

CORPUS DE APRENDIZES DE TRADUÇÃO: UMA INVESTIGAÇÃO SOBRE O EMPREGO DE COLOCAÇÕES NA TRADUÇÃO DE TEXTOS JORNALÍSTICOS

TRANSLATION LEARNER CORPUS: AN INVESTIGATION OF COLLOCATIONS IN THE TRANSLATION OF NEWSPAPER ARTICLES

Emanoel Henrique Alves

emanoel.alves@unesp.br

Universidade Estadual Paulista

<https://orcid.org/0000-0002-3179-5120>

Adriane Orenha-Ottaiano

adriane.ottaiano@unesp.br

Universidade Estadual Paulista

<https://orcid.org/0000-0001-8417-5120>

Resumo

O objetivo deste artigo é investigar problemas no emprego das colocações no *Corpus* de Aprendizes de Tradução 2 (CAT 2), um *corpus* paralelo composto por textos jornalísticos traduzidos por aprendizes de tradução na direção português brasileiro-inglês. O referencial teórico abarca estudos da Linguística de *Corpus* (MCENERY; HARDIE, 2012; LAVIOSA *et al.*, 2017; ORENHA-OTTAIANO, 2012a, 2012b, no prelo) e dos estudos da Fraseologia (GRANGER; MEUNIER, 2008; ORENHA-OTTAIANO, 2004, 2009; TAGNIN, 2013). Para a compilação, extração e verificação das colocações, adotamos o gerenciador de *corpus on-line Sketch Engine* (KILGARRIFF *et al.*, 2004), bem como o *corpus* de referência *enTenTen 2013* (JAKUBÍČEK *et al.*, 2013), inserido nesta mesma plataforma. Ademais, alguns dicionários on-line (*Cambridge, Longman, Macmillan*) também foram usados, com o propósito de analisar e discutir os contextos das colocações nos textos traduzidos e buscar por equivalentes em inglês. Para este artigo, apontamos, na análise de dados, um exemplo a partir da colocação em português “arrumar confusão” e suas respectivas traduções observadas no *subcorpus* de textos traduzidos para o inglês. Como base neste dado e nos resultados de nossa pesquisa, foi possível evidenciar a necessidade de elaboração de propostas de atividades pedagógicas com vistas ao desenvolvimento da competência colocacional dos tradutores em formação.

Palavras-chave: *Corpus* de aprendizes de tradução; Colocações; Fraseologia; Linguística de *Corpus*.

Abstract

This paper aims to identify problems in the use of collocations based on the Translation Student Corpus 2 (CAT 2), a parallel translation corpus composed by newspaper articles originally written in Brazilian Portuguese and translated into English by translation students. The theoretical concepts were supported by the Corpus-based Lexicon and Translation Pedagogy (MCENERY; HARDIE, 2012; LAVIOSA et al., 2017; ORENHA-OTTAIANO, 2012a, 2012b, in press) and also Phraseology (GRANGER; MEUNIER, 2008; ORENHA-OTTAIANO, 2004, 2009; TAGNIN, 2013). With a view to verify possible collocational problems, a corpus manager and text analysis software (KILGARRIFF et al., 2004) was adopted. The reference corpus EnTenTen 2013 (JAKUBÍČEK et al, 2013) and some online dictionaries (Cambridge, Longman and Macmillan) were used, in order to analyze and discuss the contexts of the collocations in the translated texts as well as search for equivalents. For this paper, we pointed out an example based on the Portuguese collocation “arrumar confusão” and its respective translations extracted from the subcorpus of translated

texts into English. Based on this data and the results of our research, it was possible to highlight the need to develop proposals for pedagogical activities in order to build learner translators' collocational competence.
Keywords: Translation Learners Corpus; Collocations; Phraseology; Corpus Linguistic.

1 Introdução

Por meio dos estudos de Baker (1992, 1993), a Tradução passou a ser uma área de grande interesse dos linguistas de *corpus*. A exploração de recursos que fazem uso do computador passou a auxiliar veementemente pesquisas científicas sobre a representação da linguagem humana dentro da área da Linguística. Nesse sentido, as novas práticas da Linguística de *Corpus* (doravante LC) promovem investigações do léxico por meio de *corpus* eletrônico, o qual pode ser definido como um conjunto de textos coletados em larga escala e analisado de forma computadorizada, a fim de proporcionar subsídios para a observação da língua utilizada pelo falante real, seja de modo oral ou escrito (MCENERY; HARDIE, 2012).

Esse avanço tem contribuído para verificar a frequência de colocações, principalmente quando se trata da tradução desse tipo de fraseologia, na busca de correspondentes em outro idioma estrangeiro, sendo este um desafio para tradutores aprendizes, devido ao grau de complexidade desse fenômeno linguístico. (ORENHA-OTTAIANO, 2012a, 2012b, no prelo). Por colocações, entende-se que são àquelas “combinações recorrentes, arbitrárias e convencionais, lexicalmente e/ou sintaticamente fixas até certo grau e que podem ter um alcance colocacional mais ou menos restrito” (ORENHA-OTTAIANO, 2017, p. 458, tradução nossa)¹.

Isso demonstra a relevância de se propor mais contribuições científicas sobre o emprego dessa fraseologia, referentes à formação de tradutores aprendizes (BERNARDINI, 2007; ORENHA-OTTAIANO, 2012a, 2012b, 2015, no prelo; etc.). Por esse motivo, este artigo trata da análise dos problemas colocacionais levantados a partir do *Corpus* de Aprendizes de tradução (CAT 2), composto por traduções de textos jornalísticos, na direção português brasileiro-inglês, realizadas por discentes dos cursos de Bacharelado em Letras com Habilitação para Tradutor, do Centro Universitário Sagrado Coração (UNISAGRADO), em Bauru, entre os anos de 2018 e 2019. O trabalho teve como objetivo verificar de que forma esses aprendizes empregaram as colocações em língua inglesa, com vistas a levantar suas dificuldades na tradução desse tipo de unidade fraseológica e propor novas sugestões.

Desse modo, a metodologia prezou pela utilização do gerenciador de *corpus Sketch Engine*

¹ Collocations are understood as pervasive, recurrent, arbitrary and conventionalized combinations, which are lexically and/or syntactically fixed to a certain degree and may have a more or less restricted collocational range. [Todas as traduções neste trabalho são de nossa responsabilidade.]

(KILGARRIFF *et al.*, 2004), para a compilação e o manuseio das colocações do CAT 2. Além desse recurso informatizado, utilizamos o *corpus* de referência (*EnTenTen English Web* 2013) da mesma plataforma e os dicionários *on-line*, *Cambridge*, *Longman* e *Macmillan*, a fim de comprovar as dificuldades para traduzir colocações durante o processo de formação na graduação. Especificamente para este artigo, apresentamos um exemplo a partir da colocação “arrumar confusão” e suas 11 traduções encontradas no *corpus* de estudo.

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

A relação entre a Tradução e a LC vem sendo foco de muitos linguistas computacionais (BERNARDINI, 2004; CAMARGO, 2007; ORENHA-OTTAIANO, 2004, 2009; TAGNIN, 2013; LAVIOSA *et al.*, 2017). Essa confluência de interesses tem como base a nomenclatura Estudos da Tradução Baseado em *Corpus*. De acordo com Laviosa *et al.* (2017), essa fase da Tradução se divide em três momentos: o chamado *the dawn*, “o amanhecer”, com a ideia de *corpus* como metodologia descritiva (BAKER, 1993, 1996); a tese de doutorado de Laviosa-Braithwaite (1996) até o final dos anos de 1990; e, o crescimento dessa área a partir dos anos 2000 marcado pelos interesses da terminologia, período de aparecimento do conceito de uso na ideia linguística e análise crítica do discurso.

Bernardini, Stewart e Zanettin (2003) acordam que, na tradução descritiva, os modelos das abordagens baseadas em *corpus* são úteis para investigar se as traduções diferem dos textos originais e a forma como esse processo tradutório acontece. Por essa razão, os instrumentos computadorizados da LC tornaram-se amplamente utilizadas e servem de “[...] metodologia que enfoca na identificação de padrões recorrentes de comportamento linguístico em dados de desempenho, fornecem a ferramenta apropriada para testar hipóteses sobre normas e regularidades em textos traduzidos” (BERNARDINI; STEWART; ZANETTIN, 2003, p. 2, tradução nossa).²

Para corroborar essa informação, Bernardini, Stewart e Zanettin (2003, p. 3, tradução nossa) acrescentam que os “estudos da tradução baseados em *corpus* descritivos levaram a um melhor entendimento do fenômeno da tradução e auxiliaram na conscientização sobre o que está envolvido nessa área”³. Logo, observa-se que a produtividade do tradutor pode sofrer mais influências positivas pelo viés da LC, permitindo-lhe compreender os padrões combinatórios do léxico, pertencentes à

² “[...] as a methodology which focuses on the identification of recurrent patterns of linguistic behaviour in actual performance data, provides the appropriate tool to test hypotheses about norms and regularities in translated texts”.

³ “Descriptive corpus-based translation studies have led to a better understanding of translation phenomena, and helped raise awareness of what is involved in translating”.

fraseologia da língua, tais como: expressões idiomáticas, provérbios, binômios, colocações, etc.

O emprego do *corpus* na Tradução pode promover reflexões de como o léxico combinado é utilizado em contextos mais gerais ou de especialidade. Esse tipo de abordagem vem contribuindo significativamente para o aprimoramento na formação de tradutores, por meio da elaboração de dicionários de colocações (ORENHA-OTTAIANO, 2017), glossários especializados (KRIEGER *et al.*, 2006; ORENHA-OTTAIANO, 2004), como também na criação de materiais didáticos (ORENHA-OTTAIANO, 2015), que podem ser aproveitados desde o contexto da tradução até o ensino de língua estrangeira e materna. Sendo assim, a abordagem baseada em *corpus* ampliou a capacidade do tradutor, conduzindo-o além de um simples manuseio de ferramentas de tradução. Baker (1993, p. 237), por sua vez, salienta a importância de analisar um texto traduzido, considerando o alto volume de textos pela perspectiva do *corpus*:

[...] 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).⁴

Além disso, Baker (1993, p. 241) também acrescenta que as coletâneas de *corpora* informatizadas, seguidas do desenvolvimento de uma metodologia de pesquisa, fornecem maneiras de superar nossas limitações humanas e minimizar a nossa dependência na intuição. Orenha-Ottaiano (2012b, p. 1), também defende a ideia de que a LC promove “subsídios para a construção de *corpora* computadorizados, bem como para o desenvolvimento de uma metodologia de pesquisa que possibilita investigações mais amplas e menos dependentes da intuição do analista”.

Essa interface entre a LC e os Estudos da Tradução passou a estimular a criação de uma série de recursos (LAVIOSA *et al.*, 2017), considerando o crescimento da pesquisa descritiva por meio do *corpus* desde os anos 80. Segundo Laviosa (2017), com base em Toury (1995/2012), trata-se de uma abordagem voltada para os estudos universais da tradução, baseada em análise probabilística.

Orenha-Ottaiano (2012b) ressalta que, do enfoque teórico, a LC colabora para a delimitação, explicação e definição de fraseologismos, como as colocações. Logo, esse tipo de investigação

⁴ “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”.

baseada em *corpus* pode auxiliar no estabelecimento de parâmetros que delimitam as colocações e outras unidades fraseológicas, bem como possibilita “descobrir quais são seus significados, seus padrões combinatórios, as características semânticas que se repetem nesses padrões [...]” (ORENHA-OTTAIANO, 2012b, p. 1).

2.1 A tradução das colocações por uma Pedagogia Baseada em *Corpus*

De acordo com McEnery e Hardie (2012), o termo *collocations* foi introduzido por Firth, no século XX, o qual explica que, em uma colocação, as partes não podem ser explicadas isoladamente, mas pela forma com que as palavras se combinam, considerando ainda a coocorrência delas por meio da frequência.

Neste trabalho, as colocações podem ser definidas como “combinações recorrentes, arbitrárias, convencionalizadas e abrangentes, ou seja, presentes em diversos contextos (*pervasive*), lexicalmente e/ou sintaticamente fixas até certo grau e que podem ter um alcance colocacional mais ou menos restrito”⁵ (ORENHA-OTTAIANO, 2017, p. 458, tradução nossa).

Devido à complexidade da estrutura de uma colocação, visando seu entendimento na prática da vida real, esta pesquisa seguiu o modelo terminológico de Hausmann (1984, 1985), que apresenta esse tipo de unidade fraseológica como a junção de dois elementos linguísticos e estruturais para sua formação: base e colocado. Orenha-Ottaiano (2004, 2009, 2015) apresenta a concepção de base e colocado da seguinte forma:

Quadro 1 – Definições de base e colocado

| Base | Quanto ao colocado |
|---|---|
| <ul style="list-style-type: none"> • um elemento independente; • semanticamente autônoma; • traduzível, independentemente de seu uso na colocação; e • determina padrões lexicais que podem combinar com ela. | <ul style="list-style-type: none"> • 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 <i>et al.</i>, 1991). |

Fonte: Orenha-Ottaiano (2015, p. 852).

Tais, base e colocado, pautados nesses significados, podem ser mais bem linguisticamente entendidos a partir da taxonomia proposta por Hausmann (1984), e ampliada por Orenha-Ottaiano (2004), com exemplos da área de negócios:

⁵ “*pervasive, recurrent, arbitrary and conventionalized combinations, which are lexically and/or syntactically fixed to a certain degree and may have a more or less restricted collocational range*”.

Quadro 2 – Taxonomia das colocações

| VERBAIS – com quatro cinco básicas | NOMINAIS – com duas formas básicas |
|---|---|
| <ul style="list-style-type: none"> • Verbo colocado + Substantivo base: <i>acquire shares</i>⁶ • Substantivo base + Verbo colocado: <i>investments dropped</i> • Verbo colocado + Preposição + Substantivo base: <i>dispose of shares</i>; • Verbo colocado + Partícula Adverbial + Substantivo base: <i>set up a business</i> • Verbo colocado + Adjetivo base: <i>grow Strong</i> | <ul style="list-style-type: none"> • Substantivo base + Substantivo colocado: <i>share subscription</i>; • Substantivo colocado + Preposição + Substantivo base: <i>holder of shares</i> |
| ADJETIVAS – com uma forma | ADVERBIAIS – com três formas básicas |
| <ul style="list-style-type: none"> • Adjetivo colocado + Substantivo base: <i>bearer shares; life</i> | <ul style="list-style-type: none"> • Advérbio colocado + Adjetivo base: <i>fully eligible</i> • Verbo base + Advérbio colocado: <i>drop dramatically</i> • Advérbio colocado + Verbo base: <i>fully paid; duly appointed</i> |

Fonte: elaborado por Orenha-Ottaiano (2004).

Tagnin (2002, p. 194) afirma que as colocações são “palavras que coocorrem em uma frequência maior do que se tratasse de uma combinação aleatória”. Dessa maneira, o levantamento dessas unidades fraseológicas com base na frequência de um *corpus*, considerando também seu *LogDice* (ver item 3), permite identificar dados mais concretos quanto à frequência de uso da colocação na língua.

Por esse motivo, analisá-las segundo os padrões de sua coocorrência no *corpus*, significa, conforme aponta McEnery e Hardie (2012), um processo de descarte de uma busca intuitiva como consulta a falantes nativos, a fim de comprovar o uso de colocações em uma língua. Sendo assim, muitos pesquisadores também entendem as colocações como “uma ampla variedade de coocorrências de diferentes padrões que podem ser extraídos de um *corpus*” (MCENERY; HARDIE, 2012, p. 123)⁷.

No caso do tradutor, existe ainda a necessidade desse profissional ter “[...] um conhecimento colocacional nas duas línguas, de modo que possa compreender o que os autores querem significar [...]” (ORENHA-OTTAIANO, no prelo). À vista disso, as práticas de tradução que envolvem a empregabilidade dessa unidade fraseológica, devem ser mais cautelosas quando o sujeito dessa ação tradutória ainda é um aprendiz, pelo fato de carregar consigo certo grau de ingenuidade (TAGNIN, 2002), podendo incorrer na escolha de uma colocação inadequada.

No caso do estudo das traduções de colocações a partir de *corpus*, Orenha-Ottaiano (no prelo) menciona que embora haja alguns interesses em pesquisas desse tipo (BERNARDINI, 2007; GREGORIO-GODEO; MOLINA, 2011; KENNY, 2001), há mais contribuições voltadas para as colocações no ensino. Conseqüentemente, torna-se relevante pensar em reflexões teórico-práticas

⁶ Os exemplos foram extraídos de um *corpus* de negócios investigado em Orenha-Ottaiano (2004).

⁷ “[...] a wide range of different co-occurrence patterns that may be extracted from a corpus”.

desse fraseologismo (ROCHA; ORENHA-OTTAIANO, 2019), por uma ótica pedagógica da língua em uso, levando em conta as dificuldades enfrentadas em práticas tradutórias no processo de formação do aprendiz de tradução (ALQAED, 2017; BERNARDINI, 2007).

De acordo com Alqaed (2017), os alunos que estudam uma segunda língua enfrentam problemas para encontrar colocações equivalentes de ordem sintática, cultural ou semântica, que sejam adequadas ao idioma de destino. Bernardini (2007, p. 1, tradução nossa) afirma que “a noção de colocação é uma das pedras angulares da LC e tem sido objeto de interesse em muitas pesquisas empíricas”⁸. Ou seja, nota-se a carga de significado que essa fraseologia pode ocupar em dado contexto e seu emprego inadequado no texto traduzido, podendo gerar problemas de incompreensão.

Logo, as pesquisas sobre tradução de colocações baseada em *corpus* podem servir de apoio para a formação do aprendiz dessa área ao ponto de fazê-lo desenvolver a competência colocacional (JUKNEVIČIENĖ, 2008; ORENHA-OTTAIANO, no prelo). Dessa forma, esse interesse em estudos sobre o ensino da tradução de colocações, que envolve a LC e a tradução, pode ser mais bem explorado por meio da “Pedagogia da Tradução Baseada em *Corpus*” (ORENHA-OTTAIANO; PINTO, 2018; ROCHA; ALVES; ORENHA-OTTAIANO, 2018). Essa nomenclatura pode propiciar reflexões científicas que destaca o ensino da tradução de colocações, em que o *corpus*, empregado como recurso linguístico, abre caminhos para a observação desse tipo de fraseologia tanto na língua de partida quanto na língua-alvo, envolvendo a participação de aprendizes.

Entretanto, segundo apontam Bernardini, Stewart e Zanettin (2003), o trabalho com *corpus* nos contextos de ensino da tradução não devem se limitar a sua mera inserção no currículo, mas serem utilizados em disciplinas sobre ensino de línguas e terminologia. Assim, abordamos um pouco acerca da importância do *corpus* e as contribuições teóricas-metodológicas dos Estudos da Tradução Baseada em *corpus*. Após essa discussão, seguimos para a etapa consoante à metodologia, apresentando o processo de formação do CAT 2 e a execução do gerenciamento das colocações, utilizando-se de uma ferramenta computacional de análise.

3 O *corpus* e os recursos metodológicos para verificação das colocações

A análise desenvolvida neste trabalho tem como base o *Corpus* de Aprendizes de Tradução 2 (CAT 2), um *corpus* paralelo, subdividido em um *subcorpus* de textos originais (TOs), com cerca de 3 mil palavras, e outro formado por publicações traduzidas (TTs) para a língua inglesa, com

⁸ “The notion of collocation is one of the cornerstones of corpus linguistics, and has been the subject of substantial speculation and empirical research.”

aproximadamente 96 mil palavras. Para as traduções, foram selecionados seis artigos em português de caráter político, contendo 500 a 800 palavras, extraídos de jornais brasileiros na modalidade *on-line*.

O processo de tradução contou com a participação dos estudantes do curso de Bacharelado em Tradução⁹, os quais foram submetidos a um teste de proficiência. Como critério de seleção dos sujeitos da pesquisa, os aprendizes deveriam apresentar um nível de proficiência entre B2 e C2, após serem avaliados por meio de teste de proficiência (ALLAN, 2004), segundo o “Quadro Comum Europeu de Referência”. Tanto a avaliação da proficiência dos alunos quanto às traduções foram realizadas durante as aulas de “Estágio Obrigatório” ou em aulas de “Prática de Tradução”, na universidade *locus* da pesquisa, Centro Universitário Sagrado Coração (UNISAGRADO), sob a supervisão da coordenação do curso e de um docente responsável. Em uma primeira etapa, em 2018, tivemos como sujeitos da pesquisa alunos do 3º ano do curso e, em um segundo momento, alunos do 2º ano. A ideia foi obter o máximo de traduções possíveis, para que os dados coletados viabilizassem mais a hipótese de identificar dificuldades de empregabilidade das colocações no CAT 2.

Sendo assim, durante o processo tradutório, optou-se por seguir alguns critérios anteriormente estabelecidos pelo CAT 1, a saber: **1.** As traduções deveriam ser realizadas em duas horas ininterruptas; **2.** As traduções não deveriam ser feitas com auxílio de outros colegas; **3.** 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.); **4.** A entrega das traduções deveria seguir os prazos estipulados pela equipe do projeto.

Depois dessa etapa de coleta de dados, adotaram-se alguns recursos metodológicos para verificar as colocações no CAT 2. O primeiro deles foi a ferramenta *Sketch Engine* (KILGARRIFF *et al.*, 2004), um gerenciador de *corpus* estatístico que agrupa em sua interface inicial mecanismos que norteiam a busca de dados linguísticos, permitindo pesquisas mais detalhadas dentro do *corpus*, além de proporcionar a criação, a compilação e a análise de um *corpus* próprio. Dentre os mecanismos disponíveis pelo *Sketch Engine*, priorizamos as ferramentas *Keywords*, *Word Sketch* e a *Concordance*. A ferramenta *Keywords* (palavra-chave) foi utilizada para criar as palavras-chave no *subcorpus* de

⁹ Alunos brasileiros matriculados no curso de Letras-Tradutor, do Centro Universitário Sagrado Coração (UNISAGRADO), em Bauru, São Paulo, entre os anos de 2018 e 2019. Participação conforme preveem os requisitos legais exigidos para esse tipo de atividade científica de coleta de dados envolvendo humanos, esse projeto foi submetido ao *Comitê de Ética* da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), campus de São José do Rio Preto, e teve sua aprovação no dia 09 de setembro de 2018, **CAAE**: 92624318.9.0000.5466, sob nossa responsabilidade.

TOs. Em seguida, selecionamos uma palavra de busca para identificar, com base na palavra de busca escolhida, se havia sido gerada alguma colocação no *subcorpus* de TTs. Baseado na colocação identificada, extraímos as opções tradutórias no *subcorpus* de TTs e suas respectivas frequências, utilizando-se do *Concordance*.

Assim sendo, depois dessa etapa de identificação das colocações traduzidas, passamos a fazer uso do *corpus* do português brasileiro (871.117,178 palavras), para verificar possíveis ocorrências dessa colocação a partir da palavra de busca, bem como o *corpus* de referência *English Web* (19.685.73,337 palavras), que serviu para comprovar se as sugestões apresentadas pelos aprendizes seriam usuais ou não em língua inglesa. Ambos os *corpora* pertencem à família *EnTenTen* (JAKUBÍČEK, 2013) e estão inseridos no gerenciador de *corpus Sketch Engine*. Para essa fase de identificação da colocação na língua-fonte e sua comprovação de uso ou não na língua-alvo, recorreremos à ferramenta *Word Sketch*, que serviu para averiguar a frequência dessas colocações, bem como indicar o *LogDice* das opções tradutórias.

O *LogDice* é uma medida estatística, cujas características auxiliam a linguística computacional quanto à observação de listas de candidatos a colocação (RYCHLÝ, 2008), sendo ainda pouco explorada em pesquisas sobre aprendizagem de línguas (GABLASOVA; BREZINA; MCENERY, 2017). De acordo com informações do gerenciador, na ferramenta *Word Sketch* esse mecanismo é um tipo de score que indica a força da colocação. Relativamente, quanto mais alto o score, mais forte é a colocação na língua, por outro lado, um score mais baixo significa que as palavras da colocação também se combinam com muitas outras, ou seja, ocorre um número maior de variação entre as palavras combinadas. Para corroborar, o apontamento do *LogDice* é relevante por tratar-se de “medidas de associações estatísticas que estabelecem a força de associação entre o nódulo e seus colocados” (BERBER SARDINHA; ACUNZO; FERREIRA, 2016, p. 179-180).

Berber Sardinha, Acunzo e Ferreira (2016) explicam que o score do *LogDice* independe do tamanho do *corpus*, pois opera segundo as frequências relativas entre as partes da colocação “nódulo” + “colocado”. No que tange à frequência e o *LogDice* das colocações analisadas no CAT 2, em comparação com o *corpus* de referência, consideramos o mínimo de 3,0 como nota de corte (FRANKENBERG-GARCIA, 2018), para ponderar qualquer sugestão empregada pelos estudantes adequadamente e/ou indicar novas colocações, sendo que 14 é o valor máximo (RYCHLÝ, 2008).

Posteriormente a esse processo, os dicionários *on-line* baseados em *corpus* (*Cambridge, Longman e Macmillan*) foram utilizados para conferir se as colocações traduzidas pelos estudantes eram realmente empregadas em língua inglesa como idioma de chegada.

4 Análise e discussão dos dados

Antes de discorrermos sobre as análises das colocações, julgamos necessário apresentar os resultados do teste de proficiência aplicado com os estudantes. Para uma observação mais clara dos dados, elencamos as informações em uma tabela: a primeira parte mostra a classificação dos alunos do 3º ano do curso (segundo semestre de 2018), enquanto que, na segunda está a classificação dos aprendizes de 2º ano (ambos semestres de 2019).

Tabela 1 – Resultados do *Placement Test* com os alunos do 2º e 3º anos

| Turma: 3º ano – 2018 | | |
|------------------------|-------|---------------|
| Níveis de proficiência | | N.º de alunos |
| Básico | A1-A2 | 0 |
| Independente | B1-B2 | 21 |
| Proficiente | C1-C2 | 08 |
| Turma: 2º ano – 2019 | | |
| Níveis de proficiência | | N.º de alunos |
| Básico | A1-A2 | 0 |
| Independente | B1-B2 | 23 |
| Proficiente | C1-C2 | 03 |

Fonte: elaborada pelo autor.

Conforme os dados da primeira parte da tabela, dos 29 participantes nenhum se classificou no nível básico, 21 alunos figuraram no nível independente, isto é, intermediário, e oito discentes no nível proficiente. Já na segunda, também observamos que, entre os 26 estudantes, não houve classificados no nível básico, tendo ocorrido uma pequena variação de 23 no nível independente, e três classificações no nível proficiente, se comparada aos aprendizes de 2018.

Como os dados mostram, em ambos os anos o número de alunos classificados no nível intermediário é maior. Uma informação importante foi que, nas duas turmas, a maioria dos aprendizes não atingiu o nível de proficiência C1-C2, de acordo com o teste aplicado, fato que corrobora com a hipótese de que os estudantes apresentam dificuldades em traduzir colocações para a língua inglesa (ORENHA-OTTAIANO, 2012b, 2015).

4.1 Estatística do CAT 2 e discussão de dados

Selecionamos algumas *Keywords* no *subcorpus* de TOs que poderiam indicar a geração de possíveis colocações no *subcorpus* de TTs. Nessa fase, realizamos também uma seleção dessas palavras, mantendo apenas aquelas de conteúdo (não gramaticais), dentre as quais descartamos lexemas que por si só não formam colocações, tais como nomes próprios, artigos, 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 artigo (HAUSMANN, 1984; ORENHA-OTTAIANO, 2004). Das palavras-chave observadas no *subcorpus* de TTs, elencamos, para este artigo, um exemplo de dificuldade tradutória a partir da colocação “arrumar confusão” (verbo *colocado* + substantivo *base*).

Em seguida, ao utilizar a ferramenta *Word Sketch*, do *Sketch Engine*, a qual indica ser ou não uma colocação por meio das medidas estatísticas empregadas pelo programa, encontramos as seguintes opções no *corpus* do português brasileiro, a partir da base “confusão” combinada com seu colocado “arrumar”, seguida de suas respectivas frequências (f.) e *LogDice* (f. 1.022 - *LogDice* 8,83). Foi possível observar outros exemplos em que o colocado sofre mudança, ou seja, em que surgem novos colocados para a base “confusão”, como em: “armar confusão” (f. 372- *LogDice*, 6,92), “arranjar confusão” (f. 307- *LogDice*, 6,71), “criar confusão” (f. 1.958 - *LogDice* 5,85), “gerar confusão” (f. 2.515 – *LogDice* 7,16), “provocar confusão” (f. 1.040 - *LogDice* 6,57).

Notamos que todos os exemplos apresentam *LogDice* maior 3,0, o valor mínimo estabelecido nesta pesquisa, podendo demonstrar a existência da colocação indicada na língua origem. Mesmo não fazendo referência direta à temática do *corpus*, essa fraseologia foi empregada no *subcorpus* de TOs em um contexto totalmente político, conforme demonstra o quadro abaixo:

Quadro 3 – Colocação “arrumar confusão” (TO)

Um presidente boquirroto, que “arruma confusão” até mesmo com sua equipe de seguranças.

Fonte: elaborado pelos autores.

De acordo com o quadro abaixo, identificamos 11 traduções da colocação “arrumar confusão” no *subcorpus* de TTs, respectivamente distribuídas entre os 22 participantes:

Quadro 4 – Colocação “arrumar confusão”: traduções, frequência e exemplos em inglês no CAT 2

| Traduções | Freq. | Exemplos |
|---------------------------|-------|---|
| <i>causes problem</i> | 01 | <i>a talebearer president, who causes problems even with his security staff.</i> |
| <i>gets in trouble</i> | 08 | <i>a chatty president, who even gets in trouble with his security team.</i> |
| <i>gets even trouble</i> | 01 | <i>a drunken president, who gets even trouble with his security team</i> |
| <i>stirs up trouble</i> | 01 | <i>he is a tactless president, who stirs up trouble even with his security team.</i> |
| <i>implicates</i> | 01 | <i>an intrusive president, that implicates even with his security staff.</i> |
| <i>messes around</i> | 01 | <i>an unpleasant president that messes around even with his security guards.</i> |
| <i>make problems</i> | 01 | <i>a big mouth president, that make problems even his safety staff.</i> |
| <i>gets even confused</i> | 06 | <i>a boquacious president, who gets even confused with his security team.</i> |
| <i>pick a fight</i> | 01 | <i>a loquacious president, who pick a fight even with his security team.</i> |
| <i>makes disturbances</i> | 01 | <i>an indiscreet president which makes disturbances even with its own security.</i> |
| <i>makes confusion</i> | 01 | <i>an indiscreet president that makes confusion even with his security team</i> |

Fonte: elaborado pelos autores com base no *subcorpus* de TTs.

A princípio, com base no quadro anterior, realizamos uma seleção de quais dessas opções tradutórias deveriam ser descartadas da análise, sob o critério de que não geravam colocações de acordo com o contexto da colocação no original. Primeiramente, observamos que a opção tradutória *messes around*, embora se trate de um *phrasal verb*, ou seja, de um fraseologismo, não forma uma colocação, conforme a estrutura proposta na taxonomia (Verbo **colocado** + Partícula Adverbial + Substantivo **base**).

Haja vista que o aprendiz não faz uso de um substantivo que formaria uma colocação verbal. Além disso, notamos, por meio dos dicionários adotados, que *mess around* carrega a ideia de “bagunça” e não de “criar uma confusão em si”. Além deste exemplo, a opção tradutória com o verbo *implicates*, por sua vez, também não selecionada, em razão de não ter sido empregado como uma colocação e, dessa forma, também foi previamente descartado, por igualmente não atender à taxonomia das colocações adotada.

Sendo assim, passamos a considerar somente as traduções que geraram colocações, no intuito de que fossem analisadas sob uma ótica fraseológica, com foco na competência colocacional dos aprendizes de tradução: *causes problem, gets in trouble, gets even trouble, stirs up trouble, make problems, gets even confused, pick a fight, makes disturbances e makes confusion*¹⁰. Por se tratar de uma colocação verbal, estas traduções para a colocação “arrumar confusão”, foram classificadas de acordo com quatro formas básicas, segundo as cinco estruturas morfossintáticas das colocações do tipo verbal, indicados na taxonomia a qual nosso estudo se fundamenta (HAUSSMAN, 1984; ORENHA-OTTAIANO, 2004), conforme a seguinte classificação:

Quadro 5 – Estruturas morfossintáticas das traduções da colocação “arrumar confusão”

- Verbo **colocado** + Substantivo **base**: *causes problem, pick a fight, makes disturbances, make confusion*
- Verbo **colocado** + Preposição + Substantivo **base**: *gets in trouble*
- Verbo **colocado** + Adjetivo **base**: *gets confused*
- Verbo **colocado** + Partícula Adverbial + Substantivo **base**: *stir up trouble*

Fonte: elaborado pelos autores.

Dentre as sugestões tradutórias produzidas pelos estudantes, houve maior prevalência de uso

¹⁰ Nas opções *make problems* e *pick a fight* identificou-se a ausência do s, que configura a 3ª pessoa do singular do tempo verbal presente simples em inglês; a opção [*even*] na sugestão tradutória *gets even confused* também não foi considerada durante a verificação dessa colocação.

de duas delas: *gets in trouble* (oito ocorrências) e *gets even confused* (seis ocorrências). Em seguida, a combinação *cause problem*, *gets even trouble*, *pick a fight*, *makes disturbances* e *makes confusion* tiveram apenas uma ocorrência no CAT 2. O próximo passo foi o de verificar, por meio da *Word Sketch*, as frequências e o *LogDice* dessas indicações no *English Web 2013*.

Tabela 3 – Estruturas morfossintáticas das traduções da colocação “arrumar confusão”

| <i>Word Sketch: frequência da tradução de “arrumar confusão” no English Web 2013</i> | | |
|--|------------------------------------|----------------|
| Traduções | Frequência no <i>English Web 2</i> | <i>LogDice</i> |
| <i>causes problem</i> | 190.397 | 9,57 |
| <i>gets in trouble</i> | 49 | 13,37 |
| <i>gets trouble</i> | 4.489 | 2,83 |
| <i>stirs up trouble</i> | 3.097 | 11,28 |
| <i>makes problems</i> | 21.558 | 4,37 |
| <i>gets confused</i> | 0,267 | 6,4 |
| <i>picks a fight</i> | 9.502 | 8,09 |
| <i>makes disturbances</i> | 0 | 0 |
| <i>makes confusion</i> | 0 | 0 |

Fonte: elaborada pelos autores.

Dada a importância do *LogDice* na pesquisa com colocações, percebe-se, de acordo com a tabela acima, que a maioria contém indicação de *LogDice*, exceto nos exemplos *makes disturbances* e *make confusion*, as quais também não co-ocorrem juntas na *Word Sketch*. Nas opções com *LogDice* e suas respectivas frequências (f.), em uma ordem decrescente dos números temos: *gets in trouble* (f. 49 - *LogDice*, 13,37), *causes problem* (f. 190.397 - *LogDice*, 9,57), *pick a fight* (f. 9.502 - *LogDice*, 8,09), *gets confused* (f. 0.267 - *LogDice*, 6,4), *makes problems* (f. 21.558 - *LogDice*, 4,37) e menor em *gets trouble* (f. 4.489 - *LogDice* 2,83). Depois dessa verificação quantitativa dos dados, notamos, a partir das palavras de base (*problem*, *disturbances*, *confusion*, *trouble*, *fight*), possíveis exemplos que correspondessem à colocação “arrumar a confusão” por meio dos dicionários (*Cambridge*, *Longman* e *Macmillan*). Quase todas as opções apresentam um *LogDice* maior que 3,0, exceto *gets trouble*. Dentre as indicações encontramos as colocações *pick a fight* (frequência, 9.502 - *LogDice*, 8,09) e *stirs up trouble* (frequência, 3.097 - *LogDice*, 11,28), relacionadas à ideia de armar um barraco em público, discutir com alguém, por meio de uma discussão, dado que seus escores apresentam uma pontuação alta.

Para a colocação *pick a fight*, há definições que indicam contextos relacionados à briga e à discussão, apenas em dois dos dicionários escolhidos *I could see he was trying to pick a fight with me* (*Longman*) e *He seemed determined to pick a fight with everyone at work that day* (*Macmillan*), exceto no *Cambridge* que, por sua vez, é o único que contém tradução da definição “começar uma

briga/uma discussão”.

Já a colocação *Stir up trouble* apresenta maior intencionalidade do sujeito da ação, figurando emoções, problemas que podem tomar novas proporções e até desconforto entre as pessoas envolvidas na situação. Em todos os três dicionários há exemplos de uso da colocação *stir up trouble*, sendo eles: *The teacher told him to stop stirring up trouble (Cambridge)*, *“John was always stirring up trouble in class (Longman)* e *Stop stirring! (Macmillan)*.

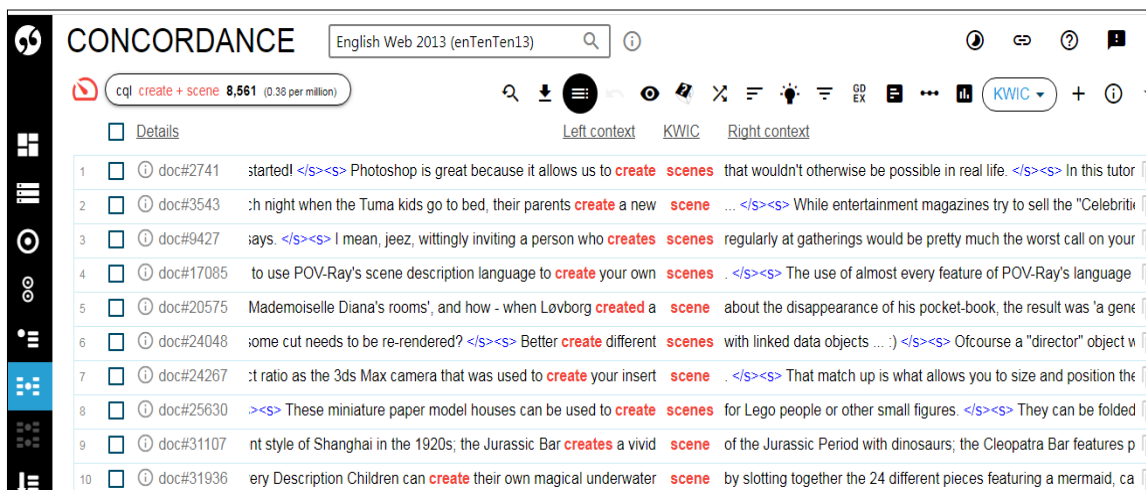
Tendo em vista a dificuldade na tradução, por parte dos aprendizes, da colocação “arrumar confusão”, *pick a fight* e *stirs up trouble*, à princípio, parecem ser opções tradutórias um pouco mais adequadas em relação ao contexto original em que foi empregada, uma vez que parecem carregar um significado mais próximo, no que diz respeito à ideia de “arrumar confusão”. Contudo, podem também remeter a certo tipo de relação física, deixando dúvidas na empregabilidade dessas duas traduções para a colocação em análise.

Nesse sentido, o próximo item trata de sugestões ou outras possibilidades tradutórias por meio do uso de *corpus*, de medidas estatísticas e com auxílio do programa *Sketch Engine*.

4.1.1 Novas sugestões tradutórias com base na colocação analisada

A fim de buscar uma opção tradutória mais adequada para a colocação investigada, seguimos com a investigação a partir de buscas mais detalhadas e pontuais com a exploração de *corpus* e uso da *Word Sketch*. Tendo em vista que os dicionários e contextos analisados nos deram pistas para a tradução de “confusão” pelo item lexical *scene*, partimos para a observação da referida palavra com base nos resultados da *Word Sketch*. A partir da análise das diversas colocações formadas, encontramos a colocação verbal *create a scene* (verbo_{colocado} + substantivo_{base}), apresentando 8.561 ocorrências e um *LogDice* de 4,94 no *English Web 2013*, mostrando-se como um equivalente adequado para a colocação em análise. Para corroborar essas informações, vejamos os exemplos oferecidos pelo *Concordance*:

Figura 1 – Linhas de concordância da colocação *create a scene*



| Doc ID | Snippet |
|-----------|--|
| doc#2741 | started! </s><s> Photoshop is great because it allows us to create scenes that wouldn't otherwise be possible in real life. </s><s> In this tutor |
| doc#3543 | :h night when the Tuma kids go to bed, their parents create a new scene ... </s><s> While entertainment magazines try to sell the "Celebrity |
| doc#9427 | ays. </s><s> I mean, jeez, wittingly inviting a person who creates scenes regularly at gatherings would be pretty much the worst call on your |
| doc#17085 | to use POV-Ray's scene description language to create your own scenes . </s><s> The use of almost every feature of POV-Ray's language |
| doc#20575 | Mademoiselle Diana's rooms', and how - when Løvborg created a scene about the disappearance of his pocket-book, the result was 'a gene |
| doc#24048 | ome cut needs to be re-rendered? </s><s> Better create different scenes with linked data objects ... :) </s><s> Ofcourse a "director" object w |
| doc#24267 | :t ratio as the 3ds Max camera that was used to create your insert scene . </s><s> That match up is what allows you to size and position the |
| doc#25630 | ><s> These miniature paper model houses can be used to create scenes for Lego people or other small figures. </s><s> They can be folded |
| doc#31107 | nt style of Shanghai in the 1920s; the Jurassic Bar creates a vivid scene of the Jurassic Period with dinosaurs; the Cleopatra Bar features p |
| doc#31936 | ery Description Children can create their own magical underwater scene by slotting together the 24 different pieces featuring a mermaid, ca |

Fonte: captura de tela do *Sketch Engine*.

Por outro lado, nos dicionários adotados como recurso para verificação da colocação analisada (neste caso, *Longman e Macmillan*) há outros dois exemplos que remetem à ideia de “arrumar confusão” a partir da palavra de base *scene*. Ambos os dicionários apresentam a colocação *make a scene* em contextos que envolvem discussão em público, conforme observamos em *There were angry scenes in parliament today* (*Longman*) e em *Stop making such a scene!* (*Macmillan*), sendo que este segundo dicionário é o único que indica a opção de usar *cause* ou *make* como colocado da base *scene*. Contudo, esses exemplos não apresentam frequência de uso na *Word Sketch*.

Pelos contextos dos dicionários, em ambos os casos existe correspondência à colocação “arrumar confusão”. Nota-se, ainda, que, nos exemplos selecionados, o uso da colocação faz jus a uma ação argumentativa ocorrida em público, de modo que demonstra uma postura de tom alterado por parte do emissor. À vista disso, averiguamos as frequências de *cause/make a scene* a partir dessas colocações no *corpus* de referência *English Web 2013*, onde foi possível verificar esse agrupamento somente por meio do *Concordance* diretamente, indicando 3.811 ocorrências para *make a scene*, e 1.303 para *cause a scene*.

Diante das situações expostas, se compararmos a frequência e o *LogDice* da colocação *create a scene*, esta parece ser uma opção que apresenta maior densidade estatística de uso por meio do *corpus* de referência. Levando em conta as informações acima apresentadas com base nos dicionários baseados em *corpus*, as colocações *cause/make a scene* também poderiam servir como equivalentes ao contexto de “arrumar confusão”. Contudo, damos preferência à opção tradutória *create a scene*, considerando seu resultado estatístico indicativo de que se trata de uma combinação que coocorre

com frequência e, desse modo, pode ser considerada uma colocação adequada.

Considerações finais

Este artigo teve como objetivo tratar das dificuldades e dos problemas no emprego das colocações extraídas do *Corpus* de Aprendizes de Tradução 2 (CAT 2), a fim de refletir sobre a empregabilidade desse tipo de fraseologia no âmbito tradutório, sob uma perspectiva descritiva e pedagógica. Observamos que as ferramentas de gerenciamento de *corpus* disponibilizadas pelo *Sketch Engine*, o *corpus* de referência (*EnTenTen, English Web 2013*), bem como os dicionários *on-line* baseados na língua em uso podem contribuir para otimizar pesquisas relacionadas ao estudo de tradução de colocações a partir de *corpora* de estudantes.

Conforme o exemplo discutido, além de identificar problemas ou dificuldades tradutórias concernentes às escolhas colocacionais dos aprendizes, julgou-se relevante indicar novas sugestões, com base nos obstáculos encontrados, na comparação entre os TOs e TTs. No exemplo, envolvendo o uso da colocação “arrumar confusão”, *a priori*, as colocações correspondentes, *pick a fight* e *stirs up trouble*, mostraram-se mais adequadas como sugestão tradutória quando comparadas às outras traduções observadas no *subcorpus* de TTs.

No entanto, identificamos que as opções tradutórias com a base *scene* (*create/cause/make a scene*) podem corresponder com mais exatidão ao contexto da colocação no TO, com preferência para a colocação *create a scene*, uma vez que esta é a única dentre as indicações sugeridas pelos autores, que aparece estatisticamente na ferramenta *Word Sketch*.

Fundamentado na ideia de averiguar possíveis dificuldades dos aprendizes de tradução ao produzirem colocações em língua inglesa como língua-alvo, bem como propor novas opções tradutórias com base nas evidências fornecidas pelo CAT 2, torna-se plausível refletir acerca de propostas de atividades futuras, a fim de suprir essas lacunas.

De certa forma, a ingenuidade presente no processo de formação desses discentes em relação à busca por colocações correspondentes em um segundo idioma é uma hipótese relevante para pesquisas com esse foco científico. Assim, além de encontrar e de fornecer dados linguísticos que corroboram essa indagação, esses pressupostos podem resultar no desenvolvimento da competência colocacional dos tradutores em formação (ORENHA-OTTAIANO, 2015), servindo de preparo para uma atuação adequada na vida profissional.

Referências

- ALLAN, D. *Oxford placement tests 2*. Oxford: OUP, 2004.
- ALQAED, M. A. Perceptions on L2 lexical collocation with a focus on English-Arabic. *Journal of Education and Practice*, v. 8, n. 6, p. 128-133, 2017.
- BAKER, M. *In Other Words*. London: Routledge, 1992.
- BAKER, M. Corpus linguistics and translation studies: implications and applications. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (org.). *Text and technology: in honor of John Sinclair*. Philadelphia: John Benjamins, 1993. p. 233-250.
- BAKER, M. Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, H. (org.). *Terminology, LSP and translation: studies in language engineering in honor of Juan C. Sager*. Amsterdam: John Benjamins, 1996. p. 177-186.
- 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 (Online)*, São Paulo, v. 18, p. 175-198, 2016.
- BERNARDINI, S.; ZANETTIN, F.; STEWART, D. *Corpora in Translator Education*. New York: Routledge, 2003.
- BERNARDINI, S. Corpora for translator education and translation practice: achievements and challenges. In: YUSTE RODRIGO, E. (ed.) Paris: *ELRA (European Language Resources Association): Proceedings of the Third International Workshop on Language Resources for Translation Work, Research & Training (LR4Trans-III)*, 2004.
- BERNARDINI, S. "Collocations in Translated Language: Combining Parallel, Comparable and Reference Corpora", en *Proceedings of the Corpus Linguistics Conference (CL2007)* (pp. 116). Birmingham: University of Birmingham, 2007.
- CAMBRIDGE DICTIONARY. Disponível em: <https://dictionary.cambridge.org/pt/dicionario/>. Acesso em: 17 de jul. 2020.
- 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. p. 65.
- FRANKENBERG-GARCIA, A. Investigating the collocations available to EAP writers. *Journal of English for Academic Purposes*, Scotland, v. 35, p. 93-104, 2018.
- 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.
- GRANGER, S.; MEUNIER, F. (Eds.). *Phraseology: an interdisciplinary perspective*. Amsterdam: John Benjamins, 2008, p. 27-49.
- GREGORIO-GODEO, E.; MOLINA, S. Collocations and the translation of news: an English-Spanish electronic dictionary of multi-word combinations as a translation tool, *Perspectives: Studies in Translation Theory and Practice*, Reino Unido, v. 19, n. 2, p. 135-152, 2011. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/0907676X.2010.544747>. Acesso em: 17 jul. 2020.
- 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.
- JAKUBÍČEK *et al.* *The Ten ten Corpus Family*, en 7th International Corpus Linguistics Conference CL. Lancaster, UK: Lancaster University. 2013, p. 125-127. Disponível em: https://www.sketchengine.eu/wp-content/uploads/The_TenTen_Corpus_2013.pdf. Acesso em: 29 set. 2020.
- JUKNEVIČIENĖ, R. Collocations with high-frequency verbs in learner English: Lithuanian learners vs native speakers. *Kalbotyra*, v. 59, n. 3, p. 119-127, 2008.
- KENNY, D. *Lexis e Creativity in Translation*. Manchester: St. Jerome Publishing, 2001.
- KILGARRIFF, A.; RYCHLÝ, P.; SMRŽ, P.; TUGWELL, D. Itri-04-08 the Sketch Engine. Information Technology, 2004.

- KRIEGER, M. da G.; MACIEL, A. M. B., BEVILACQUA, C. R.; FINATTO, M. J. B.; REUILLARD, P. C. R. *Glossário de Gestão Ambiental*. São Paulo: Disal, 2006.
- LAVIOSA-BRAITHWAITE, S. *The English Comparable Corpus (ECC): a resource and a methodology for the empirical study of translation*. Tese de doutorado inédita. Manchester: UMIST. 1996.
- LAVIOSA, S. *et al. Textual and contextual analysis in empirical translation studies*. Singapore: Springer, 2017. p. 73-128.
- LONGMAN DICTIONARY. Disponível em: <https://www.ldoceonline.com/>. Acesso em: 17 jul. 2020.
- MACMILLAN DICTIONARY. Disponível em: <https://www.macmillandictionary.com>. Acesso em: 17 jul. 2020.
- MCENERY, T.; HARDIE, A. *Linguistics: method, theory and practice*. Cambridge: Cambridge University Press, 2012.
- 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, v. 1, p. 127-144, 2018.
- ORENHA-OTTAIANO, A. The compilation of an Online Corpus-Based Bilingual Collocations Dictionary: motivations, obstacles and achievements. In: *Proceedings of E-Lex Conference 2017*. Leiden, The Netherlands, p. 458-473, 2017
- 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.
- 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.
- 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: *Abralin em Cena*, 2012, Cuiabá. Anais [...]. Abralin, 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. Proceedings... Leiden, 2017. p. 458-473.
- ROCHA, J. M. P.; ALVES, E. H.; ORENHA-OTTAIANO, A.: *Corpus de Aprendizes de Tradução – CAT: perspectivas de pesquisa do léxico fraseológico*, em Trabalho apresentado durante V Congresso Internacional de Fraseologia e Paremiologia & IV Congresso Brasileiro de Fraseologia, 2018.
- ROCHA, J. M. P.; ORENHA-OTTAIANO, A. Ensino de colocações: contribuições do grupo Pedagogia do Léxico e da Tradução a partir de corpora. In: MONZÓN, Andrea J. B.; FADANELLI, Sabrina B. (org.). *Ensino de Línguas e Formação Profissional E*. 1ed. Araraquara: Letraria, 2019, p. 11-47.
- 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.
- 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. *O jeito que a gente diz: combinações consagradas em inglês e português*. Barueri: DISAL, 2013.

Submissão: setembro de 2020
Aceite: dezembro de 2020