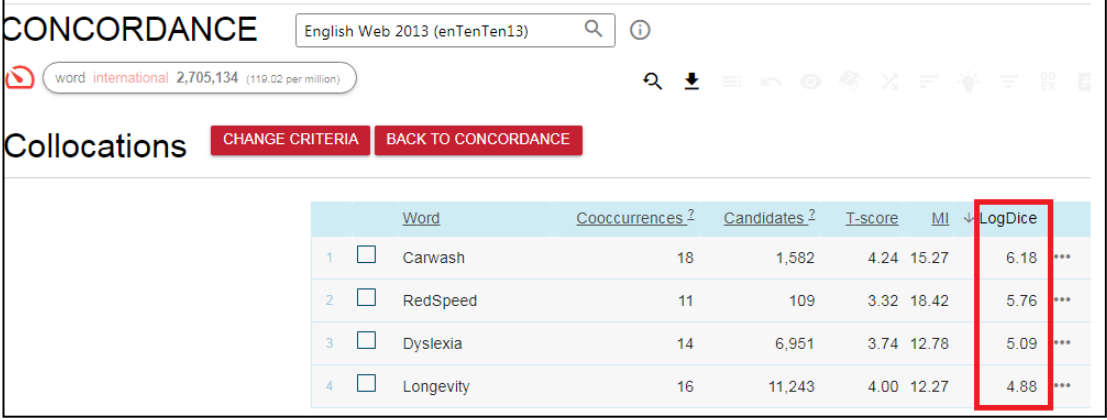
Figura 16 – *Collocations* na ferramenta *Concordance*



Fonte: captura de tela do *Sketch Engine*.

Depois, ao clicar no item *Collocations*, é gerada uma tela que mostra o item *LogDice* na sexta coluna:

Figura 17 – Logdice no item Collocations da ferramenta Concordance



The screenshot shows the 'CONCORDANCE' tool interface. At the top, it displays 'English Web 2013 (enTenTen13)' and a search bar with the word 'international' and its frequency '2,705,134 (119.02 per million)'. Below this, the 'Collocations' section is active, with buttons for 'CHANGE CRITERIA' and 'BACK TO CONCORDANCE'. A table lists four collocations, with the 'LogDice' values highlighted in a red box:

	Word	Cooccurrences ²	Candidates ²	T-score	MI	LogDice
1	<input type="checkbox"/> Carwash	18	1,582	4.24	15.27	6.18 ***
2	<input type="checkbox"/> RedSpeed	11	109	3.32	18.42	5.76 ***
3	<input type="checkbox"/> Dyslexia	14	6,951	3.74	12.78	5.09 ***
4	<input type="checkbox"/> Longevity	16	11,243	4.00	12.27	4.88 ***

Fonte: captura de tela do *Sketch Engine*.

Posteriormente a esse processo de análise estatística da colocação, os dicionários *on-line* baseados em *corpus Macmillan* e o *Online Oxford English Collocation* foram utilizados para conferir se as colocações traduzidas pelos acadêmicos eram realmente empregadas em língua inglesa como idioma de chegada.

3.3.7 Estratégia de atividade: O corpus de referência English Web 2013, ferramenta SKELL e a Online English Collocations Platform empregados para fins pedagógicos.

Em meio aos avanços tecnológicos da LC em áreas afins, como comprovado no decorrer deste trabalho, neste subcapítulo, tratamos dos recursos midiáticos como proposta de atividades sobre o uso das colocações com base em *corpus*. Trata-se da exploração pedagógica de três recursos: o *corpus* de referência *English Web 2013*, a ferramenta *SKELL* e a *Online English Collocations Platform* (ORENHA-OTTAIANO, 2015), com vistas para auxiliar aprendizes brasileiros no estudo formativo de tradução, assim como outros falantes, a alcançarem a competência léxico-fraseológica das colocações.

Tais indicações são fruto dos estudos desenvolvidos no projeto guarda-chuva, COMADEC e contam com o apoio do grupo de pesquisa intitulado Fraseologia e Colocações Baseado em *Corpora* (FRASCORP), devidamente registrado no Diretório de Grupos de Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

3.3.8 Descrição da *Online English Collocations Platform* e suas atividades com jogos

A proposta de exercícios empregando a *Online English Collocations Platform* teve como base as dificuldades colocacionais discutidas durante a análise desta pesquisa. Elaboramos propostas conforme os modelos da referida plataforma, contando ainda com o apoio do *English Web 2013* e dicionários *on-line* para extração dos termos a serem inseridos. Pedagogicamente, avaliamos situações mais pontuais observadas na análise dos dados (problemas combinatórios, combinatória inexistente no *corpus* de referência ou inadequação ao contexto).

Na etapa das atividades, elencamos dois exemplos com aporte nos modelos disponíveis na *Online English Collocations Platform*, sendo o primeiro conforme o modelo *Memory Game*, que abrange colocações especializadas da área política e um segundo a partir do padrão *Gap Fill* da língua geral. A escolha das colocações ocorreu a partir do *English Web 2013*. Em relação à inserção direta das colocações na referida plataforma, seguimos os mesmos passos apresentados por Orenha-Ottaiano (2015) e Caldas (2017), que, de antemão, introduziram uma sequência detalhada de como inserir os itens, bem como formas de utilização por parte dos partícipes.

Uma vez que já possuíamos essas fontes, que podem ser consultadas, decidimos apresentar brevemente o percurso de inclusão dos termos inseridos na categoria *Politics*. Vejamos a sequência de preparação e de inserção das atividades por meio de algumas capturas de telas:

Figura 18 – Imagem de apresentação da plataforma



Fonte: captura de tela do *Online English Collocations Platform*.

O sistema permite dois tipos de usuários, um com permissão para inserir dados, alimentando o ambiente virtual e outro que realiza apenas as atividades como visitante.

Contudo, ambos devem permitir o registro, conforme a Figura a seguir:

Figura 19 – Login da plataforma

The screenshot shows a login form with the following elements:

- Email address:** A text input field containing "name@example.com".
- Password:** A text input field containing "*****".
- Forgot your password?:** A blue link below the password field.
- Log in:** A blue button centered below the password field.
- Don't have an account? Please Sign Up:** A link at the bottom of the form.

Fonte: captura de tela da *Online English Collocations Platform*.

A tela de inserção da atividade do *Memory Game* possui campos para inclusão das colocações por meio das linhas de concordância e, também, indica a fonte pesquisada no item “referências”, conforme Figura abaixo:

Figura 20 – Inserção da atividade *Gap Fill*

The screenshot shows the 'Registers' form with the following elements:

- MEMORY GAME:** A sidebar menu with 'Categories' and 'Collocations' links.
- GAP FILL:** A sidebar menu with 'Categories' and 'Sentences' links (the latter is highlighted).
- Level of Difficulty:** A dropdown menu set to 'Easy'.
- Category:** A dropdown menu set to 'Politics'.
- Taxonomy:** A dropdown menu set to 'VERBAL COLLOCATIONS'.
- Sentence:** A text input field containing the placeholder sentence: "The brand is unlikely to meet a \$[Target] of selling 800,000 vehicles in the U.S. by 2018, according to IHS Automotive.".
- References:** A large empty text input field.
- Buttons:** 'Reset' and 'Save' buttons at the bottom.

Fonte: captura de tela do *Online English Collocations Platform*.

Para alimentar a parte de atividades de *Memory Game* não é preciso usar as linhas de concordância. Com a permissão estabelecida, o usuário precisa apenas inserir as colocações extraídas de alguma fonte, de acordo com a próxima Figura.

Figura 21 – Inserções da atividade *Memory Game*

Category	Taxonomy	Level difficulty	of	Node	Collocate
Politics	VERBAL COLLOCATIONS	Hard		TO BOOST	VOTER TURNOUT
Politics	VERBAL COLLOCATIONS	Easy		TO AFFECT	TURNOUT

Fonte: captura de tela do *Online English Collocations Platform*.

Nos dois modelos de exercícios é preciso indicar o nível de dificuldade (fácil, médio ou difícil), o tipo de colocação (verbal, nominal, adjetiva ou adverbial) e a categoria, que pode ser de língua geral ou especializada. Entretanto, apenas na atividade de *Memory Game*, que se identifica um campo para indicar qual será a palavra nódulo, também intitulado como base.

Sendo assim, elaboramos, também, um quarto capítulo com propostas de atividades de acordo com os modelos da *Online English Collocations Platform*, tomando como base o *corpus* de referência *English Web 2013*. Aproveitamos ainda este mesmo levantamento para introduzir outro modelo de atividade por meio da ferramenta *Word Sketch*, bem como a plataforma *on-line SKELL*.

Portanto, agora, o Capítulo 3, discorre sobre a metodologia empregada antes da etapa de análise de dados, na qual será abordada a verificação descritiva e explicativa das colocações à luz da linguagem estatística computacional.

3.3.9 Breve descrição de uso da ferramenta SKELL

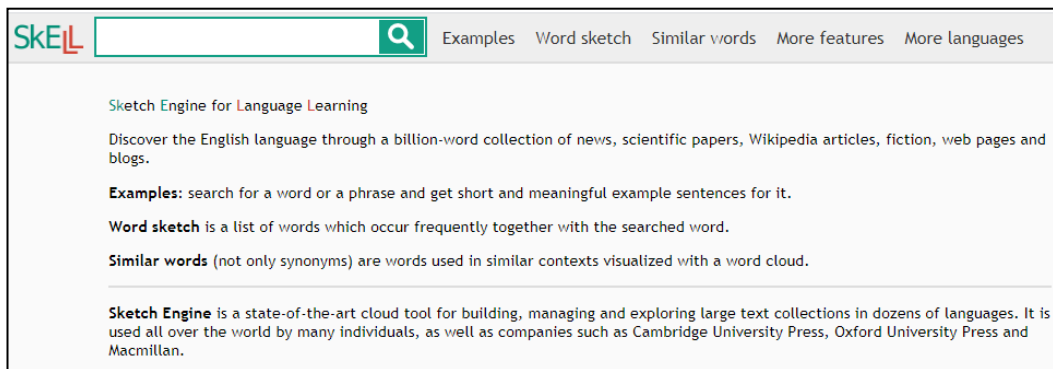
A ferramenta *SKELL Sketch Engine* tem como público-alvo atender professores e estudantes de inglês como língua estrangeira, sendo ainda uma ferramenta gratuita *on-line* e de fácil acessibilidade. Basicamente, de acordo com Baisa e Suchomel (2014), a *SKELL* tem como base o gerenciador de *corpus Sketch Engine*, capaz de compilar dezenas de idiomas a partir da *web*.

Esse processo ocorre por meio de um recurso chamado *Sketch*, que rastreia textos da internet por meio de recursos do *Sketch Engine* (*SpiderLing*, *tokeniser*, *unitok.py* e

TreeTagger) utilizados para a construção de *corpus* de referência, como os da família *Em Ten Ten*.

Por conseguinte, apresentamos a tela inicial da plataforma *SKELL*, que conta com três itens linguísticos: o *Examples*, que gera as linhas de concordância; o *Word sketch*, que apresenta relações gramaticais do termo de busca e o *similar words*, que apresenta outras palavras similares à palavra de busca.

Figura 22 – Tela inicial da plataforma *SKELL*



Fonte: captura de tela da plataforma *SKELL*.

Mais detalhes em relação ao uso da *SKELL* estão dispostos no Capítulo 4, no qual apresentamos as atividades a partir do uso desta ferramenta. A partir de agora, damos sequência a etapa da análise e discussão das colocações no CAT2.

4. ANÁLISE E DISCUSSÃO DE DADOS

Neste capítulo, apresentamos a análise quantitativa e qualitativa, em que são avaliados os termos escolhidos entre o *subcorpus* textos originais (TOs) e o *subcorpus* com os textos traduzidos (TTs), com base nas respectivas palavras-chave de ambos os *corpora*. Depois, esses conceitos elementares foram utilizados para verificar os possíveis problemas no uso das colocações, por meio da comparação entre as ocorrências no TTs e nos TOs. Como critério de seleção, durante a observação, optou-se por analisar somente os itens de conteúdo (substantivo, verbo, adjetivo e advérbio).

Os exemplos com mais problemas na tradução de colocações foram elencados para esta pesquisa. A discussão foi feita a partir da utilização de tabelas, linhas de concordância e quadros de resumos, para observar o uso das colocações traduzidas com mais precisão. Por se tratar de uma proposta com aporte na investigação de erros colocacionais, decidimos por não excluir as recorrências com menores incidências no *corpus*, pois, em conceitos que apresentaram baixa frequência, também identificamos problemas de emprego dessa fraseologia.

4.1 Os resultados do teste de proficiência

O teste de nivelamento de proficiência foi aplicado em dois momentos com turmas diferentes, porém seguimos o critério de selecionar alunos do nível intermediário ao avançado. Vejamos, a seguir, os primeiros resultados das avaliações aplicadas com tradução nas duas turmas trabalhadas:

Tabela 1 – Resultados do *Placement Test* com os alunos do 3º ano

Turma: 3º ano - 2018		
	Níveis de proficiência	N.º de alunos
<i>Básico</i>	<i>A1 – A2</i>	0
<i>Independente</i>	<i>B1 – B 2</i>	21
<i>Proficiente</i>	<i>C1 – C2</i>	08

Fonte: elaborada pelo autor.

Tabela 2 – Resultados do Placement Test com os alunos do 2º ano

Turma: 2º ano – 2019		
	Níveis de proficiência	N.º de alunos
<i>Básico</i>	<i>A1 – A2</i>	0
<i>Independente</i>	<i>B1 – B2</i>	23
<i>Proficiente</i>	<i>C1 – C2</i>	03

Fonte: elaborada pelo autor.

Segundo a Tabela 2, observamos que embora o número de participantes do segundo ano tenha sido maior, os estudantes do terceiro ano, no nível proficiente, tiveram melhores resultados, possivelmente devido à quantidade de tempo em contado com as práticas tradutórias.

4.1.1 Informações do CAT 2 e colocações selecionadas para análise e discussão dos dados

Por meio das informações gerais fornecidas pelo gerenciador de *corpus Sketch Engine*, apontamos alguns dados do CAT 2, os quais foram divididos em números de palavras, frases e documentos.

Tabela 3 – Frequência geral de contagem do Subcorpus de TOs e TTs

Informações no CAT 2	Subcorpus de TOs	Subcorpus de TTs
Palavras	3.832	95.221
Frases	153	4.532
Documentos	6	148

Fonte: elaborada pelo autor com dados extraídos do *Sketch Engine*.

Tendo observado as informações gerais do CAT 2, selecionamos algumas *Keywords* no *subcorpus* de TOs que poderiam indicar possíveis colocações no *subcorpus* de TTs. Nessa fase realizamos também uma triagem, mantendo apenas aqueles de conteúdo, dentre as quais descartamos alguns lexemas, que por si só não formam colocações, tais como: nomes próprios, preposições isoladas e pronomes. Sendo assim, foram preservadas somente as de classe verbal, substantiva, adjetiva e adverbial, por se enquadrarem na taxonomia adotada para este estudo (HAUSMANN, 1985; ORENHA-OTTAIANO, 2004). Das palavras-chave observadas no *subcorpus* de TTs, elencamos apenas 16 como amostragem, as quais evidenciam até mesmo a temática política presente no *corpus*, demonstrando ainda mais a representatividade entre os textos que compõem o CAT 2.

Em seguida, selecionamos alguns termos do CAT2 para compor a análise desta pesquisa, depois de percebermos àquelas que contam com problemas colocacionais, por meio

da comparação entre o *subcorpus* de TOs e o TTs. Vejamos as terminologias extraídas do *subcorpus* de TOs que foram discutidas na etapa de análise dos resultados:

Quadro 12 – Colocações nos TOs e suas traduções no CAT 2

Colocações nos TOs
“arrumar confusão”; “protagonismo internacional”; “traçar estratégia”; “projetar a vida”, “fechamento de fronteiras; “amparar imigrantes”; “sucessivos atritos”; “fazer segurança”

Fonte: elaborado pelo autor.

Como apresentado no quadro acima, selecionou-se sete colocações do *subcorpus* de TOs, cujas opções tradutórias serão discutidas na próxima etapa.

4.2 Descrições quantitativa, qualitativa e estatística das colocações da pesquisa

Com base nas candidatas à colocação levantadas anteriormente, discutimos alguns erros colocacionais tanto da língua geral e quanto da especializada, identificados anteriormente.

Exemplo 1: colocação “arrumar confusão”

Para esta primeira análise, escolhemos investigar a palavra-chave “confusão”, após perceber, por exemplo, problemas de tradução no emprego da expressão “arrumar confusão” (verbo *colocado* + substantivo *base*), conforme o contexto a seguir:

Quadro 13 – Colocação “arrumar confusão” (TO)

<i>Um presidente boquirroto, que “arruma confusão” até mesmo com sua equipe de seguranças.</i>
--

Fonte: elaborado pelo autor.

De acordo com o quadro abaixo, identificamos 11 traduções no *subcorpus* de TTs, respectivamente distribuídas entre os 22 participantes:

Quadro 14 – Colocação “arrumar confusão”: traduções, frequência e exemplos em inglês no CAT 2

Traduções	Exemplos	Frequência
<i>causes problem</i>	<i>A talebearer president, who causes problems even with his security staff.</i>	01

<i>gets in trouble</i>	<i>A chatty president, who even gets in trouble with his security team.</i>	08
<i>gets even trouble</i>	<i>A drunken president, who gets even trouble with his security team.</i>	01
<i>stirs up trouble</i>	<i>He is a tactless president, who stirs up trouble even with his security team.</i>	01
<i>implicates</i>	<i>An intrusive president, that implicates even with his security staff.</i>	01
<i>messes around</i>	<i>An unpleasant president that messes around even with his security guards.</i>	01
<i>make problems</i>	<i>A big mouth president, that make problems even his safety staff.</i>	01
<i>gets even confused</i>	<i>A boquacious president, who gets even confused with his security team.</i>	06
<i>pick a fight</i>	<i>A loquacious president, who pick a fight even with his security team.</i>	01
<i>makes disturbances</i>	<i>An indiscreet president which makes disturbances even with its own security.</i>	01
<i>makes confusion</i>	<i>An indiscreet president that makes confusion even with his security team.</i>	01

Fonte: elaborado pelo autor com base no *subcorpus* de TTs.

A princípio, com base no quadro anterior, excluímos no decorrer da análise dessa colocação a opção *messes around*, por não corresponder ao que a colocação propõe e o verbo *implicate*, por não gerar uma colocação no contexto da tradução. Sendo assim, passamos a considerar somente as traduções: *causes problem*; *gets in trouble*; *gets even trouble*; *stirs up trouble*; *make problems*; *gets even confused*; *pick a fight*; *makes disturbances* e *makes confusion*²⁰. Dentre as sugestões tradutórias produzidas pelos estudantes, houve maior prevalência de uso em duas delas: *gets in trouble* (f. 8) e *gets even confused* (f. 6). Em seguida, as combinatórias *cause problem*, *gets even trouble*, *pick a fight*, *makes disturbances* e *makes confusion* tiveram apenas uma ocorrência para a colocação “arrumar confusão” no *subcorpus* de TTs. O próximo passo foi avaliar, por meio da *Word Sketch*, as frequências e o *Log Dice* dessas indicações no *English Web 2013* no *subcorpus* de TTs, a partir das palavras de base (*problem*, *disturbances*, *confusion*, *trouble* e *fight*), e possíveis correspondentes por meio do dicionário *Macmillan* (2020). Além disso, decidimos incluir a palavra *scene* nessa pesquisa e encontramos a colocação *create a scene*. Vejamos a Tabela a seguir:

²⁰ Nas opções *make problems* e *pick a fight*, identificou-se a ausência do “s”, que configura, em inglês, a 3.ª pessoa do singular do tempo verbal presente simples; a opção [*even*] na sugestão tradutória *gets even confused* também não foi considerada durante o estudo dessa verificação.

Tabela 4 – Formas verbais, traduções, frequências e LogDice da colocação “arrumar confusão”

Formas verbais	Traduções	Frequência	LogDice
Verbo <i>colocado</i> + Substantivo <i>base</i>	<i>Picks a fight</i>	9.502	8,09
	<i>Create a scene</i>	8.561	4,94
Verbo <i>colocado</i> + Partícula Adverbial + Substantivo <i>base</i>	<i>Stirs up trouble</i>	3.097	11,28

Fonte: elaborada pelo autor.

Somente as expressões *pick a fight* e *stirs up trouble* estavam relacionadas à ideia de armar um barraco em público ou discutir com alguém, dado que seus escores apresentaram uma pontuação alta.

Ao observarmos o *English Web 2013*, por meio da *Word Sketch*. Percebemos que as duas traduções mantidas do *subcorpus* TTs: a *pick a fight*, teve frequência (9.502) e *LogDice* 8,09 elevados; enquanto *stirs up trouble* contou com frequência (f. 3.097 mais baixa e *LogDice* (*LogDice*, 11,28) mais alto. Contudo, ambas apresentaram um *LogDice* maior que o mínimo de 3.0. Enquanto isso, a colocação *create a scene* apresentou frequência. 8.561, ou seja, maior do seu *LogDice*, 4,94. Além do mais, independente das medidas estatísticas, todas dizem respeito ao conceito de debater de maneira acalorada com um indivíduo.

Quadro 15 – Linhas de concordância extraídas do English Web 2013

<i>I mean, jeez, wittingly inviting a person who creates scenes regularly at gatherings would be pretty much the worst call on your part.</i>
<i>The government has picked a fight with an already embattled and angry generation who will not tolerate their loans being handed over to debt collection companies.</i>
<i>They can stir up trouble without much fear of retribution.</i>

Fonte: *Corpus* de referência *English Web 2013*.

A partir dos dicionários do *Macmillan* (2020) e o *Oxford Collocation* (2020), extraímos exemplos com as colocações *picks a fight*, *stirs up trouble* e *create a scene*.

- **Busca na primeira entrada do Macmillan: pick a fight/quarrel with someone.**
To start a fight or argument with someone.
He seemed determined to pick a fight with everyone at work that day.
- **Busca na primeira entrada do Macmillan: cause/make trouble for someone/something:**
She can make a lot of trouble for you.
Cause someone/something trouble:
It would just cause me more trouble in the long run.

- **Busca na primeira entrada do Macmillan: a noisy argument or a strong show of feelings in a public place.**
Make/cause a scene: Stop making such a scene!
- **Busca no Collocations do Macmillan: verbs frequently used with trouble.**
Cause **trouble:** bring (someone), cause (someone), foment, give (someone), incite, lead to, make, provoke.
The virus can cause trouble by consuming your computer's storage space and memory.
Her eyes were giving her trouble that summer.

Nos três exemplos elencados anteriormente, observa-se que *pick a fight*, *cause/make a trouble* e *make/cause a scene* foram identificadas por meio da pesquisa geral no *Macmillan* (2020) e todas remetem à noção de “arrumar confusão”, isto é, algo que foi provocado por uma pessoa. Já as indicações dadas no último contexto, por meio do item *collocations*, embora empreguem os verbos *cause* e até fomentem alguma problemática, esta não é diretamente relacionada a ação de um sujeito.

O *Oxford Collocation* (2020) também faz indicações a partir das palavras *fight*, *trouble* e *scene*, ressaltando que os exemplos deste recurso linguístico são colocações:

- **VERB + FIGHT - pick, start** *He tried to pick a fight with me. I don't know who started the fight. | be looking for, be spoiling for, want. Andy was drunk and spoiling for a fight. | be in, get into, get involved in, get mixed up in, have. Don't get into any more fights! | break up, stop. The fight was broken up by a teacher. | win | lose | watch.*
- **VERB + TROUBLE - cause, make.** *He had a reputation for making trouble in the classroom. | be asking for, be looking for, court, stir up. He was asking for trouble when he insulted their country. Fans wandered the town after the match looking for trouble.*
- **VERB + SCENE - cause, create, make.** *Quiet, now! Don't make a scene! | have.*

Como bem mostra o *Oxford Collocation* (2020), nos três contextos podemos visualizar: *pick/start a fight*, no primeiro exemplo; *cause/make trouble*, no segundo e, no terceiro, *cause/create/make a scene*.

Entretanto, se olharmos os detalhes, a colocação *pick a fight* é empregada quando é mencionado algum tipo de confronto, isto é, uma briga entre duas pessoas e não exatamente situação iniciada por uma das partes. Em português, se usássemos a palavra “briga” para remeter à colocação em questão, não haveria problemas, pelo simples fato de o idioma

permitir essa substituição. Porém, no inglês, a expressão não faz referência a alguém, ou seja, ao sujeito que “arruma confusão”, no sentido apenas de uma discussão verbal.

Por outro lado, o item *stirs up trouble* parece se aproximar ao contexto original, uma vez que carrega um significado bastante próximo à língua fonte, conforme mostraram as definições dos dicionários empregados neste trabalho. Porém, essa colocação está mais relacionada à intencionalidade do sujeito em provocar, ou seja, causar problema, do que arrumar uma confusão inesperada, como é o caso do contexto em que a colocação “arrumar confusão” está inserida.

Diante das situações expostas, o termo com base em *scene*, por exemplo, é o único que se aproxima como um equivalente de “arrumar confusão”. Embora *cause* e *make* não tenham demonstrado ocorrências na ferramenta *Word Sketch*, por meio do *English Web 2013*, os consideramos, juntamente com o colocado *create*, devido as suas respectivas indicações no *Oxford Collocation* (2020).

Em seguida, apresentamos o quadro resumo da análise de alguns aspectos, segundo a taxonomia adotada para esta pesquisa.

Quadro 16 – Quadro resumo da colocação “arrumar confusão”

Colocação no TO	“arrumar confusão”
Sugestão tradutória com base na tradução de aprendizes.	X
Nova sugestão tradutória	<i>Cause a scene, make a scene, create a scene</i>
Taxonomia	Verbal
Forma	verbo <i>colocado</i> + substantivo <i>base</i>
Tipo de colocação	Geral

Fonte: elaborado pelo autor.

Exemplo 2: colocação “protagonismo internacional”

O próximo exemplo extraído do CAT 2, refere-se ao “protagonismo internacional”, conforme o contexto abaixo:

Quadro 17 – Colocação “protagonismo internacional” (TO)

“O Brasil vem perdendo **protagonismo internacional**, principalmente, pelos inúmeros erros da política externa dos governos lulopetistas, pautada muito mais pela “solidariedade” a governos amigos do que pela inarredável defesa dos interesses nacionais”.

Fonte: elaborado pelo autor.

Trata-se de uma colocação com base no TT 1 do CAT 2, que apresentou quatro opções de tradução:

Tabela 5 – Traduções, exemplos e frequências da colocação “protagonismo internacional”

Traduções	Exemplos	Frequência
<i>International prominence</i>	<i>Brazil has been losing international prominence, mainly due to the innumerable mistakes of the foreign policies of the lulopetistas (related to the government of Luiz Inácio Lula da Silva) governments, governed more by “solidarity” with friendly governments than by the indispensable defense of national interests.</i>	17
<i>International significance</i>	<i>Brazil has been losing international significance, specially due to the multiple mistakes of the Workers Party's foreign policy, which is based more on “sympathy” towards close countries than on the solid defense of the country's interests.</i>	01
<i>International protagonism</i>	<i>Brazil has been losing international protagonism, mainly, because the innumerable mistakes of the extern politic of the lulopetistas government, guided much more by the “solidarity” to partner governments than the ineligible defense of national interests.</i>	06
<i>Internacional role</i>	<i>Brazil has been losing its international role, mainly due to the innumerable mistakes of the external policy of the "lulopetistas" governments, based much more by the “solidarity” for friendly governments than by the inevitable defense of the national interests.</i>	02
<i>International importance</i>	<i>Brazil has been losing international importance, mainly, due to the innumerable mistakes of the foreign policies of lulapetistas governments, based on “solidarity” to friendly governments than by the irreversible defense of national interests.</i>	02

Fonte: elaborado pelo autor com base no CAT 2.

Porém, destaca-se que, a priori, não consideramos para o prosseguimento da avaliação do verbete “protagonismo internacional”, a sugestão *international protagomism* (porque não apresentou ocorrências de relevância na *Word Sketch*) e a proposta *international role*, por não remeter a ideia original da colocação em questão, a qual também não é combinada com *international* na mesma ferramenta de busca.

Já as combinatórias *International promincence*, por sua vez, aparece escrita dessa forma, uma vez que realizamos a busca e a encontramos pela *Word Sketch*. Agora, as combinações *International importance* e *International significance* não se apresentaram combinadas dessa forma, porém, recorreremos a outra ferramenta, o nível avançado do *Concordance* no *Sketch Engine*, sendo possível identificá-las como colocações, de acordo com a frequência e *LogDice*. Ademais, incluímos juntamente a esses itens a colocação

international recognition, extraída diretamente da *Word Sketch*.

Tabela 6 – Taxonomia, traduções, frequência e *LogDice* da colocação “protagonismo internacional”

Taxonomia			Traduções	Frequência	<i>LogDice</i>
Adjetivo <i>colocado</i> Substantivo <i>base</i>	+		<i>International recognition</i>	10.171	6,88
			<i>International importance</i>	2.449	3,81
			<i>International prominence</i>	1.011	3,72
			<i>International significance</i>	1.400	3,24

Fonte: *English Web 2013*.

Por meio da tabela anterior, nota-se que a indicação *international recognition* teve maior frequência 10.171 e maior *Log Dice* 6,88, em relação às sugestões: *International importance* (f. 2.449; *LogDice*, 3,81); *International prominence* (f. 1.011; *LogDice*, 3,72) e *International significance* (f. 1.400; *LogDice*, 3,24). Todas foram levantadas a partir do *subcorpus* de TTs.

A vista disso percebe-se, ainda, que a por meio desses recursos estatísticos e metodológicos que *international prominence/importance* podem ser mantidas, porém com ressalva para a indicação de *international recognition*, cuja frequência e *LogDice* foram maiores no *corpus* de referência em comparação às propostas do *subcorpus* de TTs. A seguir, apresentamos um exemplo de linhas de concordância extraídas desde mesmo *corpus*.

Quadro 18 – Linhas de concordância extraídas do *English Web 2013*

<i>The recent Palestinian bid for international recognition has failed to secure the backing that the Palestinian leader Mahmoud Abbas was hoping for, and it looks like the United States will not have to use its veto at the United Nations Security Council.</i>
<i>Afghanistan's international importance is similar to Burma's: the world should have the moral responsibility to help out populations that are oppressed by brutal regimes.</i>
<i>Yet despite the undoubted success of such operations in bringing the Palestinian issue to international prominence, it was difficult to see how this could be turned into an overall victory over Israel.</i>

Fonte: *Corpus* de referência *English Web 2013*.

Tomando como base o dicionário *Macmillan* (2020), observamos no tópico *Collocations*, no item *adjectives*, no qual se indica o seguinte contexto a partir da base *prominence, importance, recognition*, exceto para *significance*:

- **Prominence.** *Adj. Involving a lot of people: global, international, national, public. Rob admitted that the speed of his rise to national prominence has stunned him.*

- **Importance.** *Adj.* In a particular geographical area: **international, national**. This is a list of sites considered to be of national importance for nature conservation.
- **Recognition** *Adj.* Among particular people or in certain places: **international, national, professional, public, social, wide, worldwide**. When she met him, Robert Graves had achieved international recognition.
- **Significance.** *Adj.* of a particular Type: **archaeological, clinical, constitutional, cultural, ecological, historical, legal, mystical, political, religious, spiritual, statistical, strategic, symbolic**. Stirling was the lowest bridging point on the Forth, giving the town a great strategic significance. *great:* **considerable, crucial, enormous, great, immense, particular, profound, special**. These collections have immense significance for the history of medicine in the region.

O dicionário *Oxford Collocation* (2020) também apresenta exemplos a partir das palavras de base *prominence, importance, recognition*, porém, não para *significance*:

- **Prominence.** ADJ. **great | growing, increasing | equal | special** a performance which gives a special prominence to the part of Hamlet's mother | **due | undue | international, local, national | political, public**.
- **Importance.** ADJ. **cardinal, central, considerable, critical, crucial, enormous, especial, extreme, fundamental, great, high, immense, key, major, outstanding, overriding, overwhelming, paramount, particular, primary, prime, profound, real, special, supreme, tremendous, vital | first, greatest, highest, utmost** This information is of the first importance. It is of the utmost importance that you arrive on time. | **growing, increasing | declining** the declining importance of manufacturing industry | **lesser, limited, marginal, minor, secondary | direct | general | added | immediate | continued/continuing, lasting | equal | relative | intrinsic | obvious | perceived** differences in the perceived importance of the different subjects in the curriculum | **potential | public | international, national | practical | theoretical | symbolic** the symbolic importance of iron in German culture | **archaeological, commercial, constitutional, cultural, ecological, economic, environmental, historical, legal, military, political, social, strategic**.
- **Recognition.** ADJ. **full | appropriate, due, proper | insufficient | greater, growing, increasing.** There needs to be a greater recognition of corporate crime as a social problem. | **clear, explicit, overt | implicit | apparent | grudging | positive | equal** equal recognition for the work women do | **mutual | immediate, instant | belated.** The award is being made in belated recognition of her services to the industry. | **individual, personal** personal recognition for your achievements | **general, universal, wide, widespread.** The young talent at the club deserves wider recognition. | **international, national, worldwide | public, social | professional | diplomatic | formal, legal, legislative, official | de facto** Twelve states have accorded de facto recognition to the new regime. | **government**.
- **Significance.** ADJ. **considerable, criminal, deep, enormous, great, immense, major, profound, real | full, general, wider.** The scientists are cautious about the wider significance of their findings. | **limited, minor | particular, special | real, true.** They failed to appreciate the true significance of these discoveries. | **functional, practical, statistical, strategic, symbolic, theoretical | constitutional, cultural, economic, moral, historical, political, religious, social**.

O que de fato fica evidente é que para todas as palavras de base, além de indicar o

colocado *international*, exceto para *significasse*, há menção de termos que se referem ao contexto político em questão. Dessa forma, todas as três primeiras opções *International recognition*, *International importance* e *International prominence* seriam as mais indicadas como equivalente para “protagonismo internacional”, devido ao sentido de destaque ressaltado.

Observa-se ainda que todas as indicações apresentaram *LogDice* maior que 3.0. Por esse motivo, apontamos todas as colocações do quadro anterior, mantendo a ordem decrescente conforme foi elencado. Como observado, optou-se por manter *international prominence* e *international importance*, dentre as quatro sugestões tradutórias levantadas por meio do CAT 2, acrescidas da recomendação tradutória *International recognition*. Segundo o quadro resumo, tratou-se da análise de uma colocação da língua especializada, que foi incluída na taxonomia adjetiva.

Quadro 19 – Quadro resumo da colocação “protagonismo internacional”

Colocação no TO	“protagonismo internacional”
Sugestão tradutória com base na tradução de aprendizes.	<i>international prominence</i> <i>international importance</i>
Nova sugestão tradutória	<i>International recognition</i>
Taxonomia	Adjetiva
Forma	Adjetivo <i>colocado</i> + Substantivo <i>base</i>
Tipo de colocação	Especializada

Fonte: elaborado pelo autor.

Exemplo 3: colocação “traçar a estratégia”

Em mais um exemplo de análise, exploramos a seguinte colocação “traçar a estratégia”:

Quadro 20 – Colocação “traçar a estratégia” (TO)

O ex-chanceler Celso Lafer, também presente ao Fórum, ressaltou a importância de o país traçar a estratégia de posicionamento global.

Fonte: elaborado pelo autor.

Foram extraídas oito opções tradutórias para a referida expressão, traduzida por 26 estudantes, a qual apresentou oito terminologias diferentes no *subcorpus* de TTs, conforme relacionado abaixo:

Quadro 21 – Traduções, exemplos e frequências da colocação “traçar a estratégia”

Traduções	Exemplos	Frequência
<i>Develop the strategy</i>	<i>The former chancellor Celso Lafer, also present at the Forum, reinforced the importance for the country to develop a global positioning strategy.</i>	02
<i>Outlining the strategy</i>	<i>The former chancellor Celso Lafer, also present at the Forum, pointed out the importance of the Country outlining a global positioning strategy.</i>	12
<i>Thinking about the strategy</i>	<i>The ex-chancellor Celso Lafer, also present at the Fórum, pointed out the importance of the country thinking about the strategy of global positioning.</i>	01
<i>Trace the strategy</i>	<i>The former chancellor Celso Lafer, who also was present in the forum, pointed out the importance of the country trace the global positioning strategy.</i>	02
<i>Draw the strategy</i>	<i>The former chancellor Celso Lafer, also present at the Fórum, highlighted the importance of the country drawing the global positioning strategy.</i>	04
<i>Chart the strategy</i>	<i>Former Chancellor Celso Lafer, also present at the Forum, reinforced the importance of the Country charting the global positioning strategy.</i>	03
<i>Mapping a strategy</i>	<i>The former chancellor Celso Lafer, also present at the Forum, pointed out the importance of the country mapping a strategy of global positioning.</i>	01
<i>Sketch a strategy</i>	<i>The ex-chancellor Celso Lafer, present on the Forum as well, pointed out the importance of a country to sketch a strategy of global positioning.</i>	01

Fonte: elaborado pelo autor com base no CAT 2.

Dentre as sugestões acima, não encontramos no *Word Sketch* a combinação *trace a strategy*, sendo que as outras apresentaram frequência e indicação de *LogDice*. Ademais, incluímos estas e outras novas proposições a partir da forma verbal, como mostra a tabela a seguir:

Tabela 7 – Sugestões, frequência e *LogDice* da colocação “traçar a estratégia” no *English Web 2013*

Taxonomia	Traduções	Frequência	<i>Log Dice</i>
Verbo colocado + Substantivo base	<i>Develop a strategy</i>	81.696	8,88
	<i>Formulate a strategy</i>	6.993	7,45
	<i>Plan a strategy</i>	9.767	7,22
	<i>Define a strategy</i>	8.665	6,93
	<i>Outline a strategy</i>	5.313	6,89
	<i>Design a strategy</i>	11.358	6,62
	<i>Build a strategy</i>	9.242	5,73
	<i>Prepare a strategy</i>	2.581	5,17
Verbo colocado + Partícula Adverbial +	<i>Articulate a strategy</i>	965	4,66
	<i>Think out a strategy</i>	808	10,15
	<i>Map out a strategy</i>	1.232	10,2
	<i>Draw up a strategy</i>	559	8,49
	<i>Chart out a strategy</i>	67	7,24

Substantivo <i>base</i>	<i>Think up a strategy</i>	133	7,8
	<i>Sketch a strategy</i>	32	5,75

Fonte: *Corpus English Web 2013*.

Para a primeira forma, verbo *colocado* + substantivo *base*, contém 13 indicações, incluindo as colocações mantidas a partir do CAT 2, são elas: *develop a strategy* (f. 81.696; *LogDice*, 8,88) e *outline a strategy* (f. 5.313; *LogDice*, 6,89). Além de novas sugestões, tais como: *formulate a strategy* (f. 6.993; *LogDice*, 7,45); *plan a strategy* (f. 9.767; *LogDice*, 7,22); *define a strategy* (f. 8.665; *LogDice*, 6,93); *design a strategy* (f. 11.358; *LogDice*, 6,62); *build a strategy* (f. 9.242; *LogDice*, 5,73); *prepare a strategy* (f. 2.581; *LogDice*, 5,17) e *articulate a strategy* (f. 965; *LogDice*, 4,66).

Quadro 22 – Linhas de concordância extraídas do *English Web 2013*

<i>In short develop an investment strategy that takes all of your investment assets into account, ideally one that is based on a financial plan.</i>
<i>We might use this meeting to collect or supplement ethnographic information being collected by SFSU or to begin discussion of gaps in city policy and services or being to outline a strategy.</i>
<i>With this out of your hands you can focus your time and effort in other things such as planning online strategies and finding ways to optimize your web pages.</i>
<i>The Martelly government is defining a clear strategy for housing and neighborhood reconstruction, and is already advancing with its implementation.</i>
<i>In “A Time For Action”, Kim outlines a specific strategy, based on her own experience, that anyone can follow to successfully lobby their school board.</i>
<i>The assessment findings have enabled the coalition to design a strategy on how to accelerate implementation of access rights at the national level.</i>
<i>Local and regional government leaders and their organizations have gathered together in the and Global Taskforce of Local and Regional Governments in order to build a joint strategy to contribute to the international policy making debates within the framework of the Post 2015 agenda, Rio+20 follow-up and towards Habitat III.</i>
<i>It helps governments prepare strategies to expand sustainable agricultural production and productivity, and helps design safety nets for the vulnerable.</i>
<i>Labour failed because it failed to articulate a credible overall strategy for the future government of the UK and its opinion poll leads in terms of preferred party policies on salient issues, on a range of image criteria and leadership criteria had declined quite significantly between 2001 and 2005 and declined even further between 2005 and 2010.</i>

Fonte: linhas de concordância extraídas do *corpus* de referência *English Web 2013*.

Na segunda forma, tivemos alguns *phrasal verbs*, de acordo com a estrutura verbo *colocado* + partícula adverbial + substantivo *base*. Entretanto, foi detectada *map out a strategy*, sendo a de maior frequência 1.232 e *LogDice* 10,2, em relação às outras sugestões, tais como: *draw up a strategy* (f. 559; *LogDice*, 8,49), *think up a strategy* (f. 133; *LogDice*, 7,8); *think out a strategy* (f. 808; *LogDice*, 10,15), *chart out a strategy* (f. 67; *LogDice*, 7,24) e *sketch a strategy* (f. 32; *LogDice*, 5,75), as quais, de certa forma, mantém o sentido da

colocação no TO, porém é estatisticamente baixa no *English Web 2013*.

Quadro 23 – Linhas de concordância extraídas do *English Web 2013*

<i>You need to plan, to map, to think out strategies and develop attack tactics to complete each quest.</i>
<i>In recent years, however, the government has begun to map out a strategy to address the competitiveness concerns.</i>
<i>We therefore need to think up a global strategy so that education leads to employment.</i>
<i>The biggest error is under-appreciating corporate culture's importance or ignoring it all together when charting out strategy.</i>
<i>The senior management at Mutual PR, Contributed with invaluable inputs which helped Jaeger chart out its business Strategy for an all India play.</i>
<i>The Building Regional Australia 2013 summit will include presentations, Q and A panels and planning sessions designed to draw out strategies to put to major political parties and governments.</i>
<i>We can sketch out a brand engagement strategy, campaign themes and effective PR stories.</i>

Fonte: *Corpus English Web 2013*.

Em outra busca realizada a partir da base *strategy*, no tópico *collocations*, por meio do dicionário *Macmillan (2020)*, observamos exemplos de colocados no sentido de desenvolver e definir, portanto, sugere-se uma estratégia:

- *Develop a strategy: **arrive at, come up with, deliver, design, determine, develop, devise, establish, formulate, identify, plan, prepare, produce, shape**. Instead of exploiting the poor, we have developed strategies to address poverty.*
- *Define a strategy: **define, outline, set out**. The Welfare Reform Green Paper outlined our strategy to enable people to overcome obstacles to work.*
- *Suggest a strategy: **propose, recommend, suggest**. The President is responsible for diagnosing problems affecting the members and proposing strategies **for** dealing with these problems.*

Além do mais o *Online Oxford Collocation Dictionary (2018)* apresentou a estrutura *verb + strategy*, para o seguinte contexto:

- ***VERB + STRATEGY - have | design, develop, devise, draw up, formulate, map out, plan, work out**. A coherent strategy for getting more people back to work needs to be developed. The council is drawing up a strategy to meet the needs of the homeless. | **explore**. It is certainly a strategy worth exploring. | **adopt, decide (on), employ, follow, implement, pursue, use**. They're pursuing a strategy of massive retaliation. | **outline, propose, set out, unveil**. The document sets out the government's new strategy.*

Sendo assim, como correspondente de “traçar a estratégia”, optou-se por manter as *develop a strategy* e *outlining a strategy*, incluindo as novas opções tradutórias, conforme o quadro:

Quadro 24 – Quadro resumo da colocação “traçar a estratégia”

Categorização	
Colocação no TO	“traçar a estratégia”
Sugestão tradutória com base na tradução de aprendiz	<i>develop a strategy</i> e <i>outlining a strategy</i>
Nova sugestão tradutória	<i>Formulate a strategy, Plan a strategy, Define a strategy, Design a strategy, Build a strategy, Prepare a strategy; Articulate a strategy.</i>
Taxonomia	verbal
Forma	Verbo colocado + Substantivo base
Tipo de colocação	Especializada

Fonte: elaborado pelo autor.

Dentre as opções formadas pela estrutura verbo *colocado* + partícula adverbial + substantivo *base*, mantivemos a opção *think about the strategy* e indicamos apenas as três mais recorrentes como sugestão tradutória. Vejamos a seguir:

Quadro 25 – Quadro resumo da colocação “traçar a estratégia”

Categorização	
Colocação no TO	“traçar a estratégia”
Sugestão tradutória com base na tradução de aprendizes.	X
Nova sugestão tradutória	<i>map out a strategy</i>
Taxonomia	verbal
Forma	Verbo colocado + Partícula Adverbial + Substantivo base
Tipo	Especializada

Fonte: elaborado pelo autor.

Exemplo 4: colocação “fechamento das fronteiras”

Dando prosseguimento a mais um exemplo, agora, tratamos do termo “fechamento das fronteiras”:

Quadro 26 – Colocação “fechamento de fronteiras” (TO)

“Antes disso, porém, uma marcha fúnebre foi marcada para sábado em Chemnitz em memória da vítima do crime – o que para os ativistas de extrema direita será uma nova oportunidade de reivindicar o *fechamento das fronteiras*”.

Fonte: elaborado pelo autor.

De acordo com o contexto, a partir do *subcorpus* de TTs, identificamos quatro traduções.

Quadro 27 – Traduções, exemplos e frequências da colocação “fechamento de fronteiras”

Traduções	Exemplos	Frequência
<i>Closure of borders</i>	<i>However, before that, a funeral march was set for Saturday in Chemnitz in memory of the victim of the crime - which for far right activists will be a new opportunity to demand the closure of the borders.</i>	20
<i>Shutting down of border</i>	<i>However, before that, a funeral march is scheduled for this Saturday in Chemnitz in memory of the murder victim - which, for the far-right activists, will be a new opportunity to demand the shutdown of Germany's borders.</i>	01
<i>Closing of the borders</i>	<i>However, before that, a funeral march was scheduled for Saturday in Chemnitz in memory of the crime's victim - what is going to be a new opportunity for the far-right activists to claim the closing of the borders.</i>	03
<i>Edges closing</i>	<i>Before that, a funeral walk was stablished to next Saturday in Chemnitz in memorium to the crime victim - which for the extreme right activists will be a new opportunity to claim the edges closing.</i>	01

Fonte: elaborado pelo autor com base no CAT 2.

Inicialmente, descartamos a opção *edges closing*, pois a palavra “edges” não atende ao conceito de fronteiras. Em seguida, passamos a verificar essas combinações na *Word Sketch* a partir da base *border* e identificamos uma alta relação combinatória formando uma colocação verbal. Sendo assim, entre as sugestões, notamos o uso de *closing of the border*, que, embora, teve apenas três ocorrências no *subcorpus* de TTs, foi a que mais se aproximou daquelas encontradas no *English Web 2013*, por meio da *Word Sketch*, como mostra o primeiro exemplo do quadro a seguir.

Quadro 28 – Sugestões, frequência e LogDice da colocação “fechamento de fronteiras” no English Web 2013

Taxonomia	Traduções	Frequência	LogDice
Verbo colocado + Substantivo base	<i>Close the border</i>	3.197	6,38
	<i>Seal the border</i>	902	6,41
Verbo colocado + Partícula Adverbial + Substantivo base	<i>Shut down the border</i>	118	5,19
	<i>Close off the border</i>	53	7,23
	<i>Seal up the border</i>	13	6,49

Fonte: elaborado pelo autor com base no *corpus English Web 2013*.

Sendo assim, conforme a frequência e o *Log Dice* maior que 3.0, como mostram os

dados da tabela, tivemos a incidência de duas formas: verbo *colocado* + substantivo *base*, com dois exemplos, *close the border* (f. 3.197; *LogDice*, 6,38) e *seal the border* (f. 902; *LogDice*, 6,41); e a segunda estrutura, verbo *colocado* + partícula adverbial + substantivo *base*, sendo: *shut down the border* (f. 118; *LogDice*, 5,19); *close down the border* (f. 37; *LogDice*, 5,65); *close off the border* (f. 53; *LogDice*, 7,23) e *seal up the border* (f. 13; *LogDice*, 6,49). Depois de discutidas essas informações, extraímos algumas linhas de concordância:

Quadro 29 – Linhas de concordância extraídas do *English Web 2013*

<i>On Aug. 15, Egypt closed its border crossing with the Gaza Strip, trying to seal off one route for Islamic militants entering Sinai, while disrupting the connection between the Muslim Brotherhood and its Palestinian ally Hamas.</i>
<i>We must seal the border and round up and deport all illegal aliens even if it takes 10 - 15 years.</i>
<i>Anti-government sentiment is growing to dangerous levels, and Egypt shutting down the border while their Arab Gazan brothers get the shit blown out of them is not playing well at home</i>
<i>Rebels seized handle of big border posts with Iraq and Turkey, top rated Iraqi officials to close off the border with Syria.</i>
<i>There is only time to get organized and demand your governor, with the assistance of your state legislature, seal up your border with Mexico.</i>

Fonte: *Corpus English Web 2013*.

O dicionário *Macmillan (2020)* não apresenta nenhum tópico de colocações para a palavra *border*, porém, percebemos na primeira consulta a indicação do exemplo *close/seal a border*:

- *The **official line separating two countries or regions** close/seal a border: Troops have sealed the border in an effort to catch the guerrillas.*

Não foram encontradas frases com os *phrasal verbs* extraídos da *Word Sketch*. Enquanto isso, no *Online Oxford Collocation Dictionary (2018)*, conseguimos os seguintes tópicos com os mesmos verbos:

- **VERB + BORDER** arrive at, reach, stop at | cross, drive across/over, slip across/over. *They slipped across the border at nightfall.* | escape across/over, flee across/over | form, mark *A river forms the border.* | draw (up), establish, fix | guard, patrol | open | close, seal

Dessa forma, pelo fato de apresentarem maior frequência no *corpus* de referência sugerimos a expressão tradutória para “fechamento de fronteiras”, a opção *close/seal the border*, classificada como uma colocação verbal de tipo especializada, como é possível

observar no quadro resumo a seguir:

Quadro 30 – Resumo da colocação “fechamento de fronteiras”

Categorização	
Colocação no TO	“fechamento (fechar) fronteiras”
Sugestão tradutória com base na tradução de aprendizes.	<i>Close(ing) the border</i>
Nova sugestão tradutória	<i>Seal the border</i>
Taxonomia	Verbal
Forma	Verbo colocado + Substantivo base
Tipo de colocação	Especializada

Fonte: elaborado pelo autor.

Destacamos que a opção indicada no *subcorpus* de TTs *closing of the border* com a preposição *of* teve 54 ocorrências, dentro da estrutura *close the border*, podendo ser empregada, uma vez que o *Sketch Engine* lematiza o verbo em mais de uma conjugação.

Exemplo 5: colocação “amparar imigrantes”

O próximo exemplo faz jus à colocação “amparar imigrantes”, conforme a problemática extraída do *subcorpus* de TO:

Quadro 31 – Colocação “amparar imigrantes” (TO)

“O Status de Proteção Temporária (TPS, na sigla em inglês) é um mecanismo criado pelo Congresso para amparar imigrantes que, em razão de desastres naturais ou de conflitos armados, não podem retornar em segurança a seus países”.

Fonte: elaborado pelo autor.

A partir dessa informação, identificamos três traduções em inglês no *subcorpus* de TTs, seguida do contexto tradutório e frequência.

Quadro 32 – Traduções, exemplos e frequências da colocação “amparar imigrantes”

Traduções	Exemplos	Frequência
<i>Support immigrants</i>	<i>The Temporary Protection Status (TPS) is a mechanism created by the Congress to support immigrants who, due to natural disasters or armed conflict, can not return safely to their countries.</i>	15
<i>Help immigrants</i>	<i>The Temporary Protected Status (TPS) is a mechanism created by the Congress to help immigrants, which due natural disasters or armed conflicts cannot return in security to their countries.</i>	04

<i>Shelter immigrants</i>	<i>The Temporary Protected Status (TPS) it's a mechanism the Congress has created to shelter immigrants that due to natural disasters or armed conflicts won't be able to return safety to their countries.</i>	02
---------------------------	--	----

Fonte: elaborado pelo autor com base no CAT 2.

Não detectamos na *Word Sketch* a estrutura *to support immigrants* quando realizamos a consulta por meio do termo base (*immigrant*) no plural, para a fim de gerar mais opções combinatórias, pois dessa forma o *Sketch Engine* já foi capaz de localizá-la. Desse modo, encontramos apenas as opções *help immigrants* e *shelter immigrants*, e, ainda, indicamos outras duas traduções, como: *welcome immigrants* e *assist immigrants*, ambas na *Word Sketch*.

Tabela 8 – Sugestões, frequência e LogDice das colocações no English Web 2013

Frequência da tradução de “amparar imigrantes” no CAT 2			
Taxonomia	Traduções	Frequência	LogDice
Verbo <i>colocado</i> + substantivo <i>base</i>	<i>Wecolme imigrants</i>	912	5,48
	<i>Assist immigrants</i>	513	4,38
	<i>Help immigrants</i>	2.197	3,92
	<i>Shelter immigrants</i>	28	0,03

Fonte: elaborado pelo autor com base no *corpus English Web 2013*.

Conforme a Tabela 8, estas expressões atendem a forma verbo *colocado* + substantivo *base* da taxonomia adotada. Das quatro traduções, as três primeiras *wecolme imigrants* (f. 912; *LogDice*, 5,48), *assist immigrants* (f. 513; *LogDice*, 4,38) e *help immigrants* (f. 2.197; *LogDice*, 3,92) apresentaram frequência mais elevada e *LogDice* maior que 3.0.

Porém, considerando ser a colocação com mais incidência e *Log Dice* maior que o mínimo estipulado, *help immigrants* poderia servir como uma forte candidata a equivalente à terminologia “amparar imigrantes”, mesmo não tendo sido a mais empregada no *subcorpus* dos alunos. A sugestão *shelter immigrant*, que apresentou baixa frequência e *Log Dice* menor que 3, não seria uma indicação para o contexto em questão.

Vejamos outras formas comentadas aqui, com base no *corpus* de referência *English Web 2013*.

Quadro 33 – Linhas de concordância extraídas do English Web 2013

*We must continue to **welcome immigrants** and foreign workers who come to our country legally, giving priority to those who can advance the nation's interests and common good.*

*Also speaking at the Jewish Center of Jackson Heights were several community groups who **assist** current-day **immigrants** with the adjustment to America.*

*The Job Search Workshop (JSW) program **helps** recent **immigrants** develop the necessary skills to find employment in the shortest possible time.*

*It was a center for **sheltering immigrants**, above all immigrants who had escaped from wars: from Kosovo, from Palestine, from Kurdistan.*

Fonte: *Corpus English Web 2013.*

Depois dessa etapa, consultamos a palavra *immigrants* no item *Collocations* no dicionário *Macmillan* (2020) e encontramos o contexto verbal a seguir:

- **Verbs** frequently used with **immigrant**
encourage or allow immigrants into a country: **attract, integrate, welcome.** *During the nineteenth century, Argentina was eager to attract immigrants **from** Europe.*

Logo, a primeira definição, indica verbos que tratam de encorajamento ou da permissão de imigrantes em um país, sendo que em um deles há menção do verbo *welcome*. Em continuidade, o *Online Oxford Collocation Dictionary* (2018) também apresenta colocações a partir da palavra de base *immigrant*:

- VERB + IMMIGRANT **accept, welcome** | **deport, return** *ships laden with would-be immigrants who were forcibly returned.*

Assim, não foram identificados exatamente grandes problemas tradutórios neste exemplo, mas devido às diferenças entre as frequências do *corpus* de estudo e o de referência quanto ao emprego da colocação *help immigrants*, percebeu-se certo grau de desconhecimento em relação ao uso do termo na língua-alvo, demonstrando, assim, a presença da ingenuidade do aprendiz na escolha do item correspondente.

Mesmo que a opção *support immigrants* tenha apresentado frequência inferior a *help immigrant*, preferimos mantê-la, devido ao fato de contêm a ideia da colocação no TO. Quanto ao resumo da colocação “amparar imigrantes”, optou-se, então, por utilizar as duas sugestões empregadas (*help immigrants* e *support immigrants*), elencadas na taxonomia verbal e classificadas como especializadas. Vejamos o quadro abaixo:

Quadro 34 – Resumo da colocação “amparar imigrantes”

Colocação no TO	“amparar imigrantes”
Sugestão tradutória com base na tradução de aprendizes.	<i>help immigrants</i>
Nova sugestão tradutória	<i>Welcome immigrants</i> <i>Assist immigrants</i> <i>House immigrants</i> <i>Shelter immigrants</i>

Taxonomia	Verbal
Forma	Verbo <i>colocado</i> + Substantivo <i>base</i>
Tipo de colocação	Especializada

Fonte: elaborado pelo autor.

Exemplo 6: colocação “projetar a vida”

Agora, tratamos da expressão “projetar a vida”, referindo-se aos imigrantes que vivem ilegalmente nos EUA, como pode ser observado no trecho exemplificado:

Quadro 35 – Colocação “Projetar a vida” (TO)

Cidadãos de cinco países que entraram legalmente nos EUA graças a visto de permanência especial em função de desastres naturais projetam a vida na clandestinidade.

Fonte: elaborado pelo autor.

Dado o contexto, encontramos seis tipos de traduções levantadas no *subcorpus* de TTs, como pode ser observado no próximo quadro, seguidas de suas respectivas linhas de concordância e de frequência empregadas pelos alunos.

Quadro 36 – Traduções, exemplos e frequências da colocação “projetar a vida”

Traduções	Exemplos	Frequência
<i>Projects life</i>	<i>Trump deportation plan must take 300,000 immigrants into illegal condition Citizens from five countries who entered the US legally, thanks to special stay visa due to natural disasters, projects life in clandestinity.</i>	09
<i>To Plan their life</i>	<i>Citizens from five countries who entered legally in the United States due special stay visa in reason of natural disasters start to plan their life in clandestinity.</i>	02
<i>Arrange their lives</i>	<i>Citizens from five countries who legally entered the US thanks to the special permanent visa due natural disasters, arrange their lives undercover.</i>	01
<i>Design life</i>	<i>Citizens from five countries, who entered the US legally, thanks to a special stay visa due to natural disasters, design life in clandestinity.</i>	01
<i>Visualize lives</i>	<i>Citizens of five countries who have entered the US legally because of a special permanent visa due to natural disasters visualize their lives in clandestinity.</i>	01
<i>Live on</i>	<i>Tree thousands of immigrants will live illegally because of trump deportation plan Citizens from five countries who entered legally on US due to the permanent visa based on natural disasters, will live on clandestinely.</i>	01

Fonte: elaborado pelo autor com base no CAT 2.

Durante a consulta por meio da palavra base *life*, no singular, percebemos que ela não combina com os verbos *project*, *plan*, *arrange*, *design* e o *phrasal verb* *live on* como

colocado. Apenas identificamos a ocorrência do *phrasal verb to plan out*, mas este apenas remete à ideia de planejar e, não, projetar, como mostra a colocação original. Por esse motivo, elencamos novas sugestões:

Tabela 9 – Sugestões, frequência e *LogDice* das colocações no *English Web 2013*

Frequência da tradução de “projetar a vida” no CAT 2			
Taxonomia	Traduções	Frequência	<i>LogDice</i>
Verbo <i>colocado</i> + Substantivo <i>base</i>	<i>Live their lives</i>	363.046	10,99
	<i>Lead lives</i>	61.852	8,16
	<i>Spend their lives</i>	43.639	7,56
	<i>Dedicate their lives</i>	20.791	7,25
	<i>Start their lives</i>	30.343	7,09
	<i>Begin their lives</i>	21.244	6,98
	<i>Put their lives</i>	26.840	6,64
	<i>Keep their lives</i>	11.022	5,16

Fonte: elaborada pelo autor com base no *corpus English Web 2013*.

Sendo assim, mesmo entre essas várias proposições, observamos que elas atendem ao conceito de projetar a vida em algo (clandestinidade). Considerando o *LogDice* maior para o menor, temos de acordo com a tabela, a sequência: *Live their lives* (f. 363.046, *LogDice*, 10,99); *lead lives* (f. 61.852; *LogDice*, 8,16); *spend their lives* (f. 43.639; *LogDice*, 7,56); *dedicate their lives* (f. 20.791; *LogDice*, 7,25); *start their lives* (f. 30.343; *LogDice*, 7,09); *begin their lives* (f. 21.244; *LogDice*, 6,98); *put their lives* (f. 26.840; *LogDice*, 6,64); *keep their lives* (f. 11.022; *LogDice*, 5,16). Para explicitar mais detalhadamente o uso dessas colocações, a seguir apresentamos cada uma delas em seus devidos contextos:

Quadro 37 – Linhas de concordância extraídas do *English Web 2013*

<i>Catherine Cassidy founded U*styled in 2008 to give professional women the tools to live their lives in style by curating a wardrobe they love that makes getting dressed easy and effortless for every morning, big meeting, presentation and special event.</i>
<i>There is also confusion between what Lenin may or may not have thought about the prospect for liberation in a free society and how he thought revolutionaries should lead their lives under capitalism.</i>
<i>After all, I don't want to spend my life shuffling pixels from one board to another.</i>
<i>People agree to do it anyway – they compete and dedicate their entire lives to giving themselves even an opportunity to do it, because there is a tremendous audience for it – and it's fun.</i>
<i>“Grown Ups 2” – A Hurt Your Sides Summer Comedy The sequel to the 2010 blockbuster, “Grown Ups 2” follows Lenny, Eric, Kirk and Marcus after they start their new lives by moving back to the town where they grew up.</i>
<i>For three years, I worked at McCann Erickson, an advertising agency in Madrid, but then I decided to move to Barcelona because I wanted to begin my life as a freelance illustrator.</i>

*Red Lion Foods is a different sort of food company and understands that many people in the UK have links to the Armed Forces or simply want to say 'thank you' to those who **put their lives on the line** to defend Queen and country*

*As the economic crisis worsens, more and more Americans struggle **to keep their lives together**.*

Fonte: *Corpus English Web 2013*.

No dicionário *Macmillan* (2020) temos a indicação dos verbos *live* e *spend* no sentido de viver a vida e *dedicate* e *devote*, com relação a gastar tempo de vida com algo, como mostram as frases:

Verbs frequently used with life.

- Live your life: **go through, live, spend.**
She spent her whole life fighting for racial equality.
- Spend a lot of your life doing something: **dedicate, devote, give.**
He was able to devote his life to intellectual pursuits.

Enquanto isso, no *Oxford Collocation* (2020) também encontramos duas indicações de colocações na estrutura verbo + *life* em duas situações, remetendo ao período entre o nascimento e a morte, com sugestões para *live, spend, dedicate* e *devote*, bem como sobre modos de vida, a partir de *lead* e *live* (novamente indicado) e *enjoy*.

- **period between birth and death**
VERB + LIFE go through, live, spend She went through life always wanting what she couldn't get. He spent his whole life in Cornwall. | end He ended his life a happy man. | shorten | prolong | dedicate, devote He devoted his life to the education of deaf children.
- **Way of living**
VERB + LIFE have, lead, live She leads a busy social life. | enjoy | change Learning meditation changed her life. | dominate, take over He never let his work dominate his life. | ruin He ruined his life through drinking.

Sendo assim, ambas foram elencadas na seguinte forma da taxonomia adotada:

Quadro 38 – Resumo da colocação “projetar a vida”

Categorização	
Colocação no TO	“projetar a vida”
Sugestão tradutória com base na tradução de aprendizagens	X
Nova sugestão tradutória	<i>Live their lives, Lead lives, Spend their lives, Deticate their lives, Start their lives, Begin their lives, Put their lives, Keep their lives</i>
Taxonomia	Verbal
Forma	Verbo colocado +Substantivo base
Tipo de colocação	Geral

Fonte: elaborado pelo autor.

Exemplo 7: colocação “sucessivos atritos”

No próximo exemplo, trabalhamos o termo “sucessivos atritos”, que envolve situações de desavenças entre líderes mundiais.

Quadro 39 – Colocação “sucessivos atritos” (TO)

Os sucessivos atritos com líderes ocidentais, suas críticas à União Europeia [...].

Fonte: elaborado pelo autor.

Quadro 40 – Traduções, exemplos e frequências da colocação “sucessivos atritos”

Traduções	Exemplos	Frequência
<i>Successive clashes</i>	<i>Successive clashes with Western leaders, the criticism about the European Union, free trade, immigration and support to nationalist causes such as Brexit, opened last week the progressive rupture between US President Donald Trump and Western Europe, Washington's privileged for the past 70 years.</i>	11
<i>Successive attrition</i>	<i>The successive attrition with Western leaders, criticism against European Union, free trade, immigration and the support for nationalist causes, like the Brexit, showed last week the progressive rupture between the United States president, Donald Trump, and the Western Europe, Washington's prime partner for the last 70 years.</i>	02
<i>Successive Frictions</i>	<i>Successive frictions with Western leaders, their criticism of the European Union, free trade, immigration and support for nationalist causes such as Brexit, has opened up last week the progressive rupture between US President Donald Trump and Western Europe, Washington's privileged partner for the past 70 years.</i>	02
<i>Successive disagreements</i>	<i>Successive disagreements with Western leaders, their criticism of the European Union, free trade, immigration and support for nationalist causes such as Brexit last week revealed the progressive rupture between US President Donald Trump and Western Europe, Washington's prime partner for the last seventy years.</i>	01
<i>Successive problems</i>	<i>Successive problems with Western leaders, their criticism of the European Union, free trade, immigration, and support for nationalist causes such as Brexit last week opened the progressive rupture between US President Donald Trump and Western Europe, Washington's prime partner for the past 70 years.</i>	01
<i>Successive Conflicts</i>	<i>Successive conflicts with Western leaders, criticism of the European Union, to free trade, immigration and support for nationalist causes such as Brexit last week opened the progressive rupture between US President Donald Trump and Western Europe, Washington's prime partner for the past 70 years.</i>	01
<i>Successive discussions</i>	<i>Successive discussions with Western leaders, critics to European Union, free trade, immigration and his support to nationalists causes such as Brexit, set wide open last week the progressive rupture among the president of United States, Donald Trump, and the Western Europe, Washington's privileged partner for the past 70 years.</i>	01

Fonte: elaborado pelo autor com base no CAT 2.

Segundo o quadro 40, em uma ordem decrescente no CAT 2, observamos uma preferência maior de uso da combinatória *successive classes* e menos ocorrência no *English Web 2013*. Em seguida, as opções de combinação *successive attrition*, *successive frictions* e *successive disagreements*, embora tenham apresentado alguma frequência (mesmo que baixa no CAT 2), elas não são usuais como podemos verificar no *English Web 2013*. Por outro lado, nas três últimas sentenças traduzidas (*successive problems*, *successive conflicts* e *successive discussions*), que também tiveram baixa ocorrência, notamos que ao contrário das primeiras, estas apresentaram maior número incidência no *corpus* de referência.

Tabela 10 – Sugestões, frequência e LogDice das colocações no *English Web 2013*

Taxonomia	Frequência da tradução de “sucessivos atritos” no CAT 2		
	Traduções	Frequência	LogDice
Adjetivo colocado + Substantivo base	<i>Ongoing conflicts</i>	4.283	6,98
	<i>Constant conflict</i>	1.317	5,34
	<i>Ensuing conflicts</i>	213	4,45

Fonte: elaborada pelo autor com base no *corpus English Web 2013*.

A tabela traz que a colocação *Ongoing conflicts* (f. 4.283; *LogDice*, 6,98) foi a mais recorrente, seguido por *Constant conflict* (f. 1.317; *LogDice*, 5,34) e *Ensuing conflicts* (f. 213; *LogDice*, 4,45).

Desse modo, conforme as colocações indicadas, apresentamos seus respectivos contextos, com base no *corpus* de referência *English Web 2013*.

Quadro 41 – Linhas de concordância extraídas do *English Web 2013*

<i>The OPEC Fund for International Development (OFID) has approved a US\$500,000 grant in humanitarian aid in support of Palestinian communities affected by the ongoing conflict in Gaza.</i>
<i>Here born, although the sea is what shapes our character, we also have a great attachment to the land, for after all, in a time of famine as was the time of the Hundred Years and the Iberian Wars in Europe, everyone who moved here was looking for a peaceful place, a place away from constant armed conflicts, simply, a patch of land to farm and feed their family.</i>
<i>Of course the gross oversimplifications, if not sheer idiocy, of the "domino theorists" are perhaps best understood by following the ensuing conflicts between China and Vietnam after the war and the ultimate overthrow of the Khmer Rouge regime by the Vietnamese.</i>

Fonte: *Corpus English Web 2013*.

Entretanto, na checagem dessas opções encontramos no dicionário *Macmillan* (2020) apenas a indicação de *ongoing*:

- *Adjectives frequently used with conflict*

- *armed, bitter, bloody, ethnic, military, ongoing, sectarian, violent. There are currently at least 300,000 children under the age of 18 participating in armed conflicts around the world.*

Em seguida, no *Online Oxford English Collocations* (2020), na classe grammatical adjetivo, foi possível verificar colocados como: *constant, continued, continuing*.

- *conflict noun - fight/argument*
ADJ. big, great, major | bitter, serious, violent | growing, increasing | constant, continued, continuing, unresolved He is in constant conflict with the authorities. | open, overt | global, internal, regional | armed, military | civil, class, cultural, ethnic, family, industrial, political, social, religious.

Não foi encontrada nenhuma evidência contextual de *ensuing conflict*, embora o *corpus* tenha apresentado essa colocação como relevante. Contudo, considerando nosso conhecimento prévio, decidimos por sugerir outra colocação de uso neste contexto, os conceitos-chaves *Ongoing conflicts, constant conflict* e *ensuing conflicts*, indicando sempre aquele de maior frequência para emprego na situação analisada aqui. Dessa forma, optou-se por manter a sugestão tradutória *successive conflicts*, indicando uma nova proposta: *ongoing conflicts*. Ambas expressões de taxonomia adjetiva e classificadas como colocação especializada, conforme o quadro resumo abaixo:

Quadro 42 – Resumo da colocação “sucessivos atritos”

Colocação no TO	“sucessivos atritos”
Sugestão tradutória com base na tradução de aprendizes.	X
Nova sugestão tradutória	<i>Ongoing conflicts, constant conflict</i> e <i>ensuing conflicts</i>
Taxonomia	Adjetiva
Forma	Adjetivo colocado + Substantivo base
Tipo de colocação	Especializada

Fonte: elaborado pelo autor.

Exemplo 8: colocação “fazer segurança”

No último exemplo, exploramos a colocação verbal “fazer segurança” (Verbo *colocado* + Substantivo *base*), a partir do seguinte contexto.

Quadro 43 – Colocação “fazer segurança” (TO)

Para evitar o risco de confrontos, centenas de policiais foram mobilizados para fazer a segurança do protesto, que acabou sem incidentes.

Fonte: elaborado pelo autor.

Em seguida, encontramos no *subcorpus* de TTs, as traduções *make/do security* de “fazer segurança”:

Quadro 44 – Traduções, exemplos e frequências da colocação “fazer segurança”

Traduções	Exemplos	Frequência
<i>Make security</i>	<i>In order to avoid the risk of confrontations, hundreds of police officers were deployed to make security of the protest, which ended without incident.</i>	15
<i>Do security</i>	<i>To prevent confrontations; hundreds of polices are mobilized to do security in the protest which ends with no incidents.</i>	02

Fonte: elaborado pelo autor com base no CAT 2.

Como pode ser observado, *make* teve maior ocorrência em relação ao verbo *do*. Contudo tais expressões indicadas no *subcorpus* para “fazer segurança” não apresentaram nenhuma incidência no *corpus* de referência, de acordo com a *WordSketch*. Por esse motivo, digitamos a palavra de base *security* nesta mesma ferramenta de busca e identificamos as seguintes colocações com base na frequência e no *LogDice*.

Tabela 11 – Taxonomia, sugestões, frequências e *LogDice* das colocações no *English Web 2013*

Frequência da tradução de “fazer segurança” no CAT 2			
Taxonomia	Traduções	Frequência	<i>LogDice</i>
Verbo colocado + Substantivo base	<i>Ensure security</i>	19.410	8,24
	<i>Guarantee security</i>	5.732	7,38
	<i>Provide security</i>	37.058	6,63
	<i>Assure security</i>	1.700	6,21

Fonte: elaborada pelo autor com base no *corpus English Web 2013*.

Logo, notamos que estas quatro sugestões tradutórias atendem ao contexto original de “fazer segurança”. Todas as indicações apresentam um *LogDice* bem maior que o mínimo de 3, demonstrando a força dessas colocações, independente das diferenças entre as frequências. Percebe-se, ainda, que o termo *ensure security* (f. 19.410; *LogDice*, 8,24) e *guarantee security* (f. 5.732; *LogDice*, 7,38) tiveram uma incidência mais baixa em relação, por exemplo, a *provide security* (f. 37.058; *LogDice*, 6,63) e diferente de *assure security* (f. 1.700; *LogDice*, 6,21), que apresentou dados mais baixos em relação aos outros elementos. Mesmo assim, considerando o *Log Dice* do maior para o menor, todas estas propostas podem ser comumente empregadas, pois atenderam à ideia da colocação original.

Para corroborar a estas indicações, selecionamos quatro linhas de concordância, que foram norteadas pela *WordSketch*:

Quadro 45 – Linhas de concordância extraídas do *English Web 2013*

<i>The Government should ensure the security of the elections throughout the country and not let residents of insecure districts be deprived from the election process.</i>
<i>The city recently launched a test program to see whether a rotating assignment would solve the problem while continuing to guarantee security at bars, clubs and other establishments.</i>
<i>And those displaced persons who do manage to procure aid from the government can only collect it for three months, which provides little security for people who are often unable to return home for years.</i>
<i>Teheran is the only country that has demonstrated it can assure the security of Afghan-Iranian border, specially for drug trafficking.</i>

Fonte: *Corpus English Web 2013*.

Para comprovar o contexto de uso, encontramos no dicionário *MacMillan* (2020) exemplos em que alguns dos verbos listados acima que mencionam o propósito de se garantir a segurança (*Make security certain*): *assure, ensure, guarantee* e *provide*.

- **Verbs frequently used with security**
make security certain: afford, assure, enforce, ensure, guarantee, maintain, protect, provide, safeguard. We have frameworks in place to safeguard the security of the information we hold. make security more.

Detectamos no *Online Oxford Collocation Dictionary* (2018) a base *security*, antecedida por verbos como: *ensure, provide (sb with)*:

- **VERB + SECURITY** *ensure, provide (sb with) | improve, strengthen, tighten (up)* *We need to tighten security around the hotel during the president's visit. | compromise, undermine* *The leaking of secrets from the Defence Ministry has compromised national security.*

Em consequência disso, temos a seguir o quadro que reúne os principais dados de análise:

Quadro 46 – Resumo da Colocação “fazer segurança”

Colocação no TO	“fazer segurança”
Sugestão tradutória com base na tradução de aprendizes.	X
Nova sugestão tradutória	<i>Ensure security, Guarantee security, provide security, assure security</i>
Taxonomia	Verbal
Forma	Verbo <i>colocado</i> + Substantivo <i>base</i>
Tipo de colocação	Especializada

Fonte: elaborado pelo autor.

A partir do estudo, foi possível comprovar a validade deste trabalho no que tange a incentivar as pesquisas com base na tradução de aprendizes, que, ainda por falta de algum conhecimento, apresentam certa credulidade na escolha de uma colocação em outro idioma, como é o caso da língua inglesa.

De modo geral, o que mais pontuamos durante os exemplos dados é a questão da ingenuidade do tradutor, em alguns momentos se prende ao texto original, apresenta dificuldades em interpretá-lo, o que também acaba confirmando seu desconhecimento sobre os aspectos convencionais de uma determinada expressão em um idioma estrangeiro. Depois de percorridas a investigação desses oito exemplos extraídos do *subcorpus* de TTs, dos quais encontramos como sugestão colocações verbais (6) e adjetivas (2), sendo que quantificamos as colocações discutidas em quatro etapas.

Quadro 47 – Quatro etapas de discussões das colocações

Número de colocações analisadas no <i>subcorpus</i> de TOs	08
Número de traduções analisadas no <i>subcorpus</i> de TTs	46
Número de traduções do <i>subcorpus</i> de TTs mantidas	06
Número de traduções novas sugeridas pelo pesquisador	33

Fonte: elaborado pelo autor.

Portanto, uma vez que encontramos as dificuldades conforme a hipótese de que aprendizes de tradução têm problemas no emprego das colocações, apresentamos, no próximo capítulo, três exemplos de atividades com base nas referidas análises.

5. PROPOSTA DE ATIVIDADES A PARTIR DA *ONLINE ENGLISH COLLOCATIONS PLATFORM* E EXERCÍCIOS COM O USO DO *ENGLISH WEB 2013* E DA FERRAMENTA *SKELL*

Neste capítulo, tratamos de questões pedagógicas a partir do levantamento das palavras-chave no CAT 2 e a análise das colocações realizadas nesta pesquisa. Desse modo, apresentamos algumas sugestões de atividades baseadas em *corpus*. A primeira delas, no subcapítulo 4.1, foi realizada a partir da plataforma *Online English Collocations Platform* (ORENHA-OTTAIANO, 2015, no prelo).

A ferramenta oferece dois modelos de tarefas, *Memory Game* e *Gap Fill*, os quais serviram de suporte pedagógico para inserirmos outras propostas de colocações, por meio dos tópicos, *Politics e General*. Sendo assim, esses dois tipos de exercícios *Online English Collocations Platform* ainda dispõem de mecanismos que nos permitiram classificar as terminologias sugeridas em três níveis de dificuldade: *Easy* (fácil), *Medium* (médio) e *Hard* (difícil). Todos esses recursos tiveram como eixo os desvios em seu uso convencional, cometidos por parte dos aprendizes, participantes desta investigação.

Em seguida, no subcapítulo 4.2, discorreremos a respeito de outra ideia, que adotou o uso de *corpus* de referência *English Web 2013* e a plataforma *SKELL Sketch Engine*. Da mesma forma que a primeira proposta pedagógica, os itens deste subcapítulo também foram elencados com base nas dificuldades dos aprendizes de tradução, tendo em vista os erros colocacionais levantados nesta pesquisa.

Do ponto de vista pedagógico as atividades foram pensadas a partir de duas abordagens já relatadas na fundamentação teórica, que são: a Abordagem Movida por Dados (*Data Driven Learning – DDL*), proposta por Johns (1991) e Boulton (2009, 2015), em que o discente passa a ocupar o centro da aprendizagem e o docente tem a tarefa de mediar o conhecimento; e a Abordagem Lexical (*Lexical Approach*), estudada por Lewis (2000), que oferece ao usuário (sendo, aqui, o aprendiz de tradução) oportunidades para que este produza e compreenda as colocações a partir do seu contexto de uso, considerando seu grau de fixação na língua.

5.1 As atividades colocacionais com base no CAT 2 inseridas na *Online English Collocations Platform*

Nesta seção, introduzimos uma proposta didática com base em *corpus*, a qual conta com duas atividades iniciais a *Memory Game* e a *Gap Fill*, oferecidas pela *Online English Collocations Platform*, as quais foram elaboradas tendo como aporte os problemas de uso colocacional, encontrados a partir da análise do material elaborado pelos aprendizes de tradução deste estudo. Primeiramente, para utilizar a plataforma *Online English Collocations Platform*, o usuário deve abrir a interface inicial, sendo possível acessá-la por meio dos campos para *Login* e senha ou *Sign Up*, caso o usuário não possua uma conta. Vejamos a imagem a seguir:

Figura 23 – Tela inicial da *Online English Collocations Platform*



Online English Collocations Workbook

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 disposal, 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.

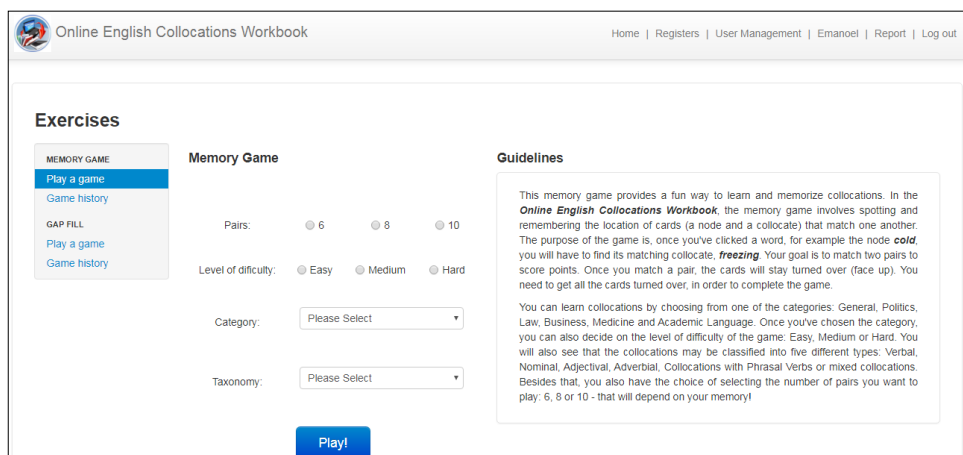
Don't have an account? Please [Sign Up](#)

Fonte: captura de tela da *Online English Collocations Platform*.

Uma vez logado, apresentamos abaixo atividades que podem ser realizadas na plataforma, usando os modelos *Memory Game* e *Gap Fill*, por meio das categorias *Politics* e *General*, como opções temáticas.

A primeira, trata-se de uma prática com colocações especializadas utilizando o modelo *Memory Game*. A seguir, temos a tela inicial na *Online English Collocations Platform* que introduz este tipo de exercício.

Figura 24 – Tela inicial do exercício *Memory Game*



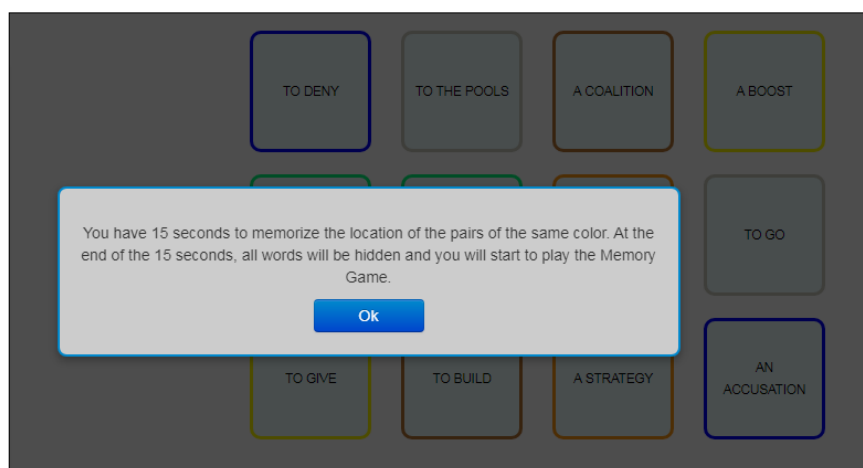
Fonte: captura de tela da *Online English Collocations Platform*.

Para dar início ao *Memory Game*, o usuário pode seguir as seguintes orientações de uso:

- No ícone *EXERCISES* à esquerda, clique na atividade *Memory Game*;
- Em seguida, como apresentado no meio da imagem, selecione os itens *Pairs*, *Level of difficulty*, *Category* e *Taxonomy*;
- Clique na tecla *Play*.

A próxima figura traz as próximas etapas de como é a atividade na plataforma. Após realizado o passo a passo acima, uma tela de aviso é aberta, com informações sobre o *Memory Game*:

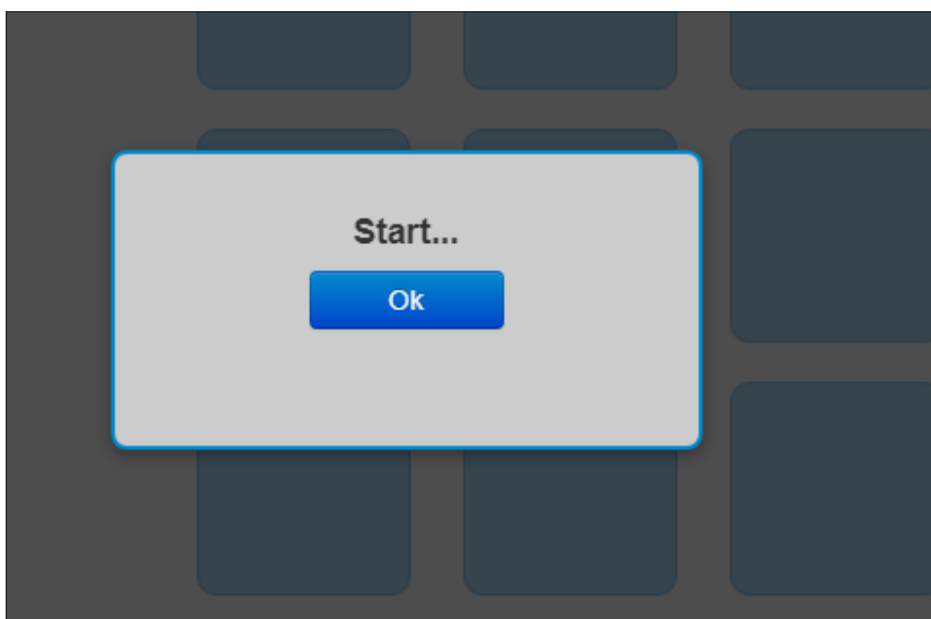
Figura 25 – Tela instrucional do exercício *Memory Game*



Fonte: captura de tela da *Online English Collocations Platform*.

Após, clicar em *ok*, outra tela aberta, porém, dessa vez, com todos os *cards* do jogo embaralhados. O exercício se inicia quando o usuário clicar em *ok* novamente, logo abaixo de *start*.

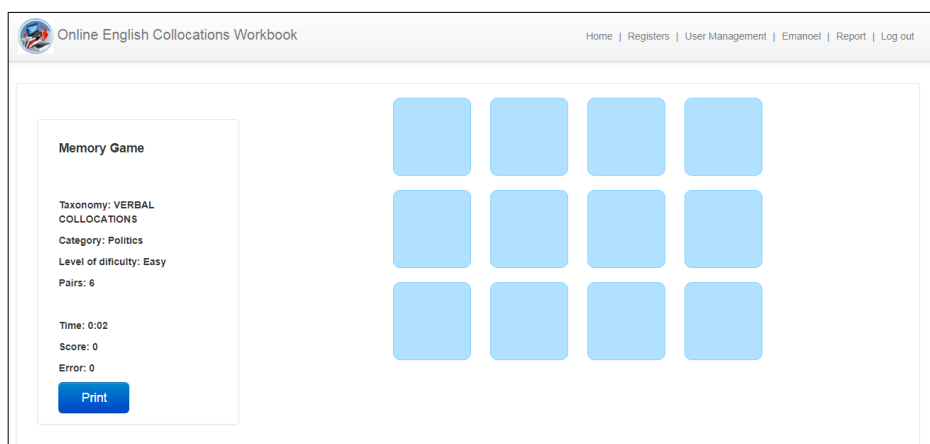
Figura 26 – Tela inicial de início exercício *Memory Game*



Fonte: captura de tela da *Online English Collocations Platform*.

A seguir, os *cards* já estão disponíveis para iniciar o jogo. Nesta etapa, o usuário deverá se atentar ao tempo de execução da atividade.

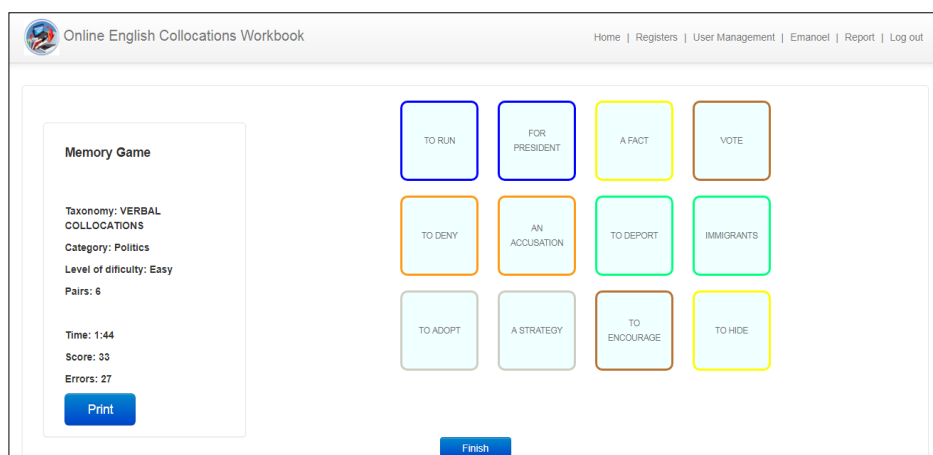
Figura 27 – Tela inicial de início exercício *Memory Game*



Fonte: captura de tela da *Online English Collocations Platform*.

Conseqüentemente, a tela do jogo estará disponível para realização da atividade, como apresentado na imagem a seguir:

Figura 28 – Tela dos *cards* embaralhados – Exercício *Memory Game*

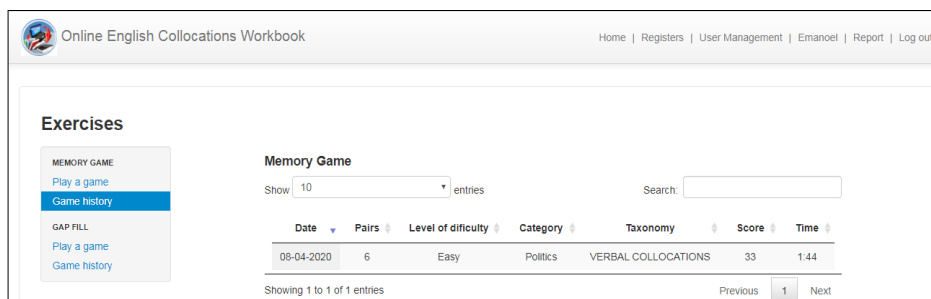


Fonte: captura de tela da *Online English Collocations Platform*.

Dos itens à esquerda, destacamos a primeira parte, em que é possível observar que se optou por explorar e estimular o uso da colocação verbal, inserida na categoria *Politics*, cujo nível de dificuldade selecionado é *Easy*, ou seja, de fácil compreensão.

Durante o jogo e ao final da atividade, o estudante perceberá cada par de cartas e, também, a indicação de seus acertos. Por fim, ao clicar no botão *Finish*, o jogador será direcionado à tela final com o resultado. Além dos itens relacionados, a interface apresenta o *Score* (“pontuação”) e o *Time* (“tempo”), conforme pode ser observado a seguir:

Figura 29 – Resultado do exercício *Memory Game*



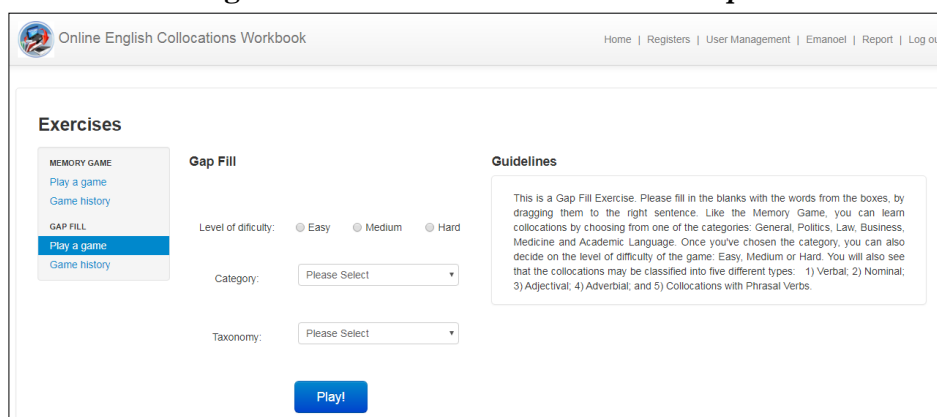
Fonte: captura de tela da *Online English Collocations Platform*.

Portanto, observou-se que a atividade de *Memory Game* acontece de modo interativo, variando os pares de colocações inseridos na ferramenta, de acordo com a categoria e nível de dificuldades escolhidos. Recomenda-se este tipo de exercício, que também pode ser

classificado como um jogo colocacional, principalmente, pelo fato de contar com um conteúdo sobre o uso das colocações a partir de uma testagem empírica, como realizada nesta pesquisa. A seguir, discutimos uma segunda proposta pedagógica a partir do uso do *Gap Fill*, que traz a dinâmica de preenchimento dos espaços.

O referido recurso tem, a princípio, a mesma interface que a *Memory Game* na *Online English Collocations Platform*, porém, traz algumas diferenças nas instruções. Segundo ilustra a figura abaixo, selecionamos as opções dadas nas três ferramentas da tela inicial: *Pairs*, *Level of difficulty*, *Category* e *Taxonomy*. Diferente da primeira, esta não possui o item *Pairs*.

Figura 30 – Tela inicial do exercício *Gap Fill*



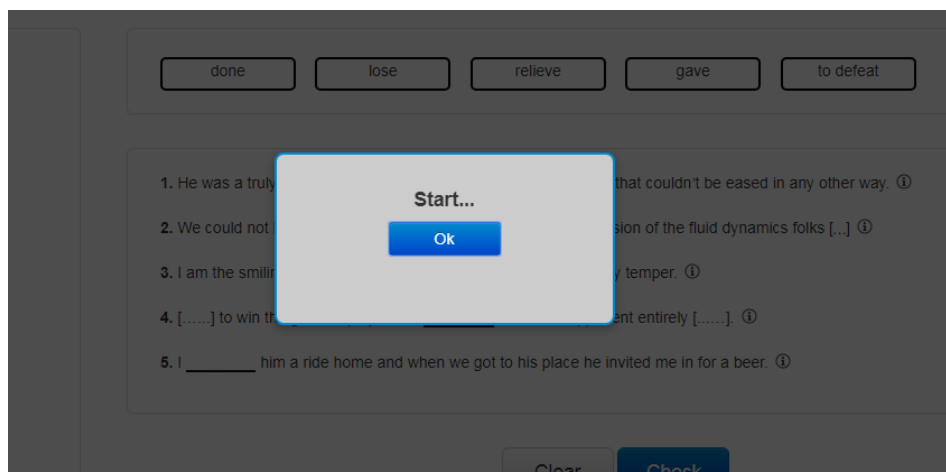
Fonte: captura de tela da *Online English Collocations Platform*.

Logo, percebe-se que, assim como no modelo anterior, este exercício de *Gap Fill* também oferece ao usuário algumas orientações de uso:

- No ícone *EXERCISES* à esquerda, clique na atividade *Play Game* no item *Gap Fill*;
- Em seguida, como apresentado no meio da imagem, selecione os itens *Level of difficulty*, *Category* e *Taxonomy*;
- Clique na tecla *Play*.

Depois de selecionar os campos acima, o estudante será remetido para uma primeira tela e, então, para dar início à atividade, deverá clicar em *Ok*. Diferente do *Memory Game*, observa-se que este tipo de exercício não conta com uma tela de espera.

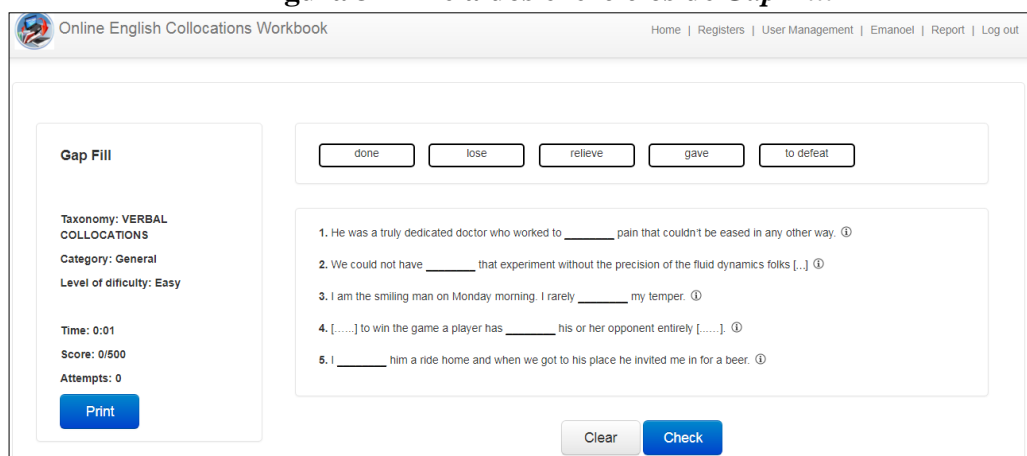
Figura 31 – Tela inicial de início exercício *Gap Fill*



Fonte: captura de tela da *Online English Collocations Platform*.

Logo depois, temos a apresentação da tela inicial do *Gap Fill*, em que os alunos podem arrastar as palavras acima das cinco frases, conforme mostra a figura abaixo:

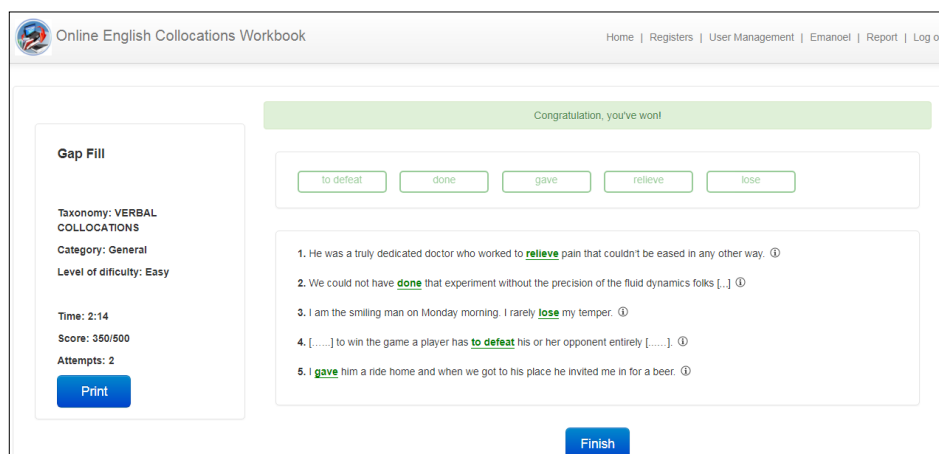
Figura 32 – Tela dos exercícios de *Gap Fill*



Fonte: captura de tela da *Online English Collocations Platform*.

Depois de clicar em *Check*, o discente poderá conferir as respostas da atividade realizada.

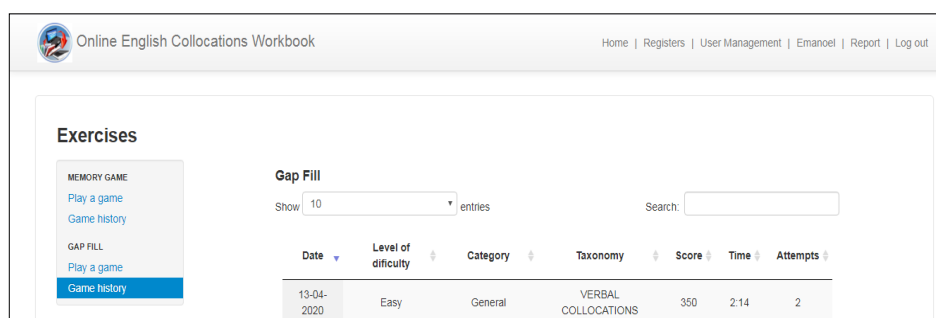
Figura 33 – Tela inicial de início exercício *Gap Fill*



Fonte: captura de tela da *Online English Collocations Platform*.

Finalmente, após selecionar a opção *Finish*, aparecerá uma tela com o resultado e a pontuação alcançada pelo usuário, conforme imagem abaixo:

Figura 34 – Resultado do exercício *Gap Fill*



Fonte: captura de tela da *Online English Collocations Platform*.

Desse modo, notamos uma grande vantagem neste exercício com base no modelo de *Gap Fill*, devido à oportunidade de aprender a usar as colocações, a partir de um contexto do uso, que consiste na utilização de linhas de concordância. Ressaltamos que este é um aspecto essencial nos estudos com LC, servindo, ainda, de ferramenta para o desenvolvimento de atividades desse tipo.

Do ponto de vista pedagógico, os modelos explorados neste subcapítulo se destacam como recursos didáticos, uma vez que conduzem a uma aprendizagem significativa do emprego das colocações. As duas propostas oferecidas pela *Online English Collocations Platform* apresentam interfaces interativas, de fácil acesso e úteis, sempre prezando por exemplos de terminologias extraídas de seu contexto real do ponto de vista da linguagem, como propõe os estudos LC e da Fraseologia com base em *corpus*.

Assim, a metodologia adotada por esta plataforma, em parceria com a área Computacional, mostra-se relevante para a exploração do léxico na comunidade científica e, também, útil para o aprimoramento da aprendizagem do uso das colocações, tanto no âmbito da linguagem específica quanto da língua geral. Nos próximos subcapítulos apresentamos outro modelo, que utiliza *corpus* de referência como recurso de pesquisa e de análise de coocorrências.

5.2 Proposta de atividade baseada no *corpus* de referência *English Web 2013*

Considerando as atividades sugeridas anteriormente, notamos a viabilidade mostrar mais possibilidades ainda com base nos dados analisados neste estudo.

No caso de tradutores aprendizes, indicamos que seja feita uma introdução ao conceito de Linguística de *Corpus* e uma explicação sobre a utilidade do *corpus* em si como um método de apoio a tradução. Dessa forma, o professor pode ainda apresentar a definição da Fraseologia, enquanto disciplina e partes da língua com implicações para as colocações, uma vez pautado pela teoria desde trabalho.

Após esta explanação inicial, sugerimos uma próxima atividade com base na proposta de Frankenberg-Garcia (2012), que por meio das linhas de concordância de *corpus* de referência disponíveis ou pagos os estudantes podem aprender a identificar a coocorrência a partir de palavras-chave. Esta tarefa pode ser facilmente realizada em aulas de “Prática de Tradução”, que comumente acontecem em laboratório de informática.

Para tanto, sugerimos o uso do *corpus* de referência *English Web 2013*, por se tratar de um recurso constantemente atualizado e de grande proporção, sendo ainda de fácil manuseio por meio do *Sketch Engine*, que tem permissão gratuita de uso por até 30 dias.

A atividade tem como base a terminologia *immigrant*, analisada nesta pesquisa. A sugestão é usar a ferramenta *Word Sketch* para pesquisar e identificar os colocados aliados ao referido termo, pois esta oferece indicações detalhadas, permitindo que o usuário se conheça as relações gramaticais em torno do item pesquisado, bem como dados estatísticos. Vejamos a imagem:

Figura 35 – Coocorrências com a palavra *Immigrants* no *Word Sketch*

The screenshot shows the Word Sketch interface for the word "immigrant" in the English Web 2013 corpus. It displays various grammatical categories and their associated word counts and frequencies.

Category	Word	Count	Freq
modifiers of "immigrant"	illegal	43,252	11.06
	undocumented	15,084	10.89
	irish	3,251	7.61
	mexican	2,648	7.5
latino	1,344	7.21	
nouns modified by "immigrant"	visa	4,152	8.7
	petition	1,226	7.5
	population	5,476	7.12
	worker	5,798	6.81
labor	1,018	6.57	
verbs with "immigrant" as object	deport	1,242	8.62
	detain	913	7.75
	arrive	1,455	6.95
	legalize	454	6.76
	smuggle	221	5.99
	assimilate	209	5.94

Fonte: captura de tela do *corpus Sketch Engine*.

A imagem da *Word Sketch* traz diversas combinações com *immigrants*. Em exemplo prático, o professor pode indicar aos alunos que cliquem em *immigrant visa*, que resultará nas linhas de concordância:

Figura 36 – Linhas de concordância da colocação *immigrant visa*

The screenshot shows the Concordance interface for the search query "cql immigrant + visa" in the English Web 2013 corpus. It displays a list of 17 search results with their respective contexts.

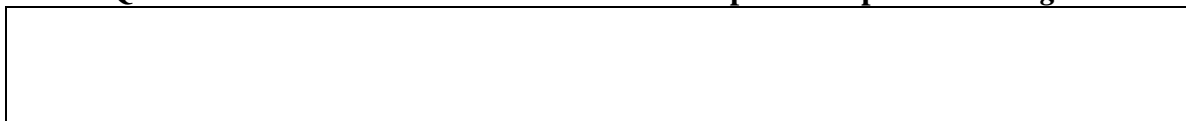
Line	Context
1	requirements Permanent resident visas are listed as non-quota immigrant visas on the consulate websites, so searching for the details can be a
2	an be a pain. You can find the the requirements for non-quota immigrant visas at the Los Angeles consulate's website. To break it down
3	ant Visa I can assist someone with consular processing for an immigrant visa . Consular processing refers to the steps one takes to obt
4	Consular processing refers to the steps one takes to obtain an immigrant visa (TV) at a United States consular post abroad. An alien is
5	United States consular post abroad. An alien is eligible for an immigrant visa if he or she satisfies certain requirements imposed on an applican
6	ligibility for an Immigrant Visa An alien is eligible to receive an immigrant visa if. The alien is the beneficiary of an approved visa petition
7	Numerical Control and Priority Dates The allocation of immigrant visas is controlled by a system of worldwide numerical limitations, base
8	isdiction With few exceptions, an alien who is applying for an immigrant visa must file his or her application at the consular office which has juri
9	or labor certification that indicates that the Beneficiary will apply for the immigrant visa abroad, the USCIS forwards the approved petition or labor certifi
10	t of Homeland Security has made in its efforts to reduce the backlog of immigrant visas . Although almost 863,000 records were "closed" in the I
11	Numerical levels and categories for high-skilled nonimmigrant and immigrant visas should be responsive to economic need and, where appropriate,
12	of these small groups of migrants. Neither bill actually creates immigrant visas , but both deal in long-term, easily renewable nonimmigrant visas
13	the brother of former Secretary of State Hillary Rodham Clinton get an immigrant visa for a Chinese investor. – Washington Times President Ob
14	ssified here. The Diversity Program is usually allotted 50,000 immigrant visas , with a few thousand reserved for other purposes than the lottery,
15	to rush to your nearest US Embassy to claim your right to apply for the immigrant visa Another thing you should be aware of, is that you will no
16	u to have a so-called "visa sponsor" in order to apply. The US immigrant visas are one of those. If you, for example, get married in the I
17	isa sponsor. With that in hand, you can go to apply for the US immigrant visa . Register for the lottery Wait If you get s

Fonte: captura de tela do *Sketch Engine*.

Depois desse percurso de pesquisa, os discentes podem acessar novamente a tela da *Word Sketch*, que contém a palavra *immigrant* e elencar, no quadro a seguir, outras coocorrências. O objetivo é aproximar o aprendiz à linguagem do *corpus*, tomando como base as relações gramaticais em torno da palavra-chave para o desenvolvimento da sua autonomia

(JOHNS, 1991; BOULTON, 2009, 2015).

Quadro 48 – Resumo das UFs identificadas a partir da palavra *immigrant*



Fonte: elaborado pelo autor.

Para essa atividade, o professor pode sugerir que os estudantes também selecionem as coocorrências com a palavra de busca, dividindo-as de acordo com os itens modificadores dispostos na tela da *Word Sketch* na imagem 35.

O fato de os alunos já estarem familiarizados com a *Word Sketch* acaba facilitando a expansão da tarefa. Então, sugerimos que o docente apresente o projeto CAT 2 e, em seguida, explique a função do *LogDice* a partir do *Sketch Engine*.

Depois da explanação teórica, indicamos uma dinâmica pedagógica, por meio da colocação “fazer segurança”, como modelo baseado em *corpus* de aprendizes.

Inicialmente, o educador pode apresentar o contexto da colocação, bem como perguntar como cada aprendiz traduziu o termo para a língua inglesa, a partir do exemplo a seguir:

Quadro 49 – Contexto para tradução

“Para evitar o risco de confrontos, centenas de policiais foram mobilizados para ‘fazer a segurança’ do protesto, que acabou sem incidentes”.

Fonte: elaborado pelo autor.

Após esta etapa inicial de discussão, indica-se a disposição das opções tradutórias sugeridas no *subcorpus* de TTs do CAT 2. É, também, o momento de refletir em grupo sobre quais foram as impressões que todos tiveram a partir da imagem proposta.

Figura 37 – Verbos com “segurança”

verbs with "security" as object			
make	15	13.02	...
do	2	9.08	...

Fonte: captura de tela do *Sketch Engine*.

Depois desse debate sobre o emprego combinatório de *make/do security*, sugerimos

questioná-los para perceberem se imaginam que estes verbos aparecem como colocados da base na palavra *security* no *English Web 2013*. Diante disso, eles podem acessar a *Word Sketch* a fim de comprovar se esse uso procede ou não. Como já relatado na etapa de análise de dados, os discentes descobrirão que *do/make security* não se combinam dessa maneira, sendo possível dialogar a respeito da questão de ingenuidade do tradutor (TAGNIN, 2002) neste momento.

Conforme os estudantes verificam a não procedência combinatória destes itens lexicais, o educador pode conduzi-los durante o manuseio da *Word Sketch*, a fim de verificarem quais verbos serviriam de sugestão.

Figura 38 – Sugestão de colocado para a base “security”

verbs with "security" as object			
ensure	<u>19,410</u>	8.24	...
back	<u>5,250</u>	7.67	...
mortgage backed securities			
compromise	<u>4,233</u>	7.67	...
threaten	<u>4,525</u>	7.5	...
enhance	<u>8,383</u>	7.4	...
guarantee	<u>5,732</u>	7.38	...
improve	<u>17,186</u>	7.18	...
strengthen	<u>3,578</u>	6.84	...
trade	<u>2,503</u>	6.71	...
traded securities			
provide	<u>37,058</u>	6.63	...

Fonte: captura de tela do *Sketch Engine*.

Uma vez identificadas as colocações para “fazer segurança”, o docente pode mostrar a aplicabilidade do *LogDice*, propor o questionamento de acordo com este item estatístico e discorrer sobre a melhor opção para o verbo “fazer” em inglês no contexto dado. Na próxima seção abordaremos outra ferramenta, propondo uma nova atividade didática com base em *corpus*.

5.3 Proposta de atividade baseada na ferramenta *SKELL*

A proposta também tem como sugestão o uso da *SKELL* para analisar as coocorrências geradas nesta plataforma a partir de palavras-chaves (ALVES; CECÍLIO; ORENHA-OTTAIANO, 2020). Como já apresentada a sua interface inicial no Capítulo 3 deste estudo, a ferramenta possui três abas de trabalho, que são: *Examples*, *Word Sketch* e *Similar Words*.

Na sequência, a título de ilustração, temos os resultados de busca a partir de *border*, permitindo um acesso mais rápido à palavra de busca. Entretanto, diferente do *English Web 2013*, a *SKELL* não conta com recursos que conduzem a uma interpretação mais detalhada das coocorrências, conforme figura a seguir.

Figura 39 – Linhas de concordância a palavra-chave *border*

SKELL borders [Examples](#) [Word sketch](#) [Similar words](#)

borders 19.07 hits per million

- 1 The same speaker might offer different **borders** .
- 2 Money recognizes neither geographical nor political **borders** .
- 3 Does anyone seriously advocate completely open **borders** ?
- 4 Developing trust across traditional **borders** makes possible future business negotiation.
- 5 Most illegal drug seizures occur along those **borders** .
- 6 A number of bilateral overlapping maritime **borders** were resolved.
- 7 Our **borders** remain open to illegal immigration.
- 8 The question of **borders** is always difficult.
- 9 This is especially true of marketing activities across geographical **borders** .
- 10 This principle justifies intervention across **borders** to rectify perceived violations.
- 11 Sometimes body parts are even shipped across **borders** .
- 12 Many girls and women are illegally trafficked across **borders** .
- 13 The alliance victory defined **borders** between the neighboring states.
- 14 His decorations include nature imagery and grotesque **borders** .
- 15 The national territory was divided along the existing internal **borders** .
- 16 Private resorts are available outside the park **borders** .
- 17 It had four lanes running 109 km along the city **borders** .
- 18 Their **borders** were meaningless and arbitrarily drawn.
- 19 Dutch dialects are varied and cut across national **borders** .

Fonte: captura de tela da plataforma *SKELL*.

A partir desta pesquisa, logo, observa-se alguns exemplos de coocorrência de termos com a busca por *border*. Em seguida, na segunda aba, a *Word Sketch*, encontramos exemplos de relações gramaticais. Essa segunda ferramenta da *SKELL* é parecida com a *Word Sketch*, que está inserida dentro do *Sketch Engine*, porém, com algumas limitações e com menor número de itens estatísticos de controle do usuário (frequência, *LogDice*, informações dos textos do *corpus* etc.).

Destacamos que a *Word Sketch* é construída com base nos mesmos critérios dos *corpora* da *Web*, adotados pelo *Sketch Engine*. Sendo assim, por meio da *Word Sketch*, os alunos podem explorar as relações gramaticais apresentadas para *border*, permitindo-lhes ainda observar o contexto do item lexical pesquisado, como mostra a Figura 41.

Figura 40 – Relações gramaticais com a palavra-chave *border*

The screenshot shows the SKELL platform interface for the word "border". At the top, there is a search bar with "border" entered and a magnifying glass icon. To the right of the search bar are links for "Examples", "Word sketch" (highlighted in red), "Similar words", "More features", and "More languages". Below the search bar, the word "border" is displayed with a "noun" tag and a "switch to border (verb)" option. A "Context" button with a circular arrow icon is also present. The main content is organized into several sections, each with a heading and a list of related words in a light grey box:

- verbs with border as subject:** cross, separate, guard, divide, surround, control, extend, close, run, lie, remain, open, change, follow, become
- verbs with border as object:** cross, straddle, patrol, guard, transcend, secure, seal, demarcate, share, close, defend, define, mark, form, delineate
- adjectives with border:** secure, open, due, such
- modifiers of border:** southern, eastern, northern, western, Mexican, Mexico, Canadian, porous, Syrian, land, Turkish, international, Afghan, Swiss, maritime
- nouns modified by border:** crossing, guard, dispute, security, patrol, fence, coltie, checkpoint, town, skirmish, region, clash, adjustment, post, control
- words and:** border, boundary, mile, security, territory, line, area, people
- or border:** border, boundary, bed, River, north, land, line, state

Fonte: captura de tela da plataforma SKELL.

Figura 41 – Relações gramaticais com a palavra-chave *border*

This screenshot provides a more detailed view of the SKELL platform's "Word sketch" for "border". It features the same search bar and navigation links as Figure 40. The word "border" is shown with "noun" and "switch to border (verb)" options, along with a "Context" button. The sections and their examples are as follows:

- verbs with border as subject:** border crossing, border separating, the border guards, border dividing, border surrounding, no border controls, border follows, border became
- verbs with border as object:** crossed the border, straddles the border, patrolling the border, guarding the border, transcend borders, secure t, defined borders, marks the border, forms the border, delineated the border
- adjectives with border:** the border is secure, borders are open, border due to, border such as
- modifiers of border:** the southern border, the eastern border, the northern border, the western border, the Mexican border, the U.S., the international border, the Afghan border, the Swiss border, maritime border
- nouns modified by border:** border crossings, border guards, border dispute, border security, border patrol, the border fence, border coltie, border post, border controls
- words and:** border , the border, borders and boundaries, border , a few miles, borders and security, borders and territories
- or border:** border , the border, boundaries and borders, beds and borders, River , the border, north and south borders, land

Fonte: captura de tela da plataforma SKELL.

Sendo assim, esta e as outras atividades podem ser aplicadas nos modelos tanto na modalidade presencial quanto *on-line*, por meio do ensino remoto e da Educação à Distância (EAD). Cabe ao educador aplicar a proposta da melhor forma, de acordo com sua realidade.

Também, salientamos que não foi nosso objetivo esgotar todas as possibilidades pedagógicas ao utilizar a metodologia de análise linguística a partir do *corpus* de aprendizes desta pesquisa. Certos de que outras contribuições vindouras surgirão, indicamos apenas modelos de como aplicar os dados da análise deste estudo para formação de tradutores, servindo, é claro, para outras situações de ensino.

Diante do exposto, nas duas partes deste capítulo, foram apresentadas algumas sugestões ao aprendiz de tradução quanto ao uso das colocações e as formas de identificar os padrões combinatórios. As estratégias discutidas no item 5.1, com a *Online English Collocations Platform*, colocam o discente em um contato inicial com as colocações, sem se aprofundar no conceito teórico delas à luz da LC. Em seguida, no item 5.2 e 5.3, destacamos as estratégias de como desenvolver a autonomia do estudante, a fim de que este aprenda a observar parâmetros, de modo que alcance a competência léxico-fraseológica necessária para sua prática tradutória futura.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa, inserida no projeto de pesquisa “Compilação de Materiais Didáticos e Dicionários Especializados de Colocações Baseados em *Corpora*” (COMADEC), teve como objetivo investigar como os aprendizes de tradução traduziram as colocações presentes nos textos que compõem o *Corpus* de Aprendizes de Tradução (CAT 2) e, assim, propor atividades colocacionais com base nas dificuldades discutidas em relação ao emprego deste tipo de UF. Esse processo ocorreu por meio da exploração e análise das colocações à luz da LC, que seguiu os mesmos parâmetros e metodologia de compilação e de estudos já realizados a partir do CAT 1 (ORENHA-OTTAIANO, 2012b, no prelo).

Para a aplicação do trabalho, por se tratar da utilização de humanos como sujeitos desta pesquisa, contamos com o apoio do Comitê de Ética em Pesquisa, da Universidade Filho de Mesquita Filho (UNESP), de São José do Rio Preto, a fim de obter a validação inicial da proposta e que nos foi permitido sob a concessão do Certificado de Apresentação de Apreciação Ética (CAAE), de número 2624318.9.0000.5466.

A fundamentação teórica abarcou temáticas sobre a aplicabilidade da LC e suas implicações nos Estudos da Tradução Baseados em *Corpus* e no ensino de LE para tradutores, bem como sobre as questões da convencionalidade do léxico, no que diz respeito à Fraseologia, suas características na língua geral e na língua de especialidade.

Nesse ponto, contamos com teorias que apresentaram um suporte tecnológico, podendo discorrer sobre como a fraseologia contemporânea, por meio das ferramentas de *corpus*, tem se ocupado dos estudos das UFs. Dentro de todo esse aparato teórico, discorreremos sobre questões referentes às colocações, como um processo arbitrário e restrito ao idioma, devido às convenções. De modo geral, seguimos os pressupostos teóricos da Linguística de *Corpus* (MCENERY; HARDIE, 2012; MEYER, 2004) e da Fraseologia (ORENHA-OTTAIANO, 2004, 2009, 2012a, 2012b, 2015, no prelo; MARTELLI, 2007; TAGNIN, 2013), para observação das colocações como dados linguísticos de forma computadorizada.

Em seguida, na metodologia, adotamos os critérios de seleção dos participantes, suas características, bem como aquelas da própria instituição *locus* da pesquisa, com atenção para o Curso de Letras – Tradutor. O CAT 2 contemplou a participação dos alunos do Curso de Letras – Tradutor, da Universidade Sagrado Coração (UNISAGRADO). Para participação efetiva na investigação, os estudantes foram submetidos a um teste de proficiência, sendo que

todos os participantes se classificaram entre os níveis B2 e C2, os quais tiveram suas traduções inseridas no CAT 2.

Em nossa discussão e análise de dados, foi possível verificar o comportamento léxico-fraseológico com foco nas colocações, assim como apontar contribuições relevantes com base em *corpus*. Quanto ao CAT 2, especificamente, comprovamos que o uso de um *corpus* de aprendizes pode servir um norteador pedagógico, o qual nos permitiu identificar dificuldades e problemas tradutórios no uso das colocações e das terminologias, bem como sugerir outras opções tradutórias baseadas em resultados contextuais e estatísticos promovidos pelo uso de *corpus* e ferramentas de análise lexical. Ademais, com base nas dificuldades identificadas durante a análise, propusemos atividades on-line e baseadas em *corpora* por meio das palavras-chave geradas pelos TTs, em comparação com seus respectivos TOs, que formaram o CAT 2.

Desse modo, ficou comprovado, por meio dos exemplos discutidos nesta pesquisa, que os aprendizes de tradução apresentam dificuldades no emprego de colocações. Para corroborar esta afirmação, no primeiro exemplo, elencamos o termo “arrumar confusão”, cujas várias traduções apenas duas delas se aproximaram do original. Porém, foi necessária, e devidamente comprovada por meio do *corpus* de referência, indicar a colocação *create a scene*, de forma que atendesse totalmente ao contexto principal. Notamos que realizar investigações como esta, no âmbito universitário, pode auxiliar de forma significativa que graduandos de curso de tradução, o público-alvo nesta dissertação, apresentem textos com repertório lexical de maneira mais fluente na língua de chegada.

Na etapa de indicações pedagógicas, realizamos uma proposta de atividade baseada nos possíveis erros colocacionais identificados no CAT 2, a partir do mesmo *corpus* de referência *English Web 2013*, que foi utilizado para comprovar a frequência de coocorrência das colocações. As sugestões elaboradas tiveram como base a *Online English Collocations Platform* e *corpora* de referência disponíveis, a partir do gerenciador de *corpus Sketch Engine* (enquanto ferramentas pedagógicas), teoricamente embasadas segundo as leituras percorridas nesta pesquisa, conforme apresentamos na fundamentação teórica nesta investigação.

Justificamos a efetividade deste projeto de mestrado, considerando os resultados alcançados, o público-alvo que não tem, a priori, contato direto com a Linguística de *Corpus* como eixo de pesquisa por meio de Programas de Pós-graduação *Stricto Sensu*. Por esta razão, indicamos a continuidade desse modelo de investigação nas atividades comuns dos cursos de graduação.

A partir dos dados já observados, percebemos que este trabalho apresentou relevantes

contribuições para a Linguística Aplicada por seu caráter empírico, tanto pelas vias teóricas da LC, como, também, por envolver a participação de alunos em uma área tão antiga como a Tradução. Sabemos que o trabalho com o ser humano demanda tempo, principalmente, por envolver uma série de burocracias para obter aprovação legal, porém, temos notado que este tipo de projeto permitiu uma interação sincrônica entre análise de dados e o grupo de estudantes.

Portanto, em possibilidades futuras, temos por certo de que nossa experiência, por meio deste estudo, abriu, e abrirá, caminhos para novas reflexões sobre o uso da LC com o propósito de indicar melhorias nas questões do ensino da tradução de colocações, relacionadas ao desenvolvimento da competência léxico-fraseológica do aprendiz de tradução.

REFERÊNCIAS

- ALEXANDER, R. J. Fixed expressions in English: a linguistic, psycholinguistic, sociolinguistic and didactic study (Part I). **Anglistik und Englischunterricht**, Heidelberg, v. 5, n. 6, p. 171-188, 1978.
- ALLAN, D. **Oxford placement tests 2**. Oxford: OUP, 2004.
- ALTENBERG, B. On the phraseology of spoken English: the evidence of recurrent word-combinations. In: COWIE, A. P. (ed.). **Phraseology: theory, analysis, and applications**. Oxford: Oxford University Press, 1998. p. 101-122.
- ALVES, E. H.; CECÍLIO, G. A.; ORENHA-OTTAIANO, A. O uso da ferramenta Skell como recurso pedagógico para a formação de professores e o ensino de língua estrangeira. In: VASCONCELOS, A. W. S. de; VASCONCELOS, T. N. S. de. (org.). **Linguística, Letras e Artes e as Novas Perspectivas dos Saberes Científicos 3**. Ponta Grossa: Atena Editora, 2020. p. 151-166. Disponível em: <https://www.atenaeditora.com.br/post-artigo/43510>. Acesso em: 10 jan. 2021.
- ALVES, E. H.; ORENHA-OTTAIANO, A. Corpus de aprendizes de tradução: uma investigação sobre o emprego de colocações na tradução de textos jornalísticos. **Caletroscópio**, v. 8, p. 14-31, 2020. Disponível em: <https://periodicos.ufop.br:8082/pp/index.php/caletroscopio/article/view/4487>. Acesso em: 10 jan. 2021.
- ANDERSON, W.; CORBETT, J. **Exploring English with Online Corpora**. Basingstoke: Palgrave MacMillan, 2009.
- AMADEU-SABINO, M. O campo árido dos fraseologismos. **Signotica** (UFG), Goiânia, v. 23, n. 2, p. 385-401, 2011.
- AMOSOVA, N. N. **Osnovui anglijskoy frazeologii**. Leningrad: University Press, 1963.
- ANTHONY, L. **AntConc 3.4.4w**. 2014.
- ARAGÃO, M. O Brasil e o mundo. **O Estado de São Paulo**, São Paulo, 6 mar. 2018. Disponível em: <https://opinioao.estadao.com.br/noticias/geral,o-brasil-no-mundo,70002290022>. Acesso em: 20 jun. 2018.
- ASTON, G. Corpus use and learning to translate. **Textus**, [Bologna], n. 12, p. 289-314, 1999. Disponível em: <https://www.sslmit.unibo.it/~guy/textus.htm>. Acesso em: 20 jun. 2019.
- AZEVEDO, L. F. Deslizes Linguísticos em colocações convencionais, 1986 (manuscrito).
- AZEVEDO, D. J. O.; SILVA, F. M. Colocações, estereótipos e clichês: definições e diferenças. **ReVEL**, Novo Hamburgo, v. 15, n. 29, p. 37-52, 2017. Disponível em: <http://www.revel.inf.br/files/b4f9129fed69407b4ec7c6e747aa1c5f.pdf>. Acesso em: 11 abr. 2019.

BAISA, V.; SUCHOMEL, V. SkELL – Web Interface for English Language Learning. In: **Eighth Workshop on Recent Advances in Slavonic Natural Language Processing**. Brno: Tribun EU, 2014, p. 63-70. ISSN 2336-4289.

BAKER, M. Corpus linguistics and translation studies: implications and applications. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (org.). **Text and technology**: in honour of John Sinclair. Philadelphia: John Benjamins, 1993, p. 233-250.

BALLY, C. **Traité de stylistique française**. Paris: Klincksieck, v. 1, 1951.

BARLOW, M. Parallel texts in language teaching. In: BOTLEY, S. P.; McENERY, A. M.; WILSON, A. (ed.). **Multilingual Corpora in Teaching and Research**. Amsterdam: Rodopi, 2000. 106-115.

BARROS, L. A. **Curso básico de terminologia**. São Paulo: Edusp, 2004.

BENNET, G. R. **Using Corpora in the Language Learning Classroom**: corpus linguistics for teachers. Michigan: University of Michigan Press, 2010.

BENSON, M. Collocations and idioms. In: ILSON, R. (ed.). **Dictionaries, lexicography and language learning**. Oxford: Pergamon, 1995. p. 61-68.

BERBER SARDINHA, T. Corpora eletrônicos na pesquisa em tradução. **Cadernos de Tradução**, Florianópolis, v. 1, n. 9, p. 15-59, 2002.

BERBER SARDINHA, T. Resenha de lexicology and corpus linguistics: an introduction. **DELTA**, São Paulo, v. 25, n. 2, p. 497-510, 2009.

BERBER SARDINHA, T.; ACUNZO, C. M.; FERREIRA, T. L. S. B. Metáforas da economia no dicionário de colocações do português brasileiro: Uma análise multidimensional baseada em corpus. **Filologia e Linguística Portuguesa**, São Paulo, v. 18, p. 175-198, 2016. Disponível em: <https://www.revistas.usp.br/flp/article/view/107209>. Acesso em: 11 abr. 2020.

BERNARDINI, S. Corpora for translator education and translation practice: achievements and challenges. In: YUSTE RODRIGO, E. (ed.). **Proceedings of the Third International Workshop on Language Resources for Translation Work, Research & Training**. Paris, ELRA (European Language Resources Association, LR4Trans-III, 2004.

BERNARDINI, S. Educating translators for the challenges of the new millenium: The potencial of parralel bi-directional corpora. In: haller, J.; MAIA, B.; ULRICH, M. (ed.) **Training the lanhuage servecies provider for the new millenim**. Porto: Universidade do Porto, 2002. P. 173-186.

BEVILACQUA, C. R. **A fraseologia jurídico-ambiental**. 1996. 121 f. Dissertação (Mestrado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1996.

BEVILACQUA, C. R. Fraseologia: perspectiva da língua comum e da língua especializada, **Revista Língua e Literatura**, Frederico Westphalen, v. 7, n. 10/11, p. 73-86, 2004/05. Disponível em: <http://revistas.fw.uri.br/index.php/revistalinguaeliteratura/article/view/40>. Acesso em: 11 abr. 2019.

BIBER, D. *et al.* **Longman Grammar of Spoken and Written English**. Harlow: Longman. 1999.

BIBER, D. **Dimensions of register variation: a cross linguistic comparison**. Cambridge: Cambridge University Press, 1995.

BIDERMAN, Maria Tereza C. As ciências do léxico. *In*: OLIVEIRA, Ana Maria P. P; ISQUERDO, Aparecida N. (org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. Campo Grande: Ed. UFMS, 2001. v. 1, p. 153-166.

BLAIS, E. Le phraséologisme. Une hypothèse de travail. **Terminologies Nouvelles**, Brussels, v. 10, RINT, p. 50-56, 1993.

BOLINGER, D. L. Meaning and memory. **Forum Linguisticum**, Lake Bluff, v. 1, p. 1-14, 1976.

BOULTON, A. Applying data-driven learning to the web. *In*: LEŃKO-SZYMAŃSKA, A.; BOULTON, A. (Eds.). **Multiple Affordances of Language Corpora for Data-driven Learning**. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015. p. 1-14.

BOULTON, A. Data-driven learning: reasonable fears and rational reassurance. **Indian Journal of Applied Linguistics**, v. 35, n. 1, p. 81-106, 2009.

BOURIGAULT, D.; JACQUEMIN, C.; L'HOMME, M. **Recent advances in computational terminology**. Amsterdam: John Benjamins Publishing, 2001.

BOWKER, L. Towards a methodology for exploiting specialized target language corpora as translation resources. **International Journal of Corpus Linguistics**, Amsterdam, v. 5, n. 1, p.17-52, ago. 2000. Disponível em: https://www.researchgate.net/publication/233565055_Towards_a_Methodology_for_Exploiting_Specialized_Target_Language_Corpora_as_Translation_Resources. Acesso em: 5 abr. 2019.

BRESSAN, I. I. F.; ORENHA-OTTAIANO, A. Colocações extraídas de um *corpus* de aprendizes de língua escrita. *In*: Congresso de Iniciação Científica UNESP, 25 ed., 2013. Barra Bonita. **Resumos** [...] São Paulo: UNESP, 2013.

BRITISH NATIONAL CORPUS. Extra downloads for WordSmith Tools. **Lexically**. Disponível em: <http://www.lexically.net/wordsmith/support/extras.html>. Acesso em: 12 nov. 2011.

BRITO, Y. S. **Levantamento das colocações verbais presentes em um corpus de aprendizes de tradução**. 2012. Trabalho de conclusão (Letras com Habilitação em Tradutor) - Universidade Estadual Paulista Júlio de Mesquita Filho, São José do Rio Preto, 2012.

BULLA, B. Plano de deportação de Trump deve jogar 300 mil imigrantes na ilegalidade. **O Estado de São Paulo**, São Paulo, 2 set. 2018. Disponível em: <https://internacional.estadao.com.br/noticias/geral,plano-de-deportacao-de-trump-deve-jogar-300-mil-imigrantes-na-ilegalidade,70002483471>. Acesso em: 11 set. 2018.

BURGER, H. *et al.* **Phraseologie/Phraseology**. Ein internationales handbuch zeitgenössischer forschung/An international handbook of contemporary research. Berlin: De Gruyter, v. 2, 2007.

BURGER, H. **Phraseologie**: eine einföhrung am beispiel des deutschen. Berlin: Erich Schmidt, 1998.

CABRÉ, M. T. Hacia una teoría comunicativa de la terminología: aspectos metodológicos. *In*: CABRÉ, M. T. **La terminología**: representación y comunicación. Barcelona: IULA, 1999. p. 109-127.

CABRÉ, M. T. La teoría comunicativa de la terminología, una aproximación lingüística a los términos. **Revue Française de Linguistique Appliquée**, Paris, v. 14, p. 9-15, 2009.

CABRÉ, M. T. **La terminología**: teoria, metodologia, aplicaciones. Barcelona: Editorial Antártida: Empúries, 1993.

CALDAS, A. D. D. R. **A identificação de colocações especializadas extraídas do Corpus CSI e do Corpus Comparável Criminal para a elaboração de atividades didáticas**. 2017. 139 f. (Dissertação Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2017.

CAMARGO, D. C. **Metodologia de pesquisa em tradução e linguística de corpus**. São Paulo/São José do Rio Preto: Cultura Acadêmica/Laboratório Editorial do IBILCE, UNESP, 2007, v. 1. 65 p.

CAMARGO, D. C. **Padrões de Estilo de Tradutores: Um estudo de semelhanças e diferenças em corpora de traduções literárias, especializadas e juramentadas**. 2005. Tese (Livre Docência em Estudos da Tradução) - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista (UNESP), São José do Rio Preto, 2005.

CARNEADO MORÉ, Z.; TRISTÁ, A. M. **Estudios de fraseologia**. Havana: Academia de Ciências de Cuba: Instituto de Literatura y Lingüística, 1983.

CARRIÓ-PASTOR, M. L. Phraseology in Specialised Language: A Contrastive Analysis of Mitigation in Academic Papers. *In*: CORPAS PASTOR G.; MITKOV, R. (ed.). Computational and Corpus-Based Phraseology. Lecture Notes in Computer Science (LNCS), **EUROPHRAS**, v. 11755, Springer, 2019. Disponível em: https://doi.org/10.1007/978-3-030-30135-4_5. Acesso em: 5 abr. 2019.

CARTER, R. **Vocabulary**: applied linguistics perspectives. London: Unwin Hyman, 1987.
CARTER, R.; MCCARTHY, M. J. (org.). **Vocabulary and language teaching**. London: Longman, 1988.

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

CHOMSKY, N. **The Minimalist Program**. Cambridge, MA: MIT Press, 1995.

COLSON, J. P. Cross-linguistic phraseological studies: an overview. *In*: GRANGER, S.; MEUNIER, F. **Phraseology**: an interdisciplinary perspective. Amsterdam: John Benjamins Publishing Company, 2008. p. 191-206.

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

CORPAS PASTOR, G. Translating English Verbal Collocations into Spanish: on Distribution and other Relevant Differences related to Diatopic Variation. **Linguisticae Investigationes**, Amsterdam, v. 38, n. 2, p. 229-262, 2015. DOI: <https://doi.org/10.1075/li.38.2.03cor>. Disponível em: <https://benjamins.com/catalog/li.38.2.03cor>. Acesso em: 11 abr. 2020.

CORPAS PASTOR, G.; ORTIZ ALVAREZ, M. L. Fraseologia e Paremiologia: uma entrevista com Gloria Corpas Pastor. Tradução de Ana Carolina Spinelli. Revisão técnica de Gabriel de Ávila Othero. **ReVEL**, v. 15, n. 29, 2017. Disponível em: www.revel.inf.br. Acesso em: 5 abr. 2019.

COWIE, A. P. The treatment of collocations and idioms in learners' dictionaries. **Applied Linguistics**, Oxford, v. 2, n. 3, p. 223-235, 1981. Disponível em: <https://doi.org/10.1093/applin/II.3.223>. Acesso em: 20 abr. 2020.

COWIE, A. P. Collocational Dictionaries-A Comparative View. *In*: THE ANGLO-SOVIET ENGLISH STUDIES SEMINAR, 22-24 de septiembre 1985, Manchester. **Anais [...]**. Manchester: [s. n.], 1985.

COWIE, A. P. Phraseology. *In*: Asher, R. E. (ed.). **The Encyclopedia of Language and Linguistics**. Oxford: Pergamon, 1994. p. 3168-3171.

COWIE, A. P. (ed.). **Phraseology**: theory, analysis, and applications. Oxford: Oxford University Press, 1998. p. 101-122.

COWIE A. P.; MACKIN, R. *Oxford Dictionary of Current Idiomatic English*. London: Oxford University Press, 1975. v. 1.

CRUSE, D. A. **Lexical semantics**: Cambridge textbooks in linguistics. Cambridge: Cambridge University Press, 1986.

ETTINGER, S. Formación de palabras y fraseología en la lexicografía. *In*: HAENSCH, G. **La lexicografía**: de la lingüística teórica a la lexicografía práctica. Madrid: Gredos, 1982. p. 233-258.

FARROKH, P. Raising awareness of collocation in ESL/EFL classrooms. **Journal of Studies in Education**, Las Vegas, v. 2, n. 3, p. 55-74, 2012.

FIALA, P. Pour une approche discursive de la phraséologie. Remarques en vrac sur la locutionnalité et quelques points de vue qui s'apportent, sans doute. **Langage et Société**, Paris, n. 42, p. 27-44, 1987.

FILLMORE, C. J. Innocence: a second idealization for linguistics. **Berkeley Linguistics Society**, Berkeley, v. 5, p. 63-76, 1979.

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

FLEISCHER, W. **Phraseologie der deutschen gegenwartssprache**. Leipzig: VEB Verlag Enzyklopädie, 1982.

FRANCIS, W. N. Language Corpora B.C. *In*: Svartvik, J. (ed.). **Directions in Corpus Linguistics**. Berlin: De Gruyter Mouton, 1992. p. 17-32. (Trends in Linguistics. Studies and Monographs [TiLSM], 65). Disponível em: <https://www.degruyter.com/document/doi/10.1515/9783110867275.17/html>. 30 jul. 2018.

FRANKENBERG-GARCIA, A. Investigating the collocations available to EAP writers. **Journal of English for Academic Purposes**, Scotland, v. 35, p. 93-104, 2018.

FRANKENBERG-GARCIA, A. Integrating corpora with everyday language teaching. *In*: THOMAS, J.; BOULTON, A. (ed.). **Input, process and product: development in teaching and language corpora**. Brno: Masaryk University Press, 2012. p. 36-53.

FUERTE-OLIVERA, P. A.; TARP, S. **Theory and practice of specialised on-line dictionaries: lexicography versus terminography**. Berlin: Walter de Gruyter, 2014.

GABLASOVA, D.; BREZINA, V.; MCENERY, T. Collocations in corpus-based language learning research: Identifying, comparing and interpreting the evidence. **Language Learning**, Michigan, v. 67, n. 1, p. 155-179, 2017.

GABRIEL, R. S. Trump e Putin são os czares modernos? **Época**, Rio de Janeiro, 26 jul. 2018. Disponível em: <https://epoca.globo.com/trump-putin-sao-os-czares-modernos-22921632>. Acesso em: 30 jul. 2018.

GARCÍA-PAGE SÁNCHEZ, M. Sobre implicaciones 142aiwanese142as: solidaridad léxica y expresión fija. **Estudios Humanísticos Filología**, León, n. 12, p. 2015-227, 1990.

GAVIOLI, L. **Exploring Corpora for ESP Learning**. Amsterdam: John Benjamins, ed. 21, Studies in Corpus Linguistics, 2005.

GILQUIN, G.; GRANGER, S.; PAQUOT, M. Learner corpora: the missing link in EAP pedagogy. **Journal of English for Academic Purposes**, Renfrewshire, v. 6, n. 4, p. 319-335, 2007.

GLÄSER, R. **Phraseologie der englischen sprache**. 2. ed. Leipzig: VEB Verlag Enzyklopädie, 1986a.

GLÄSER, R. Idiomatik und Sprachvergleich. **Sprache und Literatur in Wissenschaft und Unterricht**, Munich, v. 16, p. 67-73, 1985.

GLÄSER, R. The translation aspect of phraseological units in English and German. **Papers and Studies in Contrastive Linguistics**, Poznan, v. 18, p. 123-134, 1984.

GOUADEC, D. Nature et traitement des entités phraséologiques. **Terminologie et phraséologie. Acteurs et aménageurs**. Actes du deuxième Université d'Automne en Terminologie. Paris: La Maison du Dictionnaire, 1994. p. 164-193.

GRANGER, S. The International of Learner English. *In*: AARTS, J.; DE HAAN, P.; OOSTDIJK, N. (ed.). **English language corpora: design, analysis and exploitation**. Amsterdam: Rodopi, 1993. P. 57-69.

GRANGER, S. A bird's eye view of learner corpus research. *In*: GRANGER, S.; HUNG, J.; PETCH-TYSON, S. (ed.). **Computer learner corpora, second language acquisition and foreign language teaching**. Amsterdam: John Benjamins Publishing, 2002. p. 3-33.

GRIES, S. T. Phraseology and linguistic theory: A brief survey. *In*: GRANGER, S. & F. MEUNIER (ed.) **Phraseology: An interdisciplinary perspective**. Amsterdam: John Benjamins.

MEUNIER, F.; GRANGER, S. (ed.). **Phraseology in foreign language learning and teaching**. Amsterdam : John Benjamins, 2008.

GRANGER, S.; PAQUOT, M. Disentangling the phraseological web. *In*: GRANGER, S. & F. MEUNIER (ed.). **Phraseology: an interdisciplinary perspective**. Amsterdam: John Benjamins, 2008. p. 27-50.

GRÉCIANO, G. **Signification et dénotation en allemand** : la sémantique des expressions idiomatiques. Paris : Klincksieck. 1983. (Recherches Linguistiques, 9).

GRÉCIANO, G. Vers une modélisation phraséologique : acquis et projets d'EUROPHRAS. **Terminologies Nouvelles**, Brussels, v. 10, RINT, p. 16-22, 1993.

HAENSCH, G. **La lexicografía**: de la lingüística teórica a la lexicografía práctica. Madrid: Gredos, 1982. p. 233-258.

HALLIDAY, M. A. K. **Lexical relations**: system and function in language: selected papers. Oxford: Oxford University Press, 1961. P. 73-83.

HALLIDAY, M. A. K. Corpus studies and probabilistic grammar. *In* : AIJMER, K.; ALTENBERG, B. (org.). **English corpus linguistics**: studies in honour of Jan Svartvik. London: Longman, 1991.

HALLIDAY, M. A. K. *et al.* **Lexicology and corpus linguistics**: an introduction. Londres: Continuum, 2004.

HAUSMANN, F. J. Wortschatzlernen ist kollokationslernen: zum lheren und lernen französischerwortverbindungen. **Praxis des neusprachlichen Unterrichts**, v. 31, p. 305-406. 1984.

- HAUSMANN, F. J. Kollokationen in deutschen wörterbuch: ein beitrag zur theorie des lexikographischen beispiels. *In* : BERGENHOLTZ, H; MUGDAN, J. (org.). **Lexikographie und grammatik**. Tübingen : Max Niemeyer, 1985.
- HEID, U.; MARTIN, W; POSCH, I. An Overview of approaches towards the description of collocations. *Feasibility of standards for collocational description of lexical items*. EUROTRA 7 - Report, Stuttgart/Amsterdam, 1991.
- HEID, U. Computational phraseology: an overview. *In*: GRANGER, S.; MEUNIER, F. (org.). **Phraseology: an interdisciplinary perspective**. Amsterdam: John Benjamins, 2008. p. 337-360.
- HEYLEN, D.; MAXWELL, K. Lexical functions and the translation of collocations. *In*: International Conference on Computational Linguistics, 13., 1994, Kyoto. **Proceedings** [...]. Kyoto, 1994. P. 298-305.
- HOWARTH, P. **Phraseology in 144aiwane academic writing**: some implications for language learning and dictionary making lexicographica. Tübingen: Max Niemeyer, 1996.
- HERNÁNDEZ, H. **Los diccionarios de orientación escolar**: contribución al estudio de la lexicografía monolingüe española. Tübingen: Max Niemeyer, 1989.
- JAKUBÍČEK, M. *et al.* The Ten Ten corpus family. *In*: **7th International Corpus Linguistics Conference CL**, July, p. 125-127, 2013.
- JOHNS, T. Data-driven learning: the perpetual challenge. *In*: KETTEMANN, B.; MARKO, G. (org.). **Teaching and learning by doing analysis**. Amsterdam: Rodopi, 1991.
- KELLER, E. Gambits: conversational satrategy signals. *In*: **Journal of ragmatics**, 3:249-238, 1979.
- KILGARRIFF, A. *et al.* The sketch engine: ten years on. **Lexicography ASIALEX**, Hong Kong, v. 1, n. 1, p. 7-36, 2014.
- KJELLMER, G. Aspects of English collocations. *In*: MEIJS, W. (org.). **Corpus linguistics and beyond**. Proceedings of the Seventh International Conference on English Language Research on Computerised Corpora. Amsterdam: Rodopi, 1987. P. 133-140.
- KNOWLES, G. The use of spoken and written corpora in the teaching of language and linguistics. **Literary and Linguistic Computing**, v. 5, n. 1, p. 45-48, 1990. Disponível: <https://doi.org/10.1093/lc/5.1.45>. Acesso em: 11 abr. 2019.
- KORNEVA, V. V. **Lexicologia y fraseologia de la lengua espanola**. Voronezh, Rússia: Voronezh State University Publishing House, 2016. 67 p. Disponível em: <https://rucont.ru/efd/635552>. Acesso em: 11 abr. 2019.
- KRIEGER, D. Corpus linguistics: what it is and how it can be applied to teaching. **The Internet TESL Journal**, [S. l.], v. 9, n. 3, 2003. Disponível em: <http://iteslj.org/Articles/Krieger-Corpus.html>. Acesso em: 11 abr. 2019.

L'HOMME, M. C. A methodology for describing collocations in a specialised dictionary. *In*: NIELSEN, S.; TARP, S. (ed.). **Lexicography in the 21st century**: in honour of Henning Bergenholtz. Amsterdam: John Benjamins, 2009. p. 237-256.

L'HOMME, M. C. Understanding specialized lexical combinations. **Terminology**: International Journal of Theoretical and Applied Issues in Specialized Communication, Amsterdam, v. 6, n. 1, p. 89-110, 2000.

LAVIOSA, S. *et al.* **Textual and contextual analysis in empirical translation studies**. Singapore: Springer, 2017. p. 73-128.

LAVIOSA-BRAITHWAITE, S. **The English Comparable Corpus (ECC)**: a resource and a methodology for the empirical study of translation. 1996. Tese (Doutorado). Instituto de Ciência e Tecnologia da Universidade de Manchester (UMIST), Manchester, 1996.

LEWIS, M. **The lexical approach**. London: Language Teaching Publications, 2000.

LORENTE *et al.* Specialized collocations in specialized dictionaries. *In*: TORNEL, S.; BERNAL, E. (org.). **Collocations and other lexical combinations in spanish**: theoretical and applied approaches. New York: Routledge, 2017. p. 200-222.

LUQUE NADAL, L. **Principios de culturología y fraseología española**: creatividad y variación en las unidades fraseológicas. Frankfurt: Peterlang, 2012.

LYONS, J. **Introduction to theoretical linguistics**. London: Cambridge, 1968.

MACHADO, M. A.; BEVILACQUA, C. R. Metodologias para a extração e identificação de unidades fraseológicas especializadas eventivas em corpora textuais. **Guavira Letras**, v. 27, p. 75-95, 2018.

MACMILLAN DICTIONARY. Disponível em: <https://www.macmillandictionary.com>. Acesso em: 17/07/2020.

MALKIEL, Y. Studies in irreversible binomials. *Lingua*, Amsterdam, v. 8, p. 113-160, 1959. DOI: [https://doi.org/10.1016/0024-3841\(59\)90018-X](https://doi.org/10.1016/0024-3841(59)90018-X). Disponível em: <https://www.sciencedirect.com/science/article/pii/002438415990018X>. Acesso em: 11 ago. 2019.

MARTELLI, A. **Lexical collocations in learner english**: a - based approach. Alessandria: Edizioni dell'Orso, 2007.

MARTÍNEZ MARÍN, J. Fraseología y diccionarios modernos de español. **Voz y Letra**, Madrid, v. 2, n. 1, p. 117-126, 1991.

MATUDA, S. **A fraseologia do futebol: um estudo bilingue português-inglês direcionado pelo corpus**. 2011. 304 f. Dissertação (Mestrado em Letras) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.

MCENERY, T.; WILSON, A. Corpora and Translation: Uses and Future Prospects. **UCREL Technical Papers**, Lancaster, p. 1-11, mar. 1993. Disponível em: <http://ucrel.lancs.ac.uk/papers/techpaper/vol2.pdf>. Acesso em: 11 ago. 2019.

MCENERY, T.; WILSON, A. **Corpus linguistics: an introduction**. Edinburgh: Edinburgh University Press, 1996.

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

MEYER, C. **English linguistics: an introduction**. New York: Cambridge University Press, 2004.

MONTEIRO-PLANTIN, R. S. ReVEL na escola: fraseologia e paremiologia: para que ensinar, se todo o mundo sabe? **ReVEL**, Novo Hamburgo, v. 15, n. 29, p. 1-16, 2017.

MOON, R. S., phraseology, and lexicography. **International Journal of Lexicography**, Oxford, v. 21, n. 3, p. 243-254, 2008.

MOON, R. Textual aspects of fixed expressions in learner's dictionaries. *In*: ARNAUD, P. Y.; BÉJOINT, H. (ed.). **Vocabulary and applied linguistics**. London: Palgrave Macmillan, 1992. p. 13-27.

MÜLLER, R. **Phraseologismen in englischen fachtexten der humanmedizin**. 6. ed. Frankfurt: Peter Lang, Leipziger Fachsprachen-Studien, 1993.

MURAD, C. R. R. O. **A tradução como mediação em contexto jornalístico: uma análise de textos de opinião da Seleções do Reader's Digest - 2014**. 191 f. Tese (Doutorado em Estudos Linguísticos) - Universidade Federal de Uberlândia, Uberlândia, 2014.

MURAKAMI, D. S. **Colocações verbais em um corpus de aprendizes brasileiros de inglês**. 2016. Dissertação (Mestrado em Estudos Linguísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo (USP), São Paulo, 2016. Disponível em: doi:10.11606/D.8.2016.tde-08082016-095903. Acesso em: 15 out. 2020.

NAKANISHI, D. S. **As colocações extraídas de um dos temas de um corpus de aprendizes de tradução: reflexões para uma pedagogia do léxico**. 2013. Iniciação Científica (Graduação em Letras - Inglês) – Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), São José do Rio Preto, 2013.

NATION, I. S. P. **Learning vocabulary in another language**. Cambridge: Cambridge University Press, 2001.

NATION, I. S. P. **Teaching and learning vocabulary**. Boston: Henley & Henley, 1990.

NESSELHAUF, N. **Collocations in a learner corpus**. Amsterdam: John Benjamins, 2005.

NETTO, A. Retórica virulenta aproxima Trump de países do Leste Europeu. **O Estado de São Paulo**, São Paulo, 15 jul. 2018a. Disponível em: <https://internacional.estadao.com.br/noticias/geral,retorica-virulenta-aproxima-trump-de-paises-do-leste-europeu,70002402501>. Acesso em: 30 jul. 2018.

NETTO, A. Extrema direita alemã amplia atos contra imigrantes. **O Estado de São Paulo**, São Paulo, 30 ago. 2018b. Disponível em: <https://internacional.estadao.com.br/noticias/geral,extrema-direita-alema-amplia-atos-contra-imigrantes,70002481074>. Acesso em: 11 set. 2018.

NETTO, A. Nova legislação migratória visa fim dos guetos na Dinamarca. **O Estado de São Paulo**, São Paulo, 30 set. 2018c. Disponível em: <https://internacional.estadao.com.br/noticias/geral,nova-legislacao-migratoria-visa-fim-dos-guetos-na-dinamarca,70002525252>. Acesso em: 11 out. 2018.

NICHOLLS, D. The Cambridge Learner Corpus: error coding and analysis for lexicography and ELT. **UCREL Technical Papers**, Lancaster, p. 572-581, 1999. Disponível em: <http://ucrel.lancs.ac.uk/publications/CL2003/papers/nicholls.pdf>. Acesso em: 11 mai. 2019.

OLIVEIRA, J. T. de. **A Linguística de Corpus na formação do tradutor: compilação e proposta de análise de um corpus paralelo de aprendizes de tradução**. 2015. Dissertação (Mestrado em Estudos Linguísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo (USP), São Paulo, 2014. Disponível em: doi:10.11606/D.8.2014.tde-26052015-104749. Acesso em: 15 out. 2020.

O'KEEFE, A.; MCCARTHY, M.; CARTER, R. **From concordance to classroom**. Cambridge: Cambridge University Press, 2007.

ONLINE ENGLISH COLLOCATIONS PLATFORM (Brasil). UNESP: GBD. Disponível em: <https://www.institucional.grupogbd.com/workbook/index>. Acesso em: 02 fev. 2020.

ORENHA-OTTAIANO, A. **A compilação de um glossário bilíngue de colocações, na área de jornalismo de negócios, baseado em 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.

ORENHA-OTTAIANO, A. **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 Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2009.

ORENHA-OTTAIANO, A. English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. **Acta Scientiarum. Human and Social Sciences** (Impresso), Maringá, v. 34, p. 241-251, 2012a.

ORENHA-OTTAIANO, A. Compilação de um *corpus* de aprendizes de tradução e análise de aspectos colocacionais. In: Abralín em Cena, 2012, Cuiabá. **Anais do Abralín**, 2012b.

ORENHA-OTTAIANO, A. *Collocations workbook*: um material de apoio pedagógico on-line baseado em para o ensino de colocações em inglês. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 23, n. 3, p. 833-881, 2015.

ORENHA-OTTAIANO, A. The compilation of an Online Corpus-Based Bilingual Collocations Dictionary: motivations, obstacles and achievements. *In: ELECTRONIC LEXICOGRAPHY IN THE 21ST CENTURY*, 2007, Leiden. **E-Lex Proceedings**. Leiden, 2017. p. 458-473.

ORENHA-OTTAIANO, A. Escolhas colocacionais a partir de um *Corpus* de Aprendizes de Tradução e a importância do desenvolvimento da competência colocacional. **Cadernos de Fraseologia Galega**, Espanha. No prelo.

ORENHA-OTTAIANO, A.; PINTO, P. T. Pedagogia do léxico e da tradução: novas práticas em pesquisa. *In: ROCHA, N.; RODRIGUES A.; CAVALARI, S. (org.). Novas práticas em pesquisa sobre a linguagem: rompendo fronteiras*. Araraquara: Cultura Acadêmica, 2018. p. 127-144. (Trilhas Linguísticas, n. 30). Disponível em: <https://culturaacademica.com.br/catalogo/novas-praticas-em-pesquisa-sobre-a-linguagem-rompendo-fronteiras/>. Acesso em: 11 jan. 2019.

ORENHA-OTTAIANO, A.; PAIVA, P. T. P. Uso de corpora em aulas de prática de redação de língua inglesa. **Revista Horizontes de Linguística Aplicada**, Brasília, DF, v. 11, n. 1, p. 189-210, 2012. Disponível em: <https://periodicos.unb.br/index.php/horizontesla/article/view/1181>. Acesso em: 11 abr. 2019.

OXFORD COLLOCATION. **Oxford Collocation Dictionary - On-line**. United States. Disponível em: <https://www.freecollocation.com/>. Acesso em: 11 out. 2018.

PAIVA, P. T. P. **Uma investigação de traduções de textos da área médica sob a luz dos estudos da tradução**. 288 f. 2009. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista (UNESP), São José do Rio Preto, 2009.

PAQUOT, M.; GRANGER, S. **Formulaic language in learner corpora**. *Annual Review of Applied Linguistics*, [s. l.], v. 32, p. 130-149, 2012. <https://doi.org/10.1017/S0267190512000098>.

PAVEL, S. La phraséologie en langue de spécialité. Méthodologie de consignation dans les vocabulaires terminologiques. **Terminologies Nouvelles**, Brussels, v. 10, RINT, p. 67-82, 1993.

PAWLEY, A. Phraseology, linguistics and dictionary. Review of COWIE, A. P. *Phraseology: theory analysis, and applications*. Oxford: Clarendon Press, 2001. **International Journal of Lexicography**, Oxford, v. 14, n. 2, p. 122-134, 2001.

PAWLEY, A.; SYDER, F. H. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. *In: RICHARDS, J. C.; SCHMIDT, R. W. (org.). Language and communication*. London: Longman, 1983. p. 191-225.

PHOOCHAROENSIL, S. Collocational errors in EFL learners' interlanguage. **Journal of Education and Practice**, Hong Kong, v. 2, n. 3, p. 103-120, 2011.

POTTIER, B. **Linguística geral: teoria e descrição**. Tradução e adaptação portuguesa de Walmírio Macedo. Rio de Janeiro: Presença: Universidade Santa Úrsula, 1978.

REBECHI, R. R.; SILVA, M. M. Brazilian recipes in portuguese and english: the role of phraseology for translation. *In: MITKOV, R. (org.). Computational and corpus-based phraseology*. Cham: Springer, 2017. p. 102-114.

REPPEN, R. **Using Corpora in the Language Classroom**. New York: Cambridge University Press, 2010.

ROBERTS, R. **The Oxford handbook of Applied Linguistics**. Ed. Robert B. Kaplan. Chapter 33: Translation. New York: Oxford University Press, 2002. p. 429-4

ROCHA, J. M. P. **Tradução de fraseologismos metafóricos do português para o inglês: um estudo de corpus de aprendizes brasileiros**. 2021. 258 f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2021.

ROCHA, J. M. P.; VIANA, V.; ORENHA-OTTAIANO, A. Tradução de fraseologismos metafóricos: contribuições teórico-metodológicas da linguística de corpus. **Revista da ABRALIN**, v. 19, n. 1, p. 1-26, 16 nov. 2020. Disponível em: doi 10.25189/rabralin.v19i1.1697. Acesso em: 15 out. 2020.

RYCHLÝ, P. A lexicographer-friendly association score. *In: Sojka P, Horák A. (Eds.). Proceedings of Recent Advances in Slavonic Natural Language Processing – RASLAN*. Brno: Masaryk University; 2008. p. 6-9.

SALAMANCA MARTINEZ, O.; SUAREZ DE LA TORRE, M.M. Pragmatic parameter for contrastive analyses of the equivalence of eventive specialized phraseological units. *In: EUROPHRAS: International Conference on Computational and Corpus-Based Phraseology*, 2. ed. **Proceedings** [...]. London, 2017. p. 115-117.

SAUSSURE, F. **Curso de lingüística geral**. 14. ed. São Paulo: Cultrix, 1988.

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

SCOTT, M. **WordSmith Tools**: version 6. Stroud: Lexical Analysis Software, 2012.

ŞEN BARTAN, Ö. Lexical collocation errors in literary translations. **Dil Dergisi: Language Journal**, Ankara, v. 170, n. 1, p. 71-86, jan. 2019. DOI: 10.33690/dilder.528981. Disponível em: <https://dergipark.org.tr/en/pub/dilder/issue/43587/528981>. Acesso em: 11 jun. 2020.

SHITU, F M. Collocation Errors in English as Second Language (ESL) Essay Writing. **International Journal of Cognitive and Language Sciences**, [S. l.], v. 9, n. 9, p. 3270-3277, nov. 2015. DOI: doi.org/10.5281/zenodo.1110888. World Academy of Science, Engineering and Technology, Open Science Index 105, Cognitive and Language Sciences. Disponível em: <https://publications.waset.org/10003328/collocation-errors-in-english-as-second-language-esl-essay-writing>. Acesso em: 11 abr. 2019.

SINCLAIR, J. **Developing Linguistic Corpora: a Guide to Good Practice Corpus and Text - Basic Principles**. AHDS - Literature, languages and linguistics: Oxford Text Archive at Oxford University, 2004. Disponível em: <http://users.ox.ac.uk/~martinw/dlc/chapter1.htm>. Acesso em: 15 out. 2020.

SINCLAIR, J.; MOON, R. E. (org.). **Collins cobuild dictionary of idioms**. London: Harper Collin, 1995.

SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

SINCLAIR, J.; MOON, R. E. (org.). **Collins cobuild dictionary of phrasal verbs**. London: Collins, 1989.

SINCLAIR, J. M. **Looking up**: An account of the COBUILD project in lexical computing. London: Collins, 1987.

SKELL. Disponível em: <https://skell.sketchengine.eu/#home?lang=en>. Acesso em: 1 maio 2020.

SKETCH ENGINE. Disponível em: <http://www.sketchengine.co.uk/>. Acesso em: 1 maio 2020.

SOBRAL, A. **Dizer o “mesmo” a outros**: ensaios sobre tradução. São Paulo: SBS, 2008.

SUCHOMEL, V.; POMIKÁLEK, J. Efficient web crawling for large text corpora. *In: Proceedings of the seventh Web as Corpus Workshop (WAC7)*, p. 39-43, 2012.

SUMMERS, D. Coverage of spoken English in relation to learners' dictionaries, especially the Longman Dictionary of Contemporary English. *In: HERBST, T.; POPP, K. (org.). The perfect learners' dictionary*. Tübingen: Max Niemeyer, 1999. p. 257-265.

TAGNIN, S. E. O. Collecting data for a bilingual dictionary of verbal collocations: from scraps of paper to corpora research. *In: Practical Applications in Language Corpora, Lodz., Proceedings [...]*. Lodz: Lodz University Press, 1999.

TAGNIN, S. E. O. Os *corpora*: instrumentos de auto-ajuda para o tradutor. **Cadernos de Tradução**, Florianópolis, v. 9, p. 191-213, 2002.

TAGNIN, S. E. O. The COMET project: comparable and parallel corpora for the english-portuguese pair. *In: International symposium on using corpora in contrastive and translation studies, 2010, Ormskirk. Proceedings [...]*. Ormskirk, 2010.

TAGNIN, S.E.O. **Linguística de Corpus no ensino e na tradução**. 2011. (ELC 2011 - Belo Horizonte).

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

TAGNIN, S. E. O. Glossário de linguística de corpus. *In: VIANA, V.; TAGNIN, S. E. O. (org.). Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2010. p. 357-361.

TAGNIN, S. E. O. **Linguística de corpus e tradução**. São Paulo: Universidade de São Paulo. Faculdade de Filosofia, Letras e Ciências Humanas, Programa de Pós-Graduação em Estudos da Tradução - TRADUSP, 2017. 73 slides. Apresentação de aula.

Disponível em: <http://comet.fflch.usp.br/publicacoes#apresenta%C3%A7%C3%B5es>. Acesso em: 15 out. 2020.

TER-MINASOVA, S. The freedom of word-combinations and the compilation of learners' dictionaries. *In: EURALEX*, 1992, Tampere. **Proceedings** [...]. Tampere: University of Tampere, 1992. p. 533-539.

TOGNINI-BONELLI, E. Working with corpora across languages. *In: TOGNINI-BONELLI, E. Linguistics at work*. Amsterdam: John Benjamins, 2001.

TRIBBLE, C.; JONES, G. **Concordancing in the Classroom**: a resource guide for teachers. Houston: Athelstan, 1997.

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

TSCHICHOLD, C. A computational lexicography approach to phraseologisms. *In: GRANGER, S.; MEUNIER, F. (ed.). Phraseology: an interdisciplinary perspective*. Amsterdam: John Benjamins, 2008. p. 361-376.

UNISAGRADO. **Letras - Tradutor - Bacharelado**. Centro Universitário Sagrado Coração, Bauru (SP), 2020. Disponível em: [Unihttps://unisagrado.edu.br/graduacao/letras-tradutor](https://unisagrado.edu.br/graduacao/letras-tradutor). Acesso em: 20 set. 2020.

VIANA, V.; TAGNIN, S. E. O. (org.). **Corpora na tradução**. São Paulo: HUB Editorial, 2015. v. 1.

VINOGRADOV, V. V. Ob osnovnuikh tipakh frazeologicheskikh edinits v russkom yazuike. *In: SHAKHMATOV, A. A. Sbornik statey i materialov*. Moscow: Nauka, 1947. p. 339-364.

ZGUSTA, L. Multiword lexical units. **Word**, New York, v. 23, n. 1-3, p. 578-587, 1967.

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

ZULUÁGA, A. **La fijación fraseológica**. Bogotá: Instituto Caro y Cuervo, 1975.

ANEXOS

ANEXO A – Carta de anuência do Centro Universitário Sagrado Coração (UNISAGRADO)



CARTA DE ANUÊNCIA PARA DESENVOLVIMENTO DE PESQUISA

Aceito o pesquisador Emanuel Henrique Alves, do Programa de Pós-Graduação em Estudos Linguísticos, Curso de Mestrado Acadêmico, da Universidade Estadual Paulista – Câmpus de São José do Rio Preto, com a pesquisa intitulada "A investigação de fraseologismos baseada em corpora a partir de um Corpus de Aprendizês formado pela tradução de textos jornalísticos de cunho político". O referido Corpus de Aprendizês de Tradução está, nesta fase da pesquisa, em sua 2ª edição e será intitulado *Corpus de Aprendizês de Tradução 2 (CAT 2)*. Figura como 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", sob coordenação da Profa. Dra. Adriane Orenha Ottaiano e será compilado com a colaboração do referido discente.

Ciente dos objetivos e da metodologia da pesquisa acima citada, concedo a anuência para seu desenvolvimento na Universidade do Sagrado Coração junto aos alunos do Curso de Letras-Tradutor, desde que me sejam assegurados os requisitos abaixo:

- O cumprimento das determinações éticas da Resolução nº466/2012 CNS/CONEP;
- A garantia de solicitar e receber esclarecimentos antes, durante e depois do desenvolvimento da pesquisa;
- Não haverá nenhuma despesa para esta instituição que seja decorrente da participação dessa pesquisa;
- No caso do não cumprimento dos itens acima, a liberdade de retirar minha anuência a qualquer momento da pesquisa sem penalização alguma.

Bauru, 09 de maio de 2018.

Profa. Dra. Ketilin Mayra Pedro

Diretora do Centro de Ciências Humanas- Universidade do Sagrado Coração

ANEXO B – Termo de Consentimento Livre e Esclarecido



Termo de Consentimento Livre e Esclarecido - TCLE
(Conselho Nacional de Saúde, Resolução 468/2012/Resolução 510/2016)

Você está sendo convidado a participar como voluntário do projeto de pesquisa CAT 2 (*Corpus de Aprendizagem de Tradução 2*), sob responsabilidade do pesquisador(a) Emanuel Henrique Alves, como continuidade da pesquisa CAT 1 (*Corpus de Aprendizagem de Tradução 1*), iniciada em 2012, na Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto, e como 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", ambos sob coordenação da Profa. Dra. Adriane Orenha-Ottalano. O estudo será realizado com um grupo experimental de alunos do curso de Tradução da Universidade Sagrado Coração, em Bauru, São Paulo, a partir da análise de versões inseridas no CAT, como proposta de dissertação de mestrado. O objetivo é analisar o comportamento linguístico das colocações utilizadas nos textos por meio de ferramenta de manuseio de corpus. Haverá um risco mínimo para a saúde emocional, caracterizado pelo desafio e trabalho na busca de termos em outro idioma, uma vez que se trata de um processo de escrita em outra língua. Durante esta pesquisa, auxiliaremos os participantes por meio de sugestões de fontes de pesquisas e, após seu término, ocorrerá uma devolutiva aos envolvidos, direta e/ou indiretamente, por meio de publicações. Você poderá consultar o pesquisador responsável em qualquer época, pessoalmente ou pelo telefone da Instituição, para esclarecimento de qualquer dúvida. Você está livre para, a qualquer momento, deixar de participar da pesquisa. Todas as informações por você fornecidas e os resultados obtidos serão mantidos em sigilo, e estes últimos só serão utilizados para divulgação em reuniões e revistas científicas. Você será informado de todos os resultados obtidos, independentemente do fato de estes poderem mudar seu consentimento em participar da pesquisa. Você não terá quaisquer benefícios ou direitos financeiros sobre os eventuais resultados decorrentes da pesquisa. Este estudo é importante, uma vez que seus resultados fornecerão informações para a comunidade científica e contribuições pedagógicas sobre o uso da Linguística de Corpus para a área da tradução no ensino superior.

Diante das explicações, se você concorda em participar deste projeto, forneça os dados solicitados e coloque sua assinatura a seguir.

Nome: _____ R.G. _____

Endereço: _____ Fone: _____

_____ de _____ de 20__

Assinatura do Participante

Assinatura do Pesquisador(a) responsável

ANEXO C – Termo de Consentimento Livre e Esclarecido: dados do projeto de pesquisa



Nome Pesquisador(a): Emanuel Henrique Alves	Cargo/Função: Pesquisador
Instituição: Universidade Estadual Paulista Júlio de Mesquita Filho, Instituto de Biociências Letras e Ciências Exatas de São José do Rio Preto	
Endereço: Rua Cristóvão Colombo, 2265 Jardim Nazareth 15054000 - São José do Rio Preto, SP - Brasil	
Telefone: (17) 32212544 –orientadora; (18) 991997918 - pesquisador	
Projeto submetido ao Comitê de Ética em Pesquisa do IBILCE/UNESP Rua Cristóvão Colombo, 2265. Bairro: Jardim Nazareth. São José do Rio Preto/SP – Fone 17-3221.2480 e 3221.2545	

ANEXO D – Parecer de aprovação

Considerações Finais a critério do CEP:

O Comitê de Ética em Pesquisa, em reunião ordinária de 31 de agosto de 2018, deliberou, por unanimidade, pela aprovação do presente Projeto de Pesquisa. Os relatórios parciais deverão ser encaminhados semestralmente, contando a partir desta data, conforme modelo em nossa página: <http://www.ibilce.unesp.br/#!/comite/etica-em-pesquisa/relatorio-projeto>.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1160213.pdf	29/06/2018 14:16:35		Aceito
Projeto Detalhado / Brochura Investigador	brochura_pesquisador.pdf	29/06/2018 14:15:26	Emanoel Henrique Alves	Aceito
Folha de Rosto	folha_rosto.pdf	29/06/2018 14:11:05	Emanoel Henrique Alves	Aceito
Outros	carta_anuencia.pdf	25/06/2018 16:13:00	Emanoel Henrique Alves	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	termo_consentimento.doc	25/06/2018 16:08:39	Emanoel Henrique Alves	Aceito

Situação do Parecer:

Aprovado



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

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, 15 de janeiro de 2021.

Assinatura do autor