



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

Hernandes dos Santos Miguel

**Análise comparativa de erros colocacionais de brasileiros e chineses
presentes em dois *corpora* de aprendizes de tradução**

São José do Rio Preto

2022

Hernandes dos Santos Miguel

**Análise comparativa de erros colocacionais de brasileiros e chineses presentes
em dois *corpora* de aprendizes de tradução**

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
2022

M636a Miguel, Hernandes dos Santos
Análise comparativa de erros colocacionais de brasileiros e chineses presentes em dois corpora de aprendizes de tradução / Hernandes dos Santos Miguel. -- São José do Rio Preto, 2022
179 p. : il., tabs., fotos

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

1. Linguística de Corpus. 2. Colocações. 3. Tradução. 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.

Hernandes dos Santos Miguel

**Análise comparativa de erros colocacionais de brasileiros e chineses presentes
em dois *corpora* de aprendizes de tradução**

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.

Financiadora: CAPES

Comissão Examinadora

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

Prof^a. Dr^a. Lídia Amélia de Barros Cardoso
Universidade Federal do Ceará

Prof. Dr. Ariel Novodvorski
Universidade Federal de Uberlândia

São José do Rio Preto
10 de março de 2022

白日依山尽，黄河入海流。

欲穷千里目，更上一层楼。

王之涣

AGRADECIMENTOS

Nenhuma trajetória de sucesso se faz sozinho. Tenho muita sorte de ter tido, ao longo da realização deste projeto, o privilégio de contar com a boa vontade e solidariedade de pessoas excepcionais.

Agradeço imensamente à Profa. Dra. Adriane Orenha-Ottaiano, minha orientadora, por ter aceitado me orientar e me apresentado os caminhos possíveis dentro dos estudos de unidades fraseológicas baseados em *corpora* e, apesar de todos os percalços que surgiram durante o período de realização desta pesquisa, por ter me apoiado e encorajado a participar ativamente da vida acadêmica em seminários, em congressos e na sala de aula de tradução.

Agradeço à Profa. Dra. Marilei Amadeu Sabino, que me orientou durante os quatro anos da minha iniciação científica de 2014 a 2017 e não apenas me introduziu na vida acadêmica como também despertou meu interesse pelos estudos sobre as convencionaisidades.

Gostaria de agradecer também aos docentes do Programa de Estudos Linguísticos da UNESP, Prof. Dr. Gentil Luiz de Faria, Profa. Dra. Solange Aranha, Prof. Dr. Edson Rosa, Profa. Dra. Suzi Marques Spatti Cavalari, Profa. Dra. Paula Tavares Pinto, Profa. Dra. Ana Paula Bocorny, Prof. Dra. Talita Serpa, Dra. Ana Frankenberg-Garcia e Dr. Timothy William Lewis, com quem pude aprender muito nas disciplinas por eles ministradas.

Agradeço também à Profa. Dra. Stella E. O. Tagnin, parecerista do meu painel no primeiro Seminário de Estudos Linguísticos da UNESP (SeLin) do qual participei, em 2019, e membro da minha banca de qualificação, ao Prof. Dr. Vander Viana, que debateu meu projeto na minha segunda participação no referido seminário, em 2020, trazendo questionamentos, fazendo observações e dando sugestões que foram de extrema relevância para que o prosseguimento deste estudo fosse dado da melhor maneira possível, e à Profa. Dra. Lúcia Amélia de Barros Cardoso, que também compôs a banca da minha qualificação e contribuiu imensamente para a versão final deste trabalho.

De março de 2019 a agosto de 2020, o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, à qual agradeço aqui, pelo apoio concedido.

Sou grato à professora Xiong, cujo nome português é Frida, titular do Departamento de Português da Universidade de Hubei, e ao professor Carlos, professor de português do Instituto de Comunicações de Hebei, por terem sido as pontes que me conectaram aos estudantes voluntários, entrando em contato com eles e encorajando-os a participar do projeto.

Sou igualmente grato aos alunos que aceitaram participar da pesquisa como voluntários, mesmo em um momento tão delicado quanto o da pandemia de COVID-19.

Não poderia deixar de agradecer à Direção e aos professores do *International College* e do *International Office* da Universidade de Hubei, por terem prestado todo o apoio necessário durante os momentos mais difíceis da epidemia de COVID-19 e da quarentena em Wuhan.

Além de agradecer aos membros da academia e à instituição de fomento à pesquisa acima, não posso deixar de expressar minha gratidão às mulheres que me rodeiam e me deram toda a força necessária para realizar este trabalho.

É imensurável o quanto sou grato à Renata Degasperi Colichio, minha amiga e parceira desde os tempos de graduação e que, mesmo à distância, tem estado ao meu lado e me dado o apoio psicológico necessário.

Sou extremamente grato às amigas que a China me trouxe, sobretudo à Carla Patrícia, de São Tomé e Príncipe, na África, e Ariane e Eliana Yeung, brasileiras de ascendência chinesa, que me acolheram e se mantiveram ao meu lado me dando apoio sempre que necessário.

Sou mais do que grato às mulheres da minha família. À minha mãe, Maria Cristina, e às minhas duas irmãs, Maria Júlia e Maria Eduarda, por entenderem minha ausência e entenderem que, de certa forma, minha dedicação aos estudos é para, futuramente, beneficiá-las também.

Por fim, gostaria de deixar meus agradecimentos póstumos à mulher que me criou; minha avó, Neusa Marcelina, imigrante nordestina, analfabeta e que, apesar de todas as dificuldades, nunca deixou faltar o essencial em nossa mesa e sempre batalhou para me dar o melhor que estivesse ao seu alcance. Apesar de nunca entender muito bem o porquê de eu estudar tanto, ela nunca se opôs às minhas escolhas. Infelizmente, ela faleceu poucos meses antes de eu me tornar bacharel em letras e, agora, não poderá acompanhar presencialmente mais essa conquista que é a conclusão deste projeto de pesquisa, porém, seguem aqui meus agradecimentos como uma forma irrisória de demonstrar gratidão a quem me deu a base para que eu pudesse chegar até aqui.

RESUMO

Colocações são casos de coocorrência léxico-sintática (TAGNIN, 2013) intrinsecamente ligadas a uma determinada cultura (ORENHA-OTTAIANO, 2020; AZEVEDO, SILVA, 2017; BIDERMAN, 2001a) e que, devido às suas características essenciais como recorrência, arbitrariedade e convencionalidade (ORENHA-OTTAIANO, 2020), dão naturalidade e precisão ao discurso de uma língua, motivo pelo qual o seu conhecimento é de extrema relevância para aprendizes de tradução, que devem demonstrar fluência, agilidade e confiança suficientes para proverem serviços com rapidez e qualidade (GOUADEC, 2002). À luz da Linguística de *Corpus* (MCENERY, HARDIE, 2012; SINCLAIR, 2004; TOGNINI-BONELLI, 2001; ZANETTIN, 2001), dos Estudos da Tradução baseados em *Corpora* (ZANETTIN, 2011, 2001; BERNARDINI, 2006; BAKER, 1993) e da Fraseologia, sobretudo dos estudos sobre colocações (ORENHA-OTTAIANO, 2021, 2020; 2019, 2017, 2009, 2004; TAGNIN, 2013; MARTELLI, 2007) compilamos um *corpus* de aprendizes chineses avançados de português para compor o *Corpus* de Aprendizes de Tradução (CAT), desenvolvido na Universidade Estadual Paulista (UNESP), e analisamos em quais tipos de colocações (ORENHA-OTTAIANO, 2009; HAUSMANN, 1985) os erros colocacionais ocorrem com mais frequência na tradução de textos na direção L1 (chinês) → L2 (português). A fim de atestarmos as similaridades e/ou diferenças nos erros de aprendizes de tradução e estudantes avançados de LE de outros pares linguísticos quando da tradução para LE, também analisamos um *subcorpus* L1 (português) → L2 (inglês), formado por traduções de aprendizes de tradução e estudantes brasileiros avançados de inglês que compõem o CAT 1, com o propósito de compararmos os achados presentes em ambos os *corpora*. A compilação de 150 traduções L1 (chinês) → L2 (português) resultou em um *subcorpus* de 95.928 palavras incorporado ao CAT 3, ou CAT chinês. As análises conduzidas no *subcorpus* do CAT 3 e no *subcorpus* do CAT 1 corroboram estudos anteriores que apontam para a menor dificuldade que estudantes avançados têm em compreender as combinações em LE e para a maior dificuldade que possuem em produzi-las ou empregá-las corretamente (ORENHA-OTTAIANO, 2017, 2015, 2004; MARTELLI, 2007), assim como demonstraram que os erros colocacionais são recorrentes em todos os tipos de colocações. Por meio deste estudo, objetivamos prestar contribuições à pedagogia do léxico fraseológico, da tradução e de áreas correlatas, e, consequentemente, contribuir para a formação de tradutores e aprendizes de inglês, português e chinês.

Palavras-chave: Linguística de *Corpus*. Colocações. Tradução. Português Brasileiro. Língua Inglesa. Língua Chinesa.

ABSTRACT

Collocations are cases of lexical-syntactic co-occurrence (TAGNIN, 2013) intrinsically linked to a particular culture (AZEVEDO; SILVA, 2017; BIDERMAN, 2001a) and which, due to their essential characteristics, such as recurrence, arbitrariness and conventionality (ORENHA-OTTAIANO, 2020) give naturalness and accuracy to the speech, which is why their knowledge is extremely relevant for learner translators, who should demonstrate fluency, agility and confidence enough to provide services with speed and quality (GOUADEC, 2002). Based on the theoretical-methodological framework of Corpus Linguistics (MCENERY; HARDIE, 2012; SINCLAIR, 2004; TOGNINI-BONELLI, 2001; ZANETTIN, 2001), Translation Studies based on Corpora (ZANETTIN, 2011, 2001; BERNARDINI, 2006; BAKER, 1993) and Collocations (ORENHA-OTTAIANO, 2021, 2020; 2019, 2017, 2009, 2004; TAGNIN, 2013; MARTELLI, 2007) we have compiled a corpus of advanced Chinese learners of Portuguese to compose the Translation Learner Corpus (CAT), developed at São Paulo State University, and we analyzed in which types of collocations (ORENHA-OTTAIANO, 2009; HAUSMANN, 1985) collocational errors occur more frequently when the direction of translation is L1(Chinese) → L2(Portuguese). In order to attest to the similarities and/or differences present in the errors of learner translators and advanced foreign language students from other language pairs when translating to a foreign language, we also analyzed CAT 1's L1(Portuguese) → L2(English) subcorpus, which is constituted by translations by Brazilian learner translators and advanced students of English to compare the findings present in both corpora. The compilation of 150 L1(Chinese) → L2(Portuguese) translations resulted in a subcorpus of 95,928 words incorporated into the CAT 3, or Chinese CAT). The analyses conducted in this CAT 3's subcorpus and in the CAT 1's subcorpus corroborate previous studies that point to the lesser difficulty advanced students have in understanding combinations in a foreign language and the greater difficulty they have in producing or using them correctly (ORENHA-OTTAIANO, 2017, 2015, 2004; MARTELLI, 2007) and demonstrated that collocational errors are recurrent in all types of collocations. With this study, we aim to contribute to the pedagogy of the phraseological lexicon, to translation studies as well as to related areas, and, consequently, contribute to the training of translators and learners of English, Portuguese and Chinese.

Keywords: Corpus Linguistics. Collocations. Translation. Brazilian Portuguese. English language. Chinese language.

摘要

搭配是与特定文化相联系(AZEVEDO; SILVA, 2017; BIDERMAN, 2001a)的词汇共现现象(TAGNIN, 2013)。搭配所具有的重复性、任意性、规约性等特征(ORENHA- OTTAIANO, 2020)可使语句更为地道、准确, 因此掌握搭配有利于翻译学习者增强信心、提高语言流畅性, 从而更迅速地呈现高质量的翻译作品(GOUADEC, 2002)。根据语料库语言学(MCENERY; HARDIE, 2012; SINCLAIR, 2005; TOGNINI-BONELLI, 2001; ZANETTIN, 2001)、语料库翻译学理论(ZANETTIN, 2011, 2001; BERNARDINI, 2006; BAKER, 1993)以及片语学尤其是搭配学的相关研究(ORENHA-OTTAIANO, 2021, 2020; 2017, 2009, 2004; TAGNIN, 2013; MARTELLI, 2007), 在圣保罗大学支持下, 我们制作了中国高级葡萄牙语学习者语料库, 用于组建翻译学习者语料库(CAT), 从而分析在汉(L1)译葡(L2)文本中最为常见的搭配错误(ORENHA-OTTAIANO, 2009; HAUSMANN, 1985)。为了探究翻译学习者与其他外语的高级学习者的外译文本中错误的相似与差异, 我们还通过收集翻译学习者和巴西高级英语学生的翻译文本创建了 CAT 1 的葡(L1)译 L2 英(L2)子语料库, 并比较两个子语料库的分析结果。L1(汉语)→L2(葡萄牙语)子语料库由 150 篇翻译文本组成, 包含 95,928 个词汇。通过对两个子语料库的分析, 我们证实了之前的研究发现, 即外语高级学习者在理解搭配方面困难较小, 但在正确运用搭配方面困难相对较大(ORENHA-OTTAIANO, 2017, 2015, 2004; MARTELLI, 2007)。此外, 这些错误在所有类型的搭配中都反复出现。本研究旨在完善词法、句法及翻译等相关领域的教学, 并为英语、葡语及汉语的翻译人才培养提供参考。

关键词: 语料库语言学, 搭配, 翻译, 巴西葡萄牙语, 英语, 汉语

LISTA DE FIGURAS

Figura 1 – Interface inicial do SkE e suas ferramentas	27
Figura 2 – Níveis de convencionalidade	31
Figura 3 – Convencionalidade, fraseologismos, colocações e expressões idiomáticas	31
Figura 4 – Exemplo de quadro explicativo sobre colocações em material didático de ECLE.....	32
Figura 5 – Captura de tela do <i>Yuketang</i>	45
Figura 6 – Conversa com voluntário pelo <i>Wechat</i>	46
Figura 7 – Trecho acrescido a uma das traduções contendo a opinião do voluntário	46
Figura 8 – Interface do <i>Wechat</i> , versão para dispositivos móveis	50
Figura 9 – Interface com as várias funcionalidades do <i>Wechat</i>	50
Figura 10 – Grupos do <i>Wechat</i> com os participantes da pesquisa	51
Figura 11 – Interface do <i>Yuketang</i> , versão para dispositivos móveis.....	52
Figura 12 – Interface do <i>Yuketang</i> , versão para navegador.....	53
Figura 13 – Interface do <i>Yuketang</i> , versão para dispositivos móveis após recebimento dos textos dos voluntários.....	55
Figura 14 – Interface do <i>Yuketang</i> , versão para navegador após envio dos TOs aos voluntários	55
Figura 15 – Interface do <i>Yuketang</i> , versão para navegador após recebimento dos TTs	56
Figura 16 – Interface do <i>Yuketang</i> , versão para navegador após recebimento dos TTs	56
Figura 17 – Organização dos TOs	58
Figura 18 – Questionário final sobre o perfil dos alunos na plataforma 腾讯调查问卷.....	59
Figura 19 – Página para carregamento de textos para compilação de corpus no SkE	67
Figura 20 – Resultado da busca por palavras-chave do ptCAT17_TOs	68
Figura 21 – Resultado da busca por palavras-chave do zhCAT20_TOs.....	69
Figura 22 – Resultado da busca por colocações de 法律.....	69
Figura 23 – Resultado da busca por colocações de “calafrio” com indicação da frequência	70
Figura 24 – Resultado da busca por colocações de “calafrio” com o <i>logDice</i> indicado	71
Figura 25 – Erro no processamento de 典型.....	72
Figura 26 – Erro no processamento de 严重.....	72
Figura 27 – Análise das linhas de concordância da palavra 疫情.....	73
Figura 28 – Visualização da caixa de busca do <i>Collocations</i> do <i>Concordance</i>	74
Figura 29 – Resultado da busca por colocações da palavra 疫情 com o <i>Collocations</i> do <i>Concordance</i>	74
Figura 30 – Lista de palavras mais frequentes do zhCAT20_TOs.....	75
Figura 31 – Exemplo de duplicata com frequência e <i>scores</i> diferentes	76
Figura 32 – Resultados da análise do <i>Word Sketch</i> com colocações de “discriminação”	79
Figura 33 – Equívoco na classificação das classes dos colocados de “fim...discriminação”	79
Figura 34 – Confirmação de equívoco na classificação das classes dos colocados de “fim...discriminação”	80
Figura 35 – Visualização de corpora paralelos alinhados no SkE.....	82
Figura 36 – Interface do alinhador automático do WFA.....	83

Figura 37 – Interface do alinhador automático do WFA após carregamento de textos para alinhamento	84
Figura 38 – Interface do alinhador automático do WFA após reconhecimento das línguas a serem alinhadas	84
Figura 39 – Interface do alinhador automático do WFA após conclusão do alinhamento.....	85
Figura 40 – Arquivo com traduções alinhadas gerado pelo alinhador automático do WFA	85
Figura 41 – Planilha com as todas os TTs alinhadas.....	86
Figura 42 – Código VBA incorporado ao <i>Excel</i> para editar células sem dois cliques.....	87
Figura 43 – Configurações da ferramenta <i>find</i> e <i>replace</i> para marcação dos TTs.....	88
Figura 44 – Código VBA incorporado ao <i>Excel</i> para extração de texto com base na cor	88
Figura 45 – Resultado das extrações dos textos de cor diferente	89
Figura 46 – Traduções das colocações após acomodação em nova planilha	89
Figura 47 – Traduções das colocações após acomodação em nova planilha	90
Figura 48 – Conversão de textos em colunas para separação das unidades tradutórias.....	91
Figura 49 – Colocações e suas traduções após conclusão do processo de identificação e extração	91
Figura 50 - Exemplo de ficha de análise das traduções	92
Figura 51 - Linhas de concordância das soluções tradutórias encontradas para 受到对待.....	100
Figura 52 - Ficha com as soluções tradutórias para "projeto de lei"	101
Figura 53 - Definição de <i>bill</i> no <i>Cambridge Dictionary</i>	102
Figura 54 - Linhas de concordância de <i>law project</i>	103
Figura 55 - Resultados da busca por <i>law project</i> no <i>Google</i>	103
Figura 56 - Ficha com as soluções tradutórias para "dar declaração"	112
Figura 57 - Linhas de concordância de <i>give declaration</i> no <i>subcorpus</i> de TTs.....	113
Figura 58 - Linhas de concordância de <i>give statement</i>	115
Figura 59 - Linhas de concordância de <i>make statement</i>	116
Figura 60 - Ficha com as soluções tradutórias para "dar educação"	117
Figura 61 - Linhas de concordância de <i>have education</i> no <i>corpus</i> de referência.....	118
Figura 62 - Linhas de concordância de <i>have upbringing</i> no <i>corpus</i> de referência.....	118
Figura 63 - Definição de <i>upbringing</i> no <i>Cambridge Dictionary</i>	119
Figura 64 - Ficha com as soluções tradutórias para “engolir a seco”	121
Figura 65 - Linhas de concordância de <i>swallow wholesale</i>	123
Figura 66 - Definição de <i>swallow</i> no <i>Cambridge Dictionary</i>	124
Figura 67 - Ficha com as soluções tradutórias para “plano inclinado”	125
Figura 68 - Ficha com as soluções tradutórias para “plano de paz”	126
Figura 69 - Ficha com as soluções tradutórias para “disseminação das drogas”	127
Figura 70 - Ficha com as soluções tradutórias para 受到不人道待遇	129
Figura 71 - Linhas de concordância de “submeter” + “tratamento desumano”	130
Figura 72 - Ficha com as soluções tradutórias para 超越法治	131
Figura 73 – Linhas de concordância de "afrontar"+"Estado de Direito"	132
Figura 74 – Linhas de concordância de "subverter"+"Estado de Direito"	133
Figura 75 – Linhas de concordância de "solapar"+"Estado de Direito".....	133
Figura 76 – Linhas de concordância de "atentar"+"Estado de Direito"	133

Figura 77 – Linhas de concordância de “pilar” + “Estado de Direito”	134
Figura 78 – Linhas de concordância de “viger” + “Estado de Direito”	135
Figura 79 - Ficha com as soluções tradutórias para 严重的暴力	135
Figura 80 - Ficha com as soluções tradutórias para 有关当局	138
Figura 81 - Ficha com as soluções tradutórias para 疫情区域	140
Figura 82 - Ficha com as soluções tradutórias para 发冷得严重	141
Figura 83 - Dinâmica dos grupos do WeChat	148
Figura 84 - Projeção de explicação sobre erros colocacionais após atividade tradutória em turma de português avançado.....	156

LISTA DE TABELAS

Tabela 1 – Resultado das buscas rápidas no Google Scholar para “fraseologia” e “terminologia” + “português” + LE	18
Tabela 2 – Tipos de <i>corpus</i> monolíngue segundo McEnery e Hardie (2012)	25
Tabela 3 – Comparação entre os diferentes tipos de colocação	35
Tabela 4 – Temas, títulos e números de palavras dos TOs do CAT chinês	47
Tabela 5 – Número de palavras por <i>corpus</i> da família <i>TenTen</i> do SkE	77
Tabela 6 Códigos usados para análise das colocações	93
Tabela 7 Número de textos coletados na compilação do CAT chinês	95
Tabela 8 - Lista de verbos que coocorrem com <i>declaration</i> como objeto.....	114
Tabela 9 - Lista de verbos que coocorrem com <i>statement</i> como objeto.....	114
Tabela 10 - Verbos que coocorrem com <i>mother</i> como sujeito.....	120
Tabela 11 - Adjetivos depois de <i>swallow</i>	122
Tabela 12 - Modificadores de <i>plane</i>	125
Tabela 13 - Frequência e score de <i>plan</i> + <i>involve</i>	127
Tabela 14 - “dissemination of”	128
Tabela 15 - Verbos que coocorrem com "tratamento desumano" em construção com sujeito da passiva impessoal	130
Tabela 16 – Verbos que coocorrem com estado de direito como objeto.....	132
Tabela 18 - adjetivo + violência	136
Tabela 19 - violência + adjetivo	137
Tabela 20 - Frequência e score de grave + violência	137
Tabela 21 - Frequência e score de violência + grave	137
Tabela 22 - Frequência e score de autoridade+adjetivo	139
Tabela 23 - Frequência e score de adjetivo + autoridade	139
Tabela 21 - Frequência e score de adjetivo + calafrio.....	143
Tabela 22 - Frequência e score de calafrio + adjetivo.....	143

LISTA DE ORGANOGRAMAS

Organograma 1 – Organização do CAT 1 – O <i>corpus</i> brasileiro	42
Organograma 2 – Esquema do CAT 1 e CAT 3	96

LISTA DE QUADROS

Quadro 1 - Linhas de concordância de 受到对待 no <i>subcorpus</i> de TOs.....	100
Quadro 2 - Linha de concordância de "dar declaração" no <i>subcorpus</i> de TO	113
Quadro 3 - Linha de concordância de "dar educação" no <i>subcorpus</i> de TO	116
Quadro 4 - Linha de concordância de <i>have education</i> no <i>subcorpus</i> de TTs	117
Quadro 9 - Linha de concordância de “engolir a seco” no <i>subcorpus</i> de TOs	120
Quadro 6 - Linha de concordância de "plano inclinado" no <i>subcorpus</i> de TOs	124
Quadro 7 - Linha de concordância de <i>end of life</i> no <i>subcorpus</i> de TTs.....	126
Quadro 5 - Linha de concordância de “submetido a tratamento desumano” no <i>subcorpus</i> de TTs	130
Quadro 8 - Linha de concordância de 有关当局 no <i>subcorpus</i> de TOs	138
Quadro 11 - Linha de concordância de 疫情区域 no <i>subcorpus</i> de TOs	140
Quadro 12 - Linha de concordância de solução tradutória para 疫情区域 no <i>subcorpus</i> de TTs	141
Quadro 10 - Linha de concordância de 发冷得严重 no <i>subcorpus</i> de TOs	142

LISTA DE GRÁFICOS

Gráfico 1 – Porcentagem de cada TO sobre o total de TOs do CAT chinês	48
Gráfico 2 – Nível de formação dos voluntários chineses	60
Gráfico 3 – Autopercepção do nível de proficiência em português dos alunos chineses.....	61
Gráfico 4 – Tempo de experiência em países de língua oficial portuguesa	62
Gráfico 5 – Destino dos alunos com experiência em países de língua oficial portuguesa	63
Gráfico 6 – Outras línguas estrangeiras faladas pelos voluntários chineses	63
Gráfico 7 – Ferramentas de tradução usadas pelos voluntários chineses	64
Gráfico 8 – Tradutores automáticos usados pelos voluntários chineses.....	65
Gráfico 9 – Outras ferramentas usadas pelos voluntários chineses durante as traduções	65
Gráfico 10 – Proporção de colocações por tipo, no <i>subcorpus</i> L1(português)	97
Gráfico 11 – Proporção de colocações por tipo no <i>subcorpus</i> L1(chinês)	98

Gráfico 12 – Total de todas as colocações traduzidas por colocações inexistentes na L2 em ambos <i>subcorpora</i> brasileiro e chinês	105
Gráfico 14 - Total de todas as colocações traduzidas por colocações inexistentes na L2 no CAT brasileiro L1(português) → L2(inglês).....	106
Gráfico 15 - Total de todas as colocações traduzidas por colocações inexistentes na L2 no CAT chinês L1(chinês) → L2(português)	107
Gráfico 16 - Total de todas as colocações traduzidas por colocações inexistentes na L2 por tipo de colocação no CAT brasileiro - L1(português)→L2(inglês)	109
Gráfico 17 - Total de todas as colocações traduzidas por colocações inexistentes na L2 por tipo de colocação no CAT chinês - L1(chinês)→L2(português).....	111

LISTA DE ABREVIATURAS E SIGLAS

BELC	<i>Brazilian English Learner Corpus</i>
BNC	<i>British National Corpus</i>
Br-Icle	<i>Brazilian International Corpus of Learner English</i>
CAT	<i>Corpus de Aprendizes de Tradução</i>
COMADEC	Compilação de Materiais Didáticos e Dicionários Especializados de Colocações Baseados em <i>Corpora</i>
COMET	<i>Corpus Multilíngue para Ensino e Tradução</i>
ECLE	Ensino de Chinês como Língua Estrangeira
enTenTen20	<i>English Web Corpus 2020</i>
ET	Estudos da Tradução
ETBC	Estudos da Tradução baseado em <i>Corpora</i>
HEBIC	Instituto de Comunicações de Hebei
HUBU	Universidade de Hubei
IBILCE	Instituto de Biociências, Letras e Ciências Exatas
ICLE	<i>International Corpus of Learner English</i>
L1	Língua materna
L2	Segunda língua, língua estrangeira
LC	Linguística de <i>Corpus</i>
LE	Língua estrangeira
MLC	<i>Usp Multilingual Corpus</i>
MT	Memórias de Tradução
MUST	<i>Multilingual Student Translation Corpus</i>
ptTenTen11	<i>Portuguese Web Corpus 2011</i>
PUC-SP	Universidade Católica de São Paulo
SkE	<i>Sketch Engine</i>
TO	Texto original
TT	Texto traduzido

UF	Unidade fraseológica
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
USP	Universidade de São Paulo
VBA	<i>Visual Basic Editor</i>
WFA	<i>Wordfast Anywhere</i>
zhTenTen17	<i>Chinese Web Corpus 2017</i>

SÍMBOLOS

→	Direção tradutória
>	Sequência de comandos no computador
+	Comandos simultâneos no computador
➔	Botão de seta para a direita no teclado QWERT

SUMÁRIO

INTRODUÇÃO	17
2 FUNDAMENTAÇÃO TEÓRICA.....	23
2.1 Linguística de <i>Corpus</i>	23
2.2 Linguística de <i>Corpus</i> e Estudos da Tradução.....	27
2.3 Unidades fraseológicas e as Colocações.....	29
3 PROCEDIMENTOS METODOLÓGICOS	39
3.1 Compilação do <i>corpus</i> de estudo	39
3.1.1 O CAT brasileiro.....	40
3.1.2 O CAT chinês.....	42
3.1.3 Alternativas metodológicas para a compilação de um corpus de aprendizes em meio a pandemia de COVID-19.....	48
3.2 Identificação, seleção e extração das colocações para análise.....	66
3.2.1. Extração das colocações dos <i>subcorpora</i> de TOs com o SkE.....	66
3.2.3 Alternativas metodológicas para a observação e extração das colocações dos <i>subcorpora</i> de TTs	83
4 ANÁLISE DOS RESULTADOS	95
4.1 Resultados da compilação do CAT chinês.....	95
4.2 Resultados do levantamento das colocações.....	96
4.2.1 Análise dos erros colocacionais no <i>subcorpus</i> brasileiro.....	112
4.2.2 Análise dos erros colocacionais no <i>corpus</i> chinês	128
5 CONSIDERAÇÕES FINAIS.....	145
REFERÊNCIAS BIBLIOGRÁFICAS	158
APÊNDICES	170
Apêndice A– Perguntas e resultados do questionário final sobre o perfil dos alunos na plataforma 腾讯调查问卷.....	170
ANEXOS	177
Anexo A– Termo de Consentimento Livre e Esclarecido	177

INTRODUÇÃO

Um dos marcos iniciais da Linguística de *Corpus* (LC) foi o lançamento do *Brown University Standard Corpus of Present-Day American English*, ou simplesmente *Brown Corpus* (1964). Anos mais tarde, a LC alcançou popularidade e passou a ocupar espaço de grande relevância nas mais diversas áreas de estudo (TAGNIN, 2018), não se limitando apenas aos estudos da língua e da linguagem, mas se expandido a áreas como biblioteconomia, ciências da informação, entre outras (BOWKER, 2020, 2018).

Dentro das ciências que se dedicam a estudar a língua e a linguagem, algumas das áreas de aplicação mais notáveis da LC são a Tradução e a Fraseologia (TAGNIN, 2018).

Encontramo-nos no ponto de convergência entre a LC, os Estudos da Tradução (ET) (mais especificamente Ensino de Tradução) e a Fraseologia, lançando um olhar atento sobre as colocações, com vistas a contribuir para o ensino destas fraseologias a aprendizes de tradução e/ou estudantes de línguas estrangeiras (LE). Consequentemente, alcançamos também a união de dois campos de pesquisa distintos dentro da LC — pesquisas com *Corpus* de Aprendizes e Estudos da Tradução Baseados em *Corpora* (ETBC).

Essa junção foi possível, pois, ao final da compilação do nosso *corpus* de pesquisa, analisamos comparativamente os erros colocacionais de alunos brasileiros e chineses, presentes em textos que compõem dois *subcorpora* distintos: um na direção L1 (português) → L2 (inglês), formado por traduções de alunos dos últimos semestres dos cursos de Bacharelado em Letras — Tradutor e Licenciatura em Letras do Instituto de Biociências, Letras e Ciências Exatas (IBILCE) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), e, o outro, na direção L1 (chinês) → L2 (português), formado por traduções de alunos dos últimos anos dos cursos de Letras-Português da Universidade de Hubei (HUBU) e do Instituto de Comunicações de Hebei (HEBIC) que, além de terem disciplinas de tradução em suas grades curriculares, manifestaram ativamente o interesse em seguirem a carreira de tradutores.

A inserção do chinês neste trabalho se dá, principalmente, por acreditarmos que nossa pesquisa realizará um pequeno, porém, fundamental passo ao se distanciar, mesmo que minimamente, da relação eurocêntrica mantida pela LC e Fraseologia brasileiras, ao investigarmos possíveis problemas que podem emergir do cruzamento entre uma língua não-europeia e o português na sala de aula de LE e na tradução, combinação ainda pouca explorada

pelas referidas áreas, que permanecem aquém dos avanços no estreitamento das relações entre Brasil e China.

Em 2018, Tagnin realizou uma busca no *Google Books* por “fraseologia *corpus*” e “terminologia *corpus*” para demonstrar como é infindável a quantidade de estudos nas áreas da Fraseologia, Terminologia e Terminologia Fraseológica, estas duas intrinsecamente ligadas à Tradução. Sua busca resultou em 4.090 e 28.000 *hits* respectivamente. Apesar do grande número e do crescimento franco de estudos na área, até setembro de 2020, os *hits* para buscas no *Google Scholar* com os termos “fraseologia” + “português” + “chinês” foram incrivelmente menores (252) do que para buscas com a combinação “fraseologia” + “português” + línguas europeias como o inglês (2.240), francês (2.030) ou espanhol (1.480). O mesmo ocorreu quando “fraseologia” foi substituída por “terminologia”, como se observa na tabela a seguir, vindo a corroborar a carência de estudos dedicados ao encontro dos idiomas português e chinês.

Tabela 1 – Resultado das buscas rápidas no Google Scholar para “fraseologia” e “terminologia” + “português” + LE

FRASEOLOGIA		TERMINOLOGIA	
Português +	Hits	Português +	Hits
Inglês	2.240	Inglês	49.400
Francês	2.030	Francês	38.200
Espanhol	1.480	Alemão	28.200
Alemão	1.440	Espanhol	26.600
Italiano	1.150	Italiano	21.100
Chinês	252	Chinês	5.560

Fonte: elaborado pelo autor.

Quando afunilamos o recorte de estudos na área e tratamos especificamente de *corpus* de aprendizes de LE no Brasil, a diferença é ainda mais gritante. Até 2006, o único *corpus* de aprendizes construído no Brasil era o *Br-Icle* (BERBER SARDINHA, 2001), uma coletânea de textos argumentativos de aprendizes brasileiros de inglês incorporados ao *International Corpus of Learner English* (ICLE). Liderado por Berber Sardinha, da Pontifícia Universidade Católica de São Paulo (PUC-SP), o projeto foi concluído em 2010 e contou com contribuições de

investigadores de diversas outras universidades. De lá para cá, várias são as iniciativas na criação de *corpora* de aprendizes no país, inclusive motivadas pelo Br-Icle, como é o caso do *USP Multilingual Learner Corpus* (MLC), parte de um projeto maior intitulado *Corpus Multilíngue para Ensino e Tradução* (COMET), ambos sob coordenação de Tagnin (2013, 2003), que contempla textos de todos os gêneros produzidos por aprendizes dos cursos de inglês, espanhol, francês, italiano e alemão da Universidade de São Paulo (USP).

Além deste, *corpora* menores têm sido criados com fins mais específicos, como o *Brazilian English Learner Corpus* (BELC) construído por Pacheco (2010) para estudar a aquisição de morfemas por estudantes brasileiros de inglês, dentre outros projetos liderados por pesquisadores do espanhol, francês, italiano e alemão. No entanto, que chega ao nosso conhecimento, nenhum *corpus* de aprendizes no Brasil contempla línguas não europeias, como é o caso do chinês. O próprio *Corpus* de Aprendizes de Tradução (CAT), projeto desenvolvido no IBILCE/UNESP sob direção de Orenha-Ottaiano (2019, 2017, 2012), ao qual nos vinculamos, contava apenas com traduções na direção português→inglês, até a inserção do chinês (*subcorpus* chinês-português).

Com o prevalecimento da hegemonia da língua inglesa, seguida de outras línguas europeias nos estudos baseados em *corpus* no nosso país, ficou a cargo de iniciativas internacionais a elaboração de *corpora* de aprendizes com a combinação chinês-português, que continuam ainda muito limitadas. Do que se chega até o nosso conhecimento, apenas o *Corpus* de Português Língua Estrangeira/Língua Segunda (MENDES et al., 2016), coordenado por pesquisadores da Universidade de Lisboa, possui um *subcorpus* composto por 323 produções de falantes de chinês como língua materna. Além deste, há somente mais uma proposta, mais recente, de construção de um *corpus* de português constituído por produções escritas de aprendizes chineses liderada por Zhang (2019), da Universidade de Macau, sem especificação quanto à variante do português a ser empregada pelos estudantes.

Observados a partir de uma perspectiva política, estes dados destoam bastante dos relacionados à aproximação das relações entre Brasil e China. Antes mesmo deste país asiático superar os Estados Unidos e se tornar o principal parceiro econômico do Brasil no ano de 2009, ambas as nações asiática e sul-americana têm coordenado manobras e concentrações estreitas dentro de blocos importantes como os BRICs¹, G20² e G5³. Nesse âmbito, em 2009, foi assinado

¹ A coordenação entre Brasil, Rússia, Índia e China.

o “Plano de Ação Conjunta entre o Governo da República Popular da China e o Governo da República Federativa do Brasil 2010-2014”, ou simplesmente “Comunicado Conjunto”, posteriormente estendido até o ano de 2021. Neste acordo sobre o desenvolvimento da parceria estratégica bilateral e de cooperação em áreas relevantes entre os dois países, há uma preocupação especial na promoção das línguas dos dois países em seus territórios, conforme explicitado no Artigo 14, sobre educação, em que os dois países dão ênfase no ensino de suas respectivas línguas, fortalecendo os intercâmbios educacionais (CHINA, 2010).

Do lado chinês, houve realmente um aumento da oferta de cursos de português, inclusive em nível universitário. Até 2004, existiam na China continental apenas três universidades com a oferta do português. Em 2010, este número passou para quinze universidades (YUQI, 2011). Entre 2010 e 2014, esse número saltou de quinze para 21 universidades (ZHONGGUO, 2014). Em 2015, a HUBU, instituição aliada à nossa pesquisa, se tornou a primeira universidade da China central a oferecer o curso universitário de português (HUBEI, 2015). Por fim, em 2021, 26 instituições de ensino superior estavam creditadas a oferecer português em nível universitário na China continental, segundo informações do site oficial do *Gaokao*⁴, o Exame Nacional Para o Ingresso no Ensino Superior da China, sem contar as instituições da antiga colônia portuguesa de Macau, como a Universidade de Macau, o Instituto Politécnico, o Instituto Português do Oriente, a Universidade de São José e a Universidade de Ciência e Tecnologia (GROSSO; ZHANG, 2018).

Do lado brasileiro, apesar da expansão do Instituto Confúcio, que consiste em uma instituição de Pequim criada pelo governo da República Popular da China, com o objetivo de promover o ensino da língua e cultura chinesa, desde seu estabelecimento na UNESP, em 2008, para outras treze universidades de todas as regiões do país, no plano dos cursos de graduação, apenas a USP oferece um curso de Bacharelado em Letras – Chinês.

Como temos tentado demonstrar, apesar da relação entre Brasil e China ser de extrema importância, as pesquisas nas duas línguas não acompanham a tendência crescente de aproximação entre os dois países e são poucas as produções que investigam os percalços que podem surgir da promoção, ainda que desproporcional, dos dois idiomas dentro da sala de aula de

² Grupo formado pelas 19 maiores economias do mundo mais a União Europeia: África do Sul, Alemanha, Arábia Saudita, Argentina, Austrália, Brasil, Canadá, China, Coreia do Sul, Estados Unidos, França, Índia, Indonésia, Itália, Japão, México, Reino Unido, Rússia, Turquia e União Europeia.

³ União de cinco dos países em desenvolvimento composta por Brasil, México, Índia, África do Sul e China.

⁴ <http://college.gaokao.com/speciality/653/>. Acesso em 25 de outubro de 2021.

LEs e de tradução.

A esse respeito, acreditamos que nossa pesquisa tem um papel fundamental, não apenas por ir ao encontro de um plano maior no qual se insere a nação brasileira e não se fixar ao eixo das línguas europeias, mas também por abrir espaço para o chinês e suas peculiaridades e prestar contribuições aos estudos fraseológicos e aos ETBC em nosso país, sobretudo no que tange às colocações.

Pesquisas baseadas em *corpus* no campo da Fraseologia, conforme aquelas desenvolvidas por Martelli (2007) e Orenha-Ottaiano (2021, 2020, 2017, 2015, 2012) alertam para a dificuldade que os estudantes de nível avançado têm em determinar se as colocações que produzem em LE são aceitáveis e demonstram que o conhecimento léxico-gramatical dos aprendizes, na maior parte das vezes, não é suficiente. Principalmente quando não existe, por parte dos mesmos, a percepção do aspecto fraseológico, natural e convencional da língua. Logo, pretendemos com este estudo, debruçando-nos sobre o assunto, contribuir com a pedagogia do léxico fraseológico bilíngue e da tradução, concomitantemente ao pequeno passo que daremos no distanciamento do eurocentrismo que permeia os estudos da área no Brasil com a inclusão do chinês em nossa pesquisa. Ao decidir não nos restringirmos ao contraste do português com línguas europeias, despertamo-nos para problemas que podem surgir no cruzamento das respectivas línguas, tanto no ensino do português brasileiro para chineses e vice-versa, como no ensino de tradução para brasileiros e chineses.

Desse modo, os objetivos deste trabalho são:

- a) Compilar um *corpus* paralelo chinês-português de aprendizes chineses de português avançado com interesse em atuar como tradutores do português e do chinês de duas universidades chinesas, a HUBU e o HEBIC, para compor o CAT;
- b) Extrair candidatas a colocações e identificar aquelas que são verdadeiramente colocações presentes nos *subcorpora* de textos originais (TO) do CAT, para, posteriormente, extrair e identificar nos *subcorpora* de textos traduzidos (TT) as soluções tradutórias encontradas pelos aprendizes.
- c) Após a identificação e extração das colocações a partir do CAT, quantificar em quais tipos de colocações (ORENHA-OTTAIANO, 2020, 2017, 2015, 2009) a ocorrência de erros colocacionais foi mais frequente quando da realização das atividades de tradução de textos na direção L1→L2, i.e., L1(português) →L1 (inglês) e L1(chinês) →L2

(português).

- d) Analisar os erros colocacionais presentes em cada *subcorpora* e compará-los, levando em consideração a língua materna dos aprendizes e o sentido da tradução, a fim de se verificar quais tipos de colocações foram mais problemáticas quando os alunos as traduziam para a LE.

Cientes dos obstáculos presentes no processo de aprendizagem e na compreensão das colocações por parte de estudantes de nível avançado, estamos confiantes de que podemos contribuir significativamente para a pedagogia do léxico fraseológico bilíngue voltada a esse público-alvo, sobretudo a aprendizes de tradução, ao despertarmos, também, para os entraves representados pelas colocações no contexto de ensino português-chinês.

Para cumprirmos com os objetivos propostos e produzirmos um trabalho de relevância, baseamo-nos em um referencial teórico que atende às necessidades desta pesquisa, sobre o qual trataremos no capítulo seguinte.

2 FUNDAMENTAÇÃO TEÓRICA

Esta dissertação fundamenta-se, principalmente, nos preceitos teóricos de três áreas distintas por nós já elencadas até aqui: a LC, os ET e a Fraseologia, mais precisamente no que diz respeito ao estudo das colocações. Iniciando pela LC, a seguir apresentaremos o arcabouço teórico dessas áreas do qual nos beneficiamos.

2.1 Linguística de *Corpus*

A LC é uma das áreas de estudo em Linguística Aplicada que mais tem se desenvolvido nas últimas décadas, sobretudo com a difusão do uso do computador e da *internet*, que torna os estudos nessa e em outras disciplinas computacionais (tais como o processamento de língua natural, linguística computacional e a terminologia) mais práticos, velozes e precisos (BARROS, 2004). Por se tratar de um campo de estudo que se ocupa principalmente da língua em seu real contexto de uso, não se limitando apenas em buscar possibilidades de construção teóricas em uma língua (BERBER SARDINHA, 2004; BIBER; CONRAD; REPPEN, 1998), a LC nos dá subsídios para analisarmos as colocações em textos autênticos produzidos em grande escala pelos estudantes avançados de língua estrangeira e/ou tradução. Sem o instrumental e o referencial dessa linha de estudo, nosso trabalho seria em muito dificultado.

Para proceder em uma investigação como a nossa, em que são observados fenômenos da língua em uso, debruçamo-nos sobre um conjunto de textos específicos; conjunto esse que forma o *corpus* linguístico. Sinclair (2005) define *corpus* da seguinte forma:

Um *corpus* é uma coleção de textos de uma língua em formato eletrônico, selecionados de acordo com critérios externos para representar, o máximo possível, uma língua ou uma variante linguística como fonte de dados para pesquisa linguística.⁵ (SINCLAIR, 2005, p. 2)

Dentro da LC, destacam-se duas abordagens principais: *corpus-based*, em que o pesquisador tem um *corpus* de pesquisa pré-selecionado que lhe servirá de fonte de exemplos “geralmente para verificar instâncias e/ou a frequência ou plausibilidade de amostras pré-

⁵ Original: *A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as possible, a language or language variety as a source of data for linguistic research.* Todas as traduções neste trabalho serão de nossa responsabilidade.[todas as traduções são de nossa responsabilidade].

selecionadas de fenômenos linguísticos”⁶ (TOGNINI-BONELLI, 2001, p. 10), e *corpus-driven*, em que o objeto de pesquisa, o *corpus*, é usado “para encontrar aspectos novos e inesperados da linguagem”⁷ (BERBER SARDINHA, 2002, p. 1), com vistas a, segundo Tognini-Bonelli (2001), validar uma hipótese ou verificar exceções em determinada teoria estudada.

Biber, Conrad e Reppen (1998) elencam quatro características fundamentais da pesquisa *corpus-based*:

- é empírica, analisando os padrões reais de uso em textos naturais;
- utiliza uma grande e fundamentada coleção de textos naturais, conhecida como "um *corpus*", como base para a análise;
- faz uso extensivo de computadores para análise, usando tanto técnicas automáticas como interativas;
- depende de técnicas analíticas quantitativas e qualitativas.⁸ (BIBER; CONRAD; REPPEN, 1998)

Os autores ressaltam que “juntas, essas características resultam em um escopo e confiabilidade que não seriam possíveis de outra forma”⁹ (BIBER; CONRAD; REPPEN, 1998, p. 4).

Mais recentemente, essa distinção vem sendo questionada por autores como McEnery e Hardie (2012), que defendem que a dicotomia entre *corpus-based* e *corpus-driven* não é, na prática, passível de divisão tão enrijecida e que toda pesquisa *corpus-driven* pode ser considerada *corpus-based*. Compartilhamos da visão dos autores, porém, de toda forma, entendemos que essas características elencadas acima são essenciais para a condução de uma boa investigação, seja ela *corpus-based* ou *corpus-driven*, e nos pautamos sobre elas para realizarmos nossa pesquisa.

Acerca do *corpus* em si, ele pode ser de diferentes tipos. A primeira distinção que pode ser feita é baseada no número de línguas de um determinado *corpus* de estudo, i. e., ele pode ser monolíngue ou multilíngue (MCENERY; HARDIE, 2012; ZANETTIN, 2011). Como os próprios nomes sugerem, um *corpus* monolíngue contém textos em uma língua. Ele pode eventualmente conter palavras estrangeiras e

⁶ Original: ...generally to check institutions and/or the frequency or plausibility of pre-selected samples of linguistic phenomena.

⁷ Original: to find new and unexpected aspects of language

⁸ Original: -it is empirical, analyzing the actual patterns of use in natural texts;

-it utilizes a large and principled collection of natural texts, known as “a corpus”, as the basis for analysis;

-it makes extensive use computers for analysis, using both automatic and interactive techniques;

-it depends on both, quantitative and qualitative analytical techniques.

⁹ Original: Taken together, these characteristics result in a scope and reliability of analysis not otherwise possible.

discursos de não-nativos, como ocorre com o *British National Corpus* (BNC), mas este não é seu propósito inicial (MCENERY; HARDIE, 2012).

Já o *corpus* multilíngue possui duas ou mais línguas. Porém, como aponta McEnery e Hardie (2012), há certa confusão na classificação dos tipos de *corpus* multilíngue, pois diferentes autores têm dado a *corpora* do mesmo tipo nomes diferentes. McEnery e Hardie classificam os *corpora* multilíngues da seguinte maneira:

Tabela 2 – Tipos de *corpus* monolíngue segundo McEnery e Hardie (2012)

TIPO	DEFINIÇÃO	EXEMPLO
A	Textos de um idioma e suas traduções em um ou mais idiomas.	➤ Hansard (BROWN et al. 1991) ➤ CRATER (MCENERY; OAKES, 1995; MCENERY et al.; 1997).
B	Pares ou grupos de <i>corpora</i> monolíngues projetados seguindo-se a mesma base de amostragem.	➤ <i>Corpus Aarhus</i> de direito contratual (FABER; LAURIDSEN, 1991) ➤ <i>Lancaster Corpus of Mandarin Chinese</i> (MCENERY et al., 2003)
C	Combinação de A e B	➤ ENPC (JOHANSSON; HOF LAND, 1994) ➤ EMILLE <i>corpora</i> (BAKER et al. 2004)

Fonte: adaptado de McEnery e Hardie (2012, p. 19).

Segundo a classificação apresentada acima, nosso *corpus* de pesquisa se encaixa no “Tipo A”, cunhado por alguns autores como *corpus* paralelo (MCENERY; HARDIE, 2012), denominação que usaremos ao longo deste trabalho.

Sobre a constituição de um *corpus* em si, Sinclair (2005) afirma que um *corpus* precisa ter representatividade e dispor de textos naturais, produções autênticas, ainda que com fins específicos, como é caso das traduções (ZANETTIN, 2011). Dessa maneira, cabe ao pesquisador

atentar-se ao objetivo de sua pesquisa para selecionar os textos que se adequem a sua proposta e atendam o requisito de representatividade. No nosso caso, contamos com um *corpus* paralelo português-inglês, chinês-português de aprendizes formado por produções autênticas de estudantes de tradução.

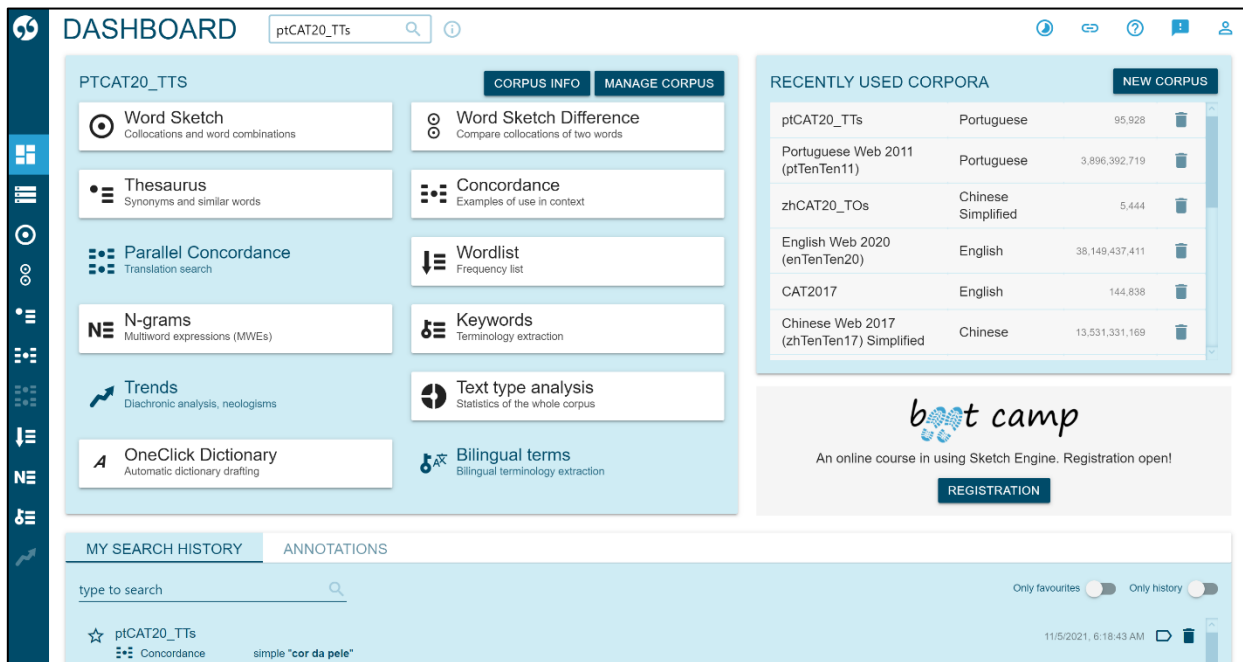
Ainda sobre a questão da representatividade, Tognini-Bonelli (2001) e Bowker e Pearson (2002) lembram que um *corpus* grande não significa um *corpus* de representatividade, e estabelecem que questões de adequação do *corpus* devem prevalecer sobre questionamentos acerca da extensão.

Para a constituição e manipulação do *corpus*, como podemos preconizar a partir das afirmações a respeito do uso do computador feitas por Biber, Conrad e Reppen (1998), lança-se mão de programas especializados com ferramentas específicas para o manuseio de *corpus*, como listas de frequência, palavras-chave e linhas de concordância (TAGNIN, 2013).

Desde os primeiros programas para análise de *corpus*, várias são as ferramentas lançadas até hoje. Entre elas, o *WordSmith Tools*, ou *WST*, que desde sua criação em 2004 foi, por muito tempo, um dos programas mais usados, e o *AntConc*, de 2014, que ganhou mais adeptos depois do *WST* (TAGNIN, 2018).

Para compilação e manipulação do nosso *corpus* de pesquisa usamos, no entanto, outro programa que também tem sido posicionado entre os líderes do setor e com número crescente de usuários, o *Sketch Engine* (SkE) (KILGARRIFF; BAISA; JAKUBÍČEK et al., 2014). Esse programa hospedado em uma plataforma totalmente *on-line* conta com diversas ferramentas, como o *Word Sketch*, que exibe diversos indicadores de palavras ou estruturas previamente selecionadas, além de outras funcionalidades que vão desde a possibilidade de se criar *corpora* paralelos até o uso dos já mencionados concordanciadores e lista de palavras. Estes recursos nos foram muito úteis, pois, com a elaboração de listas de palavras, observação e análise das linhas de concordância e leitura dos dados do *Word Sketch*, pudemos localizar as colocações presentes no *corpus* e identificar os erros tradutórios cometidos pelos estudantes.

Figura 1– Interface inicial do SkE e suas ferramentas



Fonte: captura de tela da página inicial do SkE.

Como lidamos com um *corpus* paralelo com vistas a contribuir com a ETBC, é inconcebível não nos pautarmos também sobre o referencial teórico surgido da junção entre a LC e os ET, sobre a qual trataremos na seção 2.2 a seguir.

2.2 Linguística de *Corpus* e Estudos da Tradução

Por muito tempo a participação da LC dentro dos ET sofreu uma grande resistência por parte dos linguistas de *corpus* (BAKER, 1996), pois *corpora* paralelos de textos originais (TOs) e textos traduzidos (TTs) eram usados apenas para crítica tradutória baseada em alguma teoria específica ou para análise de traduções.

Hoje são várias as disciplinas, dentro da Tradução, que se beneficiam da LC, entre elas, além das já citadas Crítica Tradutória e Análise de Traduções, a Tradução Técnica, a Terminologia, a Tradução Literária (TAGNIN, 2018, 2015; ZANETTIN, 2011) e a Tradução Automática (CASELI, 2015; LU, ZHOU, 2004).

Dentre as vantagens da união entre a LC e os ET está o uso de metodologias que permitem investigações menos intuitivas e prescritivas e mais amplas e descritivas, rompendo com a tradição perpetuada por muito tempo por estudiosos dos ET de estudar instâncias

individuais de um texto fonte comparando-as com um único texto traduzido, passando, agora, a estudar estas instâncias em um grande número de textos, por meio de *corpora* computadorizados que possibilitam o estudo de questões teóricas dantes difíceis em *corpora* muito restritos. (ORENHA-OTTAIANO, 2012a, 2012b; LAVIOSA et al., 2017; BAKER, 1993, 1996).

A junção entre essas duas áreas, no entanto, não se deu num instante e Laviosa et al. (2017) a divide em três períodos:

1. De 1993 a 1996 – O que os autores chamam de “O amanhecer” (*the Dawn*). Tem seu marco inicial quando Baker (1993) introduz a ideia de usar o *corpus* como metodologia para estudos descritivos da tradução;
2. De 1996 ao final dos anos 1990 – Este período tem início na publicação da tese de doutorado de Laviosa-Braithwaite (1996) e perdura até o final dos anos de 1990;
3. A partir dos anos 2000 – Período marcado pela expansão de estudos na área.

Nas últimas décadas, a expansão da LC nos ET é tamanha que os ETBC já configuram uma área de estudo independente (TAGNIN, 2004). O que não é de se espantar, tendo em vista que os *corpora* contribuem de maneira significativa na observação de questões extremamente relevantes à tradução, como “equivalência, convencionalidade, idiomatidade, criatividade lexical, terminologia, fraseologia”, dentre várias outras (VIANA; TAGNIN, 2015). Graças a essas possibilidades é que, além das disciplinas da Tradução mencionadas anteriormente, outra que já despertou para as vantagens do uso de *corpus* linguístico em seus estudos é o Ensino de Tradução (ORENHA-OTTAIANO, 2021, 2019, 2017, 2016, 2012a, 2012b; BOWKER, 2003).

A discussão sobre o uso de *corpora* no ensino de tradução já não é mais tão recente. Entre os estudiosos da área é unânime que, com *corpora* de textos autênticos, é possível não apenas observar regularidades linguísticas e as variáveis socioculturais a elas associadas, mas também procedimentos comuns aos tradutores — “universais da tradução” (TAGNIN, 2004).

Tognini-Bonelli (2001), citando Trosborg (1997), já demonstrava que a análise de *corpus* de textos especializados ajuda na identificação de traços estilísticos, idiosincrasias e registros, e evita que estruturas da língua de partida que não são convencionadas na língua de chegada se repliquem na tradução. Pearson (2003) também apontava para o fato de que *corpora* paralelos original-tradução aumentam a consciência sobre diferentes estratégias de tradução. Ademais,

Zanettin (2001), alguns anos antes, atentava para os benefícios do uso de *corpora* comparáveis, aqueles formados por textos sobre o mesmo assunto e no mesmo gênero do texto de língua de partida, entre os quais o autor destaca o suporte na escrita mais natural na língua de chegada.

Oliveira (2019) ressalta que os preceitos teóricos da LC também podem ajudar a desenvolver métodos de ensino para a sala de aula de tradução que, com uso de *corpora* de aprendizes, privilegiem a implementação de aulas voltadas para as necessidades e os interesses específicos dos estudantes, propiciando, assim, o ambiente ideal para aplicação de metodologias ativas de aprendizagem e, por consequência, o aumento da empatia dos alunos pelas aulas e pelo ofício de tradutor.

Mas não são apenas os ET e os ETBC que tem a ganhar com a LC. Zanettin (2011) defende a inclusão de traduções nos estudos baseados em *corpora*, sejam eles baseados em *corpora* monolíngues ou linguística contrastiva baseada em múltiplos *corpora*.

Sob a premissa de que “as traduções representam uma proporção às vezes substancial de toda a produção linguística em uma determinada cultura. Mesmo assumindo que os textos traduzidos constituem uma variedade textual específica de uma língua”¹⁰ (ZANETTIN, 2011, p. 21), o autor defende a inclusão de traduções nos grandes *corpora*. Para ele, a exclusão destes tipos de texto, na verdade, vai contra a norma padrão de referência e não pode ser justificada pela teoria.

Porém, como vimos no início desta seção ao citarmos Tagnin (2018), não foi somente a LC e os ET que se uniram; a Fraseologia também se desenvolveu muito nos últimos anos graças às ferramentas de manuseio de *corpus* e ao arcabouço metodológico da LC. Discorreremos sobre o estudo das Unidades Fraseológicas (UF) e o nosso objeto de estudo, as colocações, na seção 2.3 a seguir.

2.3 Unidades fraseológicas e as Colocações

Nos últimos anos, em várias partes do mundo, graças às ferramentas que possibilitaram a análise de grandes *corpora* textuais, capazes de identificar e evidenciar recorrências lexicais, a Fraseologia é uma das áreas dentro do estudo da língua que mais se desenvolveu (TAGNIN,

¹⁰ translations represent a sometimes substantial proportion of all linguistic production in a given culture. Even taking for granted that translated texts constitute a specific textual variety of a language.

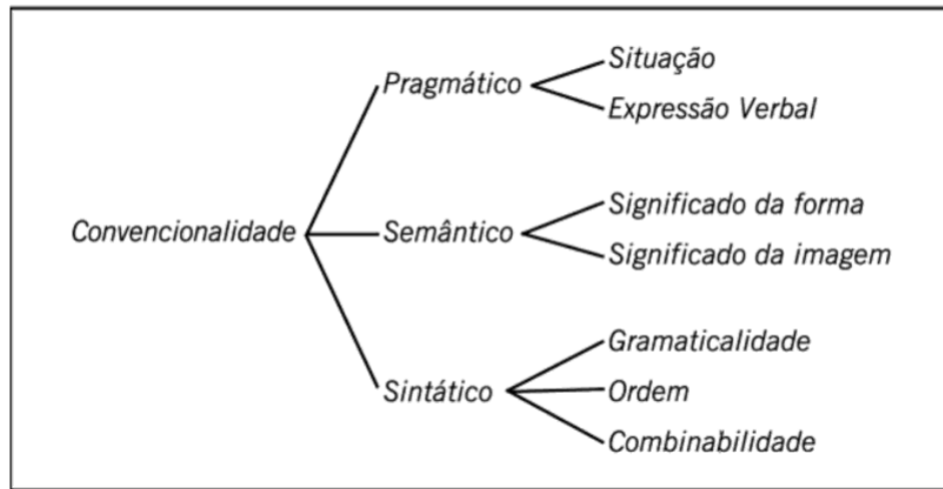
2018). Esta área tem gerado uma vasta gama de estudos baseados em *corpora*, incluindo *corpora* de aprendizes, e tem prestado enorme contribuição para o ensino de línguas estrangeiras e, consequentemente, para o Ensino de Tradução. Em nível nacional, Orenha-Ottaiano (2021, 2020, 2017, 2015, 2012a, 2012b, 2004) é uma das investigadoras que tem se dedicado especificamente à pedagogia do léxico e da tradução baseada em *corpora*, lançando um olhar atento sobre as fraseologias, com diversos estudos na compilação de glossários e dicionários a partir da extração de colocações e outras fraseologias de *corpora* de aprendizes. Um de seus projetos a cabo no IBILCE/UNESP é o *Corpus* de Aprendizes de Tradução - CAT, ao qual nossa pesquisa está vinculada, e o MUST, *Multilingual Student Translation Corpus*, sob sua coordenação no Brasil e direção de Granger e Lefer, da *Université Catholique de Louvain*, a qual uma parte do CAT, o CAT 1, cujas traduções são na direção L1 (português)—L2 (inglês), está vinculado.

As UFs são fenômenos linguísticos diretamente ligados às culturas nas quais se inserem e são consideradas parte do conjunto das convencionalidades.

As convencionalidades estão presentes em vários níveis de uma língua, e abarcam costumes e hábitos de uma sociedade que guiam as combinações lexicais produzidas por falantes nativos de uma língua (ORENHA-OTTAIANO, 2004).

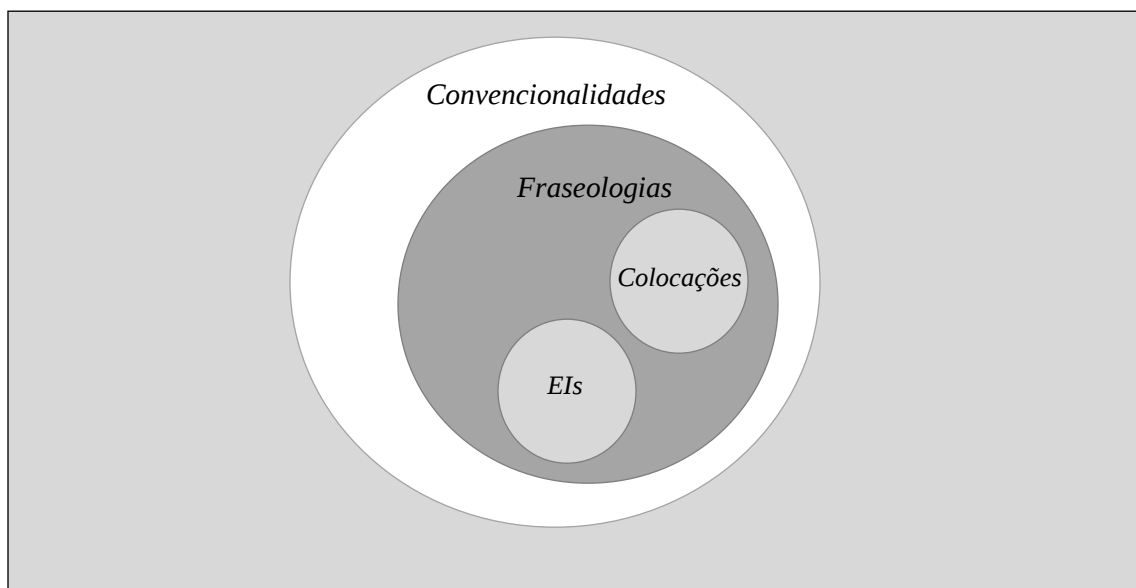
Segundo Tagnin (2013), a convencionalidade pode ocorrer em três níveis distintos relacionados às expressões convencionais e idiomáticas: o nível sintático, o nível semântico e o nível pragmático.

Conforme vemos sintetizado na Figura 2 abaixo, sobre o nível sintático, Tagnin (2013) afirma que ele diz respeito à combinabilidade, à ordem e à gramaticalidade. A convencionalidade pode ocorrer também no nível semântico, no qual “observa-se a convencionalidade na relação não motivada entre uma expressão e seu significado” (TAGNIN, 2013). A autora cita os exemplos das expressões *kick the bucket* e *bater as botas*, que significam “morrer”, UFs opacas, cuja soma dos constituintes não justifica o significado da expressão idiomático (TAGNIN, 2013). O último nível, o pragmático, diz respeito ao uso da língua em situações em que há interação entre falantes de uma língua e compreende a situação e a expressão verbal.

Figura 2 – Níveis de convencionalidade

Fonte: Tagnin, 2013.

Orenha-Ottaiano (2015, 2012, 2004) sintetiza o que é convencionalidade na língua da seguinte forma:

Figura 3 – Convencionalidade, fraseologismos, colocações e expressões idiomáticas

Fonte: adaptado de Orenha-Ottaiano (2012, p. 243).

Como podemos depreender a partir do esquema, as fraseologias fazem parte do conjunto

das convencionalidades e abarcam, dentre outras combinações convencionais, o objeto de estudo deste trabalho – as colocações, que, portanto, ocorrem nos três níveis elencados anteriormente na Figura 2 e estão intrinsicamente ligadas a uma cultura e se prestam a dar naturalidade e precisão ao discurso dos falantes de uma língua (ORENHA-OTTAIANO, 2020; AZEVEDO, SILVA, 2017; BIDERMAN, 2001). Por este motivo, apesar dos inúmeros obstáculos que se apresentam na sua compreensão e aprendizagem, o conhecimento das colocações é essencial para falantes avançados de um idioma, como tradutores e/ou outros profissionais linguísticos, que devem demonstrar fluência, agilidade e confiança suficientes para decodificar e reproduzir esses tipos de fraseologias e, assim, prover serviços com rapidez e qualidade (GOUADEC, 2002). No caso do Ensino de Chinês como Língua Estrangeira (ECLE), as colocações não apenas têm sua importância reconhecida como também são ensinadas explicitamente em materiais didáticos para ensino de língua geral ou voltados para fins específicos, como preparação para exames de proficiência, por meio de quadros e exercícios inteiramente focados nas combinações de palavras fixas da língua chinesa (LIU, 2017; JIANG, 2015; GUIMEI, 2011, XU; CUI; MU, 2011).

Figura 4 – Exemplo de quadro explicativo sobre colocações em material didático de ECLE¹¹

(二) 词语搭配

动词	+	宾语
抱怨		别人/妻子/餐厅的菜不好吃
爱护		环境/花草树木/公物/学生
定语	+	中心语
电影(的)/小说(的)/ 生活(的)		细节
电台(的)		记者/广播/新闻

5

Fonte: Jiang (2015, p. 5).

Em 1957, Firth (apud ORENHA-OTTAIANO, 2009), definiu colocações como palavras

¹¹ Tradução: 2. Colocações

Verbo + objeto

Reclamar + de alguém, da esposa, que a comida da cantina não está boa;

Proteger, cuidar + (d)o meio ambiente, (d)a natureza, (d)os animais, (d)os alunos;

Modificador + palavra principal

Filme, romance, vida + detalhes;

Estação de rádio ou TV + jornalista, transmissão, notícia;

que mostram uma coocorrência habitual sob o ponto de vista sintagmático, “palavras que usualmente andam juntas”. Halliday (1961, apud ORENHA-OTTAIANO, 2009), discípulo de Firth (1957), se aproxima bastante da definição do primeiro, afirmando que “colocação é uma associação sintagmática de itens lexicais...”. Orenha-Ottaiano (2020, p. 61) sintetiza a definição de colocações como “combinações ubíquas, recorrentes, arbitrárias e convencionais, lexicalmente e/ou sintaticamente fixas até certo grau e que podem ter um alcance colocacional mais ou menos restrito”¹². É sobre esta definição que nos baseamos para a condução deste trabalho.

Assim como seu mestre, Halliday (1961) também ressaltou a importância de um nível lexical diverso daquele do gramatical para dar conta de elucidar aspectos dantes incapazes de serem explicados apenas com o estipêndio da gramática. Isto porque, como vimos anteriormente, as combinações das colocações são geralmente lexicais e regidas pela convenção que lhes foi imposta, o que Bejoint (1994), ao tratar do Fenômeno das Afinidades Lexicais, chama de *contrainte de signe*, ou “restrição de signo”.

Sinclair (2001, 1991, 1987) se soma aos autores anteriores na definição das colocações e propõe dois princípios que regem a formação desse tipo de estruturas fraseológicas: o princípio da livre-escolha (*the open choice principle*) e o princípio idiomático (*the idiom principle*). O primeiro consiste num modelo de *slot-and-filler*. O texto seria uma série de *slots* (lacunas) a serem preenchidas por quaisquer unidades lexicais no eixo paradigmático, que satisfaçam as restrições da gramaticalidade. Contudo, esse princípio não foi capaz de explicar alguns fenômenos e, para preencher as lacunas deixadas por ele, Sinclair propõe o segundo princípio, o idiomático, com o intuito de elucidar as restrições que o princípio anterior não abarca.

Segundo esse princípio, o usuário da língua tem a seu dispor várias opções únicas de expressões “semi-pré-construídas”. Sinclair (1991) defende que, sem esse princípio não seríamos capazes de produzir textos naturais, uma vez que nem toda escolha é abarcada pela gramaticalidade. Este princípio é uma das maneiras pelas quais podemos constatar a organização da linguagem, que se manifesta por meio de um conjunto de formas convencionais que exprimem modos preferenciais de se dizer algo. Um exemplo é a preferência em inglês por dizer *pay a compliment* (fazer um elogio) em vez de *make a compliment*, no qual o princípio da livre-escolha não se aplicaria (ORENHA-OTTAIANO, 2004). Do chinês, podemos citar como exemplo a

¹² *collocations are 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.*

preferência por 浓茶 *nóng chá* (chá forte) em vez de 烈茶 *liè chá* (XU; LU; LI, 2006). Apesar de 烈 *liè* também significar “forte”, assim como 浓 *nóng*, ele coocorre mais frequentemente com 光 *guāng*, que significa “luz”.

Tagnin (1999) sintetiza algumas características, além das acima elencadas, necessárias para a pré-existência de uma colocação. Segundo a autora, colocações devem satisfazer as máximas de:

1. **Recorrência.** A combinação precisa ser recorrente;
2. **Não-idiomaticidade.** O significado da combinação é composicional, ou seja, o sentido da combinação pode ser deduzido do significado de cada um de seus elementos;
3. **Coesão.** É necessário que haja uma ligação muito forte entre os elementos da combinação, muito mais forte do que de uma combinação qualquer;
4. **Restrição contextual.** Deve haver a probabilidade de que a combinação ocorra dentro de um contexto específico;
5. **Coocorrência arbitrária dos elementos.** Não há razão semântica que explique a coocorrência entre os elementos da colocação.

Para alguns autores, colocações podem incluir *phrasal verbs*, nomes próprios, nomes compostos, conjunções correlativas e outras formas de combinações (LIU et al., 2009; MCKEOWN; RADEV, 2000). Não nos dedicamos à análise desses fenômenos neste trabalho, e nos atemos apenas às colocações que apresentam as características levantadas por Tagnin (1999).

Além de esforços na delimitação da definição do que é colocação, esforços importantes na classificação dos diferentes tipos de colocações vêm de diferentes autores.

Dentre as várias propostas para a classificação das colocações, no caso do chinês, por exemplo, na tentativa de apresentar uma proposta restritiva na classificação de colocações, Xu, Lu e Li (2006), para a construção de um banco de colocações daquela língua, lançam uma classificação que divide as colocações em quatro tipos, com base na combinação de características em termos de composicionalidade, substitucionalidade, mutabilidade e associação interna. Os tipos e os requisitos para classificação de cada tipo de colocação seguem esquematizados na Tabela 3 a seguir.

Tabela 3 – Comparação entre os diferentes tipos de colocação

	TIPO 0	TIPO 1	TIPO 2	TIPO 3
Composicional	Não	limitado a sim	sim	sim
Substituição por sinônimos	Não	não	muito limitado	limitado
Alteração da ordem	Não	não	sim	sim
Modificável	Não	não	muito limitado	limitado
Exemplo	缘木求鱼 <i>Yuánmùqiúyú</i> (subir na árvore para pegar peixe – abordagem infrutífera)	外交豁免权 <i>wàijiāo huòmiǎn</i> <i>quán</i> (imunidade diplomática)	缔结同盟 <i>dìjié</i> <i>tóngméng</i> 缔结联盟 <i>dìjié</i> <i>liánméng</i> (formar aliança)	全面合作 <i>quánmiàn hézuò</i> (ampla cooperação, cooperação abrangente)

Fonte: adaptado de XU, LU, LI (2006).

Os autores nomeiam o Tipo 0 de Colocação Idiomática, o Tipo 1 de Colocação Fixa, o Tipo 2 De Colocação Forte e o Tipo 3 de Colocação Fraca. Eventualmente retomaremos essa classificação na busca por melhor delimitar as causas que levaram aos erros na tradução de colocações de modo mais sistemático.

Ainda no que se refere à nossa pesquisa, usamos a classificação proposta por Orenha-Ottaiano (2009) com base em uma proposta anterior de Hausmann (1985). Este autor divide as colocações em uma base, palavra cuja marca semântica é maior, mais frequentemente um substantivo, e em um colocado. Baseado nessa divisão, Hausmann propõe uma taxonomia na qual quem empresta o nome à colocação é o colocado. Desse modo, uma colocação de verbo + substantivo em que o verbo é o colocado será uma colocação verbal, um adjetivo + substantivo em que o adjetivo é o colocado será uma colocação adjetiva, e assim por diante. Orenha-Ottaiano (2009) revisou e aperfeiçoou a classificação proposta por Hausmann (1985), identificando mais uma estrutura que não fora prevista pelo autor, em que o adjetivo é o colocado e o verbo é a base. Exemplificamos os diferentes tipos de colocações a seguir, com exemplos extraídos do nosso *corpus* de pesquisa.

Considerando que os exemplos extraídos são de LE, cabe ressaltar que a estrutura das colocações varia de acordo com a língua, pois algumas estruturas presentes em uma língua podem não existir outra, desse modo a tipologia pode variar de um idioma para o outro. A título de exemplificação, como veremos a seguir, em inglês e chinês temos colocações verbais do tipo “verbo (colocado) + partícula adverbial + substantivo (base), o que não encontramos no português, em que não há partículas adverbiais como *out* ou *down*, do inglês. As traduções são meramente para auxiliar o leitor no entendimento dos significados das colocações em LE.

Verbal

Verbo **colocado** + Substantivo **base**

PT: votar projeto de lei

EN: *address the issue* (tratar a questão)

ZH: 受到种族歧视 *shòudào zhǒngzú qíshì* (sofrer racismo, ser discriminado)

Substantivo **base** + Verbo **colocado**

PT: Doença se espalha

EN: *violence escaltes* (violência aumenta)

ZH: 抗议活动升级 *kàngyì huódòng shēngjí* (protesto se intensifica)

Verbo **colocado** + Preposição + Substantivo **base**

PT: Afastar-se do cargo

EN: *be/lie at the root of the problem* (estar na raiz do problema)

Verbo **colocado** + Partícula Adverbial + Substantivo **base**

EN: *hang by a thread* (estar na corda bamba)

ZH: 做出行动 *zuò chū xíngdòng* (tomar uma atitude)

Verbo **colocado** + Adjetivo **base**

PT: deixar abalado

EN: *stay strong* (seguir firme, seguir de pé)

Nominal

Substantivo **base** + Substantivo **colocado**

EN: *Peace plan* (Plano de Paz, no contexto do conflito entre Israel e Palestina)

ZH: 法治轨道 *fǎzhì guǐdào* (caminho do Estado de Direito)

Substantivo **colocado** + Preposição + Substantivo **base**

PT: combate às drogas

EN: *escalation of violence* (aumento da violência)

Adjetival

Adjetivo **colocado** + Substantivo **base**

PT: Transformação social

EN: *Palestinian issue* (questão palestina)

ZH: 非法活动 *fēifǎ huódòng* (atividade ilegal)

Adjetivo **colocado** + Verbo **base**

Este tipo de colocação não apareceu em nosso *corpus* de pesquisa, mas a título de exemplificação, apresentamos dois exemplos dados por Orenha-Ottaiano (2009) após a identificação deste tipo de colocação em uma investigação de colocações especializadas em contratos sociais e estatutos sociais traduzidos no modo juramentado e não-juramentado.

EN: *entitled to vote* (com direito a voto)

eligible to vote/to be elected (apto para voto/ser eleito)

Adverbial

Advérbio **colocado** + Adjetivo **base**

PT: extremamente violento

EN: *aesthetically beautiful* (esteticamente belo)

ZH: 十分严重 *shífēn yánzhòng* (extremamente grave)

Verbo **base** + Advérbio **colocado**

PT: denunciar formalmente

EN: *Look from head to toe* (encarar de cima a baixo)

ZH: 发冷得严重 *fā lěng dé yánzhòng* (fortes calafrios)

Advérbio **colocado** + Verbo **base**

EN: *formally report* (denunciar formalmente)

ZH: 极有可能 *jí yǒu kěnéng* (muito provavelmente)

Na seção 3 a seguir, tratamos dos procedimentos metodológicos por nós usados.

3 PROCEDIMENTOS METODOLÓGICOS

Podemos dividir os procedimentos metodológicos de nossa pesquisa em dois momentos: um primeiro para a compilação do CAT 3 e, um segundo, para extração e análise das colocações. Iniciaremos por tratar da compilação do nosso *corpus* de estudo e, em seguida, detalharemos o processo de extração e seleção das colocações para análise.

3.1 Compilação do *corpus* de estudo

O nosso *corpus* de estudo é um *corpus* multilíngue (português, inglês e chinês) formado por dois *subcorpora* paralelos, cada um formado por dois *subcorpora*, um de TOs e outro de TTs. Todos fazem parte do CAT – *Corpus* de Aprendizes de Tradução, coordenado por Orenha-Ottaiano, do IBILCE/UNESP (ORENHA-OTTAIANO, 2021, 2020, 2017, 2015).

Parte do projeto guarda-chuva “Compilação de Materiais Didáticos e Dicionários Especializados de Colocações Baseados em Corpora (COMADEC)”, o CAT conta com duas fases – O CAT 1 e o CAT 2.

Nossa pesquisa está vinculada ao CAT 1, cujo *corpus* é fruto da compilação de traduções do português para o inglês de aprendizes de tradução. Além de estar vinculado ao COMADEC, o CAT 1 também faz parte do *Multilingual Student Translation Corpus* (MUST), projeto desenvolvido em colaboração com a *Université Catholique de Louvain*, sob direção de Granger e Lefer (GRANGER; LEFER, 2018), que reúne vários *corpora* de diversas direções tradutórias com textos de tradutores aprendizes de todo o mundo. No Brasil, o MUST é coordenado por Orenha-Ottaiano.

O CAT 1 é o *corpus* brasileiro usado em nossa pesquisa, compilado por Orenha-Ottaiano (2019; 2012a), tendo tido também a colaboração de Rocha (2020) na segunda fase da compilação. Essa versão do CAT, que alimentou o MUST, teve alguns parâmetros, além daqueles previamente determinados pela Direção do MUST, definidos pela Coordenadora do projeto na UNESP (ORENHA-OTTAIANO, 2019), tais como: direção tradutória (português-inglês), gênero textual (textos jornalísticos) e participantes (aprendizes de tradução). Além de usarmos um recorte do CAT 1, também nos debruçamos sobre um *subcorpus* com traduções feitas por estudantes chineses avançados de português do CAT 3. Porém, como até a nossa inserção no projeto, o CAT contava apenas com os *corpora* na direção L1 (português) → L2 (inglês), ficou

sob nossa responsabilidade a compilação da versão chinesa do projeto, na direção L1 (chinês) → L2 (português), para formar o CAT 3.

Feita com o fim específico de inicialmente servir a esta pesquisa, apresentaremos os pormenores do processo de sua compilação mais adiante. Mas antes, discorramos sobre a compilação do CAT brasileiro, que forneceu os parâmetros para o nosso trabalho.

3.1.1 O CAT brasileiro

Originalmente, o CAT 1, cuja direção de tradução é L1 (português) → L2 (inglês), conta com vinte textos originais que foram traduzidos pelos aprendizes de tradução voluntários, conforme abaixo especificado:

- **Direção tradutória**

Como anteriormente mencionado, o projeto MUST prevê a construção de um grande *corpus* paralelo formado por traduções de aprendizes de tradução. A direção tradutória do componente brasileiro do *corpus* foi L1(português) → L2 (inglês). Essa escolha, também baseada nos parâmetros para a compilação do *corpus* pela Coordenação do projeto, se deu pelo fato de os alunos encontrarem mais dificuldade na produção de textos em língua estrangeira, sendo, portanto, a direção mais complexa (ROCHA, 2020).

- **Gênero textual**

A respeito do gênero, todos os textos são textos jornalísticos, sejam eles notícias ou artigos de opinião publicados em portais de comunicação *on-line* atuais, como Estadão¹³, IstoÉ¹⁴, Terra¹⁵, entre outros, com registro padrão do português brasileiro contemporâneo. Os temas são variados, tratando de política nacional e internacional, ativismo LGBTQIAP+, meio ambiente, educação, terrorismo e arte (ROCHA, 2020; ORENHA-OTTAIANO, 2019, 2012a, 2012b).

¹³ <https://www.estadao.com.br/>. Acesso em 1 de novembro de 2021.

¹⁴ <https://istoe.com.br/>. Acesso em 1 de novembro de 2021.

¹⁵ <https://www.terra.com.br/>. Acesso em 1 de novembro de 2021.

- **Participantes**

Os textos em língua portuguesa foram traduzidos para o inglês por alunos universitários brasileiros, regularmente inscritos no terceiro ano (5º- 6º semestre/período) dos cursos de Bacharelado em Letras com Habilitação de Tradutor – Inglês e Licenciatura em Letras do IBILCE/UNESP, no noroeste paulista. Para satisfazer as exigências do MUST e do CAT, o nível de proficiência dos voluntários deveria estar entre intermediário (B1) e avançado (C1). Segundo o Quadro Europeu Comum de Referências para Línguas (*Common European Framework of Reference for Languages* – CEFR), amplamente difundido no ensino de línguas estrangeiras ao redor do mundo, esses níveis caracterizam um usuário do idioma estrangeiro como de independente a proficiente, capaz de compreender textos complexos, longos e exigentes, além de se expressar fluente e espontaneamente com nativos e produzir textos sobre temas variados para diferentes fins sociais. (COUNCIL OF EUROPE, 2001)

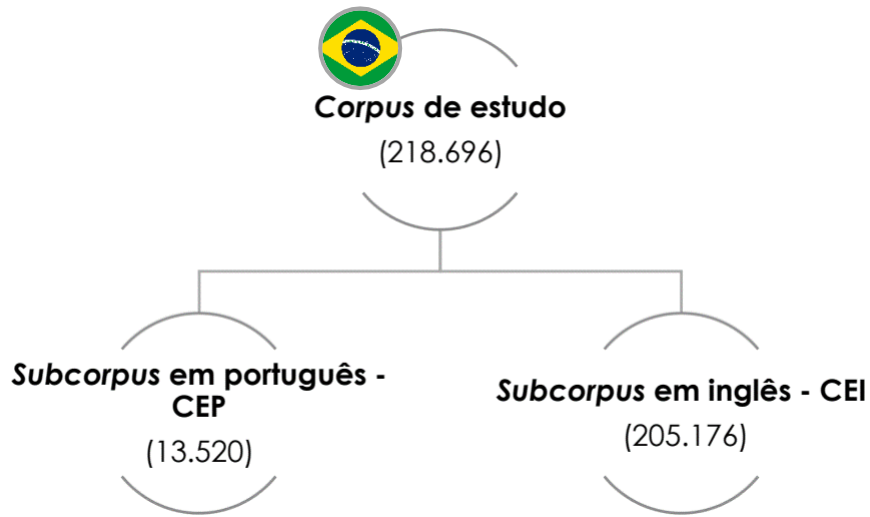
O número total de participantes do lado brasileiro foi de 21 voluntários, porém, nem todos concluíram todas as traduções, cujo tempo máximo para conclusão de cada uma era duas horas.

Para a realização das atividades de tradução, os voluntários puderam usar recursos como dicionários, memórias de tradução, buscadores *on-line*, entre outros.

A inclusão dos alunos de Letras se deu graças ao grande número de desistências dos alunos de Tradução, acrescido ao fato de o MUST também incluir na sua base de dados textos de estudantes de línguas estrangeiras, não apenas de estudantes de tradução.

A compilação do *corpus* brasileiro foi concluída em 2017 e resultou em dois *subcorpora*, um em português, de TOs, e um inglês, de TTs, totalizando 218.696 palavras em 302 traduções, como esquematizado a seguir.

Organograma 1 – Organização do CAT 1 – O *corpus* brasileiro



Fonte: Rocha (2020, p.67)

Não utilizamos o CAT 1 em toda sua extensão, mas apenas um recorte com as traduções para dez dos vinte TOs. Feito esse recorte, obtivemos um *subcorpus* de TOs em português de 5.028 palavras e um *subcorpus* de TTs em inglês de 144.388 palavras.

Sabendo como foi feita a compilação do *corpus* brasileiro usado em nossa pesquisa, vejamos agora como procedemos na compilação da versão chinesa do CAT.

3.1.2 O CAT chinês

Como já mencionamos, para mantermos o rigor nas comparações, pautamo-nos pelos mesmos parâmetros usados para a compilação do *corpus* brasileiro (CAT 1), que compõe o MUST, para fazermos a compilação do CAT 3, o *corpus* chinês. Seguindo os mesmos procedimentos para a compilação daquele *corpus*, o primeiro passo seguido foi a seleção dos TOs para compor o *subcorpus* de textos originais, respeitando-se as exigências em relação à direção tradutória, ao gênero textual, e aos participantes.

- **Direção tradutória**

Inicialmente, tínhamos a intenção de não apenas analisar as colocações no sentido L1(chinês) → L2 (português), mas também na direção L2 (português) → L1(chinês). Por esse motivo, compilamos também um *corpus* em que os voluntários chineses traduziram os TOs em português para sua língua materna, o chinês mandarim. No entanto, por questões de adequação metodológica, neste trabalho, tomamos a decisão de analisar apenas o *corpus* em que os alunos traduziram na direção inversa, da L1 (chinês) para a L2 (português).

- **Gênero textual**

Para o CAT 3 L1(chinês) → L2(português), foram selecionados, no total, dez textos com extensão média de 600 palavras cada, extraídos de portais *on-line* de grandes veículos de comunicação da China, que deveriam ser traduzidos para o português. O texto mais curto desse grupo continha 586 palavras e o mais longo, 1386.

- **Participantes**

Os textos foram traduzidos por 18 alunos - seis deles dos últimos semestres do curso de Letras – Português do Instituto de Comunicações de Hebei (HEBIC), no norte da China, e 12 do curso de Letras – Português da Universidade de Hubei (HUBU), na China central. Em consonância com os parâmetros de compilação do *corpus* brasileiro, cada texto do chinês foi traduzido para o português em no máximo duas horas.

Visando manter uma diversidade dos temas dos textos a serem traduzidos, os textos em chinês, assim como os textos em português, apresentam uma grande variedade de temas abordados. Buscamos selecionar textos que possivelmente despertariam o interesse dos participantes da pesquisa, seja pelo elemento novidade, seja pela proximidade com a cultura e realidade brasileiras.

Além dos parâmetros apresentados na seção anterior, no tocante aos temas, especificamente, a seleção dos textos do *subcorpus* de TOs seguiu critérios como:

- **Atualidade do tema**

No período da seleção dos TOs em português para compor o CAT 1, em 2017, os artigos selecionados, segundo os parâmetros do projeto (ORENHA-OTTAIANO, 2019), deveriam tratar de assuntos atuais àquela época (ROCHA, 2020; ROCHA; VIANA; ORENHA-OTTAIANO, 2020). Para seguirmos o mesmo padrão e mantermos uma uniformidade na compilação do *subcorpus* de TOs em chinês, selecionamos dez matérias sobre alguns dos temas mais comentados, veiculadas em portais de grande circulação na China, como o *Xinhua*¹⁶, *China Daily*¹⁷, *People's Daily*¹⁸, entre outros.

- **Diversidade dos temas tratados**

Os primeiros textos selecionados no *corpus* brasileiro tinham como temas principais política nacional e internacional, meio ambiente, cultura e sociedade. Seguindo a mesma linha, optamos por selecionar em chinês textos que também tivessem como tema política, meio ambiente e cultura e sociedade. Sem deixar de lado os assuntos mais comentados no período da compilação, foram também selecionados textos dos cadernos de guerra e assuntos militares que, por um lado, se enquadrariam na proposta de apresentar “política nacional e internacional”, embora não fosse um tema tão recorrente nos noticiários brasileiros, não da forma explícita como é no país asiático, onde há cadernos e colunas só para o assunto. Outro tema muito recorrente na ocasião era a pandemia da recém-descoberta COVID-19. Sob a premissa de selecionar matérias atuais, não pudemos deixar de selecionar textos sobre o assunto.

- **Restrições da censura**

Alguns dos textos do *corpus* brasileiro tratavam sobre homossexualidade, drogas, racismo e outros temas considerados controversos pelo público chinês. Ainda que houvesse a intenção de apresentar textos sobre assuntos pouco conhecidos pelos alunos a fim de observar como lidariam com as colocações nesses casos, optamos por não selecionar textos que tratassem de alguns desses temas; mesmo alguns textos que acreditávamos não apresentar problema à censura, acabaram por ter seu acesso restringido pelos aplicativos usados durante o processo de compilação.

¹⁶ <http://news.cn/>. Acesso em 1 de novembro de 2020.

¹⁷ <https://cn.chinadaily.com.cn/>. Acesso em 1 de novembro de 2020.

¹⁸ <http://www.people.com.cn/>. Acesso em 1 de novembro de 2020.

Apesar de todos os TOs em chinês terem sido extraídos de meios de comunicação da China, alguns deles, posteriormente, tiveram seu acesso bloqueado pelo *Yuketang*, aplicativo que usamos para disponibilizar os TOs a serem traduzidos, conforme trataremos mais adiante, sobretudo aqueles que tinham como assunto principal racismo e violência. Cabe observar que alguns alunos que adiaram a tradução dos textos sobre esses temas não puderam realizá-la em um segundo momento, devido a essa censura e ao bloqueio. No mais, textos sobre racismo, em especial aqueles que tratavam de racismo na China, foram recusados por alguns alunos, que não concordaram com o conteúdo das matérias. Além disso, houve alguns casos em que os alunos que se dispuseram a traduzir os textos, mesmo que contrariados, acrescentaram notas ou parágrafos inteiros afirmando não concordarem com a opinião dos autores do texto, pois, segundo eles, elas não eram verdadeiras, como nos exemplos extraídos do *corpus* abaixo.

Figura 5 – Captura de tela do *Yuketang*¹⁹



Fonte: Captura de tela do *Yuketang*, versão para dispositivos móveis.

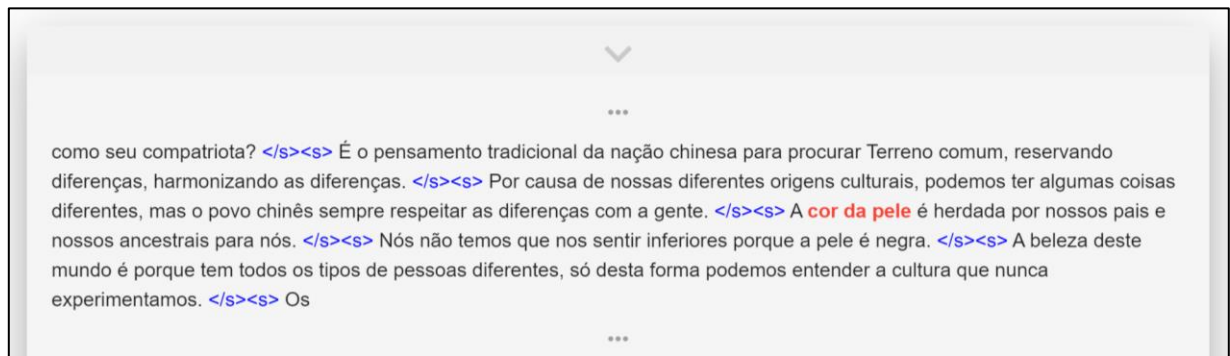
¹⁹ Tradução: O conteúdo que você está tentando acessar contém palavras sensíveis: aplicação violenta da lei. Por favor, verifique-as e tente novamente.

Figura 6 – Conversa com voluntário pelo Wechat



Fonte: captura de tela do *Wechat*, versão para dispositivos móveis.

Figura 7 – Trecho acrescido a uma das traduções contendo a opinião do voluntário



Fonte: captura de tela de expansão de linha de concordância no SkE.

A seguir, apresentamos uma tabela com os temas, títulos e número de palavras dos textos originais traduzidos do chinês para o português.

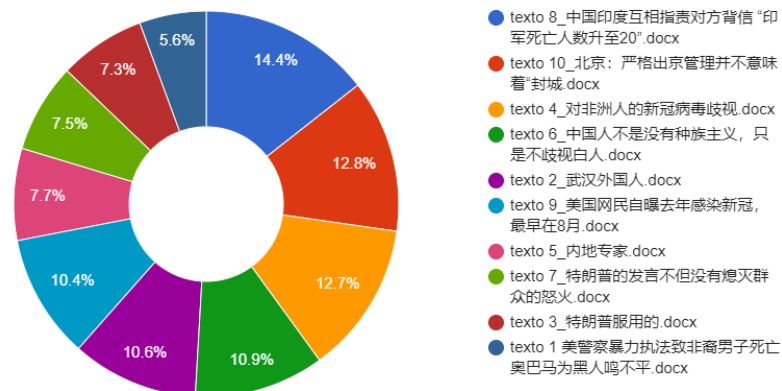
Tabela 4 – Temas, títulos e números de palavras dos TOs do CAT chinês

NÚMERO DO TEXTO	TEMA	TÍTULO ORIGINAL EM CHINÊS	TÍTULO EM PORTUGUÊS (TRADUÇÃO NOSSA)	NÚMERO DE PALAVRAS
Texto 11	Cultura e sociedade, violência policial contra negros nos EUA	美警察暴力执法致非裔男子死亡奥巴马为黑人鸣不平	Violência policial americana mata afro-americano, Obama se pronuncia em favor dos negros	371
Texto 12	Cultura e sociedade, estrangeiros na China	武汉的外国留学生 为什么选择留在这里?	Estudantes internacionais de Wuhan: por que escolheram ficar aqui?	702
Texto 13	Política internacional, Governos Trump e Bolsonaro e a pandemia de COVID-19	特朗普服用的“抗疫神药”被叫停后 超 200 万片将发往巴西	Após suspensão, mais de 2 milhões de comprimidos do "medicamento milagroso contra Covid" de Trump serão enviados ao Brasil	479
Texto 14	Cultura e sociedade, racismo seletivo na China	对非洲人的新冠病毒歧视	Racismo contra africanos causado pelo novo coronavírus	841
Texto 15	Política nacional e assuntos militares, revoltas em Hong Kong	内地专家: 解决当前香港问题的唯一途径是回到法治轨道	Especialistas do continente: a única maneira de resolver os problemas atuais em Hong Kong é retornar ao Estado de Direito	508
Texto 16	Cultura e sociedade, racismo seletivo na China	中国人不是没有种族主义, 只是不歧视白人	Não é que não haja racismo na China, só não há contra os brancos	721
Texto 17	Política internacional, Governo Trump e pandemia de COVID-19	美国以色列不宣而战, 大批战机展开空袭, 白宫真要开战转移危机?	Estados Unidos e Israel não declaram guerra, mas grande contingente lança ataques aéreos. A Casa Branca realmente entrará em guerra para desviar a crise?	493
Texto 18	Política internacional, conflitos entre China e Índia	中国印度互相指责对方背信 “印军死亡人数升至 20”	China e Índia se acusam mutuamente de traição na fronteira. "O número de mortos do exército indiano sobe para 20"	953
Texto 19	Cultura e sociedade, Estados Unidos e a pandemia de COVID-19	美国网民自曝去年感染新冠, 最早在 8 月	Internautas dos EUA revelam que foram infectados com o novo coronavírus em agosto do ano passado	688
Texto 20	Cultura e sociedade, Nova onda de infecções por COVID-19 em Pequim	北京: 严格出京管理并不意味着“封城”	Pequim: uma gestão rigorosa na entrada e saída não significa "fechar a cidade"	846
TOTAL				6602

Fonte: elaborado pelo autor.

A porcentagem de cada texto no total de TOs pode ser observada no Gráfico 1.

Gráfico 1 – Porcentagem de cada TO sobre o total de TOs do CAT chinês



Fonte: captura de tela do SkE.

Depois de delimitado o primeiro *subcorpus* da versão chinesa do projeto, o de TOs em chinês, pudemos iniciar a compilação do *subcorpus* formado pelas traduções para o português. Devido ao período crítico de isolamento social em razão da crise do COVID-19, tivemos de buscar alternativas para recrutar voluntários e disponibilizar e receber textos que mantivessem os parâmetros e exigências do projeto. Discorremos a esse respeito na subseção 3.1.3.

3.1.3 Alternativas metodológicas para a compilação de um corpus de aprendizes em meio a pandemia de COVID-19

Iniciamos a compilação do *corpus* chinês em abril de 2020, data que coincidiu com um período de grandes dificuldades em escala global causadas pela propagação do novo coronavírus, o SARS-CoV-2, causador da infecção COVID-19, que forçou governos e organizações a tomarem medidas restritivas em todos os níveis. Se em contextos ditos normais, quando não se está em sala de aula, a compilação de um *corpus* de aprendizes já apresenta grandes problemas como, por exemplo, congregar um número razoável de alunos voluntários e mantê-los motivados até o final do projeto (OLIVEIRA, 2019; ROCHA, 2020), em um contexto como o nosso, que requer isolamento social e quarentena em muitos locais, essa tarefa se tornou ainda mais complicada, principalmente porque uma das universidades chinesas onde os nossos alunos estão inscritos, a HUBU, se localiza em Wuhan, cidade chinesa onde foram registrados os primeiros casos da COVID-19 em dezembro de 2019 e primeira cidade a ser posta em quarentena pelo

Governo Chinês em 23 de janeiro de 2020. À universidade em si, que abrigou pacientes infectados pelo novo vírus, foram impostas restrições de entrada até setembro de 2020.

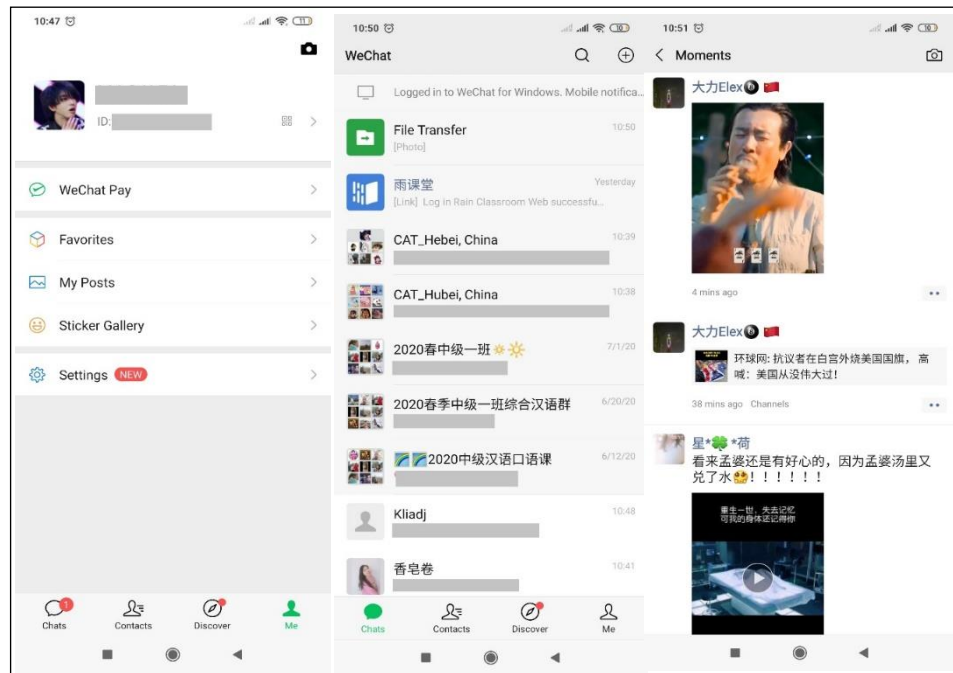
Dadas as circunstâncias a nós impostas, o processo de compilação do *corpus* foi inteiramente remoto, desde a assinatura dos termos de consentimento até o preenchimento de um questionário final, elaborado com o intuito de conhecer melhor o perfil dos participantes da pesquisa, com os quais nunca tivemos contato pessoalmente nem antes e nem durante a pesquisa devido às medidas de combate ao vírus. Para se obter um processo padronizado, em cada etapa dos trabalhos de construção do *corpus*, usamos ferramentas *on-line* específicas de baixo ou nenhum custo e disponíveis na *Internet* chinesa, já que muitas das ferramentas gratuitas conhecidas no Ocidente e que se mostraram eficazes em iniciativas similares, como as oferecidas pela *Google* (OLIVEIRA, 2019), tem seu uso proibido na China.

A seguir, nos itens de A a G, detalhamos as etapas que percorremos para proceder com a compilação deste *corpus* e as ferramentas utilizadas em cada uma delas.

A) Recrutamento dos alunos pelo *Wechat*

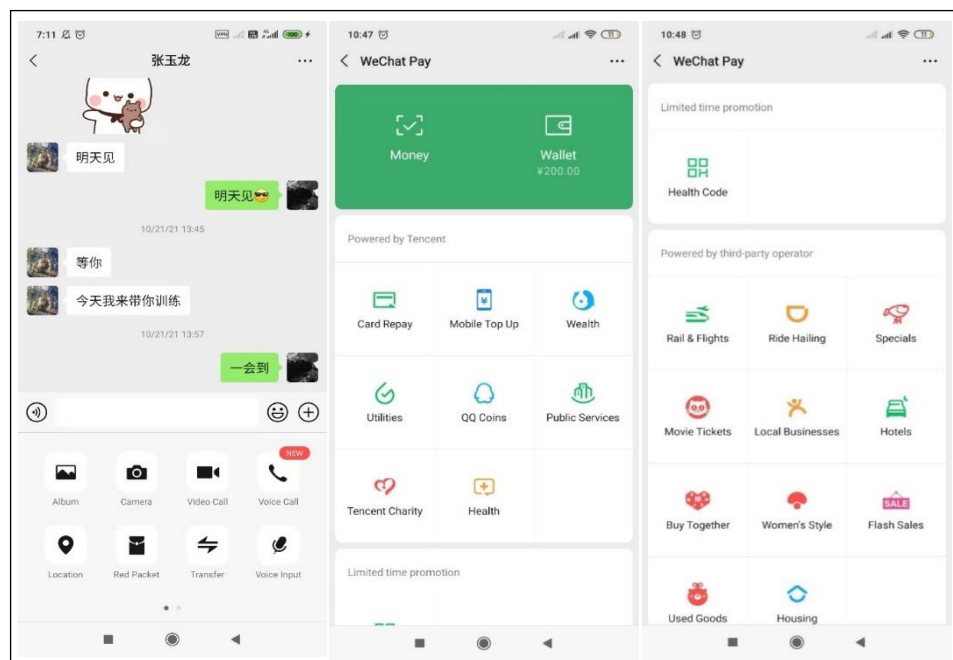
Após o contato do pesquisador deste trabalho com professores titulares dos departamentos de português do HEBIC e da HUBU, foram obtidos os contatos dos alunos voluntários na rede social chinesa *WeChat* (Figura 9), que é, de modo geral, um aplicativo de celular híbrido entre as conhecidas redes sociais ocidentais *WhatsApp*, *Facebook* e *Instagram*, com funcionalidades que vão desde chamadas de vídeo a pagamento e transferências bancárias (Figura 10). Uma vez obtidos os contatos de todos os alunos, foram criados grupos para cada universidade, de forma a permitir a difusão das informações e dos comandos com mais facilidade (Figura 11).

Figura 8 – Interface do *Wechat*, versão para dispositivos móveis



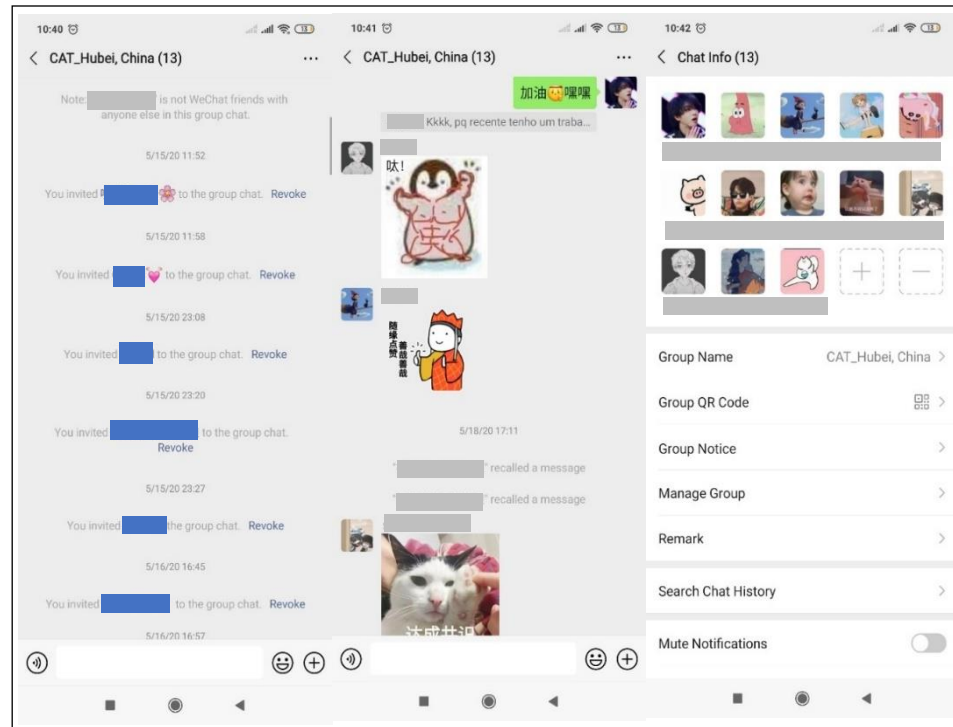
Fonte: capturas de tela do *Wechat* gerada pelo autor.

Figura 9 – Interface com as várias funcionalidades do *Wechat*



Fonte: capturas de tela do *Wechat* gerada pelo autor.

Figura 10 – Grupos do *Wechat* com os participantes da pesquisa



Fonte: capturas de tela do *Wechat* gerada pelo autor.

Depois de reunirmos todos os alunos nos grupos do *WeChat*, foi lhes solicitada a assinatura do termo de consentimento, como vemos no item B a seguir.

B) Assinatura do termo de consentimento

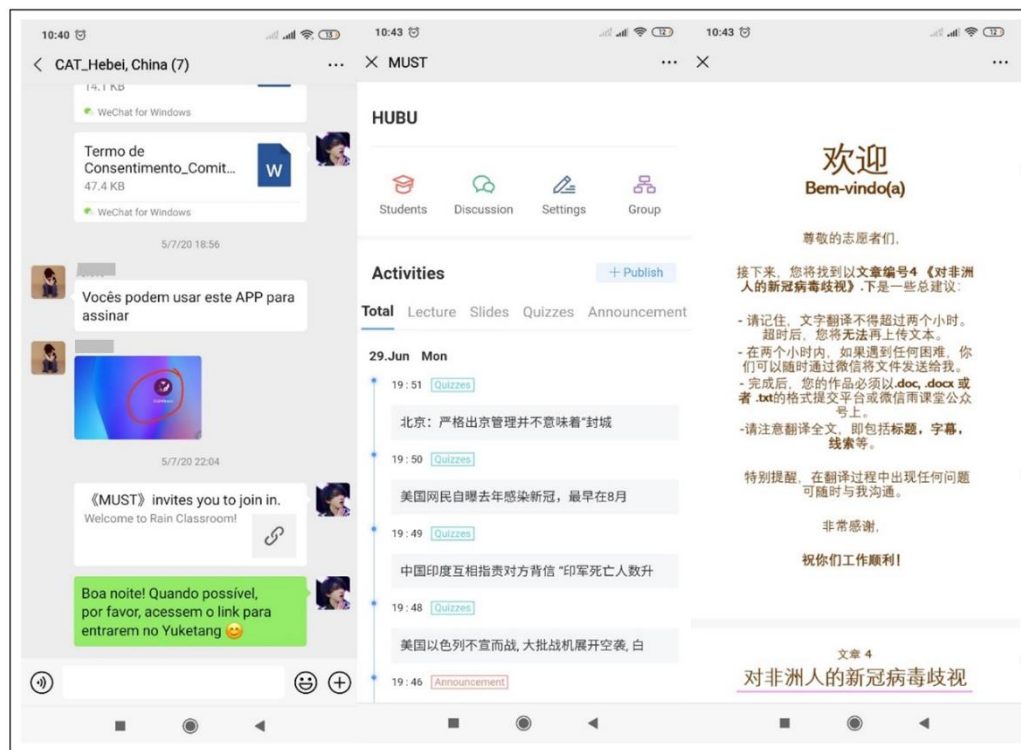
Com todos os alunos cientes do tema da pesquisa e de suas obrigações como voluntários, solicitamos a assinatura do Termo de Consentimento Livre e Esclarecido – TCLE, documento padrão para a compilação do CAT segundo foi estabelecido pela coordenação do projeto e encaminhado ao Comitê de Ética da UNESP²⁰. O termo segue no Anexo A.

²⁰ O corpus foi aprovado pelo Comitê de Ética em Pesquisa do IBILCE/UNESP, em 01/04/2019, CAAE nº 94053718.4.0000.5466, intitulado *Corpus de Aprendizes de Tradução (CAT) e Multilingual Student Translation Corpus - MUST*, sob responsabilidade da Prof^a. Dr^a. Adriane Orenha-Ottaiano. Do lado chinês, segundo fomos informados em conversas informais com docentes da HUBU, o consentimento dos alunos bastava para a compilação do *corpus*.

C) Inscrição na sala de aula virtual *Yuketang*

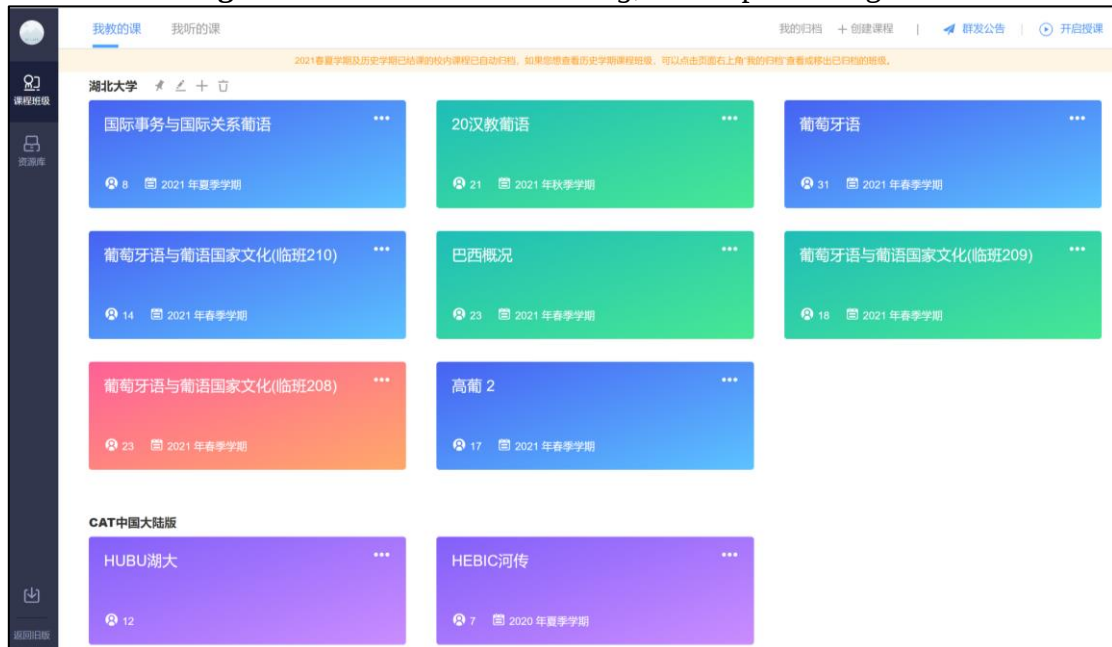
Com todos os TCLE assinados, os alunos foram convidados a se inscreverem na sala virtual *Yuketang* (雨课堂 *Yǔ kètáng*, em chinês, também conhecido pelo nome inglês *Rain Classroom*) (Figuras 11 e 13), uma plataforma hospedada dentro do *WeChat*, que permite aos usuários ministrarem e assistirem aulas por meio de apresentações no estilo *PowerPoint* personalizadas com registro de voz. No *Yuketang*, ainda é possível criar turmas, aplicar provas e dialogar com os alunos em tempo real. O professor ainda tem à disposição um painel de controle que permite controlar a assiduidade dos alunos, publicar anúncios e receber atividades.

Figura 11 – Interface do *Yuketang*, versão para dispositivos móveis



Fonte: capturas de tela do *Yuketang* gerada pelo autor.

Figura 12 – Interface do *Yuketang*, versão para navegador



Fonte: captura de tela do *Yuketang* gerada pelo autor.

Após todos os participantes da pesquisa terem se inscrito na sala virtual *Yuketang*, os TOs começaram a ser disponibilizados, como vemos no item D.

D) Envio dos TOs a serem traduzidos

Os textos a serem traduzidos foram carregados semanalmente, de três em três, no *Yuketang* no formato de provas (Figura 12), para que fosse possível controlar o tempo de duas horas para cada tradução. Os textos foram disponibilizados na plataforma por tempo indeterminado, para que os alunos pudessem acessá-los quando lhes fosse mais conveniente, dentro do prazo de doze semanas. Porém, uma vez aberto, o relógio de cada texto iniciava a contagem do tempo e, completadas as duas horas, os alunos perderiam a visualização do texto.

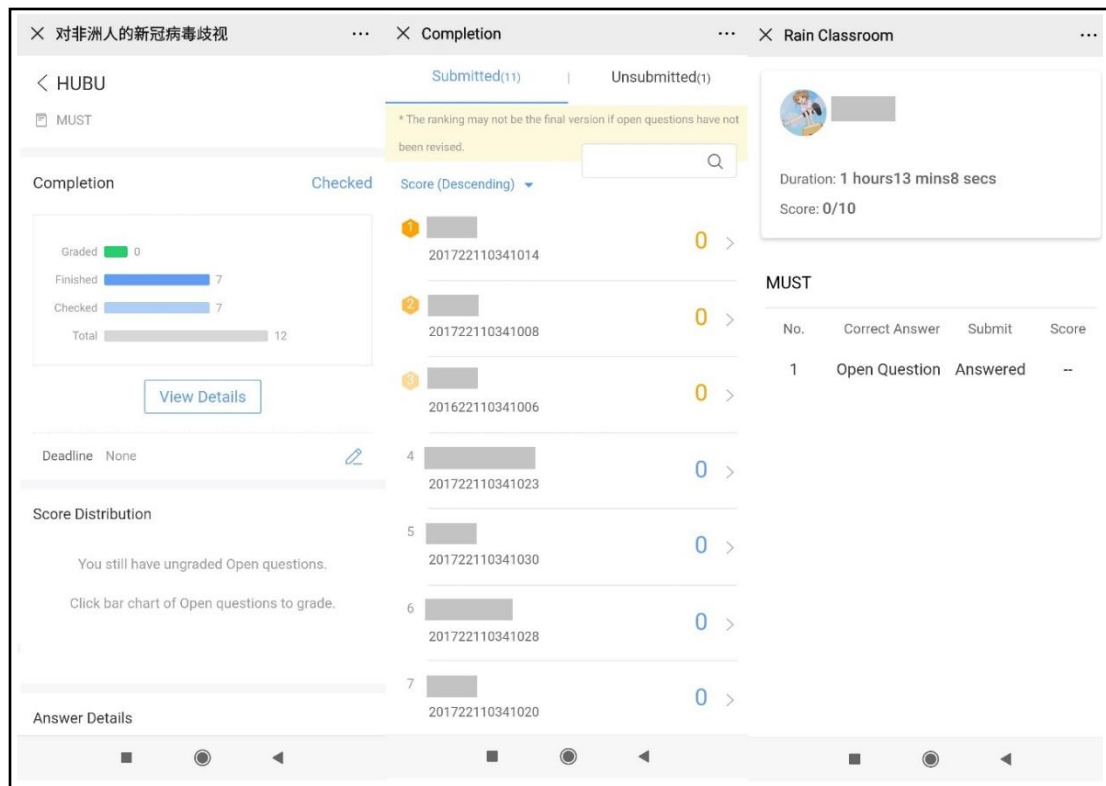
Como nosso intuito é avaliar o desempenho real dos aprendizes no tangente às fraseologias, e como sabemos que, em um contexto de trabalho remunerado, o tradutor se valerá de todos os recursos que encontrar disponíveis para realizar seu trabalho dentro do período estipulado, entendemos que permitir o uso de ferramentas de tradução nos daria um *feedback* mais fidedigno do que pode ser encontrado nas traduções que os alunos produzem cotidianamente. Dessa maneira, seguindo os parâmetros estabelecidos para a compilação do *corpus* português→inglês

(ROCHA, 2020; ORENHA-OTTAIANO, 2019, 2012a), os alunos chineses ficaram livres para usar quaisquer ferramentas que julgassem necessárias, ou seja, memórias de tradução, memórias de tradução neurais, dicionários, buscadores, entre outros. Os dados acerca do número de alunos que usaram tais ferramentas e, mais especificamente, quais delas encontram-se no Gráfico 6, na subseção 3.1.4, sobre o perfil dos voluntários chineses.

E) Recebimento das traduções no *Yuketang* e armazenamento no *Tencent Cloud*

Ao término de cada tradução, foi solicitado que os alunos as carregassem no formato *.doc* ou *.docx* na área destinada à resposta no final de cada “prova”. Como alguns dos alunos se encontravam isolados em outros países, sem acesso a um computador, oferecemos a possibilidade de digitar o texto traduzido diretamente na área de respostas, sem o intermédio de um editor de textos. Como as respostas dos alunos podem ser visualizadas somente pelo professor na versão do *Yuketang* para computador, a fim de facilitar a visualização dos textos tanto no dispositivo móvel como no computador, os textos foram redirecionados para o *Tencent Cloud* (腾讯云 *Téngxùn wēi yún*), alternativa chinesa para os armazenadores na nuvem.

Figura 13 – Interface do *Yuketang*, versão para dispositivos móveis após recebimento dos textos dos voluntários



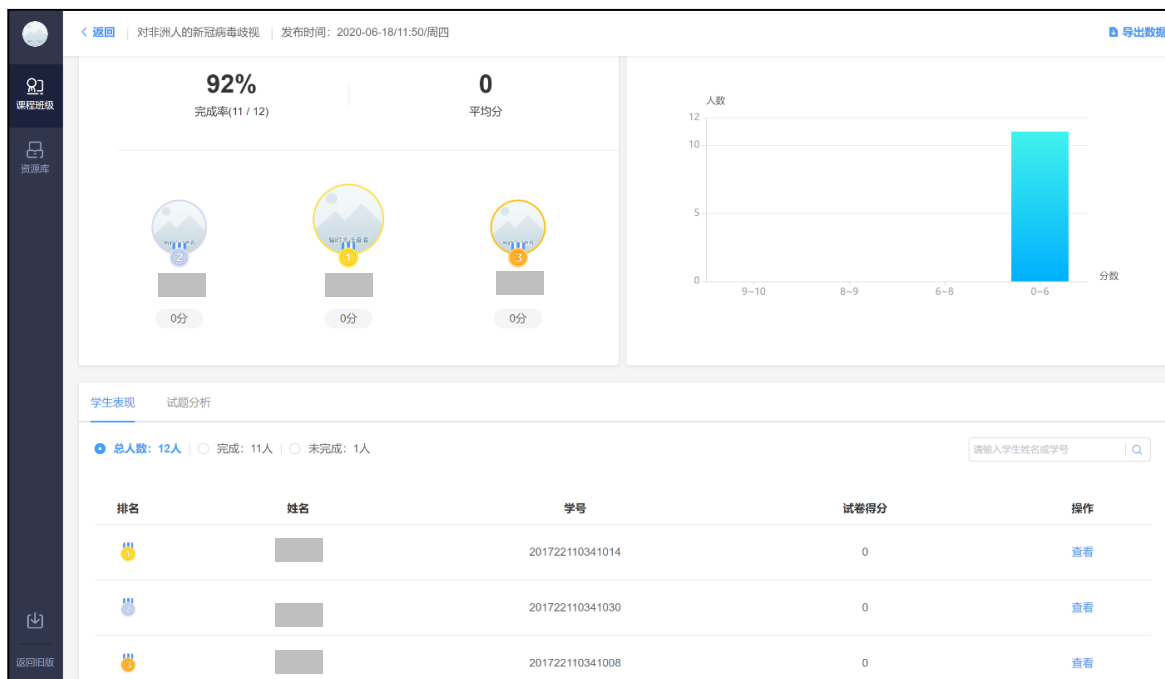
Fonte: capturas de tela do *Yuketang* gerada pelo autor.

Figura 14 – Interface do *Yuketang*, versão para navegador após envio dos TOs aos voluntários



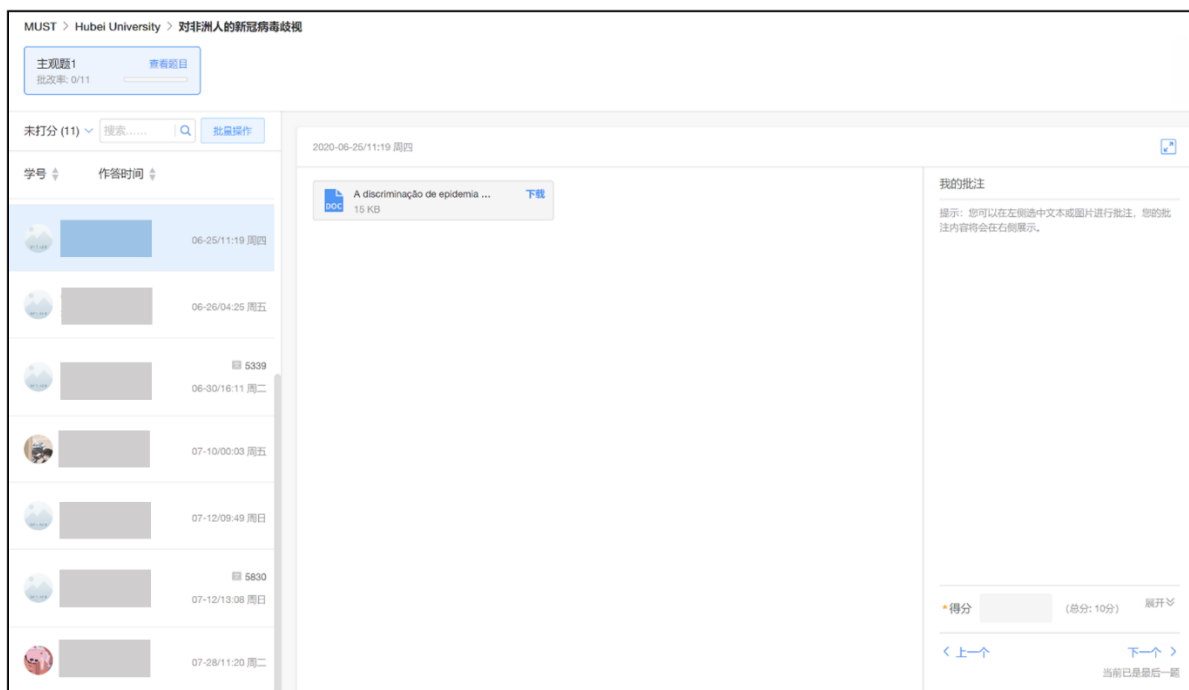
Fonte: capturas de tela do *Yuketang* gerada pelo autor.

Figura 15 – Interface do Yuketang, versão para navegador após recebimento dos TTs



Fonte: capturas de tela do Yuketang gerada pelo autor.

Figura 16 – Interface do Yuketang, versão para navegador após recebimento dos TTs



Fonte: capturas de tela do Yuketang gerada pelo autor.

Depois de coletados, todos os textos passaram por um tratamento, conforme segue no item F.

F) Tratamento dos arquivos

Após o recebimento, os textos passaram por um tratamento para posteriormente serem carregados no SkE.

O primeiro passo foi a ocultação dos nomes dos participantes e alteração dos nomes dos arquivos .doc e .docx. Para tanto, renomeamos todos os arquivos com um código, como o do exemplo abaixo:

1 2 3 4
[HEBIC_ZH-PT_16_T6]

A primeira parte (1) do código sinaliza de qual universidade é o participante, em que:

- UNESP= Universidade Estadual “Júlio de Mesquita Filho”
- HUBU= Universidade de Hubei
- HEBIC= Instituto de Comunicações de Hebei

A segunda parte (2) refere-se à direção tradutória, sendo possível duas combinações:

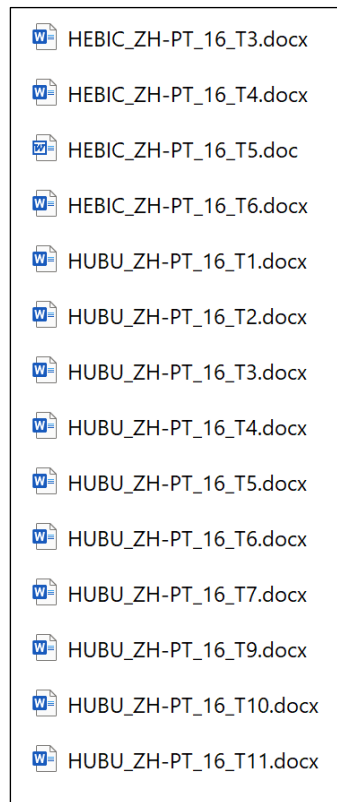
- PT-EN
- ZH-PT

Em que “PT” refere-se ao português, “EN” ao inglês, e “ZH” ao chinês.

A terceira parte (3) é um número de identificação de 1 a 20 atribuído a cada TO. Os números de 1 a 10 correspondem aos TOs em português e de 11 a 20 aos TOs em chinês.

A última parte é um código de identificação atribuído arbitrariamente a cada voluntário, em que “T” é “Tradutor” seguido de um número de 1 a 18.

Por fim, obtivemos as traduções dispostas da seguinte forma:

Figura 17 – Organização dos TOs

Fonte: elaborado pelo autor.

Finalizada a coleta e o tratamento de todos os textos, seguimos para a última etapa do processo de compilação.

G) Conclusão das atividades e questionário final com auxílio do 腾讯调查问卷 (Formulários Tencent)

Depois de transcorridos quase três meses, os alunos concluíram todas as traduções e, para se obter mais informações sobre o perfil dos alunos, suas formações e a forma como realizaram as traduções, aplicamos um questionário com 28 perguntas na plataforma 腾讯调查问卷 *Téngxùn diàochá wènjuàn*, (Formulários Tencent – tradução nossa). Além de perguntas sobre a quantidade de anos dedicados ao estudo do português e vivência em países de língua portuguesa, perguntamos se, em algum momento de sua formação, eles tiveram contato com a LC e/ou *corpora*, se já haviam estudado Teorias da Tradução, entre outras perguntas. O questionário completo está disponível no Apêndice A, na parte final deste trabalho.

Vale ressaltar aqui que, apesar de o curso dos alunos chineses ser de Letras, todos têm em sua grade curricular uma disciplina de prática de tradução (笔译 *bǐyì*) e outra de prática de interpretação (口译 *kǒuyì*) nos anos finais.

Figura 18 – Questionário final sobre o perfil dos alunos na plataforma 腾讯调查问卷

Questionário final

Prezado voluntário, chegamos à conclusão dos trabalhos de tradução! Obrigado pela sua participação. Antes de encerrarmos, gostaria de pedir que você...

01 Você é aluno(a) de qual universidade? *

☐ Universidade de Hubei

☐ Instituto de Comunicações de Hebei

02 Você é aluno(a) de qual semestre do curso de graduação? *

☐ Primeiro semestre do terceiro ano

☐ Segundo semestre do terceiro ano

☐ Primeiro semestre do quarto ano

☐ Segundo semestre do quarto ano

☐ Já me formei

03 Quantos anos você tem? *

Please select

04 Você se considera como pertencente ao sexo? *

☐ Masculino

☐ Feminino

☐ Não-binário

☐ Prefiro não responder

Next

Copyright © 1998-2020 Tencent
The system is provided by Tencent Questionnaire
Disclaimer Service agreement Privacy policy Report the questionnaire

Fonte: captura de tela do 腾讯调查问卷

Com a aplicação do questionário, pudemos traçar melhor o perfil dos participantes voluntários do CAT chinês, sobre o qual tratamos mais detalhadamente a seguir.

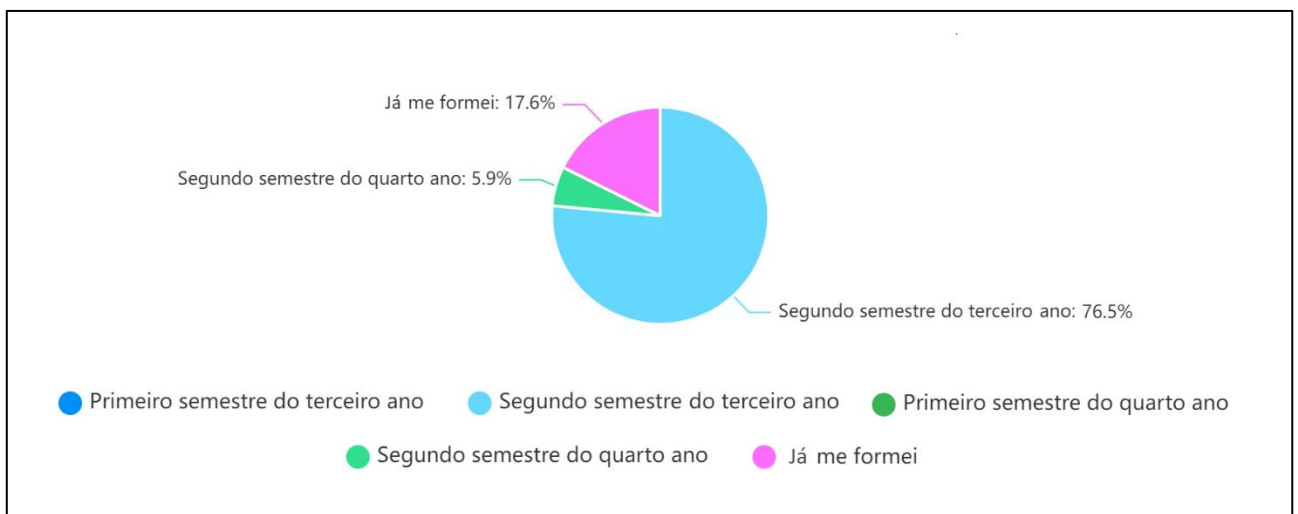
3.1.4 O perfil dos voluntários chineses

Para melhor entendermos a dinâmica de tradução dos participantes, e já pensando em nos precaver diante de resultados inesperados, aplicamos um questionário de 28 questões na plataforma chinesa 腾讯问卷 (Formulários *Tencent*, tradução nossa), da gigante *Tencent*.

Foram convidados 22 alunos para participarem do projeto, mas, ao final, 18 seguiram na pesquisa até a conclusão e 17 responderam ao questionário. Destes, 11 (64,7%) eram alunos do Curso de Letras – Português da HUBU e 6, do HEBIC (35,3%).

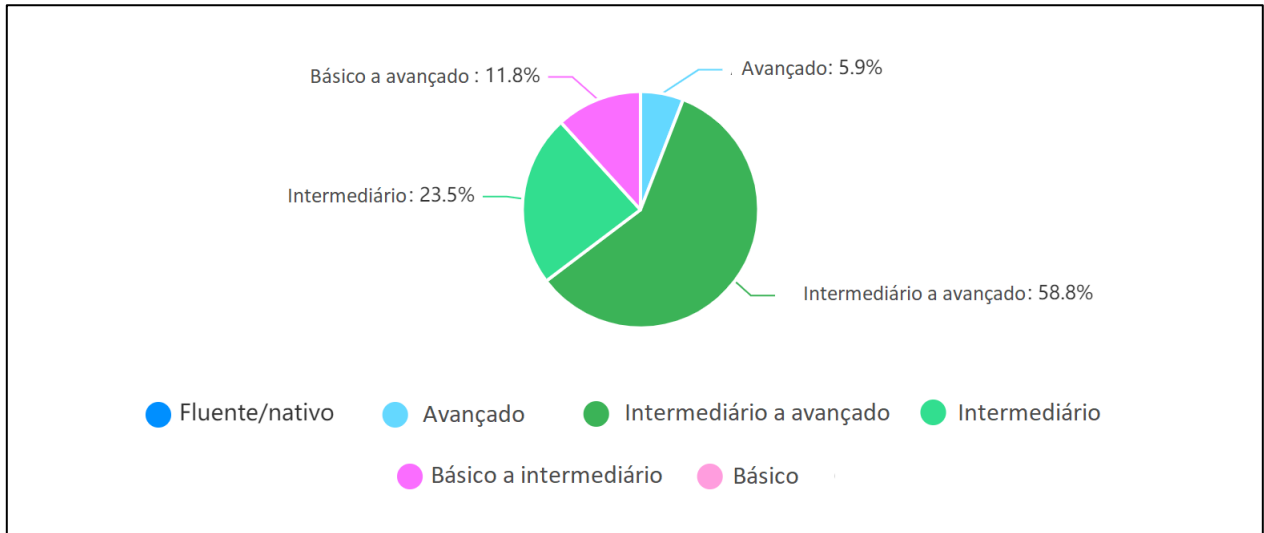
Do total de alunos, todos eram alunos do segundo semestre do terceiro (76,5%) ou quarto ano (5,9%), além de três alunos recém-formados (17,6%). Essa composição atende às exigências quanto à proficiência dos alunos, pois, de acordo com a proposta de classificação da Embaixada de Portugal em Pequim na promoção dos Prêmios Tomás Pereira – 2020, alunos dos terceiros e quartos anos da graduação em português teriam nível de proficiência entre B1 e C2 segundo o Quadro Europeu, o que condiz com a forma como os alunos reconhecem seu próprio nível de proficiência. Em relação aos voluntários, 88,2% deles, o que corresponde a 15 alunos, classificam seu português como “intermediário”, “intermediário avançado” ou “avançado”. Apenas dois alunos, por falta de segurança, mencionaram possuir nível “básico a intermediário”, mesmo que o currículo que cursavam na universidade os apontava como, no mínimo, avançado.

Gráfico 2 – Nível de formação dos voluntários chineses



Fonte: elaborado pelo autor.

Gráfico 3 – Autopercepção do nível de proficiência em português dos alunos chineses



Fonte: elaborado pelo autor.

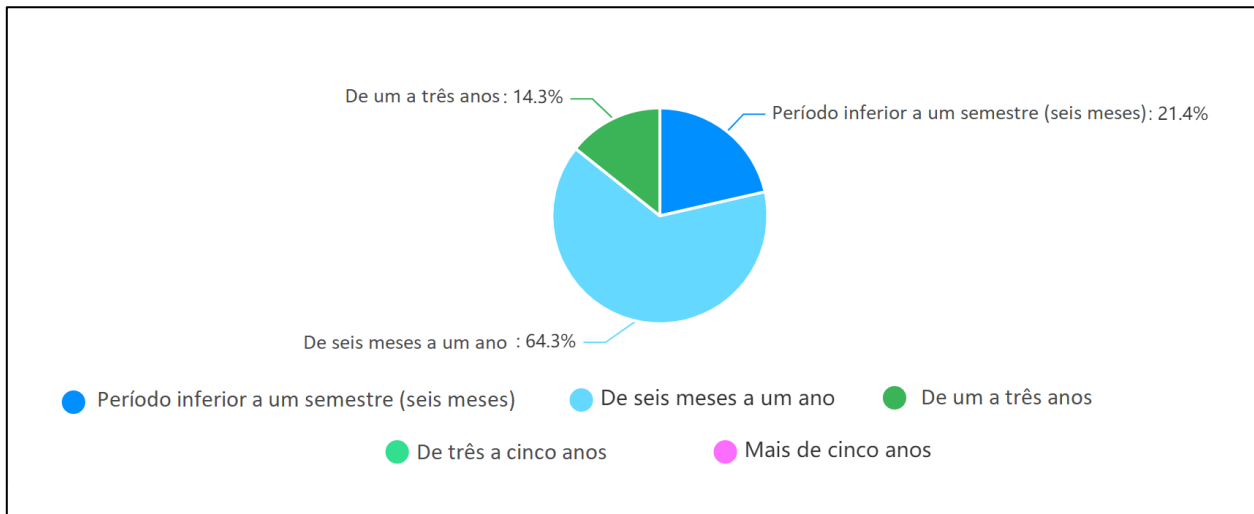
A maior parte dos alunos da pesquisa possuíam entre 21 e 22 anos (88,2%), apenas dois tinham 23 ou mais (11,8%).

O grupo de voluntários foi composto majoritariamente por indivíduos do sexo feminino. Juntaram-se ao projeto 13 alunas (76,5%), ao passo que apenas quatro alunos (23,5%) fizeram o mesmo.

No que tange à língua materna, apenas um aluno (5,9%) não tem chinês-mandarim como língua materna, mas, sim, o cantonês, língua falada principalmente na província do Cantão e em Hong Kong. No entanto, este aluno define sua proficiência em mandarim como avançada.

Dos nossos voluntários, a grande maioria – 13 deles (76,5%) já havia morado em países ou territórios de língua oficial portuguesa. Desses, 9 (64,3%) residiram em uma dessas localidades por mais de seis meses, e dois (14,3%) por mais de um e menos de três anos.

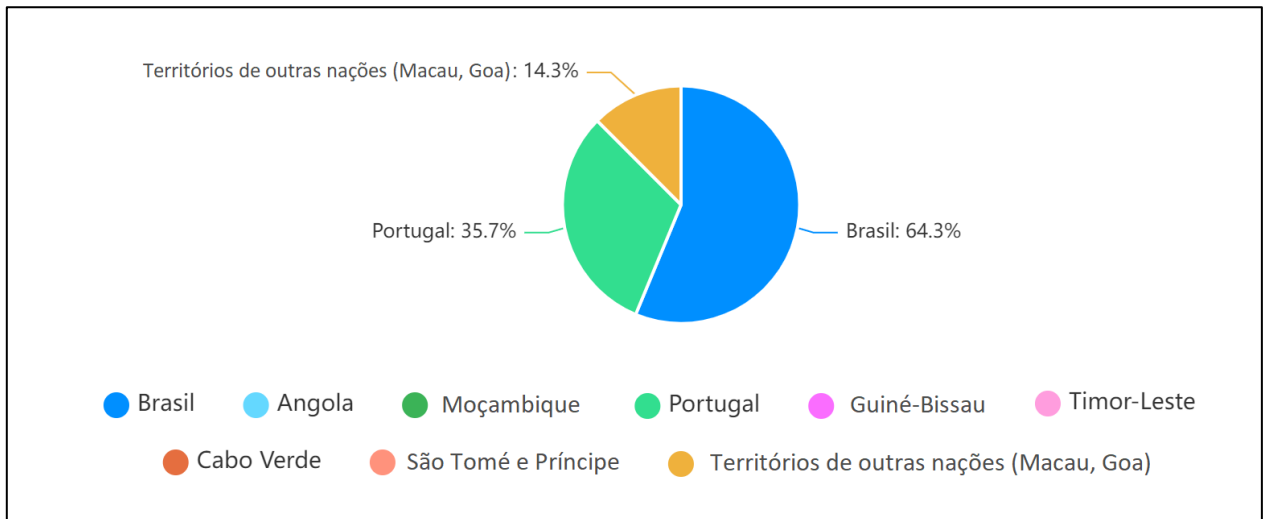
Gráfico 4 – Tempo de experiência em países de língua oficial portuguesa



Fonte: elaborado pelo autor.

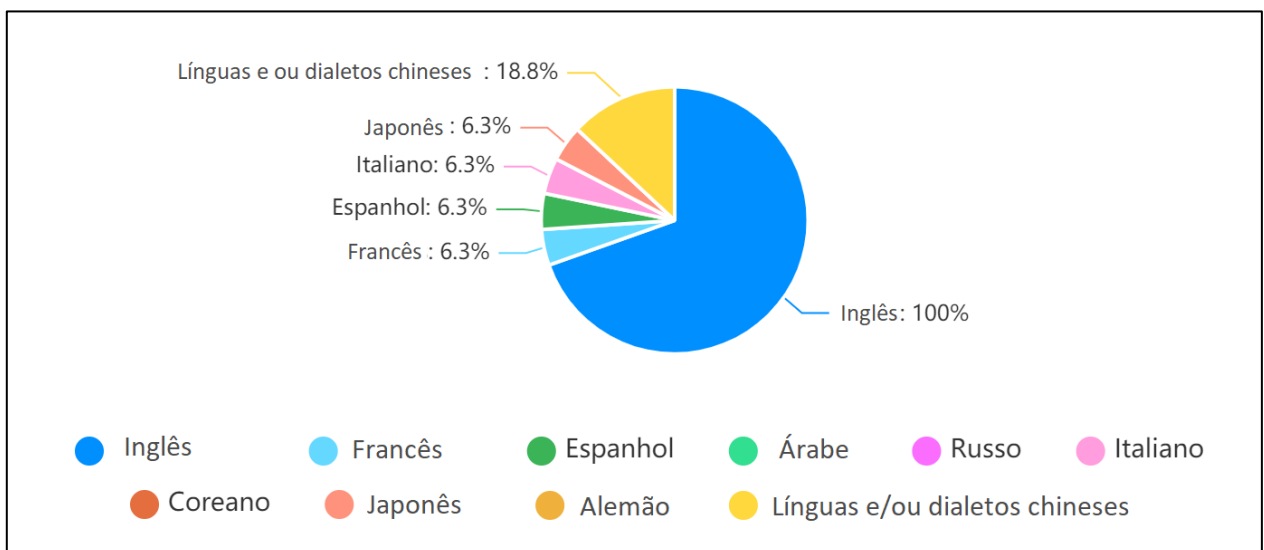
Dos alunos que já estiveram em países de língua portuguesa, 13 (92,9%) já haviam viajado para países e territórios de língua portuguesa para estudarem português, e apenas um deles viajou a turismo (7,10%).

Acerca dos destinos dos alunos, nove (64,3%) residiram no Brasil, cinco (35,7%) em Portugal e dois (14,3%) em Macau, região administrativa especial de língua oficial portuguesa na China, sendo que alguns dos alunos residiram em mais de um desses países e/ou territórios. Interessantemente, nove alunos (52,9%) estavam nesses territórios durante o período que compreendeu a compilação do nosso *corpus* de pesquisa.

Gráfico 5 – Destino dos alunos com experiência em países de língua oficial portuguesa

Fonte: elaborado pelo autor.

Do total de alunos, 16 (94,1%) falavam outros idiomas além do português e mandarim. Desses, todos afirmaram ter domínio do inglês, e um (6,2%) também do francês, um do espanhol, um do italiano, um do japonês e três (18,8%) de línguas ou dialetos chineses, como o *Wuhandua* (dialeto de Wuhan), entre outros.

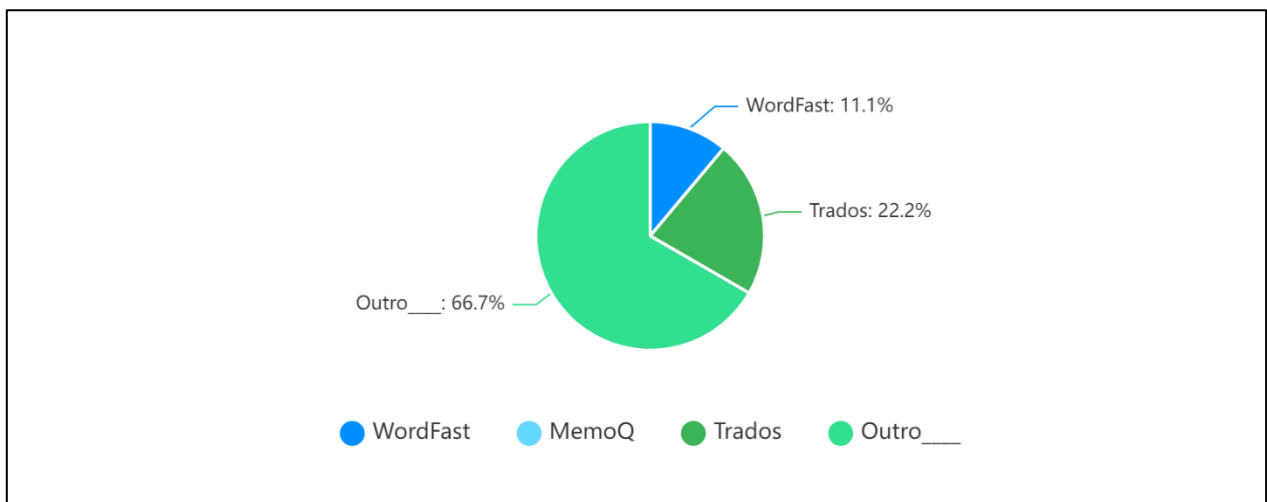
Gráfico 6 – Outras línguas estrangeiras faladas pelos voluntários chineses

Fonte: elaborado pelo autor.

Quanto à experiência profissional, apenas quatro (23,5%) deles possuíam alguma experiência. Três (75%) atuaram na tradução e revisão e um (25%) na dublagem e legendagem.

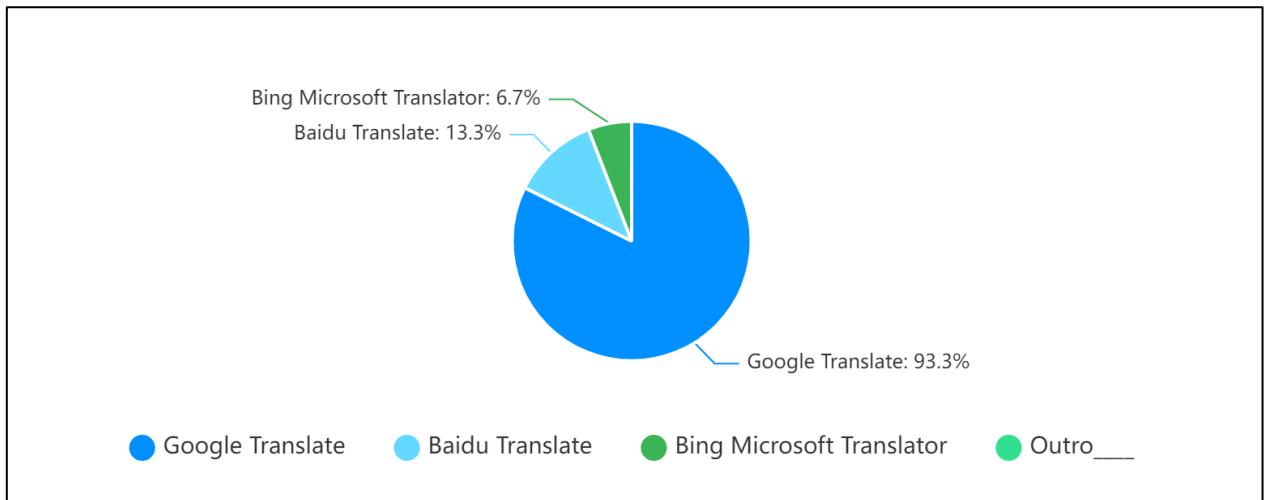
A respeito do uso de ferramentas de tradução, nove alunos (52,9%) afirmaram terem usado Memórias de Tradução (MT); sistema de base de dados que armazena segmentos textuais já traduzidos e os recupera na tradução de segmentos similares, o que pode auxiliar na tradução de documentos técnicos com terminologia repetitiva. Desses, dois (22,2%) usaram o Trados (KURNIAWATI et al., 2016), um (11,1%) o *Wordfast*, e outros seis (66,7%) outras MT.

Gráfico 7 – Ferramentas de tradução usadas pelos voluntários chineses



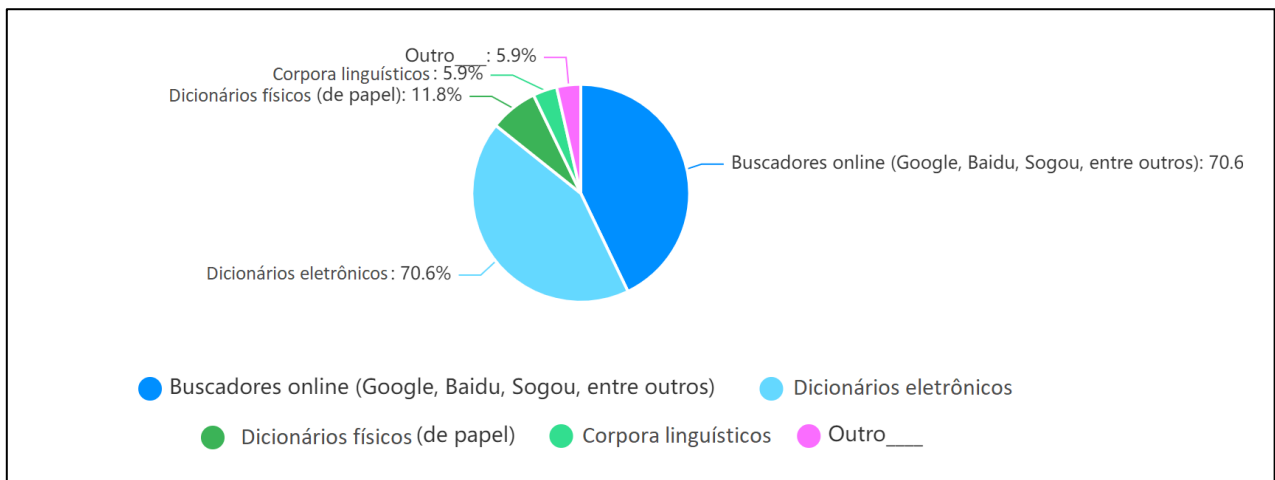
Fonte: elaborado pelo autor.

Outra ferramenta utilizada pelos alunos foram os tradutores automáticos: 15 alunos (88,2%) usaram ferramentas do tipo para consultas pontuais. Desse número, 14 (93,3%) usaram o *Google Translate*, dois (13,3%) o *Baidu Translate* e um (6,7%) o *Bing Microsoft Translate*.

Gráfico 8 – Tradutores automáticos usados pelos voluntários chineses

Fonte: elaborado pelo autor.

Por fim, 12 alunos (70,6%) relataram também terem realizado buscas em buscadores *on-line* como o *Google*, *Baidu* ou *Sogou*; dois (11,8%) usaram algum dicionário eletrônico; um usou algum dicionário físico (de papel) (5,9%) e um usou *corpora* linguísticos (5,9%).

Gráfico 9 – Outras ferramentas usadas pelos voluntários chineses durante as traduções

Fonte: elaborado pelo autor.

Interessantemente, nove alunos (52,9%) afirmaram terem usado uma língua intermediária para realizarem suas traduções, ou seja, quando defrontados com palavras desconhecidas, inicialmente procuraram por elas em uma outra língua antes de traduzir para o português. Todos eles relataram terem usado o inglês como língua intermediária.

Finalmente, apesar de todos os alunos terem interesse em seguir a carreira de tradutor, apenas sete deles (41,2%) tiveram aulas de Teorias da Tradução em algum momento de sua formação. Quatro (23,5%) tiveram aulas sobre o uso de ferramentas de tradução, possivelmente os que possuem experiência profissional.

Sete alunos (41,2%) afirmaram já ter ouvido falar de *corpus* linguístico e cinco (29,4%) já ouviram falar sobre colocações e a importância do seu estudo.

Passamos agora às ferramentas e procedimentos metodológicos usados para a extração e seleção do nosso objeto de estudo, as colocações.

3.2 Identificação, seleção e extração das colocações para análise

Podemos dividir a etapa de identificação, seleção e extração das colocações a serem analisadas em dois momentos: um momento voltado para a identificação das colocações nos TOs e, outro, para a localização dos seus correspondentes tradutórios, segundo as traduções dos aprendizes.

Nas subseções abaixo, detalhamos cada etapa, apresentando as ferramentas utilizadas e os procedimentos seguidos, bem como as dificuldades encontradas.

3.2.1. Extração das colocações dos *subcorpora* de TOs com o SkE

Depois de concluído o processo de coleta e tratamento dos TTs e o encerramento das atividades com os alunos voluntários chineses, tanto os TTs como os TOs foram carregados na plataforma SkE (KILGARIFF, 2014).

Figura 19 – Página para carregamento de textos para compilação de corpus no SkE

Fonte: captura de tela do SkE.

Na imagem acima, vemos a página para carregamento de textos para compilação de um *corpus* no SkE. Cada um dos nossos *subcorpora* foi carregado separadamente como um *Single language corpus* (*corpus* monolíngue), isso porque, apesar de o nosso *corpus* de estudo ser paralelo, multilíngue, o SkE só habilita o emparelhamento de um TO a um ou mais TT, desde que eles sejam de línguas diferentes. No nosso estudo, temos um TO, mas vários TTs na mesma língua, e o SkE não oferece suporte para esse tipo de emparelhamento.

Para cada *subcorpus* carregado, selecionamos a língua correspondente (português, inglês ou chinês) e lhes demos um nome, também na forma de código, como no exemplo:

1 2 3 4
[zhCAT20_TOs]

Em que a primeira parte (1), formada pelas duas primeiras letras, corresponde à língua; a segunda parte (2), ao nome do projeto; a terceira parte (3), ao ano de compilação do *corpus*; e a última parte (4), precedida por uma *underline* “_”, à natureza dos textos, se eram TOs ou TTs. Dessa forma, temos os seguintes *subcorpora* hospedados na plataforma:

ptCAT17_TOs

enCAT17_TTs

zhCAT20_TOs

ptCAT20_TTs

Depois de carregar os textos e compilar nosso *corpus*, passamos a manuseá-lo com as várias ferramentas disponíveis no SkE, dentre as quais destacamos a lista de palavras-chave, usada para identificação de palavras com potencial a apresentar combinações colocacionais. Para se extrair as palavras-chave de um *corpus*, é antes necessária a seleção de um *corpus* de referência, pois a ferramenta irá compará-lo com o *corpus* de estudo para gerar uma lista com as palavras mais típicas dentro do *corpus* enfocado, evidenciando as palavras mais significativas dentro deste em relação ao *corpus* de referência.

Para gerarmos as nossas listas de palavras-chave, usamos como *corpus* de referência o *Portuguese Web 2011 (ptTenTen11)*, de 3.896.392.719 de palavras, e o *Chinese Web 2017 (zhTenTen17) Simplified*, de 13.531.331.169 de palavras. Ambos os *corpora* de referência estão disponíveis para usuários registrados do SkE.

Figura 20 – Resultado da busca por palavras-chave do ptCAT17_TOs

SINGLE-WORDS ✓

MULTI-WORD TERMS ✓

reference corpus: Portuguese Web 2011 (ptTenTen11)

(items: 1,463)

Word	Word	Word	Word	Word
1 trump	11 agata	21 repressor	31 anti-irã	41 trudeau
2 janot	12 hyperlink	22 chanceler	32 polónes	42 reuven
3 netanyahu	13 chavistas	23 des-viados	33 macron	43 joesley
4 melania	14 empoderamento	24 pedofilianãoéarte	34 transfóbico	44 shinzo
5 schwartz	15 machista	25 wailliam	35 delcy	45 retroalimenta
6 faurie	16 sunita	26 br-homepage	36 heteronormativo	46 posição-chave
7 rivlin	17 díaz	27 utm_hp_ref	37 sapatona	47 tarek
8 tsj	18 ushuaia	28 cisgêneros	38 heteronormativa	48 heideggeriana
9 ortega	19 xiita	29 loizaga	39 astrini	49 binyamin
10 jamanxim	20 aloysio	30 minorizados	40 anti-arte	50 não-convencional

Rows per page:

50

1–50 of 1,000

<

<

1

/ 20

>

>>

Fonte: captura de tela do SkE.

Figura 21 – Resultado da busca por palavras-chave do zhCAT20_TOs

SINGLE-WORDS ✓

MULTI-WORD TERMS ✓

reference corpus: Chinese Web 2017 (zhTenTen17) Simplified (Items: 1,991)

Word	Word	Word	Word	Word
1 非洲人 ...	11 出京 ...	21 Ladakh ...	31 时艰 ...	41 基本法 ...
2 种族主义 ...	12 克什米尔拉达克 ...	22 印方 ...	32 冠 ...	42 边境线 ...
3 羟氯喹 ...	13 博索纳罗 ...	23 张革 ...	33 印军 ...	43 白人 ...
4 氯喹 ...	14 潘绪宏 ...	24 非洲籍 ...	34 京外 ...	44 神药 ...
5 早些时候 ...	15 郝运 ...	25 封城 ...	35 梁启超 ...	45 肺炎 ...
6 加勒万 ...	16 韩大元 ...	26 歧视 ...	36 康有为 ...	46 塔利班 ...
7 抗疫 ...	17 无症状 ...	27 弗洛伊德 ...	37 俄新社 ...	47 国徽 ...
8 印中 ...	18 白种人 ...	28 抗疟药 ...	38 黑人 ...	48 五十六 ...
9 黄种人 ...	19 茹晨曦 ...	29 拉达克 ...	39 对峙 ...	49 动武 ...
10 疫情 ...	20 赵立坚 ...	30 羟 ...	40 接触者 ...	50 感染者 ...

Rows per page:

50

1-50 of 1,000

1 / 20

Fonte: captura de tela do SkE.

Como podemos observar a partir das imagens, os primeiros resultados para a busca por palavras-chave não foram tão prolíficos em termos de palavras com potencial colocacional, pois, em sua maioria, eram nomes próprios. Por essa razão observamos as primeiras 500 palavras-chave de cada *subcorpora*, em português e chinês. Em seguida, selecionamos aquelas que julgamos ter maior potencial para combinações colocacionais para, posteriormente, buscarmos pelos seus possíveis colocados com a ferramenta *Word Sketch*, na qual pudemos visualizar todas as combinações mais frequentes de uma dada palavra, como ilustramos abaixo com exemplos extraídos do nosso *corpus*, com a palavra-chave 法律 *fǎlǜ* (lei).

Figura 22 – Resultado da busca por colocações de 法律

法律 as noun 5x ...			
Modifies	Object_of	Direct-Object_of	Measure
后果 法律 后果 ...	触犯 触犯了 香港 多项 法律 ...	触犯 触犯了 香港 多项 法律 ...	项 项 法律 ...
规范 法律 规范 ...			
规则 法律 规则 ...			
秩序 法律 秩序 ...			

Fonte: captura de tela do SkE.

Além de nos apresentar as palavras que combinam mais frequentemente com uma determinada palavra, o *Word Sketch* também apresenta a frequência com que as combinações aparecem no *corpus*, como ilustrado abaixo com a palavra-chave “calafrio” como exemplo.

Figura 23 – Resultado da busca por colocações de “calafrio” com indicação da frequência

verbo + calafrio				sintagma preposicional				calafrio + verbo				preposição+infinitivo			
sentir senti um calafrio	571	4.9	...	calafrio em substantivo	9.3%	...		arrepiar acompanhar-me , Um calafrio arrepia-me E de	6	5.2	...	...calafrio a Vinf	1.8%	...	
causar causar calafrios	356	4.9	...	...de calafrio	4.4%	...		percorrer calafrio percorreu	88	5.0	...	...calafrio de Vinf	1.7%	...	
provocar provoca calafrios	189	4.3	...	calafrio de substantivo	3.4%	...		sacudir calafrio sacudiu	6	4.1	...	...calafrio em Vinf	0.4%	...	
tremar tremar com calafrios	5	3.4	...	calafrio a substantivo	2.4%	...		gelar calafrio gelou-lhe	4	3.9	...				
sentar sente calafrios	8	2.1	...	...com calafrio	2.2%	...		perpassar Um calafrio perpassa	4	3.5	...				
dar dá calafrios	476	2.0	...	...por calafrio	1.6%	...		descer calafrio desce	7	1.1	...				
produzir produzem calafrios	22	0.5	...	calafrio por substantivo	1.3%	...		subir calafrio que sobe	13	0.5	...				
				...até calafrio	1.0%	...		correr calafrio corre	12	0.4	...				
				calafrio com substantivo	0.8%	...									
				...em calafrio	0.5%	...									
				...a calafrio	0.3%	...									
				calafrio para substantivo	0.3%	...									

Fonte: captura de tela do SkE.

Outra funcionalidade muito relevante para a realização da nossa pesquisa foram as estatísticas do SkE. Quando usamos o *Word Sketch*, temos acesso ao *logDice*, uma versão complexa e sofisticada do algoritmo *Dice* que gera um *score* indicando a força de atração de uma colocação (LEXICAL COMPUTING CZ S.R.O, 2021; RYCHLÝ, 2008). Quanto maior a força de atração uma colocação, maior a probabilidade de seus componentes ocorrerem juntos (TAGNIN, 2013). Quanto mais alto o *score logDice* de uma colocação, maior é a força de atração. Essa estatística baseia-se somente na frequência da base e do colocado e na frequência de toda a colocação, ou seja, base+colocado juntos e, por isso, não pode ser afetada pelo tamanho do *corpus*, podendo, portanto, ser usada em diferentes *corpora*.

Figura 24 – Resultado da busca por colocações de “calafrio” com o *logDice* indicado

verbo + calafrio			sintagma preposicional			calafrio + verbo			preposição+infinitivo		
sentir	571	4.9 ...	calafrio em substantivo	9.3% ...		arrepia	6	5.2 ...	...calafrio a Vinf	1.8% ...	
senti um calafrio			...de calafrio	4.4% ...		acompanhar-me . Um calafrio arrepi-me			...calafrio de Vinf	1.7% ...	
causar	356	4.9 ...	calafrio de substantivo	3.4% ...		percorrer	88	5.0 ...	...calafrio em Vinf	0.4% ...	
causar calafrios			calafrio a substantivo	2.4% ...		calafrio percorreu					
provocar	189	4.3 ...	...com calafrio	2.2% ...		sacudir	6	4.1 ...			
provoca calafrios			...por calafrio	1.6% ...		calafrio sacudiu					
tremar	5	3.4 ...	calafrio por substantivo	1.3% ...		gelar	4	3.9 ...			
tremar com calafrios			...até calafrio	1.0% ...		calafrio gelou-lhe					
sentar	8	2.1 ...	calafrio com substantivo	0.8% ...		perpassar	4	3.5 ...			
sente calafrios			...em calafrio	0.5% ...		Um calafrio perpassa					
dar	476	2.0 ...	...a calafrio	0.3% ...		descer	7	1.1 ...			
dá calafrios			calafrio para substantivo	0.3% ...		calafrio desce					
produzir	22	0.5 ...				subir	13	0.5 ...			
produzem calafrios						calafrio que sobe					
						correr	12	0.4 ...			
						calafrio corre					

Fonte: captura de tela do SkE.

Para a nossa pesquisa, consideramos para investigação mais aprofundada a respeito de sua natureza colocacional combinações cujo *score* era igual ou superior a três (FRANKENBERG-GARCIA, 2018) em uma escala de 0 a 14 (RYCHLÝ, 2008), salvo situações em que colocações reconhecidamente típicas das línguas de referência não estavam bem-posicionadas no ranqueamento gerado pela ferramenta. Em ocasiões como essa, recorreremos à frequência da combinação no *corpus* de referência. Quando o *corpus* de referência principal não apresentava resultados satisfatórios, como *logDice* muito baixo, ou simplesmente não apresentava resultados para colocações que sabíamos existir, recorreremos a outro *corpus* de referência, o que ocorreu com colocações referentes à Questão Palestina e ao conflito entre Palestina e Israel. Nesses casos, usamos o BNC como *corpus* de referência.

Eventualmente, também recorreremos ao uso da *web* como *corpus*, avaliando os *hits* para as buscas pelas combinações nos buscadores *on-line* Google²¹ e Baidu²²; e consultas às versões *on-line* dos dicionários *Cambridge Dictionary*²³, do inglês, e *Michaelis*²⁴, do português; além da nossa experiência como pesquisador da área.

Nosso olhar atento frente às possíveis combinações e o uso de critérios alternativos também se fez necessário porque, repetidas vezes, o SkE apresentou o que nos pareceu se tratar

²¹ www.google.com

²² www.baidu.com

²³ <https://dictionary.cambridge.org/pt/>

²⁴ <https://michaelis.uol.com.br/moderno-portugues/>

de falhas no processamento dos dados, como por exemplo duplicatas com frequências e *scores* diferentes e erros no reconhecimento das classes de palavras, como demonstramos abaixo.

Figura 25 – Erro no processamento de 典型



Fonte: captura de tela do SkE

Como vemos acima, ao buscarmos as possíveis colocações para a palavra 典型 *diǎnxíng* (típico), o SkE a trata como verbo, o que não condiz com a real classe dessa palavra, que pode ser, menos frequentemente, um substantivo, ou um adjetivo, como era o caso do exemplo no *corpus*. Esse não foi um caso isolado; pelo contrário, falhas como essa ocorreram não apenas no *corpus* em chinês, mas também em português, ainda que a frequência maior seja no da língua chinesa, e, por esse motivo, ilustramos com mais um exemplo abaixo.

Figura 26 – Erro no processamento de 严重



Fonte: captura de tela SkE

Na imagem acima, vemos 严重 yánzhòng (grave) ser tratado como verbo, o que na realidade não é possível, pois 严重 pode ser usado apenas como adjetivo ou advérbio. Nos aprofundamos a respeito deste problema na seção 3.2.2.

Em outras situações também recorreremos ao uso de outra ferramenta, as linhas de concordância, ou seja, linhas que permitem observar estruturas convencionais recorrentes em uma língua na forma de concordância. Com essa ferramenta, são geradas linhas nas quais uma palavra-chave aparece em seu contexto natural de uso, centralizada, o que permite a observação das palavras ao seu redor. Em alguns casos, o uso dessa ferramenta foi necessário para checarmos a ocorrência de algumas candidatas a colocação, atentos ao requisito de restrição contextual (TAGNIN, 1999).

Figura 27 – Análise das linhas de concordância da palavra 疫情

Left context	KWIC	Right context
1	① doc#0 以 确保 " 正义 能够 实现 " 。 </s><s> 奥巴 马 还 表 示 , " 新 冠 肺 炎	疫情 和 经 济 危 机 颠 覆 了 我 们 周 围 的 一 切 , 希 望 生 活 恢 复 正 常 是 很 自
2	① doc#0 议 活 动 。 </s><s> 北 京 : 严 格 出 京 管 理 并 不 意 味 着 " 封 城 " 北 京	疫情 防 控 处 于 紧 要 关 头 , 又 一 次 进 入 战 时 状 态 。 </s><s> " 6 月 18 日
3	① doc#0 入 战 时 状 态 。 </s><s> " 6 月 18 日 , 在 北 京 市 新 型 冠 状 病 毒 肺 炎	疫情 防 控 工 作 第 124 场 新 闻 发 布 会 上 , 北 京 市 委 宣 传 部 副 部 长 、 市 政
4	① doc#0 新 闻 办 主 任 、 市 政 府 新 闻 发 言 人 徐 和 建 表 示 。 " </s><s> 为 严 防	疫情 向 京 外 扩 散 , 北 京 严 格 出 京 管 理 , 坚 决 阻 断 疫 情 传 播 渠 道 , 但 并
5	① doc#0 </s><s> 为 严 防 疫 情 向 京 外 扩 散 , 北 京 严 格 出 京 管 理 , 坚 决 阻 断	疫情 传 播 渠 道 , 但 并 不 意 味 着 封 城 。 </s><s> " 北 京 市 公 安 局 副 局 长
6	① doc#0 员 出 京 北 京 市 公 安 局 副 局 长 、 新 闻 发 言 人 潘 绪 宏 表 示 , 为 严 防	疫情 向 京 外 扩 散 , 北 京 严 格 出 京 管 理 , 坚 决 阻 断 疫 情 传 播 渠 道 , 三 类
7	① doc#0 宏 表 示 , 为 严 防 疫 情 向 京 外 扩 散 , 北 京 严 格 出 京 管 理 , 坚 决 阻 断	疫情 传 播 渠 道 , 三 类 人 员 禁 止 出 京 : 一 是 全 市 确 诊 病 例 、 疑 似 病 例
8	① doc#0 s> 加 强 社 区 (村) 封 闭 式 管 理 防 控 措 施 发 布 会 上 , 北 京 新 冠 肺 炎	疫情 防 控 工 作 领 导 小 组 社 区 防 控 组 副 组 长 、 市 委 组 织 部 副 部 长 张 革
9	① doc#0 小 区 同 一 单 元 两 户 以 上 出 现 病 例 的 单 元 , 高 风 险 街 道 乡 镇 发 生	疫情 的 小 区 及 高 风 险 区 域 周 边 经 评 估 后 确 有 必 要 的 小 区 进 行 封 闭

Fonte: captura de tela do SkE.

O uso avançado das linhas de concordância ainda apresenta outra possibilidade que auxilia no reconhecimento de colocações. Por meio da função *Collocations* do *Concordance* é possível visualizar as palavras que coocorrem com mais frequência até a quinta posição de distância em relação à palavra centralizada, além de obter o *logDice* das combinações, entre outras medidas estatísticas, como o *T-score*, que indica “a certeza com a qual podemos argumentar que existe uma associação entre as palavras, i.e., sua coocorrência não é aleatória”²⁵, ou o *Mutual Information score*, o *MI score*, que “expressa a extensão em que as palavras

²⁵ certainty with which we can argue that there is an association between the words, i.e. their co-occurrence is not random.

coocorrem em comparação com o número de vezes que aparecem separadamente”²⁶ (LEXICAL COMPUTING CZ S.R.O, 2020), como podemos observar a seguir.

Figura 28 – Visualização da caixa de busca do *Collocations* do *Concordance*



Fonte: captura de tela do SkE.

Figura 29 – Resultado da busca por colocações da palavra 疫情 com o *Collocations* do *Concordance*

	Word	Cooccurrences?	Candidates?	T-score	MI	LogDice ↓
1	☐ 冠	11	36	3.26	5.81	12.29 ...
2	☐ 新	10	42	3.09	5.45	12.04 ...
3	☐ 防控	6	12	2.42	6.52	12.00 ...
4	☐ 肺炎	4	10	1.97	6.20	11.48 ...
5	☐ 传播	3	5	1.72	6.78	11.23 ...
6	☐ 病毒	3	23	1.66	4.58	10.70 ...
7	☐ 的	13	268	3.20	3.15	10.45 ...
8	☐ 美国	3	52	1.57	3.40	10.13 ...
9	☐ .	9	248	2.55	2.73	10.02 ...
10	☐ ,	12	435	2.78	2.34	9.71 ...

Fonte: captura de tela do SkE.

²⁶ expresses the extent to which words co-occur compared to the number of times they appear separately.

Outra ferramenta usada foi a lista de frequência, uma lista que elenca todas as palavras em um *corpus* com base em sua frequência, ou seja, das palavras mais recorrentes dentro do *corpus* para as menos recorrentes. Como nos alerta Tagnin (2013), palavras gramaticais geralmente são as mais frequentes e, por essa razão, geralmente aparecem no topo dessas listas. A título de exemplificação, trazemos as palavras mais frequentes de um dos *subcorpora* do nosso *corpus* de estudo.

Figura 30 – Lista de palavras mais frequentes do zhCAT20_TOs

Word	Frequency ?	Word	Frequency ?	Word	Frequency ?	Word	Frequency ?	Word	Frequency ?
1 ,	435 ...	11 有	48 ...	21 人员	29 ...	31 非洲人	22 ...	41 没有	18 ...
2 的	268 ...	12 一	47 ...	22 中	27 ...	32 与	22 ...	42 上	18 ...
3 。	248 ...	13 了	45 ...	23 表示	26 ...	33 (	22 ...	43 死亡	18 ...
4 "	142 ...	14 新	42 ...	24 我	26 ...	34 个	22 ...	44 还	18 ...
5 、	91 ...	15 不	41 ...	25 病毒	23 ...	35 人	22 ...	45 例	17 ...
6 在	86 ...	16 疫情	36 ...	26 说	23 ...	36 两	20 ...	46 北京	16 ...
7 是	69 ...	17 冠	36 ...	27 也	23 ...	37 都	20 ...	47 巴西	16 ...
8 中国	63 ...	18 就	34 ...	28)	23 ...	38 印度	20 ...	48 会	16 ...
9 和	53 ...	19 对	33 ...	29 他们	23 ...	39 歧视	20 ...	49 政府	16 ...
10 美国	52 ...	20 这	29 ...	30 但	22 ...	40 国	19 ...	50 病例	16 ...

Fonte: captura de tela do SkE.

Dentre as palavras mais frequentes no *subcorpus* de TOs em chinês como um todo, temos 的 *de*, partícula usada para indicar posse, como o 's referente ao *genitive case* em inglês, ou ainda para atribuir características a um substantivo, geralmente ligando-o a um adjetivo; 在 *zài*, que funciona como a preposição “em” do português e, às vezes, exerce a função do verbo “estar”; 和 *hé*, a conjunção “e”; 了 *le*, partícula adicionada depois de verbos para indicar ação passada ou mudança de estado; e 不 *bù*, que significa não. O primeiro verbo a aparecer na lista é 是 *shì*, “ser”, um verbo de ligação, seguido de 有 *yǒu*, “ter”.

Depois de selecionadas, as colocações presentes nos *subcorpora* de TOs foram listadas, organizadas e separadas por tipo, segundo a classificação de Orenha-Ottaiano (2009). Porém, como chegamos a um número muito elevado de colocações, usamos os colocados mais

frequentes como critério para realizar um corte na amostragem. Os resultados da divisão são apresentados no capítulo “Análise dos Resultados”.

Na subseção a seguir, apresentamos os procedimentos metodológicos para a localização das opções tradutórias dos aprendizes de tradução para as colocações elencadas.

3.2.2 Extração das colocações dos subcorpora de TTs

Depois de realizarmos o processo de extração das palavras-chave e seleção das colocações nos *subcorpora* de TOs, buscamos realizar um processo semelhante nos *subcorpora* de TTs no levantamento dos correspondentes tradutórios, o que se mostrou uma tarefa bem mais trabalhosa e demorada do que tínhamos inicialmente em mente. Algumas dificuldades que surgiram nessa etapa fizeram com que buscássemos alternativas para algumas das limitações do SkE. A primeira delas diz respeito à identificação de possíveis colocações, devido a algumas falhas que mencionamos anteriormente na subseção 3.2.1 deste trabalho, e que demonstramos abaixo.

Figura 31 – Exemplo de duplicata com frequência e scores diferentes

...de discriminação		
	✓ Show frequencies	✓ Show scores
formas de Todas as Formas de Discriminação contra	648	9.5 ...
eliminação Eliminação da Discriminação Racial	887	8.5 ...
proibição proibição de discriminação	362	7.5 ...
forma as formas de discriminação	3,714	7.3 ...
vítima vítimas de discriminação	769	7.1 ...
alvo alvo de discriminação	334	6.4 ...
observatório Observatório da Discriminação Racial e da	108	6.2 ...
crime crime de discriminação	263	6.2 ...

Fonte: captura de tela do SkE.

Inicialmente, acreditávamos que essas falhas haviam ocorrido devido ao fato de o *corpus* de referência que tínhamos à disposição, o *ptTenTen11*, com 3.896.392.719 palavras, apesar de à primeira vista parecer grande, ser na verdade um *corpus* limitado, sobretudo quando comparado com *corpora* de outras línguas.

Tabela 5 – Número de palavras por *corpus* da família *TenTen* do SkE

CORPUS	NÚMERO DE PALAVRAS
<i>English Web 2020 (enTenTen20)</i>	38.149.437.411
<i>Spanish Web 2018 (esTenTen18)</i>	16.951.839.897
<i>Russian Web 11 (ruTenTen11)</i>	14.553.856.113
<i>Chinese Web 2017 (zhTenTen17) Simplified</i>	13.531.331.169
<i>Portuguese Web 2011 (ptTenTen11)</i>	3.896.392.719

Fonte: elaborado pelo autor.

Porém, como podemos depreender a partir de Kuhn (2017), falhas como essas ocorrem devido ao *sketch grammar*, que ainda precisa ser aprimorado em português.

O *sketch grammar* é um arquivo com as relações gramaticais e diretivas de processamento que o SkE usa para computar os diferentes tipos de relações por meio de cálculos estatísticos e formam a base do *Word Sketch*, entre outras ferramentas do SkE (KUHN, 2017).

Os *sketch grammars* não podem ser compartilhados em *corpora* de línguas diferentes, mas diferentes *corpora* de um mesmo idioma podem usar o mesmo *sketch grammar*. *Corpora* de línguas sem um *sketch grammar* tem à sua disposição um *sketch grammar* universal, com estatísticas básicas que não levam em conta a gramática da língua (LEXICAL COMPUTING CZ S.R.O, 2021).

Em seu estudo, que envolveu a compilação de um *corpus* de português brasileiro e português europeu escritos em periódicos para o Dicionário de português para estudantes universitários, Kuhn (2017) avaliou os quatro *sketch grammars* disponíveis em português no SkE na época de seu trabalho: o *Sketch Grammar for Portuguese, FreeLing tagset version 1.1183*; o *Palavras parsed*, usado no *PtTenTen*, pensado principalmente para o *Oxford Portuguese Dictionary* (2015); o *Compatible Portuguese Sketch Grammar definition*, desenvolvido para a família de *corpora Aranea (Web) Corpora Family* (Benko, 2014); e o *Portuguese word sketches*

(Lingueca *parsed data*) *version 1*, desenvolvido para o *corpus* Cetenfolha, Cetempublico, formado por textos dos jornais Folha de S. Paulo, do Brasil, e Público, de Portugal (KUHN, 2017).

No caso do *Sketch Grammar for Portuguese, FreeLing tagset version*, por exemplo, Kuhn identificou que todas as palavras iniciadas por letra maiúscula, independentemente de sua classe gramatical, foram classificadas como nomes próprios, e que formas no particípio foram sempre classificadas como verbos, além de alguns *sketches* que apresentavam resultados vazios ou resultados incorretos, semelhante ao que apresentamos na Figura 32.

Outro fator que retornou resultados incorretos foi a presença de registros não-padrão contendo muitos erros de português do ponto de vista da gramática normativa, prejudicando a obtenção de alguns resultados. Os textos que constituem o *ptTenTen11*, membro da família de *corpus* *TenTen* do SkE, são textos de língua geral recolhidos na *Internet* com o uso de tecnologia especializada em coletar conteúdo da *web* com valor linguístico. Os dados foram rastreados pelo *SpiderLing web spider* entre março de 2011 e agosto de 2012 (LEXICAL COMPUTING CZ S.R.O, 2021), e eventualmente erros acabaram sendo incorporados ao *corpus*, como ocorreu quando buscávamos pelas colocações da palavra-chave “discriminação” e ilustramos nas Figuras 33 a 35, em que podemos ver o registro do verbo “pôr” sem acento circunflexo.

Figura 32 – Resultados da análise do *Word Sketch* com colocações de “discriminação”

crime	135	8.7 ...
que torna crime a discriminação		
combate	2,013	8.2 ...
combate à discriminação		
incitação	40	7.6 ...
incitação à discriminação		
incitamento	34	7.6 ...
e contra qualquer incitamento a tal discriminação		
pele	57	7.5 ...
na pele a discriminação		
incitar	31	7.1 ...
que incitem à discriminação		
vedação	25	6.9 ...
a vedação à discriminação		
fim	112	6.5 ...
fim à discriminação		
proibição	19	6.4 ...
a proibição à discriminação		
enfrentamento	38	6.3 ...
no enfrentamento à discriminação		
preconceito	14	6.0 ...
do preconceito à discriminação		
repúdio	26	6.0 ...
repúdio à discriminação		

Fonte: captura de tela do SkE.

Figura 33 – Equívoco na classificação das classes dos colocados de “fim...discriminação”

WORD SKETCH Portuguese Web 2011 (ptTenTen11)

fim ... discriminação 112x ...

discriminação noun > fim noun

☒ fim ... discriminação + verbo
 ☒ fim ... discriminação + adjetivo participial
 ☒ adjetivo + fim ... discriminação
 ☒ adjetivo participial + fim ... discriminação
 ☒ sujeito da passiva impessoal
 ☒ fim ... discriminação + adjetivo

verbo + fim ... discriminação

pôr 76 3.1 ...

sintagma preposicional

"discriminação"		"fim"
...a fim ... discriminação	100.0% ...	fim ... discriminação a substantivo 100.0% ...
fim ... discriminação contra substantivo	9.8% ...	...por fim ... discriminação 3.6% ...
fim ... discriminação de substantivo	8.9% ...	
fim ... discriminação em substantivo	8.0% ...	
fim ... discriminação a substantivo	1.8% ...	
fim ... discriminação por substantivo	1.8% ...	
fim ... discriminação entre substantivo	1.8% ...	
fim ... discriminação com substantivo	0.9% ...	
fim ... discriminação por+os substantivo	0.9% ...	

Fonte: captura de tela do SkE.

Figura 34 – Confirmação de equívoco na classificação das classes dos colocados de “fim...discriminação”



Fonte: captura de tela do SkE.

No mais, nosso *subcorpus* de textos originais conta com alguns artigos que tratam de assuntos específicos e que, apesar de não se tratar de textos técnico-científicos, trazem, muitas vezes, termos e colocações específicas das áreas referentes aos assuntos tratados como, por exemplo, representatividade LGBTQIAP+, política, assuntos militares e saúde pública. Às vezes, a possível restrição do *corpus* de referência se somava à questão da temporalidade do assunto dos textos.

Acerca do aspecto temporal, como mencionado anteriormente, o *ptTenTen11* não só possui um número de palavras inferior aos *corpora* de outras línguas hospedados no SkE, como os do inglês e chinês, mas também é mais antigo, tendo sido compilado há quase dez anos. A depender da área enfocada, muitas mudanças de caráter linguístico ocorreram no decorrer desses anos. No que tange a questões sobre representatividade LGBTQIAP+, um dos temas presentes no *subcorpus* de TOs do português, por exemplo, acreditamos que, devido ao fato de o assunto ter ganhado mais força nos últimos anos, combinações que se tornaram mais frequentes como “sistema machista” foram mais difíceis de serem cheçadas tendo apenas os *corpora* do SkE à disposição.

Na altura da compilação do *ptTenTen11*, entre 2011 e 2012, nem mesmo a sigla “LGBTQIAP+” era amplamente difundida. Naquela época, o termo mais usado pelos meios de comunicação e que se popularizava na população geral para se referir a esse conjunto de minorias ainda era “LGBT”, sucessora da antiga sigla usada em português, “GLS” (gays, lésbicas e simpatizantes), e que em inglês significa “*lesbian, gay, bisexual, and transgender*” (lésbica, gay, bissexual e transgênero). A letra “Q” só foi incluída oficialmente pela GLAAD, organização não-governamental americana que monitora como a mídia retrata a comunidade LGBTQIAP+, na

décima edição do seu *Media Reference Guide* (Guia de Referência para a Mídia) em outubro de 2016 (GLAAD, 2016), momento em que outras discussões, como a transfobia, começavam a ganhar mais força, e colocações como “sistema transfóbico”, presente em nosso *corpus*, começava a ser empregada mais recorrentemente.

No CAT chinês, tínhamos textos que tratavam da “epidemia do coronavírus” (新冠疫情 *Xīnguān yìqíng*, em chinês), porém a frequência de “coronavírus” em todos os *corpora* do SkE foi baixíssima. O *zhTenTen17 Simplified*, nosso *corpus* de referência naquela língua, foi compilado em 2017, três anos antes da eclosão da que veio se tornar uma pandemia, época em que o assunto era bem menos explorado pela mídia como um todo.

Em 2020, o SkE disponibilizou em sua plataforma o *Covid-19 corpus*, um *corpus* de 224.061.570 de palavras que compila textos parte do *COVID-19 Open Research Dataset (CORD-19)*. No entanto, este é um *corpus* monolíngue em inglês.

No caso do inglês, cujo *corpus* de referência para checar as opções dos alunos foi o mais recente, o *enTenTen20*, compilado em 2020, muitas colocações relacionadas a guerra e política, como as que se referem à histórica disputa entre Israel e Palestina, simplesmente não aparecem. É o caso de *Palestinian issue* (Questão palestina), por exemplo. Acreditamos que isso tenha ocorrido em razão da menor frequência com que o assunto passou a ser tratado na mídia global com o surgimento de novos conflitos nos últimos anos.

Eventualmente, para checar as escolhas tradutórias dos alunos, trocamos os *corpora* referência. Pudemos confirmar a colocação *Palestinian issue* com o BNC, que contém textos de épocas em que o tema “Palestina” esteve mais em voga. No caso dos *corpora* de língua portuguesa, acreditamos que eles são, de modo geral, ainda muito limitados.

Outro complicador foi a localização e seleção das alternativas tradutórias dos estudantes nos TTs. Encontrá-las seria uma tarefa em muito facilitada se pudessemos eventualmente visualizar todas os TTs alinhados por segmentos com o TO. A bem da verdade, como mencionamos anteriormente, o SkE permite a compilação de *corpora* multilíngues paralelos, o que permite a observação de equivalentes tradutórios mais pontualmente graças a um concordanciador paralelo, em que TOs e TTs aparecem lado a lado.

Figura 35 – Visualização de corpora paralelos alinhados no SkE

	EUROPAPARL7, German	EUROPAPARL7, French
①	form of a visa. And so it is the "two key " principle that applies here. A single	Es gilt also das Prinzip der zwei Schlüssel . Künftig soll ein einziger Schlüss
①	' principle that applies here. A single key is to suffice in future. The financial	chlüssel. Künftig soll ein einziger Schlüssel genügen. Der Finanzkontrolleur
①	internal audit service - " the second key " as Herr Bösch said, established in	nen Prüfdienstes - " den zweiten Schlüssel " -, wie Herr Bösch sagte, ergänz
①	enerous deal with them remains the key to lasting peace in the Middle East. I	erzielt wird, ist nach wie vor der Schlüssel zu dauerhaftem Frieden im Nahe
①	presidency on 12 January holds the key . We must revise the broad guideline	ntschaft vom 12. Januar wird der Schlüssel dazu geliefert. Die Grundzüge de
①	f of social welfare. Mr President, the key to the development of the less devel	iftigt werden. Herr Präsident, der Schlüssel für den Fortschritt der Entwicklun
①	n altered, it is extremely simple. The key to success lies, to a very large degr	ändert, ist alles sehr einfach. Der Schlüssel zum Erfolg liegt im hohem Maße
①	of sanctions. But let us be clear - the key to the lifting of sanctions lies with the	noch einmal deutlich gesagt der Schlüssel zur Aufhebung der Sanktionen lie
①	Portuguese presidency gave us the key to unlock a decade of sustained inn	tschaft gab uns in Lissabon den Schlüssel in die Hand zu einem Jahrzehnt
①	imate right to the Kurds. Indeed the key to reform is ending the war against t	senen Recht für die Kurden. Der Schlüssel zum Reform liegt in der Beendig
①	ve that reciprocal commitment is the key to the Pact ' s success and I feel tha	wechselseitige Engagement der Schlüssel für den Erfolg des Paktes und ict
①	programmes funded by the EU The key to the development of the developin	Umweltforschungsprogrammen Schlüssel für den Fortschritt in den
①	n is a universal human right and the key to sustainable human development.	reines Menschenrecht, sie ist der Schlüssel für eine nachhaltige mensche
①	aign language learning is indeed the key to integration. A third series of amen	er Fremdsprache ist nämlich der Schlüssel zur Integration. Eine dritte Reihe

Fonte: Site do SkE (disponível em: https://www.sketchengine.eu/wp-content/uploads/parallel_concordance.png. Acesso em 4 de novembro de 2021).

Para se compilar um *corpus* no SkE, o pesquisador deve carregar as duas traduções previamente alinhadas e tabuladas em um arquivo do *Excel*, por exemplo. O programa faz o reconhecimento das línguas e transforma cada versão em uma língua em *subcorpus*; no entanto, essa opção só é possível para TTs de línguas diferentes. No nosso caso, tínhamos vários TTs na mesma língua.

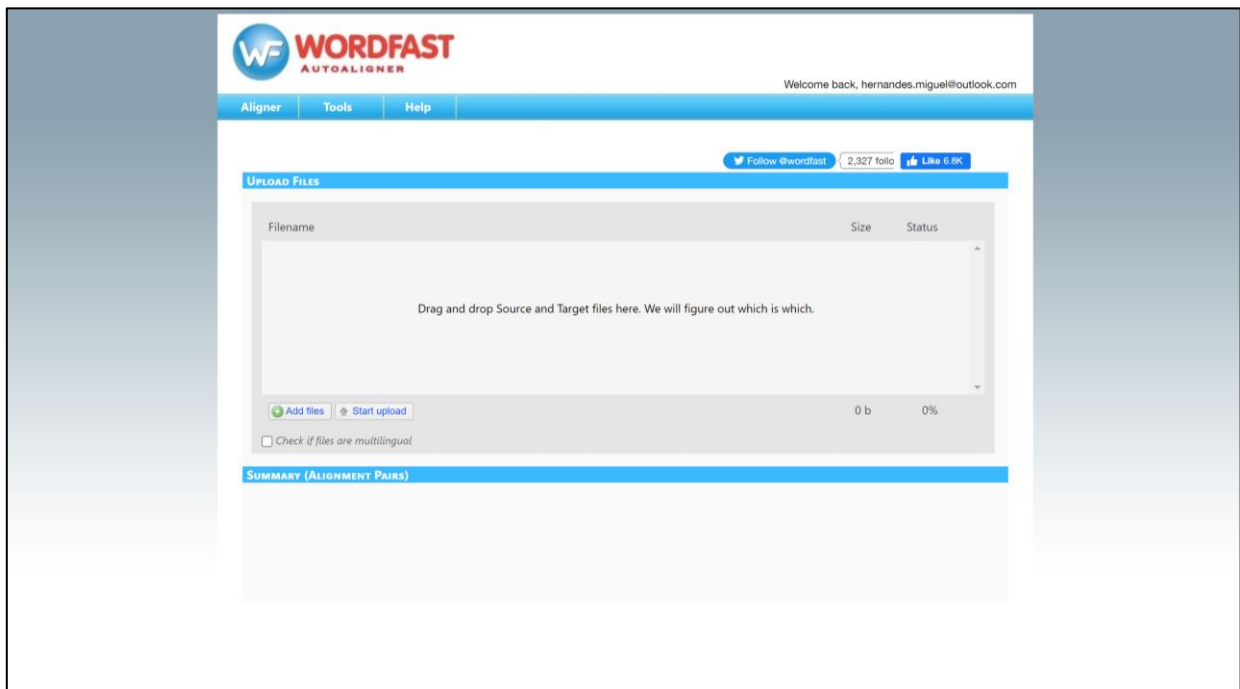
Mesmo que tivéssemos à nossa disposição o concordanciador monolíngue para consultar opções isoladas quando necessário, em muitas ocasiões, porém, a visualização das traduções alinhadas nos daria uma visão mais ampla do *corpus* de estudo e faria com que ganhássemos um tempo precioso na realização da pesquisa. Por essa razão, decidimos prosseguir com o alinhamento independente dos *subcorpora* de TTs com os *subcorpora* de TOs para, posteriormente, combinarmos o resultado das observações dos alinhamentos ao resultado das buscas com as ferramentas do SkE. Para obtermos sucesso nesta etapa da pesquisa sem que, para isso, perdêssemos muito tempo, lançamos mão de ferramentas alternativas às convencionalmente usadas na LC, conforme será apontado na subseção 3.2.3 a seguir.

3.2.3 Alternativas metodológicas para a observação e extração das colocações dos *subcorpora* de TTs

Para conseguirmos uma visão mais ampla de todos os TTs alinhados, recorreremos ao alinhador automático disponível no *WordFast Anywhere* (WFA) (2019). O WFA é uma ferramenta de MT totalmente *on-line* e gratuita, que permite que seus usuários usem um alinhador automático para carregar traduções prontas para alimentar as suas memórias pessoais, isto é, a base de dados com segmentos de tradução que podem ser recuperados quando da tradução de segmentos similares. A diferença é que no nosso caso, após o alinhamento, não alimentamos uma MT, mas baixamos os alinhamentos na forma de planilhas do *Excel*.

Após fazermos o *login* no WFA, na aba *Wordfast Anywhere*, selecionamos *autoaligner* e somos direcionados para o alinhador, onde fazemos o carregamento dos textos em qualquer um dos formatos aceitos pelo programa, que incluem *.pdf*, *.ppt*, *.xism*, dentre vários outros. Os formatos que nós usamos foram *.doc* e *.docx*.

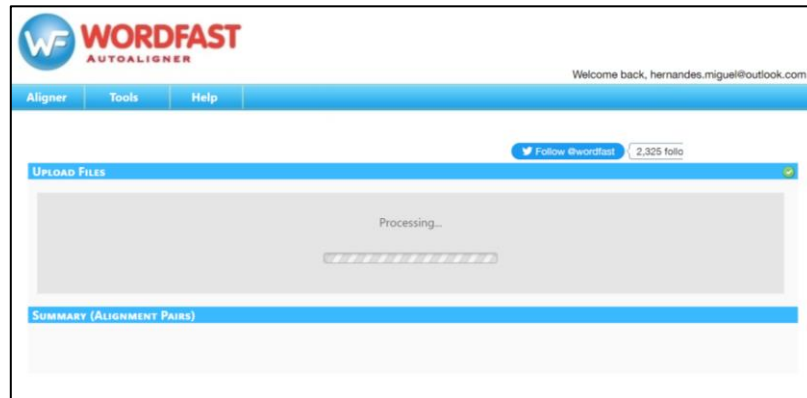
Figura 36 – Interface do alinhador automático do WFA



Fonte: captura de tela do WFA.

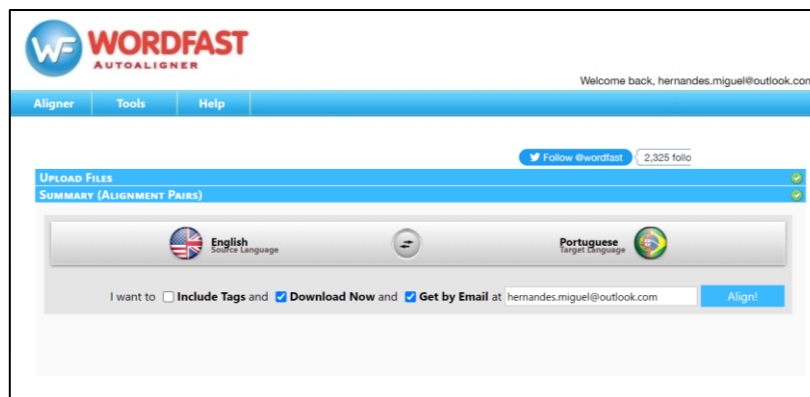
O programa faz o reconhecimento automático das línguas e sugere uma direção de tradução, que pode ser alterada pelo usuário.

Figura 37 – Interface do alinhador automático do WFA após carregamento de textos para alinhamento



Fonte: captura de tela do WFA.

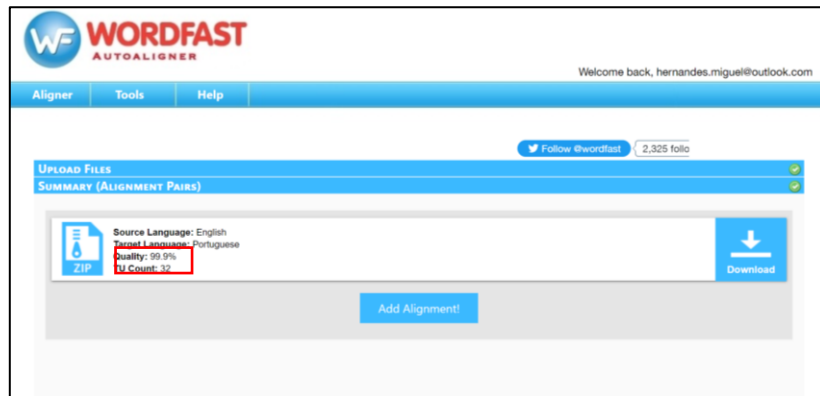
Figura 38 – Interface do alinhador automático do WFA após reconhecimento das línguas a serem alinhadas



Fonte: captura de tela do WFA.

Em seguida, o programa, assistido por tradutores automáticos e dicionários, realiza o alinhamento automático dos textos e gera um arquivo *.xism* para ser baixado. O programa ainda apresenta um *score* para a qualidade do alinhamento em porcentagem e o número total de segmentos alinhados, como vemos abaixo.

Figura 39 – Interface do alinhador automático do WFA após conclusão do alinhamento



Fonte: captura de tela do WFA.

Esse processo precisa ser repetido com cada TT, porém, este alinhador automático pode ser aberto em múltiplas abas de uma só vez, o que permite o alinhamento de vários textos simultaneamente.

Depois de concluído o alinhamento, o arquivo baixado apresenta a seguinte formatação (Fig. 40).

Figura 40 – Arquivo com traduções alinhadas gerado pelo alinhador automático do WFA

INSTRUCTIONS:			
EN	PT	Score	Notes
Brazilian President Michel Temer is at the end of his rope	Michel Temer: um presidente à beira do abismo	90	
The Brazilian President Temer's already shattered government suffered another setback when, this Monday, the Prosecutor General of the Republic formally reported him for corruption.	O governo já abalado de Temer sofreu mais um revés quando, nesta segunda-feira, o procurador-geral da República o denunciou formalmente por corrupção	92	
Michel Temer always occupied the backstage of the power, until he supplanted his partner Dilma Rousseff, a little more than one year ago.	Michel Temer sempre ocupou os bastidores do poder, até derrubar sua companheira Dilma Rousseff, há pouco mais de um ano.	96	
But, since then, nothing played out as the veteran strategist had calculated, and his struggle for political survival has been constant.	Mas desde então, nada saiu como este estrategista veterano calculou, e sua luta pela sobrevivência política tem sido constante.	94	
A final setback came this Monday, when the Prosecutor General of the Republic, Rodrigo Janot, formally reported him for corruption, leaving the President at the end of his rope.	O último revés veio nesta segunda-feira, quando o procurador-geral da República, Rodrigo Janot, o denunciou formalmente por corrupção, deixando o presidente à beira do abismo.	99	
Now a lawsuit is being filed at the Supreme Federal Court that might force him to resign, although it has to be previously validated by two thirds of the Chamber of Deputies, where Temer, currently 76 years old, started his rise to power three decades ago.	Agora se abre um processo no Supremo Tribunal Federal que poderá afastá-lo do cargo embora, antes, deva ser validado por dois terços da Câmara dos Deputados, onde Temer, hoje com 76 anos, começou sua escalada ao poder há três décadas.	97	
His government is skating on thin ice since the newspaper "O Globo" revealed, on the 17th of May, a compromising recording of a conversation with the businessman Joesley Batista in which the President seems to giving permission to pay the ex-deputy Eduardo Cunha his hush money.	Seu governo está na corda bamba desde que o jornal O Globo revelou, no dia 17 de maio, uma comprometedora gravação de uma conversa com o empresário Joesley Batista em que o presidente parece dar seu aval à compra do silêncio do ex-deputado Eduardo Cunha.	94	
"I will not resign."	"Não renunciarei."	99	
I repeat, I will not resign", he guaranteed after the SFC decided to investigate him.	Repito, não renunciarei", garantiu depois que o STF decidiu investigá-lo.	91	
And he did not resign.	E não renunciou.	99	
But, for many people, the recording of JBS's owner was the beginning of the end of a mandate that never got rid of corruption, since the impeachment of the former President Dilma Rousseff.	Mas para muitos a gravação do dono da JBS foi o início do fim de um mandato que nunca se livrou de conturbações, desde o impeachment da ex-presidente Dilma Rousseff.	94	

Fonte: elaborado pelo autor.

Como podemos notar a partir da imagem acima, o alinhador automático ainda apresenta um *score* em porcentagem para a qualidade do segmento. Depois de alinharmos todos os TTs com seus respectivos TOs, fizemos a junção de todas as colunas de textos traduzidos lado a lado em um único arquivo do *Excel* e fizemos os ajustes necessários manualmente quando ocorria diferença no número de segmentos dos TTS, o que foi raro, dada a elevada qualidade dos alinhamentos.

Figura 41 – Planilha com as todas os TTs alinhadas

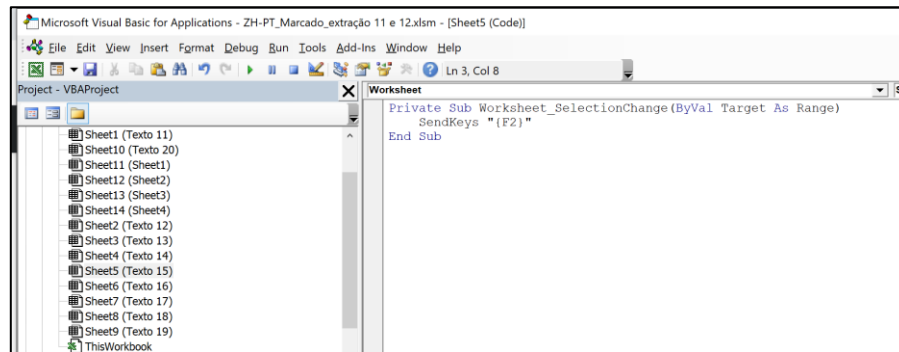
[illegible]

Fonte: elaborado pelo autor.

Depois de alinharmos todos os textos, realizamos algumas adaptações no programa *Excel* no modo de desenvolvedor para agilizar o trânsito por entre as diferentes células das planilhas geradas.

A primeira delas foi tornar possível a edição do conteúdo das células sem a necessidade de clicar nelas duas vezes ou ter que usar a tecla F2, localizada na fileira mais distante das mãos no teclado. Esta adaptação se deu com uso de um código aberto específico com o editor *Visual Basic Editor* (VBA) do *Excel*.

Figura 42 – Código VBA incorporado ao *Excel* para editar células sem dois cliques

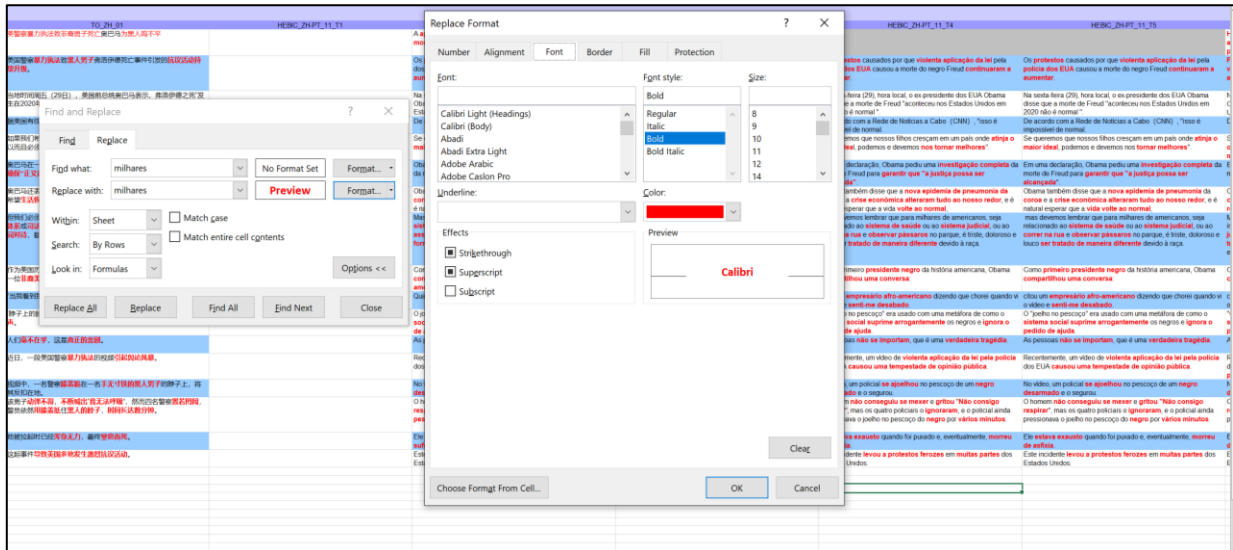


Fonte: elaborado pelo autor.

Essa mudança foi feita com o propósito de facilitar e agilizar a marcação das opções tradutórias em cada linha, pois dessa forma, com apenas um dos dedos, poderíamos saltar de um segmento de tradução para o outro, e com os outros dedos, com o comando shift+→, selecionar as traduções das colocações, sem ter, para isso, que usar o mouse ou mudar a posição das mãos sobre a mesa ou teclado.

Optamos por marcar as opções de tradução com uma coloração diferente e um asterisco “*” no final da unidade tradutória, por motivos que serão explicados adiante. Sem a necessidade de clicar duas vezes nas células para editá-las, muitas vezes, o processo de mudança de cor foi possível com o uso da opção avançada dos comandos *find* (encontrar) e *replace* (substituir). Quando uma opção notadamente se repetia várias vezes, o que foi possível de ser observado graças ao arranjo de todas as traduções uma do lado da outra, as buscávamos com o *find* e mudávamos sua cor com o *replace*.

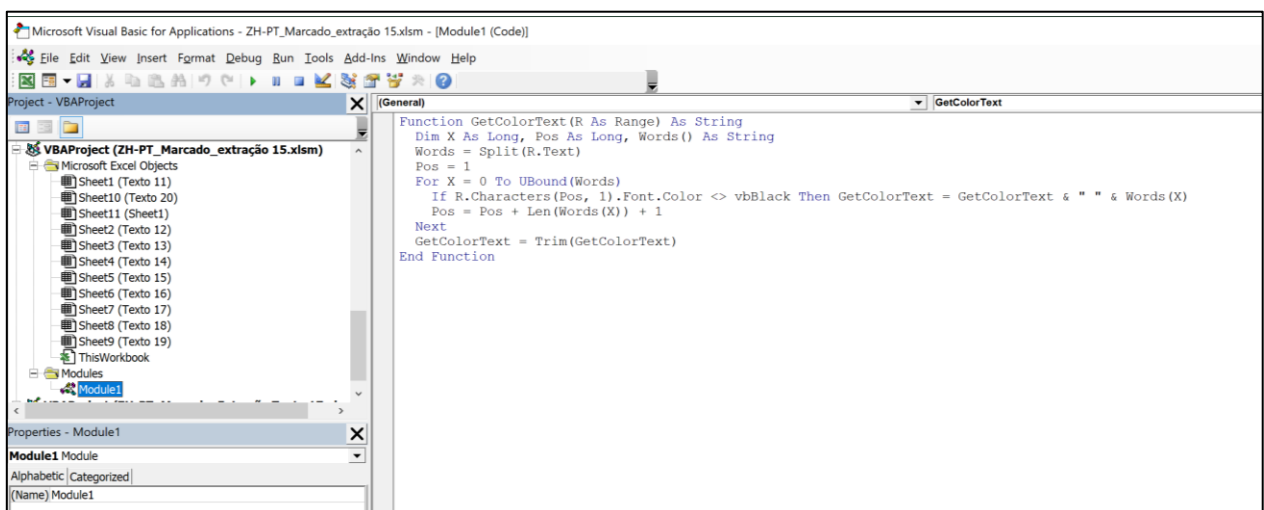
Figura 43 – Configurações da ferramenta *find e replace* para marcação dos TTs



Fonte: elaborado pelo autor.

A opção por marcar as traduções com uma cor diferente tem um propósito bem específico. Depois de marcarmos todas as traduções, criamos colunas vazias entre elas que receberiam o resultado da extração dos textos de cor diferente, i. e., as opções tradutórias, graças à edição de Macros com um código aberto no VBA e a aplicação de uma fórmula nessas colunas recém-criadas. O código VBA abaixo tornou a extração desses textos possível.

Figura 44 – Código VBA incorporado ao *Excel* para extração de texto com base na cor



Fonte: elaborado pelo autor.

	HEHC_24PT_11,11	HEHC_24PT_11,12
788法 田嶋男子 田嶋仁 田嶋平 田嶋		A notificação da lei eleitoral, pela apólice americana, contém

Fonte: elaborado pelo autor

Após a extração dos textos marcados das colunas, apenas as traduções foram copiadas (ctrl+c) e coladas sem formatação (botão direito> *paste options>match destination formatting*) em outra planilha.

A horizontal timeline diagram illustrating the sequence of events in the study. The timeline is divided into five segments labeled A, B, C, D, and E. Segment A is the longest and contains a sub-segment A1. Segment B is a shorter segment following A. Segments C, D, and E are the shortest and follow B in sequence.

Fonte: elaborado pelo autor

feitas pelos alunos. Cada coluna correspondia a um texto do *corpus* e permaneceu devidamente identificada com o código de identificação na primeira linha.

Concluída essa etapa, o próximo passo foi rearranjar as colunas, transformando-as em linhas e vice-versa, de modo que na primeira linha da planilha se encontrassem as colocações em L1 e na primeira coluna à esquerda, a identificação do texto.

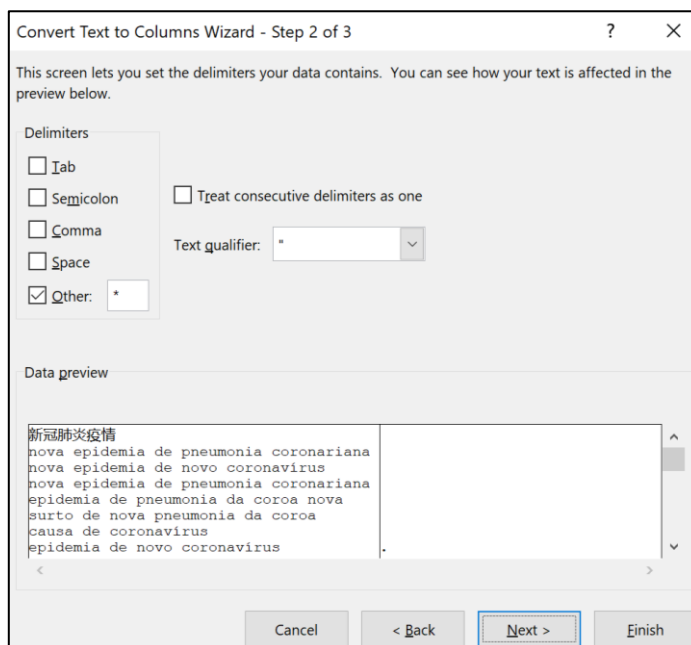
Figura 47 – Traduções das colocações após acomodação em nova planilha

	A	B	C	D
1	TO_PT_04	sofrendo os efeitos cada vez mais devastadores uso de drogas	doença apenas das grandes cidades se espalhou todo o País,	zona rural recebem o pagamento sema
2	UNESP_PT-EN_04_T1	produzidas e distribuídas tipo de gente	big city disease has spread countrywide,	rural zone get paid go get
3	UNESP_PT-EN_04_T2	is suffering the more and more devastating effects kind of people	disease only in the big cities has spread entire country,	rural zone receive their daily payment n
4	UNESP_PT-EN_04_T3	is suffering from the effects each time more devastating of drug use produced and distributed type of people	disease only of big cities spread all over the country,	countryside receive the weekly payment
5	UNESP_PT-EN_04_T4	is suffering increasingly devastating effects of drug use are produced and distributed kind of people-traffickers.	disease only from the big cities now it spread whole country,	countryside receive the weekly payment
6	UNESP_PT-EN_04_T5	is suffering effects each more devastating from the use of drugs produced and distributed kind of people:	disease of big cities has spread whole country,	rural area receive weekly payment go a
7	UNESP_PT-EN_04_T6	devastating effects of consumption of drugs produced and distributed kind of people	disease big cities has spread whole country,	rural zones, are paid run after
8	UNESP_PT-EN_04_T7	is suffering the each time more devastating effects of using drugs that are produced and distributed type of people	disease of major cities is now spread whole country,	rural areas receive their salaries run fas
9	UNESP_PT-EN_04_T8	has been suffering from the increasing devastating effects of drugs produced and distributed kind of person	problem of only big cities has spread all the country,	rural area receive the weekly payment n
10	UNESP_PT-EN_04_T9	has been suffering the effects increasingly devastating by drugs produced and distributed worse kind of people	big city disease has spread throughout the country,	rural zone receive their weekly pay run
11		has been suffering from every time more devastating effects of drug use - drugs who is produced type of people:		
12				
13				
14				
15				

Fonte: elaborado pelo autor.

Após o rearranjo das linhas e colunas, foi feita a separação, também automática, das unidades tradutórias por meio da função *text to columns*, do próprio *Excel*. Usamos como delimitador o asterisco “*” colocado no final de cada opção de tradução.

Figura 48 – Conversão de textos em colunas para separação das unidades tradutórias



Fonte: elaborado pelo autor.

Depois de divididas, cada instância foi checada e ajustes foram feitos, quando necessário, para que as opções tradutórias para uma determinada colocação permanecessem alinhadas.

Figura 49 – Colocações e suas traduções após conclusão do processo de identificação e extração

	A	B	C	D	E	F	G	H
1	新冠病毒歧视	新冠病毒	(非洲人)歧视待遇	新冠病毒疫情	有关当局	非洲裔人	就业歧视	住房歧视
2	Nova discriminação de Covid-19	Covid-19	tratamento discriminatório dos africanos	epidemia do novo coronavírus	autoridades relevantes	pessoas de África	discriminação em emprego, discriminação	em habitação
3	nova discriminação do vírus da coroa	coronavírus	tratamento discriminatório dos africanos	nova epidemia de coronavírus	autoridades relevantes	afrodescendentes	discriminação no emprego	discriminação no
4	A discriminação do coronavírus	coronavírus	discriminação contra os africanos	epidemia do coronavírus	autoridades relevantes	afrodescendentes	discriminação no emprego	discriminação no
5	discriminação do novo coronavírus	novo coronavírus	discriminar os africanos	nova epidemia do coronavírus	autoridades	pessoas de ascendência africana	discriminação no emprego	discriminação no
6	Discriminação do coronavírus	coronavírus	tratamento discriminatório dos africanos	epidemia de coronavírus	autoridades relevantes	afrodescendentes	discriminação no emprego	discriminação no
7	discriminação de novo coronavírus	novo coronavírus	tratamento discriminatório dos africanos	nova epidemia de coronavírus	autoridades relevantes	afrodescendentes	discriminação no emprego	discriminação na
8	discriminação de epidemia	epidemia	tratamento discriminatório para os africanos	nova epidemia de coronavírus	autoridades relevantes	afrodescendentes	discriminação no emprego	discriminação no
9	A Discriminação Ligado ao Coronavírus	Coronavírus	discriminação contra os africanos	epidemia do Coronavírus	autoridades relevantes	afrodescendentes	discriminação no emprego	discriminação no
10	Discriminações de novo coronavírus	novo coronavírus	tratamento discriminado de africanos	epidemia de novo coronavírus	autoridades relevantes	pessoas de ascendência africana	discriminações no emprego	discriminações n
11	A discriminação para novo coronavírus	novo coronavírus	tratamento de discriminação para os africanos	novo coronavírus	autoridades relativas	africanos	discriminações na área do emprego	discriminações n
12	A discriminação sobre o novo coronavírus	novo coronavírus	tratamento discriminatório dos africanos	surto do novo coronavírus	autoridades relevantes	afrodescendentes	discriminação em emprego	discriminação em
13	A discriminação de Covid-19	Covid-19	tratamento discriminatório dos africanos	surto de Covid-19,	autoridades	pessoas de origem africana	discriminação no emprego	discriminação na
14	Discriminação do coronavírus	coronavírus	tratamento discriminatório dos africanos	coronavírus	autoridades relevantes	afrodescendentes	discriminação no emprego	discriminação na
15	nova coronavírus discrimina	nova coronavírus	tratamento discriminatório dos africanos	surto de novel coronavírus,	autoridades	pessoas de origem africana	discriminação no emprego	discriminação na
16	Discriminação do coronavírus	coronavírus	tratamento de discriminação contra os africanos	situação do coronavírus	autoridades competentes	descendentes africanos	discriminação na área de emprego	discriminação na
17	Nova discriminação do Coronavírus	Coronavírus	tratamento discriminatório dos africanos	nova epidemia de coronavírus,	autoridades	pessoa de ascendência africana	discriminação no emprego	discriminação na

Fonte: elaborado pelo autor.

Uma vez concluída essa etapa do trabalho, fizemos o recorte das colocações que seriam de fato analisadas, devido ao elevado número de colocações e traduções levantadas, e criamos uma espécie de ficha para cada uma delas, também no *Excel*, como observamos abaixo.

Figura 50 - Exemplo de ficha de análise das traduções

<i>LogDice</i>	Frequência	Dar educação	
0	1491	3 give education	
nc	nc	1 educate	
nc	nc	1 teaches	
7.2	120401	1 provide education	
0	3662	1 have education	
Total		7	
c-Col		14% Sign. Dif.	0%
c-nCol		57%	
nc Certo		14%	
nc Errado		14%	
of		0%	
OpM			

Fonte: elaborado pelo autor.

Na primeira linha da ficha, em negrito, encontra-se a colocação cujas traduções serão analisadas. Na primeira coluna da parte superior da ficha são indicados os *scores logDice* para cada opção apresentada, seguida da coluna com as frequências dessas opções no *corpus* de referência. Na quarta coluna temos a frequência de cada opção dentro do *corpus* de estudo.

O preenchimento completo da ficha só pôde ser feito depois que a frequência de cada opção dentro do *subcorpus* de TTs e o *score* e a frequência dentro do *corpus* de referência (*enTenTen20* e *ptTenTen11*) foram obtidas com o SkE para verificarmos se elas correspondiam a colocações na L2.

De maneira padronizada, primeiramente buscamos pela base de determinada opção tradutória apresentada pelos aprendizes no *Word Sketch* para nos certificarmos de que o colocado usado na tradução realmente coocorria frequentemente com a base e que a combinação preenchia os requisitos para ser considerada uma colocação, ou seja, *logDice* acima de 3. Quando essa busca não apresentava resultados satisfatórios, buscávamos o colocado dado como opção pelos aprendizes para nos certificarmos de que ele coocorria com a base. Caso essa busca também não apresentasse resultados satisfatórios ou não apresentasse resultado algum, buscávamos pela combinação no *Concordance* para atestar a frequência no *corpus* de referência.

Para nos assegurarmos de que as soluções tradutórias significavam o mesmo que a colocação do TO, analisamos a ocorrência de cada opção dentro de seu respectivo contexto com o *Concordance*. Dessa forma, demos início à análise das colocações, e registramos os resultados

na parte inferior de cada ficha, em que o significado de cada código é apresentado na Tabela 6 adiante.

Tabela 6 - Códigos usados para análise das colocações

Total	Total de Traduções	Total de opções tradutórias apresentadas no <i>subcorpus</i> de TTs/L2 para uma colocação dos TOs/L1.
c-Col	Combinação-Colocação	Combinação lexical que é uma colocação existente na L2.
Sig. Dif.	Significado Diferente	Quando a opção atende os requisitos para ser considerada uma colocação da L2, ela teve seu significado analisado, tanto com o <i>Concordance</i> como com buscas adicionais na <i>internet</i> para verificar se o seu significado é realmente aquele da colocação no TO. Quando o significado não era o mesmo do TO, a colocação era considerada “c-Col”, porém com significado diferente, “Sig. Dif.”
c-nCol	Combinação-Não-Colocação	Combinação lexical que não é uma colocação na L2.
nc Certo	Não-combinação Certo	“Não-combinação”, i.e., unidade lexical isolada que traduz a colocação da L1, seja por ser a única opção possível, seja devido ao emprego de alguma estratégia de tradução alternativa que resultou no uso de uma unidade lexical isolada.
nc Errado	Não-combinação Errado	“Não-combinação”, i.e., unidade lexical isolada que não traduz a colocação da L1, caracterizando-se como um erro.

of	Outros fraseologismos	Usada para marcar traduções que não são colocações na L2 nos termos definidos na seção 2.5 deste trabalho, mas são outros fraseologismos, como <i>phrasal verbs</i> , expressões idiomáticas, <i>chengyu</i> ²⁷ , etc.
OpM	Melhor opção	Sugestão que potencialmente melhor traduz a colocação da L1, fruto de consulta ao <i>corpus</i> de referência.

Fonte: elaborado pelo autor

A escolha do *Excel* como local para hospedar as fichas se deu pela facilidade de analisar os dados posteriormente com as fórmulas do programa. A opção por registrar os resultados das análises de cada colocação em porcentagem foi devida ao fato de cada TO ter sido traduzido por um número diferente de alunos e, naturalmente, devida ao fato de cada colocação poder aparecer mais de uma vez ao longo do *corpus*. Dessa forma, usar números inteiros dificultaria o processamento dos dados num segundo momento.

Os resultados da análise serão apresentados no próximo capítulo.

²⁷ *Chengyu*, do chinês 成语 *chéngyǔ*, é um tipo de expressão idiomática tradicional chinesa e um dos quatro tipos de expressões formulaicas em chinês.

4 ANÁLISE DOS RESULTADOS

Este capítulo divide-se em duas seções, uma referente aos resultados da compilação dos *corpora* e outra, aos resultados do levantamento das colocações e soluções tradutórias. Na seção 4.1 a seguir, iniciamos pelos resultados obtidos na compilação do CAT chinês.

4.1 Resultados da compilação do CAT chinês

No total, foram compiladas 150 traduções na direção L1 (chinês) → L2 (português), i.e., LM→LE, que passaram a compor o CAT 3, o CAT chinês.

Dos 18 alunos que iniciaram as traduções, houve apenas três desistências, duas no HEBIC e uma na HUBU. As principais alegações dos alunos foram a falta de tempo devido a estágios obrigatórios, a busca por um emprego para depois da formatura e algumas complicações causadas pela pandemia da COVID-19. No período de compilação do nosso *corpus*, muitos voluntários encontravam-se em outros países. Com o fim de seus intercâmbios coincidindo com a auge da pandemia de COVID-19, muitos deles foram forçados a voltar para seu país natal, a China, e cumprir longos períodos de quarentena no trajeto e no destino final. Além disso, alguns alunos relataram ansiedade dada a situação caótica do período de epidemia.

Ao final, mesmo assim, as traduções no sentido L1→L2 do HEBIC totalizaram 42 e da HUBU, 108, como podemos ver na Tabela 7.

Tabela 7 Número de textos coletados na compilação do CAT chinês

	TO11	TO12	TO13	TO14	TO15	TO16	TO17	TO18	TO19	TO20
HEBIC	5	5	4	4	4	4	4	4	4	4
HUBU	10	10	11	12	11	10	11	11	11	11

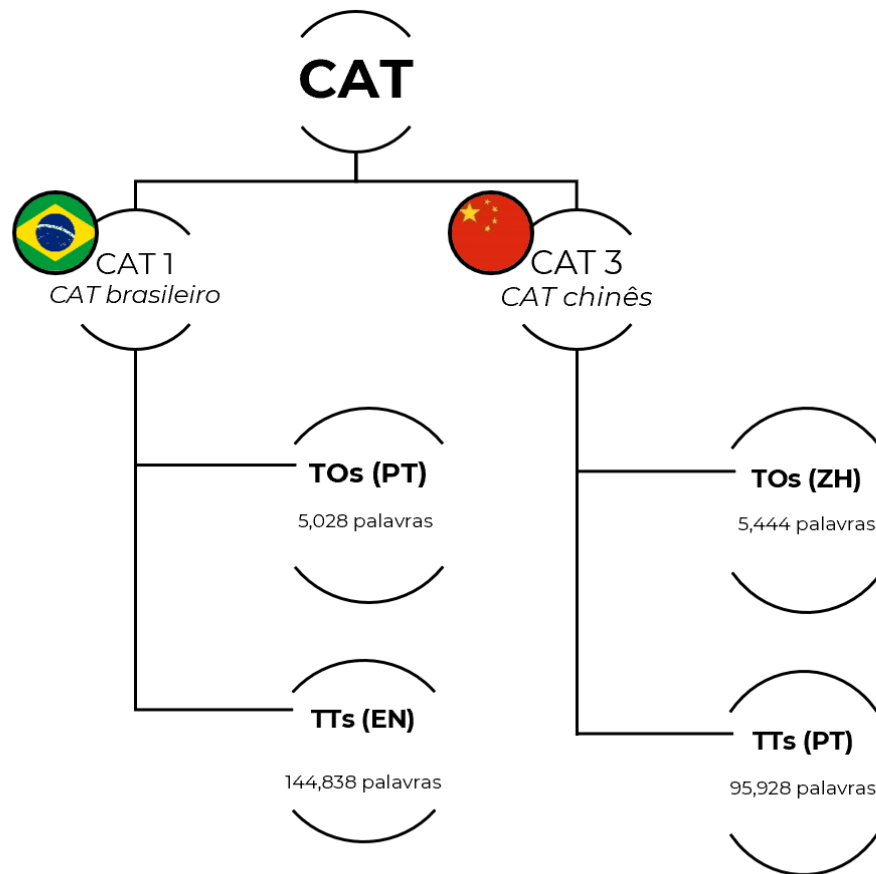
Fonte: elaborado pelo autor.

A Tabela 7 tem a contagem dos textos iniciada em “11” porque os textos de 1 a 10 foram escritos originalmente em português brasileiro para serem traduzidos para LE.

Juntos, segundo contagem do SkE, os TTS geraram um *subcorpus* de 109.914 palavras, ao passo que o *subcorpus* de TOs em português conta com 5.894 palavras.

No organograma abaixo, sintetizamos esses e os dados do corpus brasileiro.

Organograma 2 – Esquema do CAT 1 e CAT 3



Fonte: elaborado pelo autor.

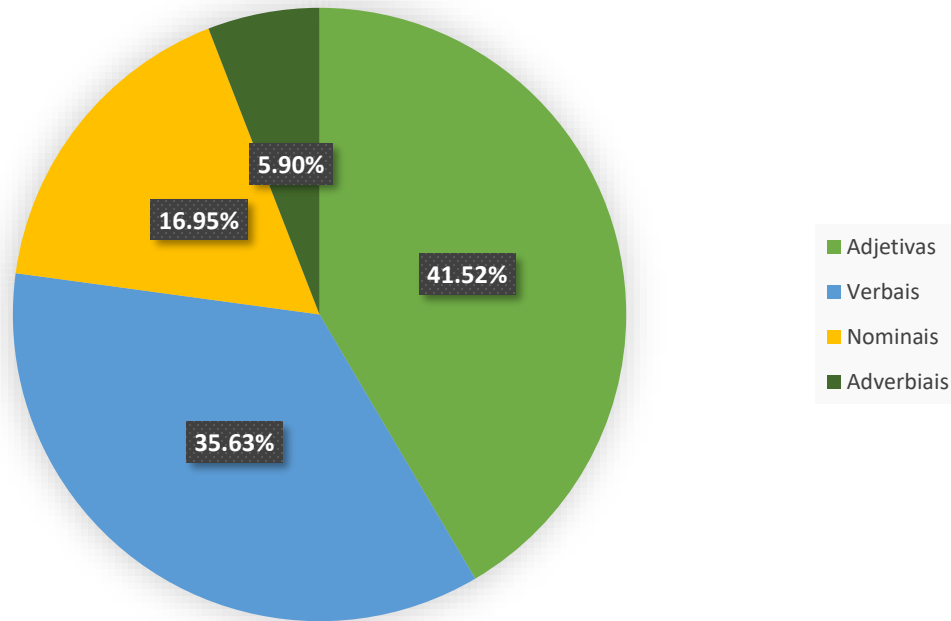
A seguir, tratemos dos resultados do segundo momento, o de levantamento e análise das soluções tradutórias das colocações.

4.2 Resultados do levantamento das colocações

Ao final da nossa meticulosa seleção, conforme mencionamos no capítulo “Procedimentos metodológicos”, subseção 3.2, chegamos a um total de 1006 colocações

levantadas em ambos os *subcorpora* de L1. Desse total, 407 foram colocações em português e 599 em chinês. Do total de colocações selecionadas em português, 169 (41.5%) são adjetivas, 145 (35.6%) são verbais, 69 (16.9%) são nominais, e 24 (5.8%) são adverbiais.

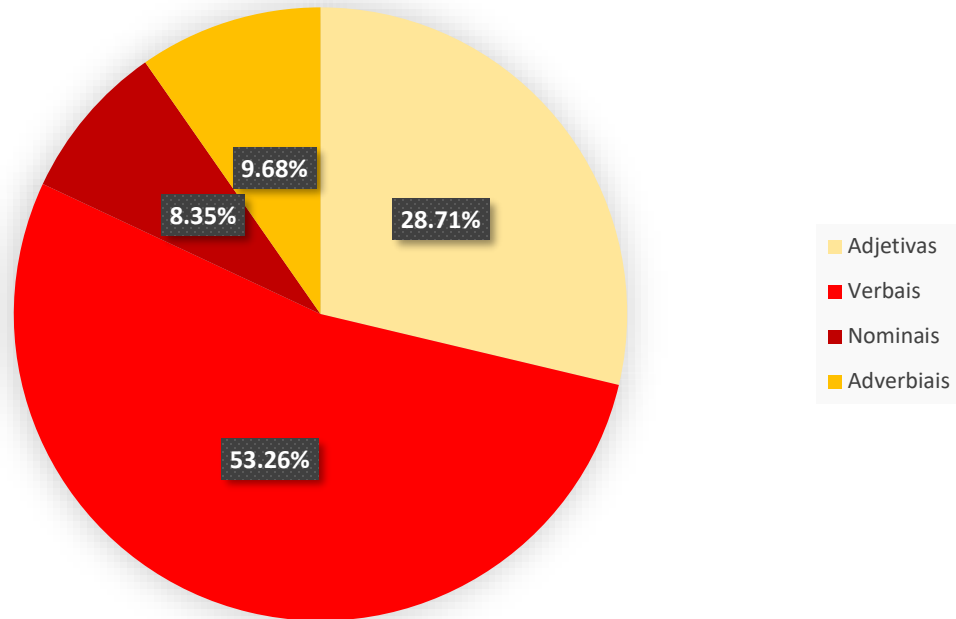
Gráfico 10 – Proporção de colocações por tipo, no *subcorpus* L1(português)



Fonte: elaborado pelo autor.

Essa proporção não se manteve nas colocações em chinês, em que, diferentemente do *corpus* brasileiro, cuja maior parte das colocações eram adjetivas, a maior parte, 319 (53,2%), são colocações verbais. As colocações adjetivas totalizam 172 (28,7%) do total de colocações do *corpus* chinês, seguidas pelas adverbiais, com um total de 58 (9,6%), e as nominais, com 50 (8,3%).

Gráfico 11 – Proporção de colocações por tipo no *subcorpus* L1(chinês)



Fonte: elaborado pelo autor.

Apesar de termos recolhido as traduções das 1006 colocações levantadas em ambos os *subcorpora* de TTs, reconhecemos que esse é um número muito elevado, que exigiria muito tempo de análise. Por essa razão, realizamos um recorte de aproximadamente 15% do total, 150 (14,9%), de colocações que seriam realmente analisadas, por entendermos ser uma quantidade representativa do *corpus* como um todo, que possibilitaria incluir na nossa amostragem um número razoável dos diferentes tipos de colocações, verbais, adverbiais, adjetivas e nominais, respeitando-se, ainda que de maneira aproximada, a proporção em que elas ocorrem em ambos os *subcorpora*, para a realização das análises dentro das limitações de tempo impostas ao projeto. Das 150 colocações selecionadas, 69 (46%) eram de colocações do português e 81 (54%) do chinês. Como critérios para seleção utilizamos a força de atração, baseada nos *scores logDice* e a frequência.

No mais, como mencionado anteriormente, no recorte feito, buscamos, mesmo que de forma aproximada, respeitar a proporção de cada tipo de colocação em relação ao total geral e, conjuntamente, balancearmos o número de colocações entre os *subcorpora*.

Das 69 colocações do português, 28 (40,5%) eram verbais, 26 (37,6%) adjetivas, 11 (15,9%) nominais e 4 (5,7%) adverbiais.

Das 81 colocações em chinês, 39 (48,1%) eram verbais, 24 (29,6%) adjetivas, 14 (17,2%) nominais e 6 (7,4%) adverbiais.

Do total de colocações, poucas foram aquelas cujas traduções não são combinações lexicais. Apenas 8% de todas as colocações levantadas foram traduzidas por apenas uma unidade lexical isolada pela maioria dos voluntários, i.e., pelo menos metade mais um.

Esse fenômeno é mais comum no *subcorpus* chinês, em que 4,9% das colocações desse *subcorpus* foram corretamente traduzidas para o português por uma única unidade lexical por todos os voluntários. É o caso de 受到对待 *shòudào duìdài* (ser tratado, lit. receber tratamento), cujo resultado das estratégias de tradução empregadas foi “tratar”, “ser tratado” e “tratamento”, como podemos observar nas linhas de concordância abaixo.

Quadro 1 - Linhas de concordância de 受到对待 no subcorpus de TOs

但我们必须记住，对于成千上万的美国人来说，无论涉及医疗保健体系或司法体系，还是在街上慢跑、在公园鸟时，因种族受到不同对待，都是可悲的、痛苦的、疯狂的。²⁸

Dàn wǒmen bīxū jì zhù, duìyú chéng qiān shàng wàn dì měiguó rén lái shuō, wúlùn shèjì yīliáo bǎojiàn tǐxì huò sīfǎ tǐxì, háishì zài jiē shàng màn pǎo, zài gōngyuán niǎo shí, yīn zhǒngzú shòudào bùtóng duìdài, dōu shì kěbēi de, tòngkǔ de, fēngkuáng de.

其他外国人群体基本上都没有受到这种对待。"中国当局声称对歧视'零容忍'，但广州对非洲人所做的正是典型的歧视行为，"人权观察中国部研究员王亚秋说。²⁹

Qítā wàiguó rén qúntǐ jīběn shàng dū méiyǒu shòudào zhè zhǒng duìdài. "Zhōngguó dāngjú shēngchēng duì qíshì' líng róngrěn', dàn guǎngzhōu duì fēizhōu rén suǒ zuò de zhèng shì diǎnxíng de qíshì xíngwéi, "rénquán guānchá zhōngguó bù yánjiùyuán wángyàqiū shuō

Fonte: elaborado pelo autor

Figura 51 - Linhas de concordância das soluções tradutórias encontradas para 受到对待

1	☐	doc#0 ericanos, seja diz respeito ao sistema sanitária ou judicial seja a respeito aos tratamentos diferentes, devido à raça desigual, quando faz o trote largo ou aprecia os pás:	
1	☐	doc#0 na rua, quando você observa pássaros no parque, é triste, doloroso e louco ser tratado de maneira diferente por causa da raça. </s><s> Como o primeiro presidente ne	
2	☐	doc#0 l, dar uma corrida na rua, observar os pássaros no parque, nesses casos foram tratados de uma forma diferente devido à etnia, tudo que é triste, angustiado, doido "	
3	☐	doc#0 a justiça, ou correr na rua ou assistir pássaros no parque, é triste e doloroso ser tratado de forma diferente por causa da raça. </s><s> Como o primeiro presidente negr	
4	☐	doc#0 la na rua ou a observação de pássaros no parque, é triste, doloroso e louco ser tratado de maneiras diferentes devido à raça." Como o primeiro presidente negro da his	
5	☐	doc#0 a saúde ou da justiça, quando corre na rua ou observa pássaros no parque, ser tratado de maneira diferente devido à raça é triste, doloroso e louco." Como o primeiro	
6	☐	doc#0 le ou o sistema de justiça, seja correr na rua ou observar pássaros no parque, é tratado de maneira diferente devido à raça diferente, que é triste, dolorosa e louca. " Cc	
7	☐	doc#0 e. </s><s> Mas temos de nos lembrar de que, para milhões de americanos, ser tratados de forma diferente por causa da raça é trágica, dolorosa e louca - quer se trate	
8	☐	doc#0 ados de forma diferente por causa da raça é trágica, dolorosa e louca - quer se trate de lidar com o sistema de saúde, quer de lidar com o sistema de justiça penal, c	
9	☐	doc#0 <s> O sistema, mesmo quando corre na rua ou observa pássaros no parque, é tratado de maneira diferente devido à corrida, que é triste, dolorosa e louca ". </s><s> (	
10	☐	doc#0 u ao correr na rua e observar pássaros no parque, é triste, doloroso e louco ser tratado de maneira diferente devido à raça. </s><s> Como primeiro presidente negro de	
11	☐	doc#0 as ruas ou ao olhar para as aves nos parques, é patético, doloroso e louco ser tratado de forma diferente de acordo com a raça." Como o primeiro presidente negro ne	
12	☐	doc#0 femia do coronavírus, é preciso lembrar que milhões dos norte-americanos são tratados de maneira diferente por causa da sua cor, tanto no sistema de saúde social ou	
13	☐	doc#0 u do sistema judicial, quando corre na rua ou observa pássaros no parque, é tratado de maneira diferente devido à raça, que é triste, dolorosa e louca". </s><s> Cor	

Fonte: captura de tela do SkE.

Em alguns casos, a tradução mais adequada é realmente uma “não-combinação”, i.e., uma unidade lexical isolada. É o que acontece, por exemplo, com a colocação nominal do português

²⁸ Mas devemos lembrar que, para milhares de americanos, seja no sistema de saúde ou no sistema judiciário, seja ao caminhar na rua ou ao observar pássaros no parque, ser tratado de forma diferente por causa da raça é triste, doloroso e enlouquecedor.

²⁹ Basicamente, nenhum outro grupo estrangeiro foi tratado desta forma. "As autoridades chinesas afirmam que há 'tolerância zero' para a discriminação, mas o que Guangzhou fez contra os africanos é um ato típico de discriminação", disse Wang Yaqui, pesquisador do Departamento de Direitos Humanos da China.

“projeto de lei”, cujo correspondente em língua inglesa é *bill*. No nosso *corpus* de estudo, 75% das traduções para “projeto de lei” são de fato *bill*, porém, outros 25% são *law project* e 4% são unidades lexicais isoladas que não traduzem adequadamente a colocação.

Figura 52 - Ficha com as soluções tradutórias para "projeto de lei"

LogDice	Frequência	Projeto de lei
nc	nc	18 bill
0	8348	5 law project
nc	nc	1 project
Total		24
c-Col		21% Sign. Dif.
c-nCol		0
nc Certo		75%
nc Errado		4%
of		0%
OpM		

Fonte: elaborado pelo autor

Na ficha acima, vemos que nas colunas de *logDice* e frequência, *bill* e *project* receberam o código “nc”, que, conforme explicado no capítulo “Procedimentos metodológicos”, subseção 3.2.3, é o código que usamos para marcar soluções tradutórias que são uma “não-combinação”, i.e., unidade lexical isolada. Cabe aqui lembrar que na parte inferior da ficha há indicação da porcentagem de “nc” que traduzem corretamente – nc Certo – e incorretamente – nc Errado – a colocação da L1. No caso em questão, *bill* classifica-se como “nc Certo” e *project*, “nc Errado”.

É válido lembrar também que “c-Col” é código usado para combinações lexicais que preenchem os requisitos para serem consideradas colocações na L2 e “c-nCol” para combinações lexicais que não se caracterizam como colocações existentes na L2, são, portanto, erros colocacionais. Já o código “of” é para traduções que não são colocações, mas sim outros fraseologismos, como *phrasal verbs*, expressões idiomáticas, entre outros.

Acerca da opção tradutória considerada a mais adequada para “projeto de lei”, *bill*, segundo o *Cambridge Dictionary*, é um documento oficial de uma nova lei que é discutido antes de ser votado no congresso ou parlamento.

Figura 53 - Definição de *bill* no *Cambridge Dictionary*

bill noun (LAW)

[C]

a formal statement of a planned new law that is discussed before being voted on:

- The bill was **amended** (= changed).
- When a bill is **passed** in Parliament it becomes law.
- informal The bill was **thrown out** (= did not go past the first stage of discussion and will not become law).

— More examples

- The bill was defeated in the Commons by 249 votes to 131.
- An amendment to the bill was agreed without a vote.
- The opposition was in full cry in Parliament last night over the proposed changes to the education bill.
- Politicians tried to dress up the bill as a bold new strategy for combatting poverty.
- He is unlikely to succeed in getting his bill through Congress, however worthy it is.

Fonte: *Cambridge Dictionary*.

Apesar de a busca por colocados de *law* com o *Word Sketch* no *enTenTen20* não apresentar *project* como resultado e vice-versa, encontramos a ocorrência de *law project* com a ferramenta *Concordance*. No entanto, como podemos depreender a partir das linhas de concordância abaixo, *law project* está, majoritariamente, associado a nome de empresas, organizações governamentais ou não-governamentais e iniciativas de universalização do direito e assistência a pessoas desfavorecidas.

Figura 54 - Linhas de concordância de *law project*

41		cov.com	Klykovo (1997). </s><s> 5 Controversial Rules In The ALI's Insurance	Law Project	</s><s> John Buchanan and David Goodwin are quoted in a Law360 a
42		gmu.edu	tries. </s><s> She is a senior associate in the Democracy and Rule of	Law Project	, a research endeavor that analyzes the state of democracy around the
43		globaljusticece...	the dangers of governmental impunity through its work with the Burma	Law Project	. </s><s> The Burma Constitution perpetuates injustice as a policy by c
44		globaljusticece...	the processes of justice, unrest is bound to follow. </s><s> Our Burma	Law Project	seeks to challenge the same constitutional suppression of women's voi
45		sfbar.org	out. tags: Politics Governance New Political Party </s><s> Business	Law Project	</s><s> The Community Organization Representation Project (CORP)
46		iicrd.org	. of conflict. </s><s> In partnership with Makerere University's Refugee	Law Project	, this doctoral research project funded by the Social Sciences and Hum
47		communitylegala...	neighborhood. </s><s> As a staff attorney for Legal Aid's Neighborhood	Law Project	– designed to breathe... </s><s> Community Legal Aid is offering free v
48		indianalegalser...	/s><s> 151 N. Delaware Suite 1800 Indianapolis, In 46204 The Senior	Law Project	of Indiana Legal Services, Inc. gratefully acknowledges the support of C
49		harvard.edu	er capita in the US. </s><s> With partners at the National Employment	Law Project	and the support of the Ford Foundation and LIFT Fund, LWP is hosting
50		webjcli.org	.116] In many ways, this difficulty is inherent to the entire mental health	law project	, as Peay has pointed out. </s><s> See J Peay, Seminal Issues in Men
51		livingwagenyc.o...	Beth Shulman, who was chair of the board of the National Employment	Law Project	and co-chair of the Fairness Initiative on Low-Wage Work. </s><s> In tl
52		essex.ac.uk	Commonwealth Students' Association (CSA) </s><s> Our Essex Street	Law project	is one of the first of its kind and is the primary pro-bono project provide
53		freerepublic.co...	unded abortion. </s><s> Moreover, on January 16, 2019, the Womens	Law Project	and the Planned Parenthood Federation of America (PPFA) filed a law
54		simple-green-ili...	xates Manual 2010-2011 Edition Produced by the Pennsylvania Utility	Law Project	</s><s> AT&T Ohio Lifeline Application Application for Lifeline Lifeline i
55		mission-launch...	24, 2017, the Center for American Progress, the National Employment	Law Project	(NELP), and Community Legal Services (CLS) hosted a two-day conve
56		touchngo.com	38-8188 Please help us support these and other worthy organizations:	Law Project	for Psychiatric Rights Soteria-alaska Choices AWAIC </s><s> Easy D
57		livejournal.com	the "experts" in the documentary. </s><s> I think that the Sylvia Rivera	Law Project's	movie, "Toilet Training" is one such documentary, and I think it is like th
58		sfpublicpress.o...	ation. </s><s> According to a 2009 report by the National Employment	Law Project	, foreign-born Latinos had an especially high minimum wage violation r
59		ccgaction.org	January 29, 2013 – Catholics for the Common Good and the Marriage	Law Project	filed an amicus brief today with the Supreme Court in support of the Ca
60		dmlp.org	ortion of this site. </s><s> Welcome to the website of the Digital Media	Law Project	. </s><s> The DMLP was a project of the Berkman Klein Center for Inte

Fonte: captura de tela do SkE.

Essa constatação se confirma com uma breve busca por *law project* no *Google*, com organizações e instituições de apoio aparecendo no topo da busca.

Figura 55 - Resultados da busca por *law project* no *Google*

Fonte: captura de tela do *Google.com* gerada em 12 nov. 2021.

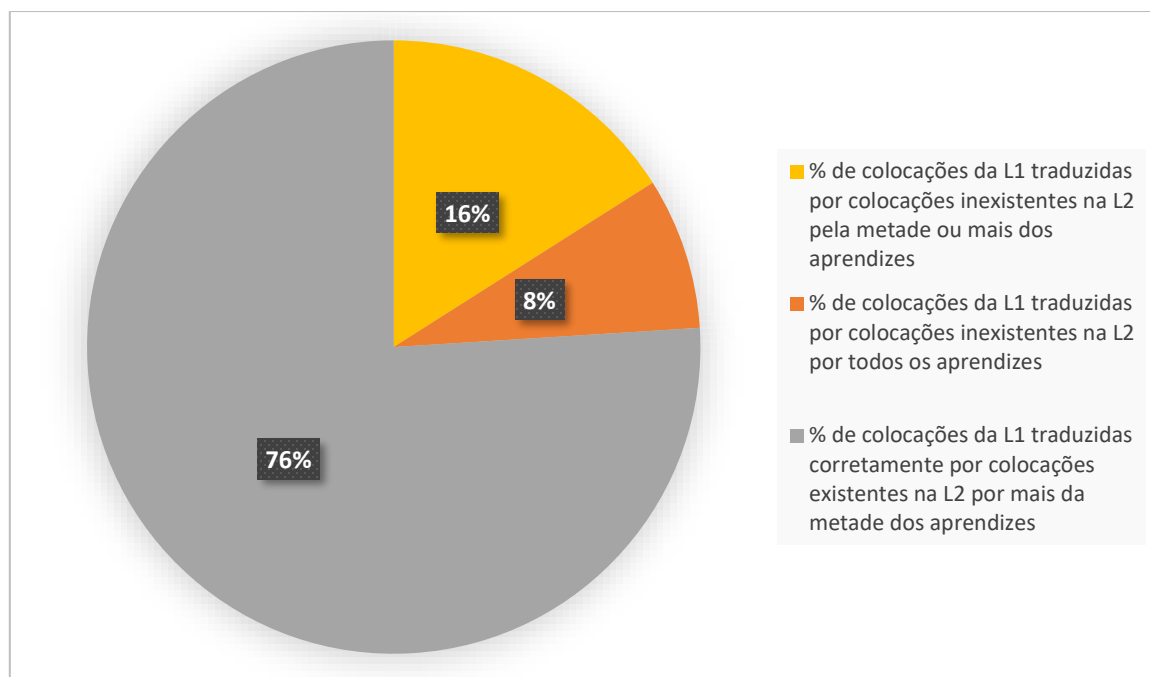
A partir das análises das linhas de concordância e dos resultados da busca no *Google*, podemos notar que *project* em *law project* é, em verdade, um projeto, uma iniciativa, uma organização, não contemplando, portanto, o significado de “projeto de lei” do português.

Porém, vale ressaltar que nem sempre a tradução de colocações por uma unidade lexical isolada resultou em traduções exitosas. Além disso, ainda houve casos em que uma colocação foi traduzida por uma frase.

Contudo, ao nos atermos às traduções que resultaram na produção de colocações, de maneira geral, dado o alto nível de proficiência dos voluntários, a maior parte delas estavam corretas, ou seja, eram colocações existentes na L2, o que não necessariamente significa dizer que o significado era o mesmo do TO e que eram adequadas a aqueles contextos, como veremos mais à frente.

Considerando-se os dois *subcorpora*, brasileiro e chinês, como um todo, 18% de todas as colocações da L1 foram traduzidas por colocações inexistentes na L2, i.e., erros colocacionais, pela metade ou pela maioria dos participantes da pesquisa; e 6% de todas as colocações de ambos os *subcorpora* foram traduzidas por colocações inexistentes na L2 por todos os participantes.

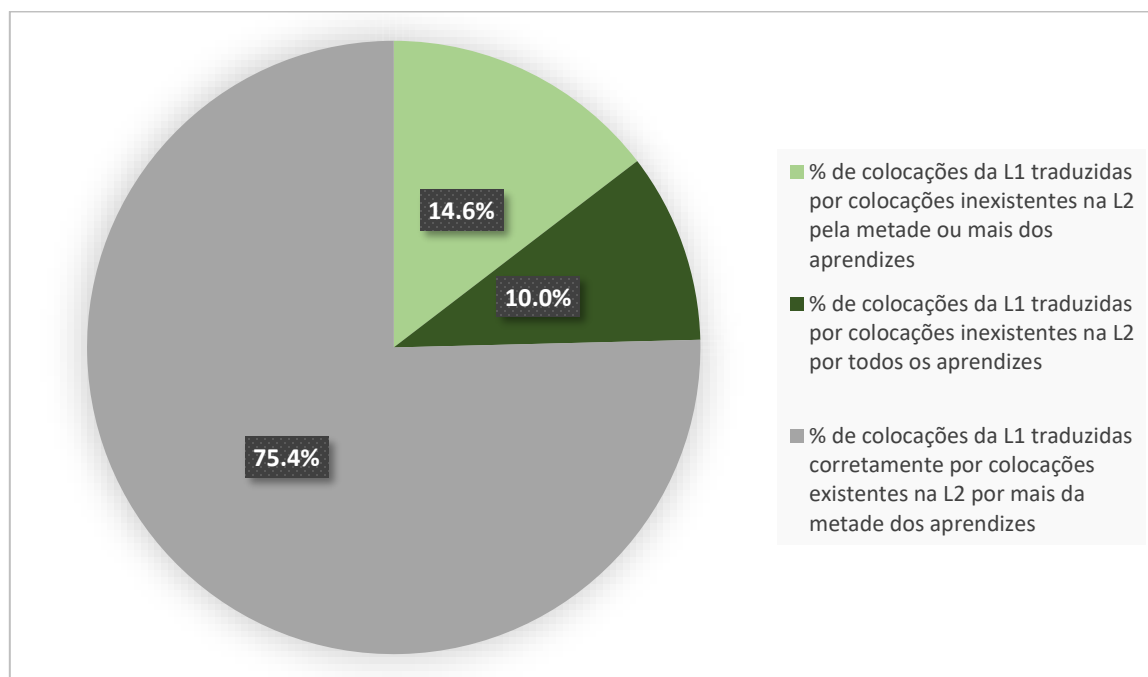
Gráfico 12 – Total de todas as colocações traduzidas por colocações inexistentes na L2 em ambos *subcorpora* brasileiro e chinês



Fonte: elaborado pelo autor.

Observando-se cada *subcorpora* isoladamente, no CAT brasileiro, 14,6% das colocações foram traduzidas por colocações inexistentes na L2 pela metade ou pela maioria dos voluntários; e 10% foram traduzidas por erros colocacionais por todos os participantes da pesquisa, como ocorreu com colocações como “aderir ao sistema” ou “ter compromisso”.

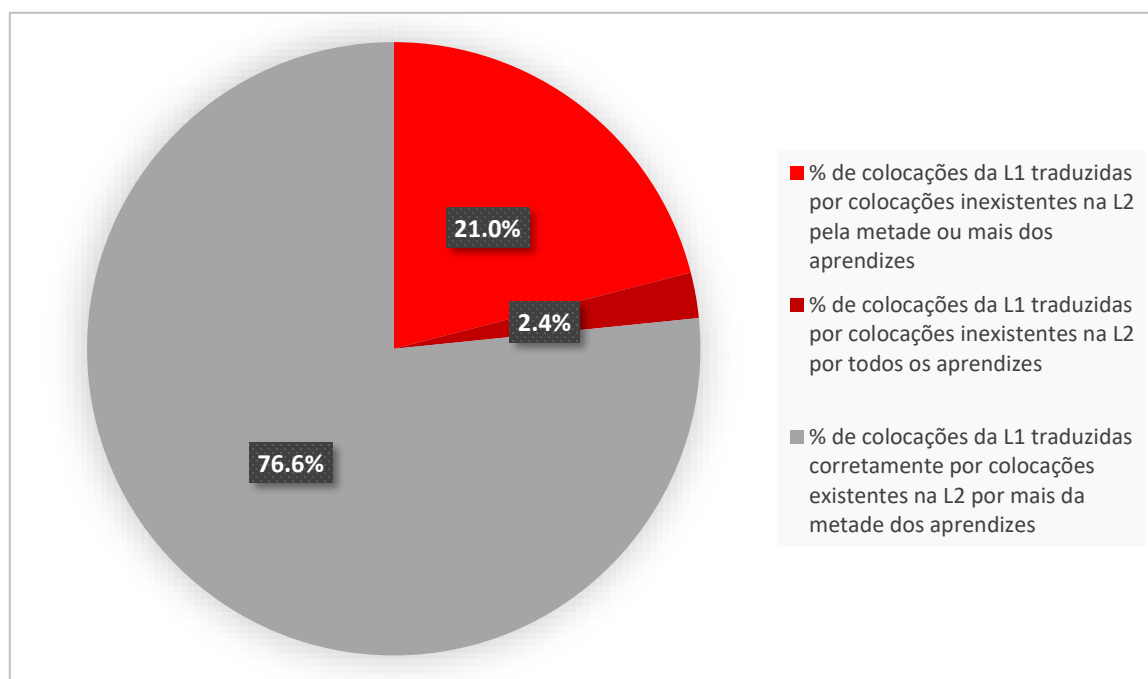
Gráfico 13 - Total de todas as colocações traduzidas por colocações inexistentes na L2 no CAT brasileiro L1(português) → L2(inglês)



Fonte: elaborado pelo autor.

Dos dados analisados no CAT chinês, 21% das colocações nesse *subcorpus* foram traduzidas por erros colocacionais pela metade ou pela maioria dos participantes; e apenas 2,4% foram traduzidas por erros colocacionais por todos eles, foi o caso de 超越法治 *chāoyuè fǎzhì* (afrontar ou subverter o Estado de Direito)

Gráfico 14 - Total de todas as colocações traduzidas por colocações inexistentes na L2 no CAT chinês L1(chinês) → L2(português)



Fonte: elaborado pelo autor.

No que concerne à tipologia das colocações, quando o *subcorpus* observado é o de estudantes brasileiros, o tipo de colocação que causou mais problema para os aprendizes foram as verbais. Dos dados analisados, do total de colocações verbais do CAT brasileiro, 21,8% foram traduzidas por colocações inexistentes na L2 pela metade ou pela maioria dos aprendizes, como ocorreu com “buscar por respeito” e “dar carta branca”; e 21% foram traduzidas por colocações inexistentes por todos os aprendizes, como ocorreu com “aderir ao sistema” e “ter compromisso”, mencionadas anteriormente, dentre outras como “sofrer bloqueio” e “deixar em segundo plano”.

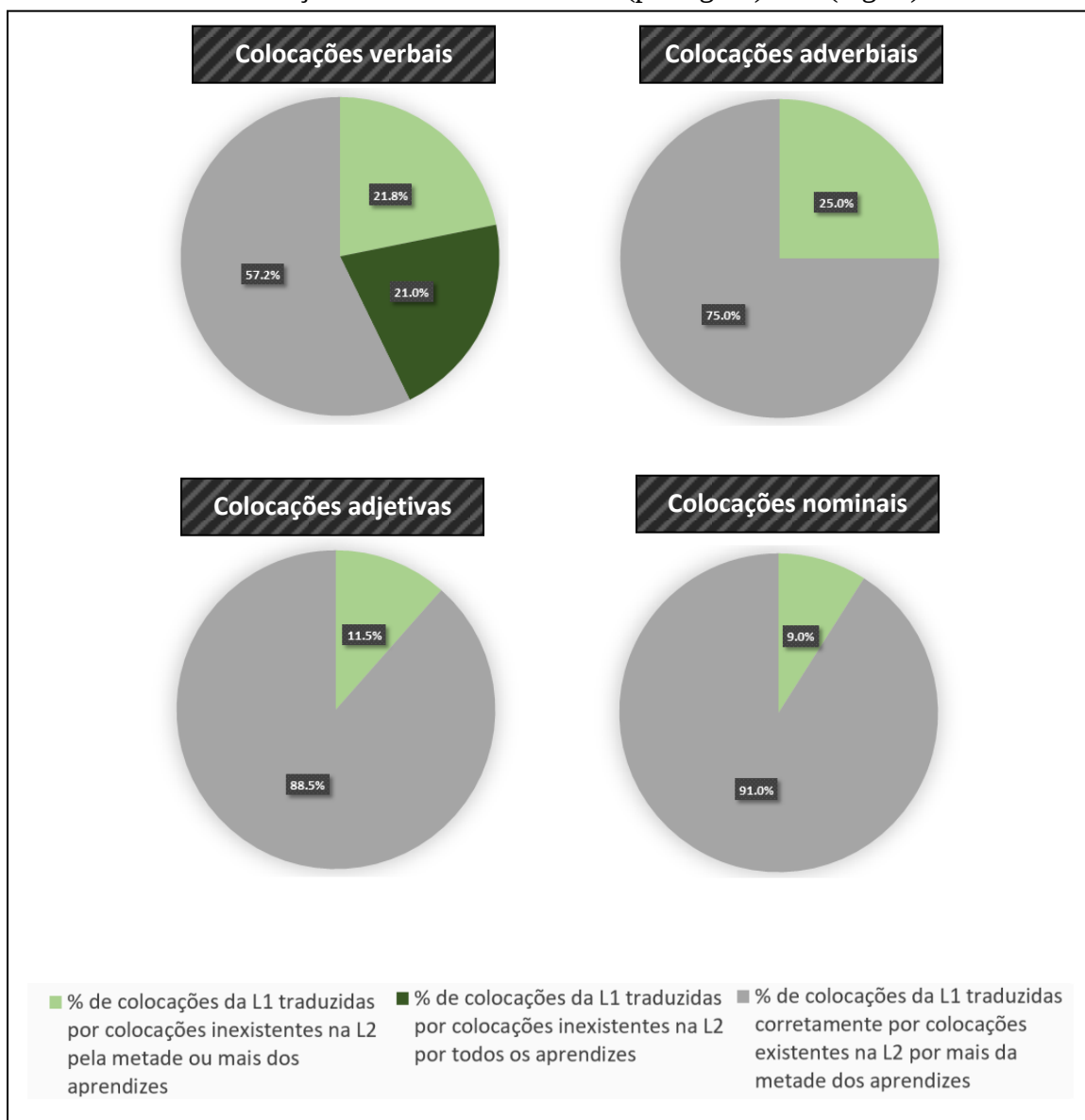
O segundo tipo de colocação no qual os alunos brasileiros cometeram mais erros foi o adverbial. Do total de colocações deste tipo no CAT brasileiro, 25% foram traduzidas por colocações inexistentes na L2 pela metade ou pela maioria dos participantes da pesquisa, como foi o caso de “engolir a seco” e “denunciar formalmente”. Porém, nenhuma das colocações adverbiais analisadas nesse *subcorpus* foi traduzida somente por erros colocacionais por todos os participantes da pesquisa.

O terceiro tipo com mais erros no CAT brasileiro foi o de colocações adjetivas. Do total de colocações deste tipo traduzidas pelos estudantes do Brasil, 11,5% foram traduzidas por colocações inexistentes na L2 pela metade ou pela maioria dos voluntários, como ocorreu com “plano inclinado”, “entrelaçamento social”, “questão palestina”, entre outras. Nenhuma das colocações deste tipo teve apenas erros colocacionais como solução tradutória.

O tipo menos problemático para os brasileiros foi o de colocações nominais – 9% do total de colocações deste tipo no CAT de L1(português) →L2(inglês) foram traduzidas por erros colocacionais pela metade ou pela maioria dos participantes da pesquisa, como ocorreu com “disseminação das drogas” e “combate às drogas”. Nenhuma colocação teve todas as soluções apresentadas por eles sendo colocações inexistentes na L2.

Os dados referentes às traduções que resultaram em erros colocacionais por tipo de colocação no CAT brasileiro seguem sintetizados no Gráfico 16 a seguir.

Gráfico 15 - Total de todas as colocações traduzidas por colocações inexistentes na L2 por tipo de colocação no CAT brasileiro - L1(português)→L2(inglês)



Fonte: elaborado pelo autor.

No caso do *subcorpus* de aprendizes chineses, as colocações que causaram mais problema aos estudantes também foram as verbais. Do total de colocações verbais analisadas no CAT chinês, 20,5% foram traduzidas por colocações inexistentes na L2, i.e., erros colocacionais, pela metade ou pela maioria dos participantes da pesquisa, como ocorreu com 停止歧视 *tíngzhǐ qíshì* (acabar com a discriminação) e 受到不人道待遇 *shòudào bùréndào dàiyù* (receber tratamento desumano, ser tratado desumanamente); e 5,1% foram traduzidas por colocações inexistentes na

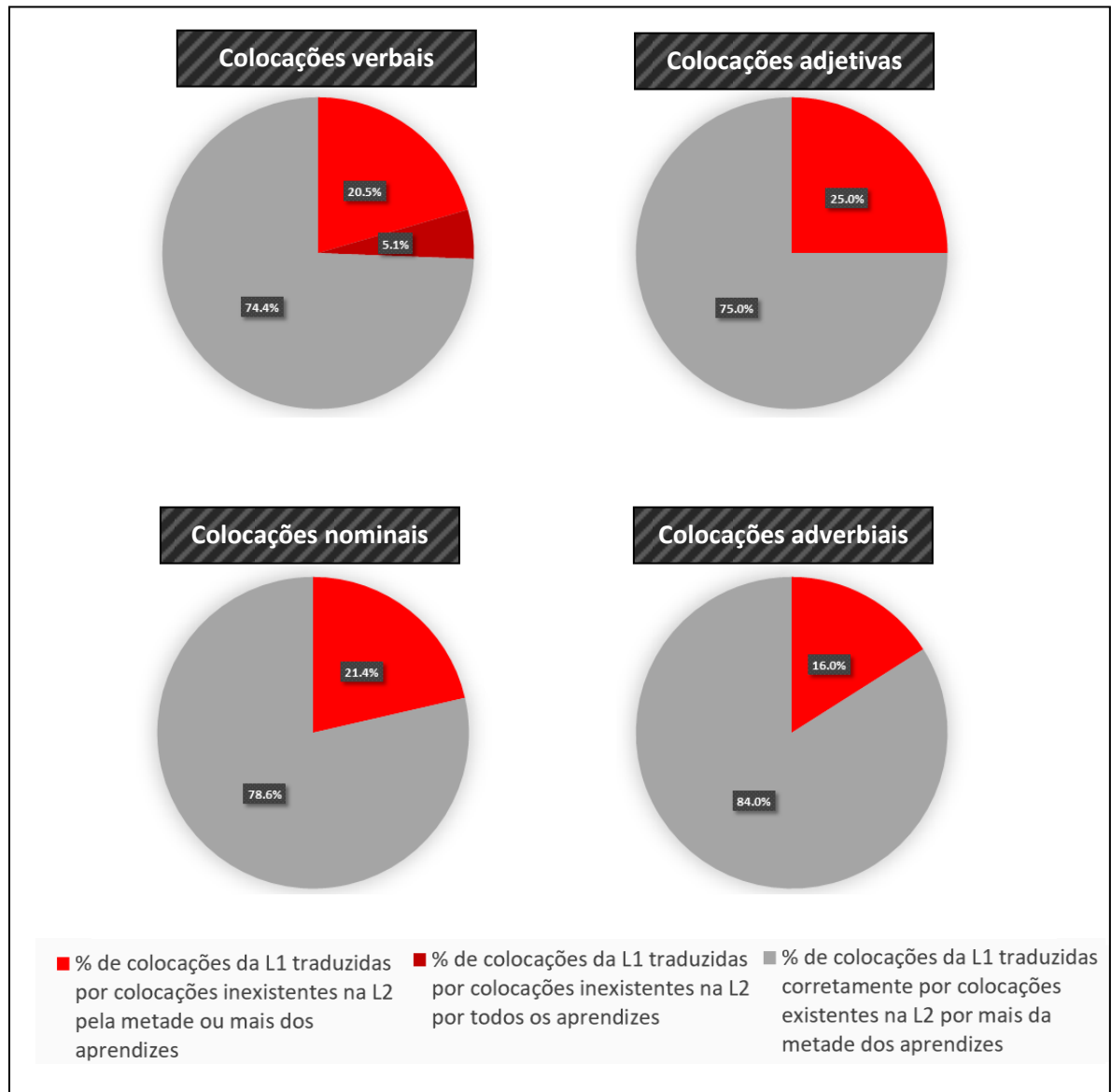
L2 por todos eles, como 超越法治 *chāoyuè fǎzhì*, mencionado anteriormente, dentre outros como 缓解疫情 *huǎnjiě yìqíng* (lit. aliviar a epidemia, reduzir o número de casos).

O segundo tipo de colocação que causou mais problema para os aprendizes chineses foi o das colocações adjetivas. 25% das colocações deste tipo presentes no CAT chinês foram traduzidas por colocações inexistentes na L2 pela metade ou pela maioria dos voluntários, como 有关当局 *yǒuguān dāngjú* (autoridade competente), 严重的暴力 *yánzhòng de bàolì* (lit. violência grave), entre outras. Porém, nenhuma das colocações adjetivas analisadas nesse *subcorpus* foi traduzida somente por erros colocacionais por todos os participantes da pesquisa.

O terceiro tipo mais problemático para os chineses foi o nominal, do qual 21,4% das colocações deste tipo foram traduzidas por colocações inexistentes na L2 pela metade ou pela maioria dos participantes da pesquisa, como ocorreu com 法律资源 *fǎlǜ zīyuán* (recurso jurídico) e 法律后果 *fǎlǜ hòuguǒ* (consequências legais). Nenhuma das colocações deste tipo teve apenas erros colocacionais como solução tradutória.

Por fim, as colocações adverbiais foram o tipo menos problemático para os chineses: 16% do total das colocações deste tipo tiveram metade ou mais das soluções tradutórias apresentadas pelos aprendizes como sendo colocações inexistentes na L2. Foi o caso de 发冷得严重 *fā lěng dé yánzhòng* (ter fortes calafrios).

Gráfico 16 - Total de todas as colocações traduzidas por colocações inexistentes na L2 por tipo de colocação no CAT chinês - L1(chinês)→L2(português)



Fonte: elaborado pelo autor.

Trazemos agora alguns exemplos destes erros e suas respectivas análises nas subseções 4.2.1 e 4.2.2. que seguem, iniciando-se pelo CAT brasileiro.

4.2.1 Análise dos erros colocacionais no *subcorpus* brasileiro

Dentro do *subcorpus* brasileiro, o tipo de colocação com mais erros por parte dos participantes da pesquisa foi o das colocações verbais. Diversas bases, como “declaração”, “educação”, “indício”, “aval”, “carta branca”, entre outras, coocorrem em nosso objeto de pesquisa com o colocado “dar”, e estas colocações foram algumas das mais problemáticas para os aprendizes. Tomemos “dar declaração” como exemplo.

Figura 56 - Ficha com as soluções tradutórias para "dar declaração"

LogDice	Frequência	Dar declaração
5.4	30752	8 give statements
4.3	29682	3 make declarations
3.8	20335	1 make assertions
0	155	6 give declarations
0	5	1 speak ally
Total 19		
c-Col	63% Sign. Dif.	100%
c-nCol	37%	
nc Certo	0%	
nc Errado	0%	
of		
OpM	speak, talk (<i>president + verb</i>) make statements	Mod Trp

Fonte: elaborado pelo autor

Das 19 traduções para esta colocação, 14 (73,6%) delas são com *give*, tradução literal em inglês para “dar”, e quatro (24%) são com *make*, literalmente “fazer”.

De início, percebemos que a tradução literal *give declaration*, que corresponde a 37% do total das opções dadas, possui uma frequência de apenas 155 ocorrências no *corpus* de referência, conforme os resultados da busca pela combinação com o *Concordance*. A busca pelas colocações tanto de *give* como de *declaration* com o *Word Sketch* não apresentou um ou o outro como resultado, pois, estatisticamente, a combinação não é significativa, de modo que não foi possível sequer obter o *logDice* da combinação com essa ferramenta. Portanto, concluímos que *give declaration* não se caracteriza como uma colocação em língua inglesa. Abaixo observemos como os aprendizes voluntários empregaram esse erro colocacional em suas traduções.

Figura 57 - Linhas de concordância de *give declaration* no subcorpus de TTs

3	☐	① doc#0	terday, Trump took part of a dinner with Netanyahu, in which he gave general declarations about his plan of peace. </s><s> According to him, the project includes "a lot of	
4	☐	① doc#0	terday, Trump participated of a dinner with Netanyahu, in which gave generic declarations about your peace plan. </s><s> According to him, the project includes "a lot of	
5	☐	① doc#0	<s> Yesterday, Trump had a dinner with Netanyahu, in which he gave generic declarations about his Peace plan. </s><s> According to him, the Project includes a "a lot of	
6	☐	① doc#0	'day, Trump participated in a dinner with Netanyahu, in which he gave generic declarations about your peace plan. </s><s> According to him, the project includes "much	
7	☐	① doc#0	'day, Trump participated in a dinner with Netanyahu, in which he gave generic declarations about his peace plan. </s><s> According to him, the project includes "lots of k	
8	☐	① doc#0	day, Trump participated of a dinner with Netanyahu, in which he gave generic declarations regarding his peace plan. </s><s> According to him, the project includes "a lo	

Fonte: captura de tela do SkE gerada pelo autor.

Apesar de *give declaration* não configurar uma colocação, 63% das outras opções de tradução preencheram os requisitos para serem consideradas colocações – as opções *give statements* (*logDice* 5.4), *make declarations* (*logDice* 4.3) e *make assertions* (*logDice* 3.8). Porém, elas não correspondem à semântica do TO. Vejamos o contexto em que a colocação aparece no TO.

Quadro 2 - Linha de concordância de "dar declaração" no subcorpus de TO

Ontem Trump participou de um jantar com Netanyahu no qual **deu declarações** genéricas a respeito de seu plano de paz.

Fonte: elaborado pelo autor.

Como podemos depreender, “dar declarações” acima refere-se a falar sobre o assunto, manifestar-se verbalmente na ocasião descrita. Porém, em inglês, *declaration* é uma produção escrita, ao passo que *statement* pode ser tanto uma forma de expressão escrita como oral. Porém, é possível, no *Word Sketch*, percebermos que os verbos que coocorrem mais frequentemente tanto com *declaration* como com *statement* no *corpus* de referência *enTenTen20* trazem consigo a acepção de escrita.

Tabela 8 - Lista de verbos que coocorrem com *declaration* como objeto

VERBO	FREQUÊNCIA	LOGDICE
sign	12,187	7.4
issue	7,241	6.6
seek	6,201	5.9
file	3,565	5.8
submit	3,486	5.6
draft	702	5.4
adopt	2,441	5.3
die	1,764	5.3
endorse	562	5.2
revoke	318	5.0
amend	551	4.7
rescind	156	4.4

Fonte: SkE.

Tabela 9- Lista de verbos que coocorrem com *statement* como objeto

VERBO	FREQUÊNCIA	LOGDICE
issue	137,486	10.0
release	82,685	8.8
make	346,418	7.8
sign	27,834	7.8
follow	83,897	7.4
mislead	10,912	7.4
prepare	20,134	7.3
file	18,274	7.3
read	34,032	7.2
submit	16,412	7.0
publish	20,475	6.8
write	32,393	6.7

Fonte: SkE.

Quando o substantivo enfocado é *statement*, notamos que, além de sua coocorrência com *give*, opção dada pelos aprendizes, conforme mencionamos anteriormente, um dos verbos que não necessariamente se refere a *statement* como um gênero escrito e que coocorre mais frequentemente com esta palavra é *make* (*logDice* 7.8). Quando observamos algumas linhas de

concordância, percebemos que *give statement* e *make statement* podem se referir a uma expressão verbal, sendo, portanto, uma melhor solução tradutória, porém, o contexto de *give statement* difere um pouco da situação descrita na TO, pois talvez não transmita a informalidade e irreverência características do ex-presidente norte-americano, Donald Trump, que se fez presente no jantar descrito no TO.

Figura 58 - Linhas de concordância de *give statement*

1			tertulian.org	ind 'Credo quia absurdum' in some depth. </s><s> He tries to give a fair statement of the positions adopted; that Tertullian is an anti-rationalist; a rationalist;	
2			tatysite.net	in discussion, it would be preferable to provide reasoning for any given statement that is on topic, in order to avoid misunderstandings. </s><s> Cleary-stal	
3			ny.gov	/ritten comments may be mailed and will be considered equally with oral statements given at the legislative hearing, provided they are received by August 1,	
4			mcnel.org	d level there was no detailed information, but three or four more specific statements were given , under which benchmark-level information was provided. </s><s>	
5			lawreform.ie	of which he or she is to act as interpreter at the ceremony, and give the statement to the registered solemniser, and </s><s> (ii) immediately after the ceren	
6			nsw.gov.au	fire safety measures assessed annually and have an Annual Fire Safety Statement given to the council, to the Fire Commissioner and prominently displaye	
7			infotel.ca	iver his sentence. </s><s> VERNON - A Supreme Court judge has ruled statements given by Curtis Sagmoen to police in the first 24 hours of his custody ca	
8			un.org	ch warns however that press freedom there remains "dire". </s><s> In a statement given to the global news agency, UN Secretary-General António Guterre	
9			softpanorama.or...	kin's lymphoma in 2009, before seeing it go into remission. </s><s> In a statement given to ABC News , Gates said he was "heartbroken by the passing of	
10			jamaica-gleaner...	urprise this morning following the revelation that a crown witness gave a statement to the police in relation to the murder case. </s><s> The witness... </s><s>	
11			jfk-online.com	President was when the second shot was fired. </s><s> They gave their statements to the Warren Commission [sic], and I suggested to the Committee inves	
12			itsaboutliberty...	hat rule of the shorter term. </s><s> How is it that an eyewitness gave a statement to the police that Zimmerman was getting pummeled on his back from ar	
13			insanejournal.c...	outside." </s><s> "All right, but I'm only staying long enough to give my statement ." </s><s> "You still going to bingo?" </s><s> Lula asks, trying to clean e	
14			voiceforthe defe...	is, or that you are involuntarily compelled to give an incomplete opening statement . </s><s> Following the testimony of each witness, assert that you were i	
15			ldolphin.org	attern of his life. </s><s> The writer goes on in verse 17 to give his last statement to the body: </s><s> "Obey your leaders, and submit to them; for they ke	
16			indiatimes.com	uncle. </s><s> However, Chandra Layout police said, "According to the statement given by the girl, only Naveen had abused her. </s><s> So, we arrested	
17			utopiastories.c...	<s> she curtly replied. </s><s> "We would appreciate it if you can give a statement down at the station. </s><s> It would be much easier than a full investige	
18			utopiastories.c...	stigation" Officer Davis sternly asked. </s><s> "I have nothing to give a statement to. </s><s> I have no idea what you're talking about." </s><s> Sam blurb	
19			reformation21.o...	I a temptation to be resisted. </s><s> Then Prince gave his most choice statement , a reply that points out the nature of church membership for any sinner.	
20			eipr.org	e while in custody at the Marg police station, to compel him to give false statements to Public Prosecution. </s><s> In addition, Marg police investigations brc	

Fonte: SkE.

Já *make statement*, que não figura entre as opções apresentadas pelos voluntários participantes da pesquisa, aparece tanto em contextos mais formais, com referência a órgãos governamentais, autoridades policiais, entre outros, como aqueles em que *give statements* geralmente ocorre, quanto em contextos mais informais, sendo, potencialmente, uma opção tradutória mais plausível para “dar declarações” no caso descrito.

Figura 59 - Linhas de concordância de *make statement*

1			philipshaw.ca	leader John Tory as "Ontario's next Premier." </s><s> Harper made this statement at a fundraiser immediately after meeting with Premier McGuinty over the	
2			countyofcolusa....	erson has knowingly, and with intent to deceive or defraud, made a false statement or representation to obtain benefits, obtain a continuance or increase of t	
3			countyofcolusa....	obtaining, continuing, or avoiding a reduction or denial of benefits, made statements which he/she did not know to be true with reckless disregard for the truth	
4			drobe.co.uk	>><s> To Michael Stubbs: Please get up to date before making very silly statements . </s><s> Check on the A75 and the potential for immediate sale of thous	
5			wrestedscriptur...	cting as His Father instructed. </s><s> We can see this by the additional statements made by Jesus in the very same chapter as the "I am" statements. </s><	
6			tvguide.co.uk	it authorized by the copyright owner, its agent, or the law; and </s><s> A statement made under penalty of perjury that the information provided is accurate s	
7			wikipedia.org	rm auctions due to security reasons. </s><s> In response, Airtel made a statement stating that it may look for alternatives for 5G as it no longer wishes to co	
8			bluefish.org	> And most of them probably know it. </s><s> But that doesn't make the statement scientifically inaccurate. </s><s> "The point is it's about as clear as it's gc	
9			jeff-goodall.co...	arrested Friday on one count each of health care fraud and making false statements relating to health care matters." </s><s> – Scott Daugherty, "Chesapeake	
10			isis-online.org	re treaty, but did not ratify it. </s><s> Of course, Argentina made several statements to the effect that we would never do anything against the purposes of the	
11			cakein15.com	compellingly drawing voice. </s><s> Her ability to make the most banal statements ("Last week I turned 24/you don't call me anymore") seem poetic and of s	
12			teachinghistory...	e each student should "read" the map in his or her own way and make a statement about the map's meaning or purpose. </s><s> Extend the lesson by editi	
13			kottke.org	, you'll like Luis," Ethan said in a creaky voice. </s><s> "He makes a big statement ." </s><s> "Turturro is coming in to play the pederast," Joel said. </s><s>	
14			thornhillcapita...	sociated with Latin American politics. </s><s> This is most likely due to a statement made by New York Times reporter Larry Rohter who characterized the el	
15			selkirk.ca	dress your current attributes and skills. </s><s> Make a strong personal statement : </s><s> Always enclose a letter to simply tell the scholarship committee	
16			meltonpriorinst...	irecting the opening of certain of Mazzini's letters in consequence of the statements made to our Government by that of Naples, to the effect that plots were t	
17			thelodownny.com	our membership not only helps defer those costs, but makes a powerful statement in support of home-grown community news, produced by and for your ne	
18			acoel.org	g releases to groundwater under the NPDES program rely principally on statements made in the preamble to a 2001 proposed rule for Concentrated Animal f	
19			ringsidereport....	e year. </s><s> Claressa "T-Rex" Shields, 5-0, 2 KO's, will try to Make a statement by defeating Hanna Gabriels, 18-1-1, 11 KO's. </s><s> Vacant WBA and	
20			ringsidereport....	of Germany, Christina Hammer, 22-0, 10 KO's, will herself try to make a statement by defeating what will become the common opponent between her and S	

Fonte: SkE.

Outra base que pode coocorrer com o colocado “dar” e gerou grande número de soluções tradutórias equivocadas foi “educação”, em “dar educação”. Observemos, abaixo, em qual contexto esta combinação ocorre no nosso *subcorpus* de TOs em português:

Quadro 3 - Linha de concordância de "dar educação" no *subcorpus* de TO

...o artista nu com consentimento da mãe. Também não acho a atitude adequada. Mas podemos pensar que não sabemos que **educação esta mãe dá** a sua criança quais códigos ela passa. Se pensarmos ser uma família adepta do naturismo percebe que a coisa...

Fonte: elaborado pelo autor.

Agora, vejamos as opções apresentadas pelos alunos:

Figura 60 - Ficha com as soluções tradutórias para "dar educação"

LogDice	Frequência	Dar educação
0	1491	3 give education
nc	nc	1 educate
nc	nc	1 teaches
7.2	120401	1 provide education
0	3662	1 have education
Total		7
c-Col		14% Sign. Dif.
c-nCol		57%
nc Certo		14%
nc Errado		14%
of		0%
OpM		

Fonte: elaborado pelo autor.

Das soluções encontradas pelos aprendizes, 57% foram erros colocacionais. Um grande número de estudantes optou pela tradução literal *give education*, cuja frequência no *enTenTen20*, segundo os resultados da busca por *give+education* com o *Concordance* é de 1.491. A busca pelos colocados de *give*, de *education* ou ainda de *give+ education* com o *Word Sketch* não apresentou resultados que comprovassem a existência dessa colocação. O mesmo ocorreu ao usarmos o *Word Sketch* para buscarmos por *have, education* ou *have+ education*. Pudemos atestar a existência dessa combinação no *corpus* de referência com o *Concordance*, que apontou para uma frequência de 3662.

Porém, uma análise mais apurada das linhas de concordância de *have education* também revela que, apesar de a combinação aparecer no *corpus*, seu uso e contexto de aplicação diferem um pouco da opção dada pelos aprendizes. Vejamos a solução apresentada pelos voluntários:

Quadro 4 - Linha de concordância de *have education* no *subcorpus* de TTs

*I also do not think the attitude was appropriate. But we cannot ignore that we do not know the **education this child is having**, the codes they is getting from his mother.*

Fonte: elaborado pelo autor.

Agora vejamos as linhas de concordância de *have education* no *corpus* de referência.

Figura 61 - Linhas de concordância de *have education* no corpus de referência

1			tradingacademy....	gement strategies, or learning how to generate short-term income, we	have education	solutions that can help them. </s><s> Visit the Education section of ou
2			com.	re both legitimate professions, however in order to be hired you must	have education	. </s><s> The courses offered online may or may not be legitimate. </
3			gifffromwithin....	xiety and overcoming traumatic memories. </s><s> A good therapist	has education	in science and the humanities, and should let you know where he or s
4			michaelzwilliam...	expected to be treated like an adult. </s><s> This means you should	have education	, presentation, a description of a problem, and a proposed solution. </
5			freecriticalthi...	ects dictate behaviour. </s><s> Foundation of the problem – we don't	have education	– we have training. </s><s> We don't draw out. </s><s> System does
6			studymode.com	><s> Because uneducated people have no skill in their hand nor they	have education	to earn money. </s><s> So it seems easy to them to commit a crime t
7			iowa.gov	vices. </s><s> Staff participating in the care of pediatric patients shall	have education	appropriate for the care of pediatric patients. </s><s> (5) Ancillary ser
8			salsburg.com	s> Job Knowledge and Skills Required The Program Manager should	have education	and/or experience in one or more of the following areas: architecture,
9			biblebb.com	ight be clothed. </s><s> Alas! she is content without her Lord, for she	has education	, oratory, science, and a thousand other baubles. </s><s> Zion's soler
10			legislation.gov...	in Australia and overseas. </s><s> 1.1 The higher education provider	has education	as a principal purpose, with governance and management of its Austr
11			mattrapf.com	oliferation can occur. </s><s> A clinical pharmaceutical scientist must	have education	in both clinical pharmacy and research. </s><s> According to their fiel
12			state.nc.us	ient; </s><s> (4) prior to or within the first three years of employment,	have education	in teaching and learning principles for adult education, including curric
13			articleweb55.co...	cellent idea. </s><s> To start with, a well-qualified weight loss trainer	has education	in fundamental nutrition and may help you choose a diet strategy that
14			indiana.edu	well as in-services professional development. </s><s> "Never before	has education	so clearly defined the pathway to success," said Hardy Murphy, study
15			dentistrytoday....	each country. </s><s> Some countries have no requirements, some	have education	requirements, and some have language requirements. </s><s> Typica
16			qisofware.com	i of CIA families- question my use of the word shrew. </s><s> Neither	have education	above high school– yet get to waste my time questioning my use of er
17			usda.gov	actice veterinarians prior to joining USDA Animal Care. </s><s> ACIs	have education	in the biological sciences and/or extensive experience in the care and
18			newcommunities....	opportunities for families. </s><s> For example, since so few parents	have education	beyond high school, Christopher House is now investigating partnersh
19			caring4you.net	PLEASE CONSULT YOUR Physician. </s><s> My kind volunteers do	have education	and medical backgrounds but are in no way engaging in any kind of a
20			creativityjour...	ery fond of communication, and she was determined that her children	had education	, and a lot of possibility. </s><s> My father always pushed us to any o

Fonte: SkE.

Com base nas linhas de concordância acima, notamos que *education*, quando acompanhado de *have*, refere-se principalmente à escolarização, formação acadêmica, e não a valores e princípios passados de mãe para filho, como é o caso do TO.

Em realidade, a ideia de “educação” como “criação dos filhos” pode ser traduzida em inglês como *upbringing*, este, sim, coocorrendo mais frequentemente com *have* (5037 vezes).

Figura 62- Linhas de concordância de *have upbringing* no corpus de referência

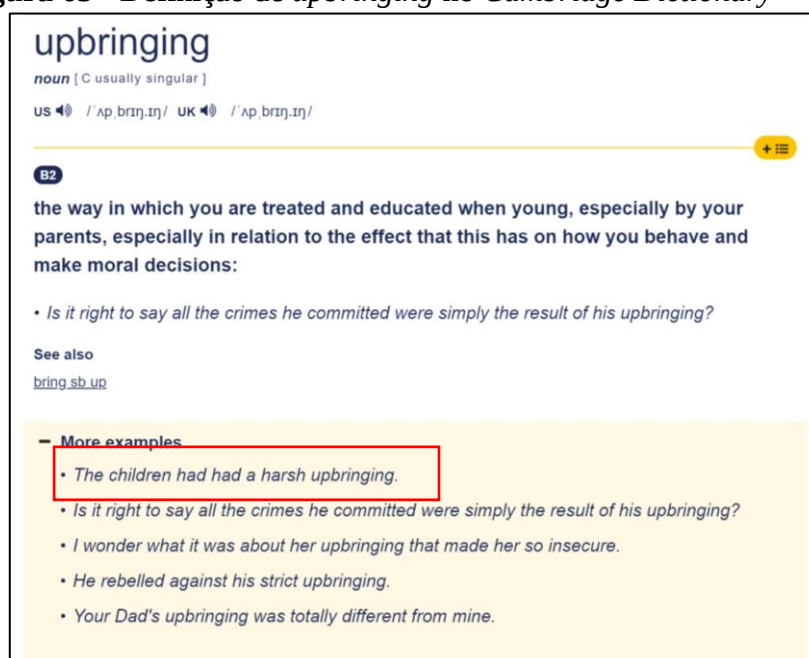
1			rys2sense.com	: </s><s> Bad Religion is what makes us so shitty. </s><s> We have a	upbringing	of invalidation. </s><s> See the first Rant You are not what you own. </s>
2			pwlpartnership....	. LEED Accredited Senior Landscape Architect </s><s> "I had an urban	upbringing	, with limited exposure to nature," says Liz, who was drawn to landscap
3			quantumleap-als...	gnized his genius it seems they made sure he had a somewhat regular	upbringing	with school and sports and chores, etc. A kid with Sam's genius could h
4			cinfilms.org	le Kidman) and stalled author Baxter (Jason Bateman) had an unusual	upbringing	. </s><s> The children of performance artists, they routinely participater
5			rgtperrybarr.co...	es :) They are very different to 'real' dogs LOL and have had a different	upbringing	and of course have been working dogs so have had different routines :)
6			nbcports.com	said 'I found the car' and closed the door." </s><s> Brown had a rough	upbringing	, often sleeping on the couches of teammates and coaches after his ste
7			juliansimon.org	: I discovered that my highest value is for my children to have a decent	upbringing	. </s><s> A depressed father makes a terrible model for children. </s><
8			johnreedbooks.c...	ge David Smith, served overseas during the war - she had a privileged	upbringing	. </s><s> She studied classics at Oxford University, where she threw he
9			bullseye.com	it of a movie or watching a show... </s><s> I mean I had a pretty rough	upbringing	. </s><s> Actually at this point I don't even speak to my parents. </s><s>
10			jewishvaluesonl...	ically. </s><s> Question: My father is Jewish, my mother had a Catholic	upbringing	but doesn't identify with a religion. </s><s> I was raised Jewish and hac
11			history.ac.uk	moving away from the idea that the majority of children had an abusive	upbringing	. </s><s> It is appropriate, then, that with a new millennium, Nicholas O
12			gradesaver.com	r bad luck after her mother... </s><s> Adeline Yen Mah had a traumatic	upbringing	. </s><s> As a child, she was raised by her cruel stepmother, abused ai
13			1st-attractive....	fer my personality to be sensitive and very kind. </s><s> I have a good	upbringing	and I believe my life views are right. </s><s> I always have limits accor
14			news.com.au	pretty sure anyone going to this movie knows Judy Garland had a hard	upbringing	. </s><s> But you'll want to see this movie because Zellweger is just an
15			pacbi.org	driven me," says Suzman. </s><s> "When I said I didn't have a Jewish	upbringing	and that I went to a convent which didn't influence me either, they said i

Fonte: SkE.

A partir destas linhas de concordância, podemos depreender que uma possível solução tradutória para este caso seria “...we do not know the **upbringing** this child has...”.

Apesar do baixo *logDice* (0.6), *have+upbringing* aparece tanto nos resultados do *Concordance*, como podemos observar na figura 61, como nos resultados das buscas pelos colocados de *upbringing* com o *Word Sketch*. Ademais, a combinação *have + upbringing* aparece dicionarizada no *Cambridge Dictionary*.

Figura 63 - Definição de *upbringing* no *Cambridge Dictionary*



Fonte: *Cambridge Dictionary*.

Ao buscarmos pelos verbos que mais coocorrem com *mother*, uma vez que no TO questiona-se a educação de uma mãe dada ao seu filho, chegamos a outras soluções, na tentativa de se transmitir a mesma ideia do TO.

Tabela 10 - Verbos que coocorrem com *mother* como sujeito

VERBO	FREQUÊNCIA	LOGDICE
die	53,583	8.8
tell	26,759	7.1
raise	11,402	7.0
teach	9,548	6.9
pass	9,503	6.4

Fonte: SkE.

Uma solução tradutória que realize um deslocamento na estrutura semântica de superfície do segmento textual e use expressões ou palavras que, a partir de um ponto de vista distinto, expressam uma mesma ideia, talvez seja possível com o uso de *raise*, *teach* ou *pass*, por exemplo.

Soluções como *we do not know how **this child has been raised** by his mother* (*mother+raise*, *logDice* 7.0); *we do not know **the principles this child is being taught*** (*teach+principle*, *logDice* de 6.4), *we do not know what **this child has been taught** by his mother* (*teach+child*, *logDice* de 9.0); *we do not know **the values this mother passes to her child*** ou *we do not know **what values this mother has passed to her child*** (*pass+value*, *logDice* 5.7) entre outras, poderiam ter sido uma saída.

De todo modo, várias poderiam ser as possibilidades de tradução que soariam mais naturais, porém, neste caso, a tradução literal não é uma delas, como pudemos depreender com a consulta ao *corpus* de referência.

O segundo tipo de colocações com mais erros no CAT brasileiro é o adverbial. Neste *subcorpus*, uma colocação reconhecidamente problemática foi “engolir a seco”, empregada com sentido figurado. Vejamos a sua ocorrência no TO:

Quadro 5 - Linha de concordância de “engolir a seco” no *subcorpus* de TOs

Muitas levarão cantadas machistas e **engolirão a seco** o vazio do respeito que não existe no dia a dia de muitas dessas profissionais.

Fonte: elaborado pelo autor.

Na situação em questão, a colocação analisada, também registrada na versão *on-line* do dicionário Michaelis como “engolir em seco”, significa “não responder a um insulto; ser obrigado a calar o que se queria dizer” (MICHAELIS). Este é um caso em que alguns alunos, ao traduzirem a colocação, usaram outros fraseologismos, sobretudo as *phrasal verbs*, como opção, como observa-se na ficha de “engolir a seco” a seguir.

Figura 64- Ficha com as soluções tradutórias para “engolir a seco”

LogDice	Frequência	engolir a seco
of	of	1 put with
of	of	1 gulp down
nc	nc	2 swallow
nc	nc	1 buy
4.8	2430	2 swallow hard
1.9	29	2 swallow dry
0	0	2 dryly swallow
0	0	1 tolerate the void
0	0	1 dry swallow hard
Total 13		
c-Col	15%	Sign. Dif. 100%
c-nCol	46%	
nc Certo	15%	
nc Errado	8%	
of	15%	
OpM		

Fonte: elaborado pelo autor.

Segundo os dados, 15% das opções de tradução para esta colocação foram *phrasal verbs* não combinadas a nenhum adjetivo ou advérbio, e 15% encontraram como solução tradutória uma colocação existente em inglês. Somados, estes números ainda são inferiores ao número de opções que são combinações lexicais e não são colocações existentes em língua inglesa, entre elas: *dryly swallow* e *swallow dry*, em que o colocado “seco” foi traduzido literalmente por *dry* ou transformado no advérbio *dryly*.

Após consulta ao *corpus* de referência, vemos que um dos adjetivos que mais ocorrem como colocado depois do verbo *swallow* e que figura entre as opções dadas é *hard*, que significa “duro”, “difícil” ou “desagradável”.

Tabela 11 - Adjetivos depois de *swallow*

ADJETIVO	FREQUÊNCIA	LOGDICE
whole	940	9.8
wholesale	76	7.3
reflex	38	7.3
tail	16	6.1
entire	22	5.9
IMOVANE	9	5.6
hard	2,430	4.8
liquid	33	4.8

Fonte: SkE.

Porém, a análise das linhas de concordância e busca pelo significado de “*swallow hard*” em dicionários nos mostra que a combinação é empregada com o significado de literalmente “engolir algo com dificuldade” ou “engolir com esforço”, significado este que “engolir a seco” também possui em português, mas não é o empregado no TO.

Quando analisamos as linhas de concordância das outras opções, vemos que as combinações mais fortes, *swallow whole*, com *logDice* de 9.8, e *swallow wholesale*, *logDice* de 7.3, podem ser candidatas à tradução da colocação de língua portuguesa, com o sentido de “aceitar sem questionar”, “calar-se sem rebater”.

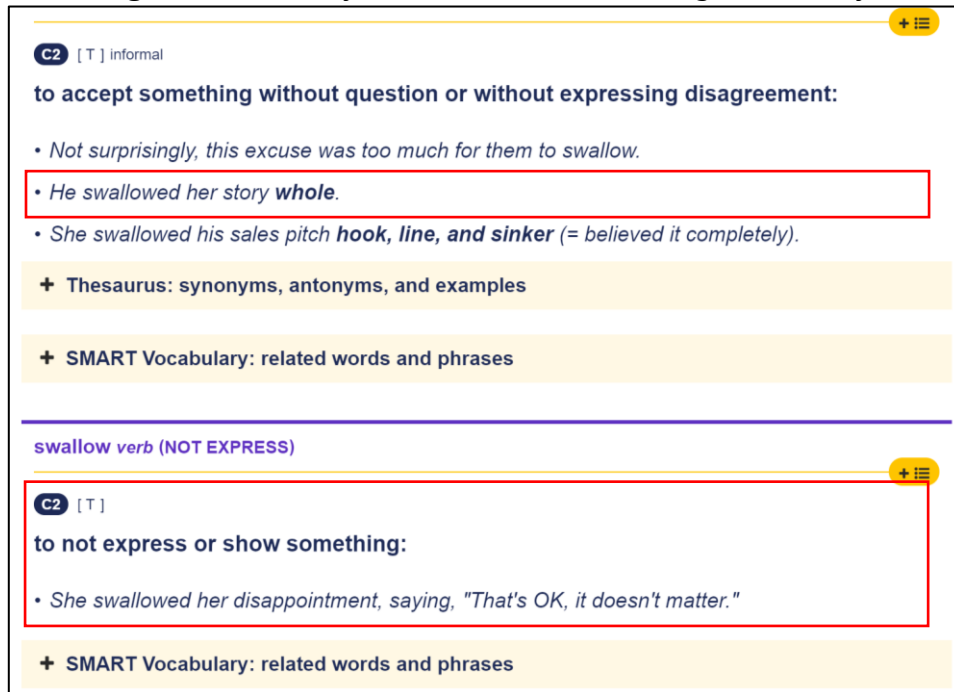
Figura 65 - Linhas de concordância de *swallow wholesale*

1			septicisle.info	: the jury instead of themselves. </s><s> Then there's the media, which	swallowed	wholesale	from the very beginning the whole idea that such an attack ir	
2			lrb.co.uk	ess, the breaking down of barriers, the ready hand of friendship – 'were	swallowed	wholesale	by the people for whom they were fashioned ... with none of	
3			theflatearthsoc...	ements of fact which are giant assumptions which you expect people to	swallow	wholesale ?	</s><s> And considering how this thread started, it would t	
4			epsr.org.uk	roletarian garbage is the shallowest (and wrong) impressionism mostly	swallowed	wholesale	from the capitalist press (and like all the CPGB-ML museum	
5			k12academics.co...	ore fully it is often revealed to be just a personal passion that has been	swallowed	wholesale	from the media. </s><s> Children with Attention Deficit Disor	
6			beginnerbikers....	wo-line soundbite, so over-simplified as to be actively misinforming, yet	swallowed	up wholesale	. </s><s> Twitter is the latest confirmation: not only can y	
7			eklesia.co.uk	><s> "This programme, like too many media stories, failed the public by	swallowing	wholesale	the evidence-free myth of a 'dependency culture' in which un	
8			theflatearthsoc...	that "proof" offered at the altar of science need not be compelling to be	swallowed	wholesale	by people eager to believe it. </s><s> Well done Ski. </s><s>	
9			ultindia.org	r Jesus nor H.P.B. lived and died in order that a book or books might be	swallowed	wholesale	, nor even that men should become disciples, but that all me	
10			ultindia.org	hat books or pamphlets lauding - and misrepresenting - him should be "	swallowed	wholesale	," but that truth should prevail... </s><s> The magazine Thec	
11			onlineopinion.c...	<s> Advertisement </s><s> But it is the Regev version of events that is	swallowed	wholesale	by politicians such as the US President, the British Foreign S	
12			indymedia.ie	is March 1st attempts of civil disobedience at Shannon was bought and	swallowed	wholesale	by the above mentioned groups. </s><s> It's been pretty mu	
13			culturematters....	<s> According to the Child Poverty Action Group, it "failed the public by	swallowing	wholesale	the evidence-free myth of a 'dependency culture' in which un	
14			hijuela.com	ormation is filtered, and historically revisionist propaganda is frequently	swallowed	wholesale	. </s><s> This information is then embellished or accommod	
15			twf.org	ust seek influence to fuel their industrial production. </s><s> Before we	swallow	wholesale	the accepted story of "good and evil" among African leaders,	
16			telus.net	icationation). </s><s> Riley's canards will no doubt be chewed up (if not	swallowed	wholesale	) by countless canines. </s><s> We all know what dogs do w	
17			kubatana.net	s endeavors, must not seek to railroad or arm-twist the participants into	swallowing	wholesale	the propositions of the people leading the consultative proce	
18			bexley-is-bonke...	n physical violence. </s><s> This would be an outrageous lie which you	swallowed	wholesale	. </s><s> Since then you have refused to provide me with ar	
19			phx-ult-lodge.o...	either Jesus nor H. P. B. lived and died that a book or books should be	swallowed	wholesale	, nor even that men should become disciples but that all mer	
20			catholicvoiceoa...	ies involving faith and morals. </s><s> We must resist the temptation to	swallow	wholesale	the latest popular cultural trend, but rather critique the cultur	

Fonte: SkE.

A confirmação de que as combinações com *whole* e *wholesale* podem ser opções tradutórias adequadas pode ser confirmada após consulta ao dicionário *Cambridge Dictionary*, em que a opção *swallow whole* é dada como exemplo no verbete de *swallow*, que por si só dá conta do significado expresso por “engolir a seco” em português e foi usada como solução tradutória em 15% dos casos.

Figura 66 - Definição de *swallow* no *Cambridge Dictionary*



Fonte: *Cambridge Dictionary*.

O terceiro tipo de colocação com mais erros no *corpus* brasileiro foi o adjetivo, e as colocações adjetivas mais passíveis de erro foram colocações com a base “plano”, como “plano inclinado”, que, no *subcorpus* de TOs, aparece no seguinte contexto:

Quadro 6 - Linha de concordância de "plano inclinado" no *subcorpus* de TOs

A maconha talvez seja a mais danosa de todas as drogas porque representa o início do **plano inclinado** na vida dos que a experimentam.

Fonte: elaborado pelo autor.

Assim como a colocação adverbial “engolir a seco”, a referida colocação é empregada figurativamente com a ideia de “decadência”. Dentre as opções apresentadas pelos voluntários, a grande maioria é resultado de estratégias de tradução que mantiveram os correspondentes diretos para cada palavra da combinação, seja a base, “plano”, ou o colocado, “inclinado”, como vemos abaixo.

Figura 67 - Ficha com as soluções tradutórias para “plano inclinado”

<i>LogDice</i>	Frequência	plano inclinado
8.1	90970	1 end of life
7.9	6415	1 inclined plane
0	17	5 inclined plan
0	0	1 bent plan
0	0	1 plan bent
Total 9		
c-Col	22%	Sign. Dif. 0%
c-nCol	78%	
nc Certo	0%	
nc Errado	0%	
of	0%	
OpM		

Fonte: elaborado pelo autor.

Ainda que empregado conotativamente, “plano inclinado” é o que podemos chamar de uma colocação especializada (ORENHA-OTTAIANO, 2009) da física, significando uma superfície inclinada para a horizontal. Em inglês, o correspondente é *inclined plane*, em que *plane*, segundo o *Cambridge Dictionary*, significa “uma superfície plana contínua em todas as direções”.³⁰

Não surpreendentemente, esta foi a única opção dada pelos voluntários com resultados na busca por colocados de *plane* com o *Word Sketch*, com um *logDice* de 7.9.

Tabela 12 - Modificadores de *plane*

MODIFICADOR	FREQUÊNCIA	LOGDICE
fighter	15,329	8.7
horizontal	13,092	8.3
focal	9,950	8.1
astral	7,383	8.0
inclined	6,415	7.9

Fonte: SkE.

³⁰ Original: *a flat surface that continues in all directions*

A opção *inclined plane* foi apresentada como solução tradutória apenas uma vez. Segundo os dados, 78% das soluções tradutórias para essa colocação continham *plan*, que significa “um conjunto de decisões sobre como fazer algo no futuro”³¹ e não coloca com *inclined* ou *bent*.

Apenas uma opção tradutória mais ousada foi bem-sucedida, que resultou em uma colocação existente em língua inglesa. A linha de concordância contendo a opção “*end of life*”, cujo *logDice* é 8.1, encontra-se abaixo:

Quadro 7 - Linha de concordância de *end of life* no *subcorpus* de TTs

*Weed may be the most harmful drug among all because it represents the beginning of the **end of life** of those who try it.*

Fonte: elaborado pelo autor.

Finalmente, as colocações menos problemáticas para os participantes brasileiros da nossa pesquisa são as nominais. No *subcorpus* brasileiro, nota-se que grande número dos erros cometidos na tradução de colocações nominais foi em decorrência de traduções literais com “*of*”, em colocações como “plano de paz”.

Figura 68 - Ficha com as soluções tradutórias para “plano de paz”

<i>LogDice</i>	Frequência	Plano de paz	
5.6	12.146	15	peace plan
0	95	2	plan of peace
0	2555	1	peace program
nc	nc	1	plan involving peace
Total		19	
c-Col		84%	Sign. Dif.
c-nCol		11%	
nc Certo		5%	
nc Errado		0%	
of		0%	
OpM			

Fonte: elaborado pelo autor.

³¹ Original: a set of decisions about how to do something in the future

Como se observa na ficha de “plano de paz” acima, a frequência de *peace plan*, cujo *logDice* é 5.6, no nosso *corpus* de referência, é muito maior (12.146) do que *plan of peace* (95), para o qual a busca por *plan* e *peace* com o *Word Sketch* não apresentou resultados.

Outra solução encontrada pelos participantes da pesquisa que também faz um deslocamento na estrutura semântica de superfície da colocação e usa palavras que expressam o mesmo significado é uma colocação verbal com *plan*, *plan involves*, cujo *logDice* é de 6.5.

Tabela 13 - Frequência e score de *plan* + *involve*

VERBO	FREQUÊNCIA	LOGDICE
involve	11,589	6.5
plan involves		

Fonte: SkE.

No entanto, não encontramos no *corpus* de referência a ocorrência de *plan involves* + *peace* com o *Word Sketch* ou *Concordance*.

Outro motivo que causou problema nos tipos de colocações analisadas anteriormente e se repetiu neste tipo foi a tradução literal do colocado da L1 em situações em que o correspondente direto na L2 não coloca com a base, como em “disseminação das drogas”.

Figura 69 - Ficha com as soluções tradutórias para “disseminação das drogas”

LogDice	Frequência	Disseminação das drogas
0	0	4 drug dissemination
0	0	3 dissemination of drugs
0	0	1 spread of drugs
Total 8		
c-Col 0 Sign. Dif.		
c-nCol 100%		
nc Certo 0%		
nc Errado 0%		
of 0%		
OpM	drug's spread	4.6 21

Fonte: elaborado pelo autor.

A maioria das soluções tradutórias para a referida colocação tem como colocado *dissemination*, que, de acordo com os resultados do *Word Sketch*, se aplica melhor à “disseminação de ideias”, “disseminação de informações”.

Tabela 14 - “dissemination of”

SUBSTANTIVO	FREQUÊNCIA	LOGDICE
finding	2,187	8.4
information	19,242	8.3
knowledge	6,085	7.8
result	3,349	7.6
news	1,019	6.9
propaganda	400	6.6
Information	509	6.5
idea	1,782	6.4
research	2,391	6.0

Fonte: SkE.

No caso de “disseminação das drogas”, o mais apropriado, segundo os resultados da consulta ao *corpus* de referência, seria *drug’s spread*, cujo *logDice* é de 4.6. O item lexical *spread* até figurou entre as opções dadas pelos voluntários, como *spread of drugs*, que não figurou nos resultados das buscas com *spread* e *drugs* no *Word Sketch*, mas apresentou uma frequência de 577 no *Concordance*.

Do lado chinês, as colocações nominais foram o terceiro tipo com mais erros. Alguns desses erros e outros estão na subseção a seguir, na qual analisamos alguns dos erros colocacionais presentes no *corpus* chinês.

4.2.2 Análise dos erros colocacionais no *corpus* chinês

No *subcorpus* chinês, assim como no *subcorpus* brasileiro, o tipo de colocação com mais erros por parte dos voluntários é o das colocações verbais. Há um grande número de colocações com o colocado 受到 *shòudào*, que pode ser traduzido por “receber”, “sofrer” e “ser vítima de”, e coocorre com várias bases, como em 受到对待 *shòudào duìdài* (ser tratado), 受到抨击 *shòudào pīngjī*

pēngjǐ (receber críticas, ser atacado verbalmente), 受到骚扰 *shòudào sāorǎo* (sofrer assédio, ser assediado), 受到仇外 *shòudào chóuwài* (ser vítima de xenofobia), entre outros.

Em média, 16,7% das soluções encontradas para as colocações com 受到 estavam incorretas. Porém, há uma colocação que acumulou o maior número de erros – 受到不人道待遇 *shòudào bùréndào dàiyù*, em que 不人道待遇 *bùréndào dàiyù* tem como significado “tratamento desumano” ou “desumanidade”. Observemos as soluções encontradas pelos alunos para esta colocação a seguir.

Figura 70 - Ficha com as soluções tradutórias para 受到不人道待遇

LogDice	Frequência	受到不人道待遇
6.1	837	1 tratamento desumano
6.6	24	1 submeter tratamento desumano
0	0	3 sofrer desumanidade
1.17	3	9 sofrer tratamento desumano
0	131	1 tratamento injusto
Total	15	
c-Col	7% Sign. Dif.	0%
c-nCol	87%	
nc Certo	7%	
nc Errado	0%	
of	0%	
OpM		

Fonte: elaborado pelo autor.

Das soluções formadas por verbo + substantivo, apenas uma estava parcialmente correta, de acordo com os resultados da consulta ao *Word Sketch*: “submeter a tratamento desumano”, em construção com sujeito da passiva impessoal, “submetido a tratamento desumano”, cujo *logDice* é 6.6. Na opção apresentada pelos voluntários, há um erro de regência, pois falta a preposição “a”. No entanto, a partir da linha de concordância desta opção tradutória abaixo, vemos que foi feito o uso da construção com sujeito da passiva impessoal.

Quadro 8 - Linha de concordância de “submetido a tratamento desumano” no *subcorpus* de TTs

...pedindo "ações corretivas imediatas" contra "o **tratamento xenófobo, racista e desumano que os africanos são submetidos** na China". Nas últimas duas décadas, a China tornou-se um parceiro económico importante da África. O investimento...

Fonte: elaborado pelo autor.

Tabela 15 - Verbos que coocorrem com "tratamento desumano" em construção com sujeito da passiva impessoal

VERBO	FREQUÊNCIA	LOGDICE
submeter	24	6.6
reservar	3	5.8
permitir	2	1.3

Fonte: SkE.

Outros exemplos da construção em questão extraídas do *corpus* de referência podem ser observadas com as linhas de concordância geradas pelo *Concordance*.

Figura 71 - Linhas de concordância de “submeter” + “tratamento desumano”

1		①	Brazil • direit...	postado no artigo 5o da Constituição Federal, ninguém será submetido a tratamento desumano . </s><s> E, obviamente, exigir que o interessado continue s	
2		①	Brazil • cimi.o...	es de proteção ao trabalho e zelar para que ninguém seja submetido a tratamento desumano , degradante ou humilhante. </s><s> Todos os empregados	
3		①	Brazil • fetrac...	onra e da imagem e aquele segundo o qual ninguém será submetido a tratamento desumano ou degradante, todos previstos na Constituição Federal. </s>	
4		①	Brazil • jurid...	tuar que, o texto constitucional, afirma que ninguém será submetido a tratamento desumano , ou seja, não importa se o indivíduo está preso ou não, o di	
5		①	Brazil • inclus...	ou deixar de fazer algo, senão em virtude da lei, nem será submetido a tratamento desumano ou degradante. </s><s> Ainda nossa lei máxima determina c	
6		①	Brazil • fetrac...	danos morais alegando ter sido duplamente atingido. "Foi submetido a tratamento desumano ou degradante no momento em que esteve na mira dos ladr	
7		①	Brazil • abgl...	si o direito à dignidade, à liberdade de expressão, sem ser submetido a tratamento desumano ou degradante. </s><s> Art. 5º - Poderão fazer uso do direit	
8		①	Brazil • direit...	ade de reparação por dano moral a detento que teria sido submetido a tratamento desumano e degradante por conta de superlotação carcerária. </s><s>	
9		①	Brazil • rs.gov...	art. 5o, inciso III, da nossa Constituição que ninguém será submetido a tratamento desumano ou degradante, esteja ele preso, privado de sua liberdade ou	
10		①	Brazil • direit...	uição Federal em seus incisos III, pelo qual ninguém será submetido a tratamento desumano ou degradante; X, que protege o direito à intimidade, à imag	
11		①	Brazil • bioeti...	peito à dignidade humana e o que reza que ninguém será submetido a tratamento desumano ou degradante. </s><s> Além disso, recorrem à resolução d	
12		①	Brazil • ambito...	o: o direito à vida, o direito a não ser torturado e a não ser submetido a tratamento desumano ou degradante, na qual o ser humano submete a outro, o dir	
13		①	Brazil • cremes...	nais de respeito à dignidade humana e que ninguém será submetido a tratamento desumano ou degradante recorrem à Resolução do Conselho Federal c	
14		①	Brazil • cnbsp...	> Em seguida, no art. 5o, III, preceitua que "ninguém será submetido a tratamento desumano ou degradante". </s><s> Segundo o Professor Oscar Vilhen	
15		①	Brazil • carcer...	normas constitucionais que determinam que ninguém será submetido a tratamento desumano , o que afronta a Lei de Execuções Penais que determina ac	
16		①	Brazil • netsab...	da dignidade da pessoa humana e o de que ninguém será submetido a tratamento desumano ou degradante. </s><s> A Carta Magna também adota, entr	
17		①	Brazil • lfg.co...	3 no inciso III do artigo 5o ao assegurar que ninguém será submetido a tratamento desumano ou degradante; </s><s> c) CR/88 no inciso III do artigo 3o q	
18		①	Brazil • espaco...	ção da sua função na colheita de maçãs, o trabalhador foi submetido a tratamento desumano , sem acesso a água para beber e sendo alvo de discriminaç	
19		①	Brazil • jurua...	s nos dogmas constitucionais de que ninguém deverá ser submetido a tratamento desumano e em atenção ao princípio da dignidade humana. </s><s> Te	
20		①	Brazil • seculo...	os fazem visitas e inauguram CDP's presos estão sendo submetidos a tratamentos desumanos , com pés e mãos algemadas, tud que estou falando é por	
21		①	Brazil • seculo...	r dentro. </s><s> Segundo Homero Mafra, o advogado foi submetido a tratamento desumano e degradante, em clara afronta à Constituição Federal, que i	

Fonte: SkE.

Porém, esta não foi a única colocação que causou problema para os chineses. Passemos para outro erro cometido em uma colocação com 法治 *fǎzhì*, uma base frequente entre as colocações verbais e nominais analisadas que, em chinês, significa “Estado de Direito”. Esta palavra aparece em colocações como 伤害法治 *shānghài fǎzhì* (ferir o Estado de Direito) e 回归法治 *huíguī fǎzhì* (voltar ao Estado de Direito), entre outras. No entanto, a colocação que mais causou problema aos voluntários chineses foi 超越法治 *chāoyuè fǎzhì*, em que 超越 *chāoyuè* significa “ultrapassar os limites de”.

Apesar de 超越 literalmente também significar “exceder”, “ultrapassar” e “transcender”, opções mais usadas pelos tradutores aprendizes, a consulta ao SkE com o *Word Sketch* e *Concordance* não apresentou resultados com essas palavras.

Figura 72 - Ficha com as soluções tradutórias para 超越法治

LogDice	Frequência	超越法治
0	0	1 exceder o Estado de Direito
0	0	2 estar além do Estado de Direito
0	0	2 ir além do estado de direito
0	0	1 passar a lei
0	0	1 superar as regras
0	0	1 transcender o Direito
0	0	5 transcender o Estado de Direito
0	0	1 transcender o império da lei
0	0	1 ultrapassar o estado de direito
Total		15
c-Col		0% Sign. Dif.
c-nCol		100%
nc Certo		0%
nc Errado		0%
of		0%
OpM		afrontar, subverter, solapar, atentar

Fonte: elaborada pelo autor

Por outro lado, a busca pelas palavras que coocorrem com “Estado de Direito” no *Word Sketch* revelou outros verbos com ideia similar que poderiam ser usados nesta tradução.

Tabela 16 – Verbos que coocorrem com estado de direito como objeto

VERBO	FREQUÊNCIA	LOGDICE
pilar	37	6.4
afrontar	37	6.3
viger	16	6.1
subverter	18	5.7
solapar	10	5.7
Atentar	58	5.6

Fonte: SkE.

Com base nos resultados apresentados acima, chegamos à conclusão de que soluções como “afrontar o Estado de Direito” (*logDice* 6.3), “subverter o Estado de Direito” (*logDice* 5.7), “solapar o Estado de Direito” (*logDice* 5.7) e “atentar contra o Estado de Direito” (*logDice* 5.6) seriam mais plausíveis e soariam mais natural.

Abaixo, nas Figuras 73 a 76, podemos observar alguns exemplos destas combinações, extraídos do *corpus* de referência com o *Concordance*.

Figura 73 – Linhas de concordância de "afrontar"+"Estado de Direito"

1			Portugal • avan...	osição dos comunistas à tomada de medidas que afrontam ainda mais o Estado de direito democrático. </s><s> Devolver à Portela o Pólo do Conservatc	
2			Brazil • bahiae...	público, ação destinada a criar-lhe constrangimento moral, que afronta o Estado Democrático de Direito , enxovalha a imagem da Polícia Federal – que, é	
3			Brazil • blogdo...	justiça, e, sem temor, enfrenta o inimigo que dilapida o erário e afronta o estado democrático de direito , então, nosso coração festeja, nossa alma se del	
4			Brazil • usp.br	s instituições universitárias, depredam o patrimônio público e afrontam o estado de direito . </s><s> Além de representar um ato extremo de agressão, a	
5			Brazil • viomun...	to com agentes do crime que tentam se impor pela violência, afrontam o Estado de Direito , cometem crimes em série e desprezam a democracia. </s><s>	
6			Brazil • manife...	idos, que medidas como essa, que amesquinham a justiça e afrontam o Estado Democrático de Direito , ao tentar retirar dos cidadãos o acesso por meic	
7			Brazil • cut.or...	a manifestação do pensamento, sendo vedado o anonimato"), afronta o Estado democrático de Direito e por esta razão é indigna de acolhimento ou def	
8			Brazil • aprema...	ual redação, constituir-se-á numa aberração jurídica, eis que afrontará o Estado Constitucional de Direito em desrespeito às regras de competência previ	
9			Brazil • aprema...	ual redação, constituir-se-á numa aberração jurídica, eis que afrontará o Estado Constitucional de Direito em desrespeito às regras de competência previ	
10			Brazil • apesp....	o serviço essencial, eles estão prejudicando a população e afrontando o Estado de Direito . </s><s> Vista Espetacular de toda a cidade 21o andar: Piscir	
11			Brazil • rj.gov...	ará igual a qualquer lugar do mundo, e não algo que chegue a afrontar o Estado de Direito a ponto de criar um território que o governo não controla", cor	

Fonte: SkE.

Figura 74 – Linhas de concordância de "subverter"+"Estado de Direito"

8		Portugal • smmp...	m existido no caso em apreciação, não têm idoneidade para subverter o Estado de Direito ".	Estado de Direito	Não faltam, neste relato, revelações desconcertante	
9		Portugal • smmp...	m existido no caso em apreciação, não têm idoneidade para subverter o Estado de Direito .	Estado de Direito	Poderão ter várias leituras nos planos político, social	
10		Portugal • smmp...	rante desvio ou abuso das suas funções (...) tenta alterar ou subverter o Estado de Direito constitucionalmente estabelecido, nomeadamente os direitos, l	Estado de Direito	constitucionalmente estabelecido, nomeadamente os direitos, l	
11		Portugal • entr...	sta esclarecer é como terá ele procurado destruir, alterar ou subverter o Estado de direito , tal como exige a norma acima citada.	Estado de direito	O truque de rac	
12		Portugal • entr...	ma: precisamente a actividade tendente a destruir, alterar ou subverter o Estado de direito .	Estado de direito	É claro que só por delírio se pode considerar que Jos	
13		Portugal • dgsl...	nosso). Quem, com intenção de destruir, alterar ou subverter o Estado de direito constitucionalmente estabelecido, em reunião pública ou por di	Estado de direito	constitucionalmente estabelecido, em reunião pública ou por di	
14		Brazil • patria...	ompromete seus governos a condenar qualquer tentativa de subverter o Estado de Direito e proclama a igualdade das pessoas com incapacidades.	Estado de Direito	e proclama a igualdade das pessoas com incapacidades.	
15		Brazil • dhnet...	combate às drogas que não respeitam os direitos humanos subvertem o Estado de Direito .	Estado de Direito	A Human Rights Watch/Americas conclama os gover	
16		Brazil • unisin...	/s><s> Por exemplo, normas clientelistas e patrimonialistas subvertem o Estado de Direito formal.	Estado de Direito	As instituições informais substitutivas são conv	
17		Brazil • idisa....	1: - Quando (a ANVISA) o faz afronta o Congresso Nacional e subverte o Estado de Direito e a própria democracia.	Estado de Direito	Segundo a resolução, os fabri	
18		Brazil • midiaa...	de rádio e TV críticas.	Estado de direito	Era parte de um esforço para subverter o Estado de direito , a partir do governo, para impedir uma sucessão democrática.	

Fonte: SkE.

Figura 75 – Linhas de concordância de "solapar"+"Estado de Direito"

1		Brazil • mre.go...	ce as instituições públicas e privadas, erodem valores sociais, solapam o Estado de Direito e distorcem a economia e a alocação dos recursos destinados	Estado de Direito	erodem valores sociais, solapam o Estado de Direito e distorcem a economia e a alocação dos recursos destinados	
2		Brazil • dhnet....	te.	Estado de Direito	Precisamos de reformas que fortaleçam a sociedade	
3		Brazil • ibase....	entanto, uma anistia indiscriminada era também inaceitável, solapando o Estado de Direito ainda frágil da nação.	Estado de Direito	Este foi o contexto da adoção de	
4		Brazil • fpa.or...	eitoral desta guerra só fortalece os bandidos e só compete para solapar o Estado de Direito .	Estado de Direito	O candidato tucano, contudo, não deixa de dar uma	
5		Brazil • ibase....	lmejadas pelo novo governo.	Estado de Direito	sobre o qual a nova situação precisava ser construída.	
6		Brazil • usp.br	i segurança hemisférica e, usados por terroristas e criminosos, solapam o Estado de Direito , engendram violência e em alguns casos impunidade, exacerb	Estado de Direito	engendram violência e em alguns casos impunidade, exacerb	
7		Brazil • mises....	cio a todos os movimentos, tal como a Freirechtsschule, que solaparam o estado de Direito .	estado de Direito	Pode-se mesmo afirmar que, para o estado de Direito	
8		Brazil • canalco...	ada vez que a menção do termo censura vem à tona já está solapando o estado de direito .	estado de direito	Imagine quando posto em operação, como no caso de	
9		Brazil • portal...	o pode levar, inevitavelmente, a uma balbúrdia, capaz de, per si, solapar o Estado Democrático de Direito .	Estado Democrático de Direito	É preferível, às vezes, "sofrer" com as re	
10		Brazil • expres...	seccional da OAB, ?tais procedimentos drasticamente solapam o próprio Estado Democrático de Direito ?.	Estado Democrático de Direito	Clicando no botão " Históricos já lanç	

Fonte: SkE.

Figura 76 – Linhas de concordância de "atentar"+"Estado de Direito"

9		Brazil • rodrig...	io as acusações a Getúlio.	Estado de Direito	Autoritários foi quem atentou contra o Estado de Direito em (1932) a oligarquia Cafeeira, e em (1935), a intentona Com	
10		Brazil • blogda...	s e afirmou também ser contrário a qualquer medida que " atente contra o estado democrático de Direito ".	estado democrático de Direito	Já o ministro Gilmar Mendes explicou a	
11		Brazil • midiaa...	o governo militar mesmo com este bando de traidores, atentando contra o estado de direito , desconheço ordens que fossem para desarmar o cidadão idón	estado de direito	desconheço ordens que fossem para desarmar o cidadão idón	
12		Brazil • pedrod...	r para descobrirem que ela, a juíza, descumpriu normas, atentou contra o estado de direito , foi policialesca e por aí vai...	estado de direito	abçs 296	
13		Brazil • blogdo...	's><s> Atentar contra a independência funcional do juiz é atentar contra o Estado Democrático de Direito .	Estado Democrático de Direito	Nenhum juiz pode ser punido apenas pc	
14		Brazil • pedrod...	<s> Que eu saiba, isso é crime previsto na Constituição - atentar contra o estado de direito .	estado de direito	Pq o cara não foi processado? Pq era visto o	
15		Brazil • pedrod...	lele shopping de botafogo.	estado de direito	e as instituições democráticas. E, além do mais, o tal d	
16		Brazil • fetrac...	ção das atividades dos servidores da Justiça Trabalhista atentou contra o Estado Democrático de Direito , a ordem pública e os princípios da legalidade, da	Estado Democrático de Direito	a ordem pública e os princípios da legalidade, da	
17		Brazil • tortur...	plentas e autoritárias, heranças da ditadura militar e que atentam contra o Estado Democrático de Direito ; 4.	Estado Democrático de Direito	Extinção da Lei de Segurança Nacion	
18		Brazil • rs.gov...	ado Marco Alba argumentou que o episódio da Aracruz atentava contra o estado de direito , acusando-o de ser um movimento orquestrado.	estado de direito	Concl	
19		Brazil • uol.co...	residência da corte, não se vislumbra uma só decisão que atente contra o Estado democrático de Direito e as garantias fundamentais do cidadão".	Estado democrático de Direito	e as garantias fundamentais do cidadão".	

Fonte: SkE.

Apesar de “pilar” e “viger” figurarem entre as três palavras que mais coocorrem com “Estado de Direito”, segundo a busca com o *Word Sketch*, após observação das linhas de concordância com o *Concordance*, logo notamos que a primeira, “pilar”, não é empregada como um verbo, mas como substantivo, como comprovado abaixo, na Figura 73, com as linhas de concordância para essa combinação.

Figura 77 – Linhas de concordância de “pilar” + “Estado de Direito”

1	☐	Portugal • mai....	anção, bem expressa a valorização da Segurança dos Cidadãos, pilar do Estado de Direito , que este Governo manifesta". </s><s> O primeiro concerto nc	
2	☐	Portugal • pnet...	indícios de factos que subvertem a liberdade de imprensa como pilar do Estado de direito quando se diz, por exemplo, que "O Sócrates perguntou-me se	
3	☐	Portugal • smmp...	e contexto a independência dos advogados é fundamental como pilar de Estado de Direito . </s><s> Um advogado que vai a tribunal com "medo" dos dor	
4	☐	Portugal • smmp...	dos pilares do Estado de direito. - Exactamente. </s><s> Como pilar do Estado de direito e como factor de desenvolvimento económico. </s><s> Mas ef	
5	☐	Portugal • amic...	o advogado António Arnaut irão abordar o tema "A Justiça como pilar do Estado de Direito " no jantar festivo de Fevereiro do Rotary Club de Coimbra que	
6	☐	Portugal • embc...	, José Maria Neves apontou o poder local como pilar fundamental de um Estado de direito democrático, uma vez que permite um maior pluralismo através	
7	☐	Portugal • sobe...	jonha), quando deixarem o governo. </s><s> A Justiça, pilar de qualquer Estado de Direito , vive os dramas da mulher de César. </s><s> Se não é cumpli	
8	☐	Portugal • setu...	tidos na "Declaração Universal dos Direitos do Homem" e pilar básico do Estado de Direito – só se atingem em alguma da sua plenitude quando existem r	
9	☐	Portugal • advo...	estão interessados numa mudança que consolide a Justiça como pilar do Estado de Direito Democrático e Social do país", disse, no final do escrutínio. </s>	
10	☐	Portugal • dgsi...	de outros direitos, o acesso ao direito e aos tribunais surge como pilar do Estado de Direito , e encontra eco nos principais textos de direito internacional c	
11	☐	Portugal • sapo...	de subordinação do Executivo ? </s><s> Como se afirma como pilar do Estado de Direito ? </s><s> Como se unem e se congregam solidariamente TOC	
12	☐	Brazil • migalh...	diretivo que a ela se outorgam, pela afronta à liberdade de ação, pilar do Estado Democrático de Direito , é de se manter a sentença que declarou a rescis	
13	☐	Brazil • direit...	s direitos humanos (entre os quais se inclui o direito à segurança, pilar do Estado Democrático de Direito). </s><s> Formou-se mais um paredão na casa c	
14	☐	Brazil • direit...	STITUCIONAL DO SÍGILLO BANCÁRIO </s><s> É tradição, porque pilar do estado democrático de direito , a proteção constitucional, em capítulo pético, do	
15	☐	Brazil • cisoja...	vernamentais, inverteu a ordem jurídica brasileira, que tem como pilar do estado de direito a presunção de inocência, e não a presunção de culpa. </s><s>	
16	☐	Brazil • direit...	005)." </s><s> Trata-se da moralidade administrativa, verdadeiro pilar do Estado Democrático de Direito , positivado em sede constitucional através da Ca	
17	☐	Brazil • direit...	ousa no artigo 1o, inciso III da Constituição Federal, pilar inarredável do Estado Democrático de Direito . </s><s> Logo, em última análise a autoridade ju	
18	☐	Brazil • ffm-br...	de conteúdo e de eventos da RELIAL baseia-se nos pilares temáticos de Estado de Direito , do sistema económico, da educação e da segurança social, o	
19	☐	Brazil • rs.gov...	nstituições, cada uma na sua função, que são verdadeiros pilares de um estado democrático de direito . </s><s> Em nome da Bancada do PDT, reitero tu	
20	☐	Brazil • pr.gov...	além de tudo, grave ameaça ao nosso modelo federativo, pilar do nosso Estado Democrático de Direito . </s><s> Além disso, os recursos dirigidos ao ST	
21	☐	Brazil • uol.co...	se confunda com a vingança. </s><s> São esses fundamentos pilares do Estado democrático de Direito , conquista da civilização humana, que não pode t	
22	☐	Brazil • catho....	elo qual o legislador constituinte elevou o trabalho à condição de pilar do Estado Democrático de Direito . </s><s> O trabalho é um direito universal e não	

Fonte: SkE.

Falhas como essa corroboram a importância do nosso olhar atento e experiência como pesquisador da área, como anteriormente mencionado nas subseções 3.2.1 e 3.2.2.

“Viger”, por sua vez, possui um significado distinto daquele expresso por 超越 no TO.

Figura 78 – Linhas de concordância de “viger” + “Estado de Direito”

1	☐	①	Brazil • blogdo...	querer. </s><s> Outra, é poder. </s><s> No Brasil, como na Itália, vige o estado de direito . </s><s> Pelo menos até segunda ordem. </s><s> Lula precisa
2	☐	①	Brazil • blogdo...	<s> Ninguém está acima de qualquer suspeita quando vige plenamente o Estado de Direito . </s><s> Por isso, indivíduos, grupos ou partidos têm sim o dire
3	☐	①	Brazil • ufsm.b...	so penal existe antes de tudo para a proteção do acusado eis que vige no Estado de Direito o primado do Princípio da Inocência. </s><s> Logo, algumas da
4	☐	①	Brazil • unisin...	tes bolsões de estado de exceção existam, para que aqueles onde vige o estado de direito possam subsistir. </s><s> É provável que a coexistência neces
5	☐	①	Brazil • multip...	os estão convocando um Estado organizado, democrático, em que vige o Estado de Direito , a se comportar como uma milícia: ou eleva o Hezbollah à conc
6	☐	①	Brazil • espiri...	, como indivíduos inteligentes, integrantes de uma sociedade onde vige o estado de direito , nos despir desses sentimentos e confiar no Estado, ente que e
7	☐	①	Brazil • direit...	guidos dos regimes totalitários. </s><s> O Brasil é um país em que vige o Estado de Direito Democrático. </s><s> Como sabido, governo é uma coisa e Est
8	☐	①	Brazil • edison...	res poderes independentes em praticamente todos os países onde vige o Estado de Direito . </s><s> No Brasil, esta inspiração está presente na arquitetura
9	☐	①	Brazil • substa...	namente constituídas, eleitas democraticamente, num país em que vige o Estado de Direito , não merece ser considerado criminoso político. </s><s> Ou se
10	☐	①	Brazil • boitem...	alidade. </s><s> O soberano é quem decide quando, como e onde vige o estado de direito e, situando-se dentro e fora do direito, é a exceção que condic
11	☐	①	Brazil • ultima...	ico-Cristã. </s><s> Bela troca! </s><s> A nação têm instituições, e vige o Estado de Direito , de modo que ela se faz representativa exatamente ali, no Legi
12	☐	①	Brazil • ultima...	uso, os artigos são publicados. </s><s> Por quê? </s><s> Por que vige o estado de direito , a democracia, e a revista tem suas normas fundadas no seu in
13	☐	①	Brazil • rdnews...	orte moral para ser uma constatação referencial dos regimes onde vige o estado de direito . </s><s> O sujeito da segunda oração (teme) é a primeira (quei
14	☐	①	Brazil • walter...	rigado a produzir provas contra si mesmo?. </s><s> No nosso país vige o estado democrático de direito , quer queiram ou não. </s><s> Todos devem se su
15	☐	①	Brazil • portal...	Procurador Jurídico do Município, faz observar que, no Brasil, onde vige o estado democrático de direito , a Constituição fixa como princípio constitucional c
16	☐	①	Brazil • nom.br	as Furnas pareceu-me demasiado estreita diante do Brasil no qual vige um Estado Democrático de Direito . </s><s> Ouvi atentamente e até gravei quase tuc

Fonte: SkE.

Depois das colocações verbais, o segundo tipo de colocação com mais erros no *subcorpus* chinês foi o das colocações adjetivas. Dentre os colocados mais frequentes neste *subcorpus*, encontra-se o adjetivo e advérbio 严重 yánzhòng (grave, sério). De modo geral, colocações com esta palavra não representaram um grande problema para os aprendizes, exceto quando combinada com 暴力 bàolì (violência).

Figura 79- Ficha com as soluções tradutórias para 严重的暴力

LogDice	Frequência	严重的暴力
3.9	167	6 violência grave
nc	nc	1 violência
0	5	1 violência séria
0	0	7 violência severa
Total		15
c-Col		40% Sign. Dif.
c-nCol	✓	53%
nc Certo		7%
nc Errado		0%
of		0%
OpM		

Fonte: elaborado pelo autor.

Como podemos depreender da ficha de 严重的暴力 *yánzhòng de bàolì* acima, os aprendizes tiveram dificuldade em determinar qual adjetivo colocaria com “violência”: “grave”, “séria” ou, ainda, uma terceira opção, “severa”.

Após consulta ao *corpus* de referência *ptTenTen11*, percebemos que, das soluções encontradas, “grave” é a escolha mais apropriada para modificar violência neste caso, com um *logDice* de 3.9. Contudo, uma consulta pelas colocações de “violência” a partir do *Word Sketch* nos permite notar que há outros adjetivos com acepções parecidas, segundo tabela abaixo, que coocorrem mais frequentemente com esta base, majoritariamente posicionadas à sua frente, e não depois, pois os adjetivos que sucedem “violência”, em grande parte, referem-se ao tipo de violência praticada.

Tabela 17 - adjetivo + violência

MODIFICADOR	FREQUÊNCIA	LOGDICE
extremo	1,003	8.2
brutal	110	7.8
tamanho	386	7.8
crescente	479	7.4
Inaudito	23	6.7
Cruel	32	6.1
Temático	36	6.0
inominável	14	6.0
excessivo	40	5.8
assustador	15	5.6
Absurdo	22	5.5
terrível	50	5.4

Fonte: SkE.

Tabela 18 - violência + adjetivo

MODIFICADOR	FREQUÊNCIA	LOGDICE
doméstico	26,484	11.7
sexual	14,415	10.3
urbano	6,738	8.6
policial	3,048	8.6
físico	5,912	7.9
psicológico	1,339	7.8
intrafamiliar	512	7.6
simbólico	810	7.4
conjugal	545	7.4
gratuito	1,142	7.0
sexista	347	7.0
verbal	515	7.0

Fonte: SkE.

Tabela 19 - Frequência e *score* de grave + violência

MODIFICADOR	FREQUÊNCIA	LOGDICE
grave	80	4.5
grave violência		

Fonte: SkE.

Tabela 20 - Frequência e *score* de violência + grave

MODIFICADOR	FREQUÊNCIA	LOGDICE
grave	167	3.9
violência grave		

Fonte: SkE.

Notamos, a partir das duas últimas tabelas, que a posição do adjetivo também influencia na construção da colocação, pois “violência” + “grave”, a opção dada pelos aprendizes, possui *logDice* de 3.9, ao passo que “grave” + “violência” possui um *logDice* mais alto, de 4.5.

Como mencionamos anteriormente, as colocações adjetivas foram o segundo tipo com mais erros no *subcorpus* chinês. Entre as opções tradutórias, outra colocação que apresentou muitos erros foi 有关当局 *yǒuguān dāngjú* (autoridade competente).

Figura 80 - Ficha com as soluções tradutórias para 有关当局

LogDice	Frequência	有关当局
0	56	10 autoridade relevante
nc	nc	4 autoridade
11.3	29.462	1 autoridade competente
0	12	1 autoridade relativa
Total	16	
c-Col	6% Sign. Dif.	0%
c-nCol	69%	
nc Certo	25%	
nc Errado	0%	
of	0%	
OpM		

Fonte: SkE.

有关 *yǒuguān* significa literalmente “ter relação” e, talvez, por este motivo, os voluntários tenham tido dificuldade em determinar qual colocado melhor coocorreria com a base “autoridade”, já que “relevante” ou “relativa” são as palavras que mais se aproximam de “ter relação”. No entanto, a busca pelos adjetivos que colocam com “autoridade” no *Word Sketch* não apresentou nenhuma destas palavras como resultado, e a partir do resultado da busca por “autoridade” + “relevante” e “autoridade” + “relativa” com o *Concordance* depreendemos que ambas as opções coocorrem muito pouco com esta base, 56 vezes a primeira e apenas 12 vezes a segunda, não caracterizando-as, portanto, como colocações em língua portuguesa.

Quando analisamos o contexto em que 有关当局 ocorre no TO, chegamos à conclusão que se trata da colocação em português “autoridade competente”.

Quadro 9 - Linha de concordância de 有关当局 no subcorpus de TOs

有关当局还应保护中国各地非洲人和非洲裔人士免于就业、住房和其他领域的歧视。
32
<i>Yǒuguān dāngjú hái yīng bǎohù zhōngguó gèdì fēizhōu rén hé fēizhōu yì rénshì miǎn yú jiùyè, zhùfáng hé qítā lǐngyù de qíshì.</i>

Fonte: elaborado pelo autor.

³² As autoridades competentes também devem proteger os africanos e os afrodescendentes em toda a China da discriminação no emprego, habitação e outros campos.

Diferentemente de “relevante” ou “relativa”, “competente” é o colocado que se liga mais fortemente à base “autoridade”, independentemente da posição, antes ou depois do substantivo.

Tabela 21 - Frequência e *score* de autoridade+adjetivo

MODIFICADOR	FREQUÊNCIA	LOGDICE
competente	29,462	11.3
policial	12,866	10.0
monetário	5,051	9.0
local	13,751	8.9
sanitário	5,640	8.8

Fonte: SkE.

Tabela 22 - Frequência e *score* de adjetivo + autoridade

MODIFICADOR	FREQUÊNCIA	LOGDICE
competente	283	7.8
diverso	5,557	7.2
alto	3,446	7.1
dito	144	7.0
máximo	218	6.5

Fonte: SkE.

A seguir, as colocações nominais foram o terceiro tipo com mais erros. Uma das colocações que mais receberam soluções tradutórias com erros colocacionais foi 疫情区域 *yìqíng qūyù*, que pode ser traduzida literalmente como “área da epidemia”. Vejamos a ficha com as opções tradutórias para essa colocação a seguir.

Figura 81 - Ficha com as soluções tradutórias para 疫情区域

LogDice	Frequência	疫情区域	
nc	nc	1 foco	Mod
ce	ce	1 regiões mais atingidas pela epidemia	Mod
4.5	74	1 foco de epidemia	Trp
0.9	2	2 áreas com epidemia	Trp
0	0	6 áreas epidêmicas	Trp
0	0	1 áreas sofrenda pela epidemia	Err
0	0	1 regiões de epidemia	Trp
0	0	1 regiões epidêmicas	Trp
0	0	1 área da pandemia	Trp
0	0	1 Ásia com epidemia	Err
Total		16	
c-Col	✓	25% Sign. Dif.	0%
c-nCol	✓	69%	
nc Certo		6%	
nc Errado		0%	
of		0%	
OpM			

Fonte: elaborado pelo autor.

Várias foram as saídas encontradas pelos aprendizes, porém, foram poucas aquelas que realmente são colocações na L2. Em verdade, após analisarmos o contexto de ocorrência de 疫情区域, acreditamos que se trata de um caso em que uma solução que não apresente necessariamente uma colocação seria a melhor opção.

Quadro 10 - Linha de concordância de 疫情区域 no subcorpus de TOs

看到一些人歧视来自疫情严重区域的亚洲人，甚至还包括亚裔巴西人，茹晨曦仗义执言，"难道这些人就毫不在乎、毫不同情人类遭受的这些苦难吗？" ³³
Kàn dào yīxiē rén qíshì láizi yìqíng yánzhòng qūyù de yàzhōu rén, shènzhì hái bāokuò yà yì bāxī rén, rú chénxī zhàngyìzhíyán, " nándào zhèxiē rén jiù háo bùzàihū, háo bùtóng qíngrén lèi zāoshòu de zhèxiē kǔnàn ma?

Fonte: elaborado pelo autor.

Dentre as opções apresentadas, a que melhor satisfaz a ideia do TO é “regiões mais atingidas pela epidemia”, cuja linha de concordância segue abaixo:

³³ Tradução: Ao ver algumas pessoas discriminando asiáticos de áreas onde a epidemia é grave, até mesmo brasileiros de origem asiática, Ru Chenxi disse com tom de indignação: "Essas pessoas não se importam ou sentem empatia pelo sofrimento dos outros?"

Quadro 11 - Linha de concordância de solução tradutória para 疫情区域 no subcorpus de TTs

Ao ver que algumas pessoas discriminam os asiáticos, e até os brasileiros de origem asiática, provenientes das **regiões mais atingidas pela epidemia**, ela disse: "essas pessoas não se importam, não têm compaixão por este sofrimento humano?"

Fonte: elaborado pelo autor.

Por fim, no subcorpus chinês, as colocações adverbiais foram as que os voluntários menos erraram, mas, mesmo assim, erros ocorreram, dentre os quais destacamos 发冷得严重 *fālěng dé yánzhòng*.

发冷 *fālěng* é um verbo chinês que significa literalmente “sentir calafrio” ou “sentir arrepios”, “arrepia-se”. Como se trata de um verbo, 发冷 pode ser modificado por um advérbio, dentre os quais, o já mencionado 严重 *yánzhòng*, que significa “grave”, “sério”.

Figura 82 - Ficha com as soluções tradutórias para 发冷得严重

LogDice	Frequência	发冷得严重
0	0	6 calafrios grave
nc	nc	1 calafrios
nc	nc	1 resfriado
8.7	1967	1 frio intenso
6.5	1797	1 sentir frio
0.8	17	1 gripe grave
0	0	1 calafrios sérios
0	0	1 escalofríos graves
0	0	1 frios muito graves
0	0	1 grave frio
Total		15
c-Col	20% Sign. Dif.	100%
c-nCol	67%	
nc Certo	13%	
nc Errado	0%	
of	0%	
OpM	forte calafrio	0.1
	enorme	1.9
	calafrio	
	violento	2.8

Fonte: elaborado pelo autor.

Algumas das opções apresentadas podem até ser consideradas colocações do português, como “frio intenso” e “sentir frio”, cujos *scores logDice* são 8.7 e 6.5, respectivamente, porém, não correspondem ao significado original da colocação, como é observável por meio da linha de concordância do TO.

Quadro 12 - Linha de concordância de 发冷得严重 no subcorpus de TOs

一名叫乔的网友表示: " 我父母于去年 12 月底、今年 1 月初得了重病而我在去年 10 月底、11 月初时也生了病。我们的症状类似咳嗽、发烧、发冷得很严重。³⁴

Yī míng jiào qiáo de wǎngyǒu biǎoshì: " Wǒ fùmǔ yú qùnián 12 yuèdǐ jīnnián 1 yuèchū dé le zhòngbìng ér wǒ zài qùnián 10 yuèdǐ 11 yuèchū shí yě shēng le bìng. Wǒmen de zhèngzhuàng lèisì kēsòu fāshāo fā lěng dé hěn yánzhòng.

Fonte: elaborado pelo autor.

A ideia de 发冷 no TO é realmente a de “calafrio”. No entanto, as opções mais literais que mantiveram esta base não foram bem-sucedidas, pois mantiveram a tradução literal de 严重, “grave” ou “sério”, opções que não colocam com “calafrio”, segundo resultados das consultas ao corpus de referência, tanto com o *Word Sketch* como com o *Concordance*.

Na realidade, quando buscamos pelos adjetivos que mais coocorrem com a base da referida colocação no *Word Sketch*, obtemos “violento” (*logDice* 2.8), “enorme” (*logDice* 0.8), “forte” e “intenso” (*logDice* 1.0), que, apesar do baixo *logDice*, pelo menos aparecem nos resultados das buscas e são os colocados mais frequentes, a depender da posição do adjetivo.

³⁴ Um internauta chamado Qiao disse: “entre o final de dezembro do ano passado e o início de janeiro deste ano, meus pais adoeceram gravemente, e eu também fiquei doente entre o final de outubro e início de novembro. Os nossos sintomas eram parecidos, tosse, febre e fortes calafrios.

Tabela 23 - Frequência e *score* de adjetivo + calafrio

ADJETIVO	FREQUÊNCIA	LOGDICE
súbito	6	4.8
certo	18	2.9
violento	4	2.8
estranho	5	1.9
enorme	12	0.8
forte	11	< 0.1

Fonte: SkE.

Tabela 24 - Frequência e *score* de calafrio + adjetivo

ADJETIVO	FREQUÊNCIA	LOGDICE
repentino	4	3.7
noturno	7	1.2
intenso	9	1.0

Fonte: SkE.

Este é um caso em que uma colocação de um tipo na L1 (colocação adverbial) é transformada em uma colocação de outro tipo na L2 (colocação adjetiva).

Mostramos até aqui, nas subseções 4.2.1 e 4.2.2 alguns exemplos dos erros colocacionais cometidos na tradução de colocações na direção tradutória L1 → L2, i.e., LM → LE, nos nossos *corpora* de estudo, o CAT 1, aqui também chamado de CAT brasileiro ou *subcorpus* brasileiro, cuja direção tradutória é L1 (português) → L2 (inglês), e o CAT 3, aqui também chamado de CAT chinês ou *subcorpus* chinês, cuja direção tradutória é L1 (chinês) → L2 (português). Demonstramos também como realizamos, com o manuseio das ferramentas de *corpus*, as análises das soluções tradutórias encontradas pelos aprendizes de tradução e estudantes avançados de idiomas a fim de determinar se elas eram colocações existentes na L2, se foram empregadas com os mesmos significados e se eram adequadas ao contexto expresso nos TOs em L1, para que chegássemos aos resultados apresentados na subseção 4.2.

Com os resultados obtidos, pudemos notar que a compreensão das colocações acontece em razão da L1, de modo que os erros colocacionais ocorreram em proporções diferentes nos diferentes tipos de colocação (ORENHA-OTTAIANO, 2009) nos diferentes *subcorpora*.

Porém, um tipo de colocação em específico se mostrou especialmente problemático para os dois grupos de voluntários: as colocações verbais.

No entanto, cabe ressaltar que, independentemente do tipo de colocação enfocado, a maior parte dos erros colocacionais é causada pela tradução literal de um dos componentes da colocação, majoritariamente o colocado, tanto no *corpus* brasileiro como no *corpus* chinês.

Tivemos uma demonstração de como a tradução literal dos colocados ocasiona erros colocacionais no *corpus* chinês com as opções dadas para colocações com colocados como 严重 *yánzhòng* (sério, grave, severo), 有关 *yǒuguān* (ter relação com, relativo a, competente), 超越 *chāoyuè* (exceder, transcender, ultrapassar), entre outros. O mesmo ocorreu no *corpus* brasileiro, como pudemos observar nas traduções com colocados como *dry* (seco) e *give* (dar). No *corpus* brasileiro, cognatos em inglês que são a tradução literal dos colocados ou bases em português, como *dissemination* (disseminação), *declaration* (declaração), *plan* (plano), *education* (educação), também causaram um número considerável dos erros na tradução de colocações por parte dos aprendizes brasileiros.

A esse respeito e a outras conclusões as quais chegamos com os resultados das análises, trataremos no capítulo a seguir, das considerações finais.

5 CONSIDERAÇÕES FINAIS

No presente trabalho, compilamos um *corpus* paralelo de traduções feitas por estudantes chineses avançados de português para ser incorporado ao Corpus de Aprendizes (CAT), projeto parte do Compilação de Materiais Didáticos e Dicionários Especializados de Colocações Baseados em Corpora (COMADEC), em desenvolvimento no Instituto de Biociências, Letras e Ciências Exatas (IBILCE) da Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), sob coordenação de Orenha-Ottaiano (2017). Este *corpus*, com traduções do chinês (L1) para o português (L2), que compõe o CAT 3, bem como outro *corpus* de traduções L1(português)→L2 (inglês), o CAT 1, previamente compilado e já incorporado ao CAT, formaram o nosso *corpus* de estudo, de onde levantamos as colocações presentes nos TOs para, posteriormente, analisarmos comparativamente os erros cometidos pelos aprendizes brasileiros e chineses quando da tradução de colocações de LM para uma LE.

Inicialmente, no primeiro capítulo deste trabalho, buscamos demonstrar como nosso projeto se insere num plano político mais amplo ao analisar o chinês mandarim. Apesar de as relações entre Brasil e China estarem cada vez mais estreitas e haver um aumento da demanda das línguas desses países em seus territórios, pouco tem sido produzido acerca do chinês nos Estudos Fraseológicos baseados em *Corpus* no Brasil, mais especificamente no que tange a trabalhos contrastivos (português/chinês). Este fato vem corroborar a relevância da nossa pesquisa para os estudos na área citada, pois realizamos um passo inédito na compilação do primeiro *corpus* de tradução de aprendizes chineses de português.

Após posicionarmos nosso projeto em um plano geopolítico maior e justificar a sua relevância para os estudos da área, movemo-nos para o referencial teórico sobre o qual nos baseamos, a iniciar pelo arcabouço teórico da Linguística de *Corpus*, que modelou nossa pesquisa desde a compilação do nosso *corpus* de estudos, seguida pelos Estudos da Tradução baseados em *Corpus*, e finalmente, pela Fraseologia.

Feita a revisão da teoria que embasa nosso trabalho, detalhamos os procedimentos metodológicos que nos levaram a chegar aos resultados obtidos, dividindo-os em duas etapas: uma de compilação do *corpus* chinês e outra de levantamento das colocações e a análise das traduções.

Para a compilação do *corpus* chinês, lançamos mão de alternativas tecnológicas para um processo de compilação das traduções 100% *on-line* e com baixíssimo custo em meio à pandemia de COVID-19. Apesar do período de grandes dificuldades impostas pelo isolamento social, mantivemos uma elevada porcentagem de participação na compilação do nosso *corpus*, com pouquíssimas desistências.

Estamos confiantes de que parte do sucesso em conseguir manter os alunos engajados na pesquisa provém da autonomia que lhes foi dada, com um processo inteiramente *on-line*. Sem que perdêssemos o controle do tempo de realização das traduções, graças as plataformas usadas, os alunos tiveram o livre arbítrio de decidirem quando começariam as traduções, dentro do prazo estipulado, e de onde traduziriam, pois para isso não foi preciso que eles se deslocassem a um local específico em um horário marcado, o que, de todo modo, não seria possível devido à pandemia, podendo, portanto, adaptar suas rotinas da maneira que lhes fosse mais conveniente.

Um outro fator que julgamos ter sido decisivo foi a proximidade estabelecida entre o pesquisador e os participantes. O *WeChat*, rede social usada no recrutamento dos aprendizes e onde a sala virtual usada para disponibilizar os TOs se encontra, é amplamente difundido e usado por pessoas de todas as idades e todos os estratos sociais em toda a China, e os alunos usam essa ferramenta em seu dia a dia para as mais diversas funções, incluindo a de se comunicar com amigos e colegas.

A China é uma sociedade que valoriza muito as estruturas hierárquicas, o que pode ser comprovado com a alta pontuação do país na dimensão “Distância do poder” na comparação de países do Hofstede Insights (INSIGHTS), grupo que realiza análises de dados culturais a partir da Teoria das Dimensões Culturais do psicólogo holandês, Geert Hofstede (HOFSTEDE, 2011). Esse professor, conhecido por desenvolver algumas das teorias mais antigas e populares para medir dimensões culturais em uma perspectiva global, em sua teoria divide a compreensão de uma cultura em seis dimensões, dentre elas a “Distância do Poder”, que é “a medida em que os membros menos poderosos da organizações e instituições (como a família) aceitam e esperam que o poder seja distribuído”.³⁵ Nessa dimensão, de acordo com os resultados das análises do

³⁵ ...the extent to which the less powerful members of organizations and institutions (like the family) accept and expect that power is distributed unequally.

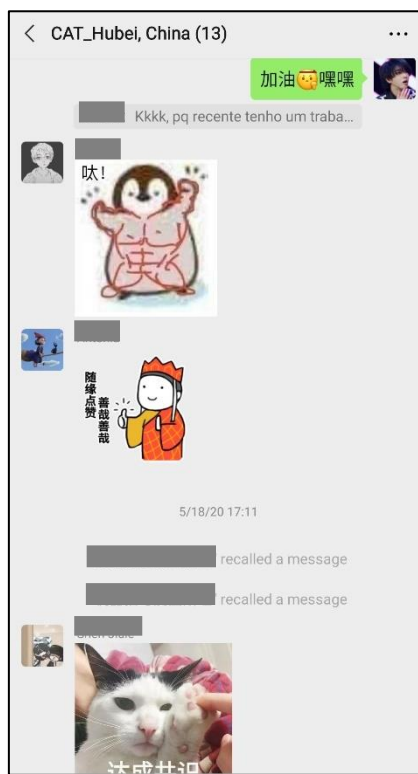
Hofstede Insights, em uma escala de 0 a 100, a China obteve pontuação de 80, o que implica dizer que a China é:

uma sociedade que acredita que as desigualdades entre as pessoas são aceitáveis. A relação subordinado-superior tende a ser polarizada e não há defesa contra abuso de poder por parte dos superiores. Os indivíduos são influenciados pela autoridade formal e sanções e, em geral, são otimistas quanto à capacidade de liderança e iniciativa das pessoas. As pessoas não devem ter aspirações além de sua posição. (INSIGHTS).³⁶

Sem a necessidade do contato físico com o pesquisador, que eles enxergam como alguém superior e, portanto, intimidador, dada a valorização da hierarquia, os alunos se sentiram mais à vontade para tirarem suas dúvidas sobre o projeto, sobre os textos e sobre as ferramentas nos grupos do *WeChat* ou por mensagens enviadas diretamente ao pesquisador. Estando mais à vontade, os alunos também falaram de suas dificuldades e eventualmente se desculparam e se justificaram por possíveis atrasos na entrega de algumas traduções. No mais, os grupos de *WeChat* foram formados por alunos que já se conheciam e, muitas vezes, eram amigos, elevando o clima de proximidade entre os membros do grupo, o que é possível notar pelas vezes em que eles compartilharam informações e 表情包 *biǎoqíng bāo*, algo como *emojis*, figurinhas e *memes* da internet.

³⁶... a society that believes that inequalities amongst people are acceptable. The subordinate-superior relationship tends to be polarized and there is no defense against power abuse by superiors. Individuals are influenced by formal authority and sanctions and are in general optimistic about people's capacity for leadership and initiative. People should not have aspirations beyond their rank.

Figura 83 - Dinâmica dos grupos do WeChat



Fonte: captura de tela do WeChat.

Outro aspecto muito interessante da cultura chinesa, além do rigor com que as relações hierárquicas são seguidas, é que os alunos são extremamente tímidos e habituados a estudarem sozinhos, e não são muito participativos em sala de aula. Porém, o que se constatou com vários relatos de professores na mídia chinesa e internacional durante o auge da epidemia do novo coronavírus na China, quando, devido a impossibilidade de retornarem às salas de aula físicas no primeiro semestre de 2020, todos os alunos chineses tiveram aula *on-line*, é que eles são mais participativos nas salas virtuais. Em conversas informais com as docentes do Departamento de Português do Instituto de Línguas Estrangeiras da HUBU, elas corroboraram essa visão, dizendo que no período em que lecionaram as aulas remotamente, obtinham mais *feedback* dos alunos, sobretudo por meio dos *chats*.

Apesar das diferenças culturais existentes entre o Brasil e a China, acreditamos que o modelo metodológico 100% *on-line* empregado no presente trabalho para a compilação do *corpus* possa ser replicado em outros estudos que também visem engajar e manter engajados

estudantes para a compilação de *corpora* de aprendizes, não apenas por dar mais autonomia aos participantes da pesquisa, mas também por ser mais conveniente e prático, podendo evitar, desse modo, desistências.

Em nossa pesquisa usamos o *Wechat* para transmitir as informações sobre como o trabalho seria realizado, pois se trata de uma ferramenta amplamente difundida e de fácil acesso. Essa é apenas uma ferramenta dentre as várias disponíveis. Na China, por exemplo, outra rede social multitarefa é o QQ. No Brasil, redes como o *Whatsapp*, *Telegram* ou até mesmo o *Facebook* poderiam ser usadas com o mesmo fim.

Para enviarmos e recebermos as traduções, usamos a sala virtual *Yuketang*, por ser hospedada dentro do *Wechat* e não exigir que os alunos baixassem nenhum aplicativo adicional, mas mesmo na China há outras salas virtuais, como o 超星学习通 *Chāoxīng xuéxí tōng*, que também poderiam ser usadas.

Após a compilação do *corpus* de estudo, na segunda etapa deste trabalho, usamos massivamente as ferramentas de manuseio de *corpus* do SkE para levantarmos as colocações dos *subcorpora* de textos originais das traduções L1(chinês) → L2 (português), cuja compilação foi de nossa responsabilidade, e traduções L1 (português) → L2 (inglês), cuja compilação foi de responsabilidade de Orenha-Ottaiano (2019) e Rocha (2020). No total, foram levantadas 1006 colocações, das quais 407 eram em português e 599, em chinês.

Para manusearmos nossos *subcorpora* paralelos de traduções, tivemos de recorrer a outras ferramentas. A partir do alinhador automático do *Wordfast Anywhere*, obtivemos uma visualização de todas as traduções lado a lado, o que nos possibilitou a observação das diferentes traduções nos diferentes textos em uma única janela. As soluções tradutórias foram selecionadas e extraídas com apoio de algumas ferramentas do *Microsoft Office Excel*, onde ficaram hospedadas para análise.

Do total de colocações levantadas, 14,9%, i.e., 150 colocações, seguiram para a etapa da análise das traduções.

Por meio das análises a partir de nosso *corpus* de estudo, verificou-se que os erros colocacionais ocorreram em todos os tipos de colocação, porém a proporção das colocações mais

passíveis a erro varia de acordo com a LM. As colocações verbais foram as mais passíveis de erro por parte dos voluntários em ambos os *subcorpora*, tanto o de brasileiros como o de chineses.

De todos os tipos de colocações, este foi o único do qual uma parcela das colocações foi traduzida equivocadamente por colocações inexistentes na L2 por todos os alunos. Ou seja, em ambas as LM analisadas, chinesa e portuguesa, os alunos tiveram maior dificuldade em produzir colocações verbais que soassem naturais na LE, denotando que, nas aulas e exercícios sobre colocações para estudantes brasileiros de inglês e estudantes chineses de português, certa atenção deve ser dada a esse tipo de colocação, por meio da explicação explícita do que se tratam as colocações e exercícios que visem desenvolver a competência colocacional dos estudantes.

No caso do *subcorpus* brasileiro, é possível notar que os erros em colocações verbais ocorrem em colocações com colocados muito frequentes, como *dar*, *ter*, *fazer*, entre outros, que podem coocorrer com uma vasta gama de bases; e em colocações que, retomando Xu, Li e Li (2006), quando dividem as colocações em idiomáticas, fixas, fortes e fracas, baseando-se nos conceitos de composicionalidade, substitucionalidade, mutabilidade e associação interna, seriam consideradas colocações do Tipo 2 e Tipo 3, i.e., colocações fortes e colocações fracas, mas sobretudo colocações do Tipo 3, em oposição às colocações do Tipo 0, idiomáticas, e Tipo 1, fixas.

Acreditamos que a tamanha naturalidade com que esses verbos são empregados na LM faça com que os alunos não reconheçam instantaneamente que estão diante de uma colocação, sobretudo quando o conhecimento do que são explicitamente as colocações não faz parte do repertório dos estudantes, e, conseqüentemente, eles não se sentem instigados a investigarem mais afundo as soluções tradutórias encontradas para esses blocos de palavras, sobretudo, dada a proximidade entre as línguas inglesa e portuguesa, que contam com diversos cognatos, como *dissemination* (disseminação), *declaration* (declaração), *plan* (plano), *education* (educação), mencionadas anteriormente nos exemplos de análise da subseção 4.2.1. Por esses motivos, possivelmente, os aprendizes acreditam que na língua alvo uma colocação resultado de uma tradução literal possa estar correta. Como demonstrado anteriormente, com o caso de *law project*, em alguns casos a tradução literal pode até resultar em uma colocação na L2, porém, nem sempre ela possui o mesmo significado expresso na L1, constatação que só poderia ser feita por parte do aluno caso ele se sentisse motivado a verificar suas escolhas.

Outro fato que nos leva a crer na hipótese levantada no parágrafo anterior é o fato de que, no *subcorpus* brasileiro, os erros colocacionais ocorreram em menor número em colocações envolvendo colocados menos frequentes e, por vezes, pertencentes a discursos mais especializados, como o das Relações Internacionais, do Direito e da Ecologia, por exemplo, ainda que presentes em textos jornalísticos de língua geral.

Ao observarmos a grade curricular dos alunos brasileiros, notamos que eles possuem em sua grade curricular, no segundo ano, um curso de Terminologia e, do segundo ao quarto ano, aulas de prática de tradução, que compreendem os mais diversos tipos de tradução, incluindo tradução técnico-especializada. Ou seja, desde o segundo ano de faculdade, mesmo que para muitos o conceito de colocações não esteja claro, eles foram preparados a lidarem com colocações do Tipo 0, i.e., idiomáticas e, sobretudo, do Tipo 1, i.e., fixas, que mais se assemelham a termos e a colocações especializadas segundo a classificação de Xu, Lu e Li (2006).

Isso não significa dizer que erros em colocações de domínios mais especializados não tenham ocorrido, como vimos com “plano inclinado”, em que uma colocação especializada foi empregada figuradamente. Em casos como este, dentre os motivos que podem ter levado ao erro, acreditamos que o fato de ter sido empregado com sentido figurado em si pode ter contribuído, pois os alunos não reconheceram plano inclinado como uma colocação especializada ou um termo de uma área específica estando presente em um texto não-técnico.

No mais, talvez por desconhecimento, muitos alunos se confundiram e se equivocaram na tradução de cognatos e homônimos, como “plano”, que em português significa tanto “superfície com uma reta qualquer que une dois de seus pontos” e “Conjunto de operações programadas para um determinado fim” (MICHAELIS). Como demonstramos anteriormente, em inglês, esses dois significados são alcançados por meio das palavras *plan* e *plane*, respectivamente.

Nesse sentido, acreditamos que ensinar explicitamente o conceito das colocações, não apenas para que os alunos adquiram consciência deste fenômeno linguístico e busquem produzir discursos mais naturais na LE, elevando, assim, sua competência colocacional na L2, mas também para que possam tomar conhecimento destas fraseologias na LM, sabendo reconhecê-las na L1, por mais frequentemente empregadas no dia a dia que sejam, poderia contribuir para evitar que, no momento da tradução, um olhar menos atento e quizá ingênuo resulte em uma tradução

literal ou em uma tradução por outros usos dos quais o aluno já tenha conhecimento, mas que não compreendem o significado original expresso no TO.

Geralmente, nas aulas de prática de tradução, presente na grade curricular dos alunos do IBILCE, existem discussões sobre trechos de tradução serem “bons” ou “ruins”, sobre “ser assim que eles falam”, em referência a forma que os falantes nativos da L2 se expressam. Discussões como essa indiretamente discorrem, em muitos casos, sobre fraseologias, porém, acreditamos que indicar aos alunos que se trata de uma convencionalidade, além de torná-los mais conscientes sobre as colocações, pode estimulá-los a lançar um olhar mais criterioso e, principalmente, estimulá-los a usar ferramentas mais apropriadas na busca por correspondentes, dentre as quais encontram-se as ferramentas de manuseio de *corpus*.

Já do lado chinês, diferentemente do que ocorreu com o CAT brasileiro, os alunos demonstraram maior dificuldade na tradução de colocações de domínios mais especializados, com colocados ou bases menos frequentes, ou que, de acordo com a classificação de Xu, Lu e Li, pertenciam aos Tipos 1 e 2, do que em colocações formadas por palavras mais frequentes na língua e colocações fracas.

Diferentemente do inglês e português, que são línguas relativamente próximas, quando observadas sob uma perspectiva global, com a existência de muitos cognatos e palavras que lembram umas às outras, o português e o chinês são línguas com estruturas e formas de se expressar muito diferentes, e, por essa razão, os alunos estão mais habituados, desde o início de seus estudos de português, à necessidade de, em muitos casos, realizar adaptações na estrutura semântica de superfície de diferentes segmentos textuais, usando expressões ou palavras que, a partir de um ponto de vista distinto, expressem um significado ou uma ideia da L1 na L2, independentemente de se tratarem de estruturas colocacionais ou não. Isso ajudaria a explicar porque 4,9% das colocações do chinês foram exitosamente traduzidas por apenas uma unidade lexical do português por todos os alunos.

Na China, os alunos têm ao longo dos quatro anos de graduação um curso de português do básico ao avançado, mas de língua geral. Então é esperado que, pelo elevado nível de proficiência, eles não tenham tantos problemas na tradução de colocações mais frequentes na linguagem do dia a dia.

Porém, quando confrontados com colocados menos frequentes ou mais usados em domínios específicos, os alunos demonstraram maior dificuldade. Dentre os motivos que possivelmente levaram a esses erros, há o simples desconhecimento das palavras da colocação e seus significados em L1 e, conseqüentemente na L2, aliados ao pouco refinamento ou domínio insuficiente no manuseio das ferramentas que auxiliam os tradutores em situações como essa. Há também a possibilidade de os alunos reconhecerem uma unidade lexical na L1, mas simplesmente não saberem como se diz na L2, novamente aliada ao pouco refinamento ou domínio insuficiente no manuseio das ferramentas. Segundo os resultados do Questionário Final, todos os alunos usaram algum tipo de dicionário, físico ou digital, além de terem realizado buscas por lexias específicas em buscadores *on-line* e/ou tradutores automáticos. Essas ferramentas são, certamente, de grande utilidade, porém se os alunos não tiverem domínio pleno de seu uso e não analisarem criticamente os resultados obtidos com suas buscas a fim de verificar se elas são realmente usuais na L2, eles podem optar por soluções tradutórias equivocadas, como a mera tradução literal do colocado ou da base, por exemplo.

Outra causa que pode ter resultado em erros colocacionais é a dificuldade em reconhecer as colocações na própria LM. De acordo com os dados do questionário final, apenas 29,4% dos voluntários chineses (cinco alunos) tinham ouvido sobre fraseologias e colocações. Para aprendizes, quando não há o reconhecimento de uma colocação na L1, produzi-la na L2 também se torna potencialmente mais difícil, especialmente se há palavras que os alunos não conheciam ou com as quais tiveram pouco contato ao longo da formação. Diante dessas circunstâncias, é natural que eles recorram à tradução literal, dada às limitações de tempo e à pouca confiança para determinar se os resultados de suas buscas nas ferramentas citadas anteriormente estão corretos ou não. É o que vimos nas colocações com 严重 *yánzhòng* (sério, grave, severo), 有关 *yǒuguān* (ter relação com, relativo a, competente), 超越 *chāoyuè* (exceder, transcender, ultrapassar), entre outros, analisados na subseção 4.2.2 deste trabalho.

Há ocasiões em que a tradução literal para cada um dos componentes da colocação está correta, mas a ocorrência das duas juntas não configura uma colocação existente na L2, mas há ainda há aquelas ocasiões em que a tradução de um dos elementos, seja a base ou o colocado, é incorreta, gerando, conseqüentemente, uma colocação existente na L2. É o caso de erros como “jato econômico”, que não figura entre os exemplos de análise da subseção 4.2.2, mas em nosso

corpus apareceu como solução tradutória para 经济危机 *jīngjìwéijī*, em que 经济 *jīngjì* foi traduzido corretamente como “econômico”, e 危机 *wéijī* foi erroneamente traduzido como “jato”, pois 危机 é “crise”, e a tradução correta para 经济危机 seria “crise econômica”.

Vale ressaltar que muito alunos - 52,9% - usaram uma terceira língua como intermediária: o inglês. Esse fato corrobora a escassez e a pouca disponibilidade de materiais didáticos de português como língua estrangeira pensados especificamente para as necessidades dos estudantes chineses, o que os levou, diante do desconhecimento de determinadas lexias, a realizar a tradução indireta via a língua inglesa. Mesmo que na China o ensino de inglês seja obrigatório do ensino primário até o ensino superior, incluindo a pós-graduação, pela nossa experiência em universidades chinesas, não há como garantir que todos os alunos tenham pleno comando do idioma, ainda que todos os participantes da nossa pesquisa tenham afirmado saber falar inglês. De todo modo, por não fazer parte do nosso escopo de pesquisa, não atestamos a proficiência em inglês dos alunos chineses, mas entendemos que a adição de uma língua a mais no processo tradutório aumenta as chances de erros colocacionais, sobretudo se o conhecimento da língua intermediária também for insuficiente e o tradutor não tiver experiência. De acordo com os resultados do questionário final, apenas quatro (23,5%) dos participantes chineses já tiveram algum tipo de experiência profissional com tradução.

Nesse sentido, esperamos que as dificuldades que os alunos apresentaram na tradução de colocações e as peculiaridades do cruzamento entre português e chinês levantadas neste trabalho possam nortear a elaboração de materiais de ensino de português como língua estrangeira voltados especificamente para estudantes chineses.

Ainda a respeito do ensino de colocações do português para os aprendizes chineses, defendemos a inclusão de aulas que ensinem explicitamente o que são colocações no conteúdo programático, pelos motivos citados anteriormente ao tratarmos do CAT brasileiro, além de aumentar a oferta de oportunidades para que os alunos entrem em contato com tradução e tradução de colocações.

Outra forma de ajudar a preparar os alunos chineses interessados em atuar como tradutores seria a disponibilização de disciplinas eletivas e/ou opcionais de prática de tradução desde o segundo ano da graduação. Apesar dos diversos benefícios do emprego da tradução na

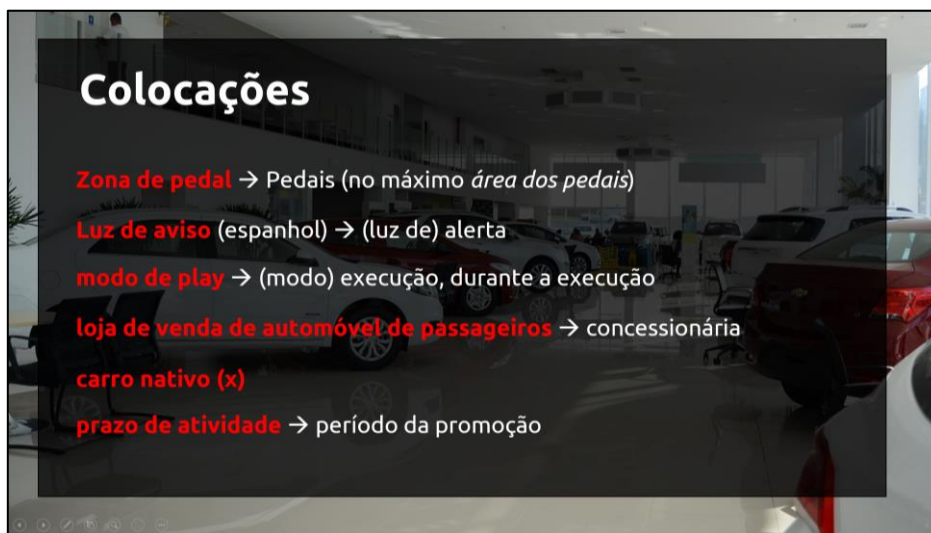
aula de língua, como possibilidade de se trabalhar língua e cultura, gênero discursivo, compreensão de equivalência, entre outros, além de se ter o aluno como autor, estimulando a reflexão sobre o idioma (SALDANHA, LAIÑO, 2020), o curso de língua portuguesa das universidades chinesas participantes da nossa pesquisa são majoritariamente elaborados para que os alunos desempenhem as mais diversas funções na língua estrangeira no futuro, atuando como professores, secretários, representantes comerciais, entre outros. Na China, que se chega ao nosso conhecimento, há apenas um curso de pós-graduação em tradução chinês-português, na Universidade de Macau, antiga colônia portuguesa devolvida ao país asiático em 1999 e com língua oficial portuguesa. Fora isso, os alunos têm prática de tradução e prática de interpretação somente no último ano, logo, é cabível dizer que eles não são expostos suficientemente aos percalços que podem surgir durante a tradução de colocações.

Para aqueles interessados em trabalhar com tradução, essas aulas seriam ideais para o ensino de boas práticas no uso de ferramentas de auxílio ao tradutor, para que os alunos possam refinar melhor suas buscas e adquirir mais autonomia no uso de ferramentas como dicionários, dicionários e buscadores *on-line*, tradutores automáticos, entre outras, incluindo ferramentas pouco exploradas, como os *corpora* linguísticos e ferramentas de manuseio de *corpus*. De acordo com o questionário final, apenas sete alunos, 41,2% dos alunos já haviam ouvido falar de *corpus* linguístico. Como podemos depreender das análises em 4.2.1 e 4.2.2, essas ferramentas são muito úteis no reconhecimento de colocações e podem ser uma ferramenta de tradução valiosa. De acordo com os resultados do Questionário Final, apenas quatro (23,5%) tiveram aulas sobre o uso de ferramentas de tradução, e esses eram possivelmente os que possuem experiência profissional.

Levando-se em consideração que nem todos os alunos da turma têm interesse em se tornarem tradutores, a inclusão de disciplinas obrigatórias de prática de tradução não seria razoável. Porém, haja visto os benefícios da tradução na aula de língua (SALDANHA, LAIÑO, 2020), julgamos plausível a inclusão de atividades de tradução nas aulas de línguas de modo não extensivo, para o benefício da turma como um todo. Acreditamos que atividades tradutórias seriam uma ótima oportunidade para que os alunos pudessem despertar para questões que não podem ser explicadas somente pela gramaticalidade, como é o caso das colocações, abrindo-se espaço para o ensino do que são ao mesmo tempo em que se discute as formas mais naturais de se expressar na L2.

No ano de 2021, foram incluídas no conteúdo programático da disciplina de Português Avançado 2 e aplicadas a alunos do terceiro ano da HUBU três atividades de tradução relacionadas aos tópicos trabalhados em sala de aula. Ao longo do semestre, eles puderam traduzir uma notícia de um portal *on-line*, um manual automotivo e documentos como atestados de antecedentes criminais e históricos escolares. Ao final de cada atividade, após o prazo de entrega já ter se esgotado, foram realizadas uma ou duas aulas para se discutir as traduções dos alunos e os erros mais frequentes. Durante essas aulas, além de questões gramaticais e estilísticas, foram explicadas explicitamente o que são colocações e projetadas as opções dos alunos e as opções existentes na L2, como podemos observar abaixo.

Figura 84 - Projeção de explicação sobre erros colocacionais após atividade tradutória em turma de português avançado



Fonte: elaborado pelo autor.

O *feedback* dos alunos foi muito positivo, tanto que na avaliação da disciplina no final do semestre, o 班长 *bānzhǎng*, algo como o líder e monitor da turma, responsável, dentre outras funções, por levar a opinião da turma acerca das disciplinas ao departamento, avaliou positivamente as atividades de tradução, inserindo-as no rol de pontos positivos do semestre.

Nessas aulas, não foram usados os textos ou os dados do CAT chinês, pois àquela altura estavam sob processamento e análise, mas estamos confiantes de que levaremos o nosso *corpus* e os resultados de nossas investigações para a sala de aula não apenas por meio de atividades de tradução, mas também com atividades sobre as colocações em si, com atenção especial às colocações verbais. Em realidade, acreditamos que nosso *corpus* possa ser usado como fonte para a elaboração de múltiplas atividades que visem o aprimoramento das mais diversas competências linguísticas, dentre as quais a colocacional.

No mais, acreditamos que estender a análise para as outras colocações levantadas no nosso *corpus* de pesquisa enriqueceria ainda mais os resultados desta pesquisa e certamente contribuiria para o contraste entre português e chinês.

Acreditamos que o *corpus* por nós compilado revela-se uma fonte muito rica para vários outros estudos de outras disciplinas linguísticas que não apenas aquelas que nortearam nosso estudo, possibilitando, por exemplo, analisar as estratégias e modalidades de tradução (AUBERT, 1998) ou procedimentos comuns aos tradutores — “universais da tradução” (BAKER, 1993); terminologia e equivalência, além de traços estilísticos, idiossincrasias, diferentes registros e criatividade lexical, com um foco específico no contraste entre português e chinês. Tivemos múltiplos exemplos de como os alunos lançaram mão da criatividade para realizarem suas traduções dentro do tempo estimado, como quando traduziram 新冠病毒 *xīnguān bìngdú* (novo coronavírus, em português) como “novo vírus da coroa”, “vírus da nova coroa”, etc.

Por fim, estamos confiantes de que contribuímos com a pedagogia do léxico fraseológico, da tradução e de áreas correlatas, e esperamos que os resultados aos quais chegamos possam ser usados como subsídio para outras pesquisas nessas áreas e possa contribuir na formação de tradutores e aprendizes de inglês, português e chinês, norteando a elaboração de atividades voltadas ao ensino de colocações. Com a análise dos erros colocacionais cometidos por estes estudantes, que não apenas têm aulas de tradução e interpretação na grade curricular de seus cursos de graduação em português na Universidade de Hubei (HUBU) e no Instituto de Comunicações de Hebei (HEBIC), mas também aspiram atuar como tradutores e intérpretes do português do Brasil, tomamos a importante iniciativa de buscar descobrir e entender quais são os percalços que surgem da interseção entre chinês e português na sala de aula de língua estrangeira e de tradução, assim como na prática tradutória de tradutores novatos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALEXANDRE, N.; PINTO, J. Aspects of relative clauses in Portuguese as a foreign language by Chinese learners. In: **20th Conference of the European Association for Chinese Studies**. 2014. p. 22-26.
- ALUÍSIO, S. M.; ALMEIDA, G. M. B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. **Calidoscópico**, v. 4, n. 3, p. 156-178, 2006.
- AZEVEDO, D. J. O.; SILVA, F. M. Colocações, estereótipos e clichês: definições e diferenças. In: **ReVEL**. Novo Hamburgo, v. 15, n. 29, p. 37-52, 2017. Disponível em: <http://www.revel.inf.br/files/b4f9129fed69407b4ec7c6e747aa1c5f.pdf>. Acesso em: 01 abr. 2020.
- BAKER, M. Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, H. (ed.) **Terminology, LSP & Translation**. Philadelphia/Amsterdam: John Benjamins, 1996. p. 175-186.
- 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.
- BAKER, P.; HARDIE, A; MCENERY, T.; XIAO, R.; BONTCHEVA, K.; CUNNINGHAM, H.; GAIZAUSKAS, R.; HAMZA, O.; MAYNARD, D.; TABLAN, V.; URSU, C.; JAYARAM, B. D.; LEISHER, M. Corpus linguistics and South Asian languages: corpus creation and tool development'. In: **Literary and Linguistic Computing**. 2004. p. 509–524.
- BARROS, L. A. O Texto Técnico, Científico e Especializado. In: BARROS, L. A. **Curso Básico de Terminologia**. São Paulo: Editora da Universidade de São Paulo, 2004. p. 44.
- BERBER SARDINHA, T. Metaphor in corpora: A corpus-driven analysis of Applied Linguistics dissertations. **International Conference on Metaphor in Language and Thought**, PUC-SP, 2002.
- BERNARDINI, S. Corpora for translator education and translation practice: Achievements and

challenges. In: **Third International Workshop on Language Resources for Translation Work, Research & Training**. 2006. p. 17-22.

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

BIDERMAN, M. T. C. **Teoria linguística: teoria lexical e linguística computacional**. 2. ed. São Paulo: Martins Fontes, 2001a.

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

BISKUP, D. “L1 influence on learners’ renderings of English collocations. A Polish/German empirical study.” In: ARNAUD, P.; BÉJOINT, H. (ed.), **Vocabulary and Applied Linguistics**. London: Macmillan, 1992.

BOWKER, L. Corpus linguistics: it’s not just for linguists! Disponível em: <https://allaboutcorpora.com/corpus-linguistics-not-just-linguists>. Acesso em: 20 jul. 2020.

BOWKER, L. Corpus linguistics is not just for linguists: Considering the potential of computer-based corpus methods for library and information science research. **Library Hi Tech**. v. 3, 2 ed. p. 358-371, 2018.

BOWKER, L. Towards a collaborative approach to *corpus* building in the translation classroom. In. BAER, B. J.; KOBAYASHI, G. S. **Beyond the Ivory Tower: Rethinking translation pedagogy**. Amsterdam: John Benjamins Publishing Company, 2003. p. 193-210.

BOWKER, L.; PEARSON, J. **Working with specialized language: a practical guide to using corpora**. London: Routledge, 2002.

BRASIL. Documentos assinados durante a visita oficial do Primeiro-Ministro da República Popular da China, Li Keqiang, ao Brasil – Brasília e Rio de Janeiro, 19 de maio de 2015. Brasília: Ministério das Relações Exteriores. Disponível em: <http://www.itamaraty.gov.br/pt-BR/component/content/article?id=9687:visita->. Acesso em: 05 jul. 2020.

BR-ICLE Brazilian Sub-*Corpus* of The International *Corpus* of Learner English (Icle) (Fase 2). São Paulo: Pesquisas PUC-SP – Seleta. Disponível em: <https://www.pucsp.br/pesquisa-seleta-2011/projetos/052.php>. Acesso em: 07 jul. 2020.

CAMBRIDGE DICTIONARY ONLINE. Cambridge: Cambridge University Press, 2020. Disponível em: <http://dictionary.cambridge.org>. Acesso em: 15 abr. 2021.

CASELI, H. M. O uso de corpora paralelos para a criação de um tradutor automático estatístico. In: VIANA, V.; TAGNIN, S. E. O. **Corpora na Tradução**. São Paulo: HUB Editorial, 2015. p. 243-267.

CENTER FOR LANGUAGE EDUCATION AND COOPERATION. Hanban, 2020. Guanyu Kong Zi Xueyuan. Disponível em: http://www.hanban.org/confuciousinstitutes/node_10961.htm. Acesso em: 03 jul. 2020.

CHINA é maior parceiro comercial do Brasil no mundo. Planalto – Presidência da República, 12 nov. 2019. Disponível em: <https://www.gov.br/planalto/pt-br/acompanhe-o-planalto/noticias/2019/11/china-e-maior-parceiro-comercial-do-brasil-no-mundo#:~:text=A%20China%20%C3%A9%20o%20maior,US%24%2029%2C2%20bilh%C3%B5es>. Acesso em: 03 jul. 2020.

CHINA. Plano de Ação Conjunta entre o Governo da República Federativa do Brasil e o Governo da República Popular da China, 2010-2014. Brasília: Embaixada da República Popular da China no Brasil, 2010. Disponível em: <http://br.china-embassy.org/por/zbngx/t682404.htm>. Acesso em: 05 jul. 2020.

COUNCIL OF EUROPE. **Common European Framework of Reference for Languages: learning, teaching, assessment**. Estrasburgo: Cambridge University Press, 2001. p.22-25. Disponível em <https://rm.coe.int/1680459f97>. Acesso em: 29 out. 2021.

DUTRA, D. P.; SILERO, R. P. Descobertas linguísticas para pesquisadores e aprendizes: a Linguística de Corpus e o ensino de gramática. In: **Revista Brasileira de Linguística Aplicada**, v. 10, n. 4, p. 909-930, 2010. DOI: <https://doi.org/10.1590/S1984-63982010000400005>. Disponível em: https://www.scielo.br/scielo.php?pid=S1984-63982010000400005&script=sci_arttext&tlng=pt#tx1. Acesso em: 03 jul. 2020.

FABER, D.; LAURIDSEN, K. The compilation of a Danish–English–French corpus in contract law. In: JOHANSSON, S.; STENSTRÖM, A. (eds.) **English Computer Corpora: Selected Papers and Research Guide**. Berlin: Mouton de Gruyter, 1991. p. 235–243.

FRANKENBERG-GARCIA, A. Investigating the collocations available to EAP writers. *Journal of English for Academic Purposes*, Scotland, v. 35, p. 93-104, 2018.

GLAAD. **GLAAD Media Reference Guide – 10th Edition**. 2016. Disponível em: <https://www.glaad.org/sites/default/files/GLAAD-Media-Reference-Guide-Tenth-Edition.pdf>. Acesso em 30 jan. 2022.

GOUADEC, D. Training Translators: Certainties, Uncertainties, Dilemmas. In: MAIA, B.; HALLER, J.; ULRICH, M. (eds.) **Training the Language Services Provider for the New Millennium**, Porto: Universidade do Porto, 2002. p. 31-41.

GRANGER, S; LEFER, M.-A. MUST: A collaborative *corpus* collection initiative for translation teaching and research. In: GRANGER, S., LEFER, M.-A., & PENHA-MARION, L. (eds). **Book of Abstracts. Using Corpora in Contrastive and Translation Studies Conference (5th edition)**. CECL Papers 1. Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain, ed. 5, p. 72-73, 2018. Disponível em <https://uclouvain.be/en/research-institutes/ilc/cecl/cecl-papers.html>. Acesso em 20 jun. 2020.

GROSSO, M. J., ZHANG, J. **A Promoção do Português em Macau e no Interior da China**. Macau: Universidade de Macau e Lidel, 2018.

HAUSMANN, F. J. Kollokationen in deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels'. In: BERGENHOLTZ, H.; MUGDAN, J. (Org.). **Lexikographie und Grammatik**. Tübingen: Niemeyer, 1985. p. 22-26.

HOFSTEDE, G. Dimensionalizing Cultures: The Hofstede Model in Context. In: IACCP. **Online Readings in Psychology and Culture: unit 2**. 2011. Disponível em: http://mchmielecki.pbworks.com/w/file/fetch/64591689/hofstede_dobre.pdf. Acesso em 20 jan. 2022.

HUANG, C.-R.; CHEN, K.-J; YANG, Y.-Y. Character-based collocation for Mandarin Chinese. In: **COLING 1994 Volume 1: The 15th International Conference on Computational Linguistics**. 1994.

HUANG, C.-R.; KILGARRIFF, A.; WU, Y.; CHIU, C.-M. Chinese Sketch Engine and the extraction of grammatical collocations. In: **Proceedings of the fourth SIGHAN workshop on Chinese language processing**. 2005.

HUBEI Daxue jiangyu 2016 nian kaishe putaoyayu zhuan. Yangguang Gaokao, 2015. Disponível em: <https://gaokao.chsi.com.cn/gkxx/yxzs/201511/20151130/1511792408.html>. Acesso em: 03 jul. 2020.

INSIGHTS. Hofstede Insights, 2022. Country Comparison – China. Disponível em: <https://www.hofstede-insights.com/country-comparison/china/>. Acesso em 30 jan. 2022.

INSTITUTO CONFÚCIO. Instituto Confúcio na UNESP, 2020. Sobre o Instituto Confúcio na UNESP. Disponível em: <https://institutoconfucio.com.br/sobre/#quem-somos>. Acesso em: 03 jul. 2020.

INTRODUCTORY Guide to the Common European Framework of Reference (CEFR) for English Language Teachers. Cambridge University Press 2013. Disponível em: <http://www.englishprofile.org/images/pdf/GuideToCEFR.pdf>. Acesso em: 29 out. 2021.

JIANG, L. **HSK Standard Course 5A**. Pequim: Beijing Language and Culture University Press, 2015.

JOHANSSON, S.; HOFLAND, K. Towards an English–Norwegian parallel corpus. In: FRIES; U.; TOTTIE; G.; SCHNEIDER, P. (eds.) **Creating and Using English Language Corpora**, Amsterdam: Rodopi, 1994. p. 25–37.

KILGARRIFF, A.; BAISA, V.; BUŠTA, J.; JAKUBÍČEK, M.; KOVÁŘ, V.; MICHELFEIT, J.; RYCHLÝ, P.; SUCHOMEL, V. The Sketch Engine: ten years on. In: **Lexicography**, ed. 1, 'p. 7-36, 2014. Disponível em: <http://www.sketchengine.eu>. Acesso em: 24 abr. 2020

KILGARRIFF, A.; KENG, N.; SMITH, S. Learning Chinese with the Sketch Engine. In: SMITH, S. **Corpus Linguistics in Chinese Contexts**. Springer, 2015.

KURNIAWATI, L. A.; RAHAJENG, D. T.; KRISTIANTO, B.; KASTUHANDANI, F. C. Introducing sdl trados to beginning translators. In: **Indonesian Journal of English Language Studies (IJELS)**, v. 2, n. 1, p. 15, 2016.

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

LAVIOSA, S. PAGANO, A.; KEMPPANEN, H.; JI, M. **Textual and contextual analysis in empirical translation studies**. Cingapura: 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.

LEXICAL COMPUTING CZ S.R.O. Sketch Engine, 2021. ptTenTen: Corpus of the Portuguese Web. Disponível em: <https://www.sketchengine.eu/pttnten-portuguese-corpus/>. Acesso em: 31 abr. 2021.

LEXICAL COMPUTING CZ S.R.O. Sketch Engine, 2021. T-score. Disponível em: https://www.sketchengine.eu/my_keywords/t-score/. Acesso em: 18 abr. 2021.

LEXICAL COMPUTING CZ S.R.O. Sketch Engine, 2021. LogDice. Disponível em: https://www.sketchengine.eu/my_keywords/logdice/. Acesso em: 18 abr. 2021.

LEXICAL COMPUTING CZ S.R.O. Sketch Engine, 2021. MI Score. Disponível em: https://www.sketchengine.eu/my_keywords/mi-score/. Acesso em: 18 abr. 2021.

LEXICAL COMPUTING CZ S.R.O. Sketch Engine, 2021. Word Sketch grammar. Disponível em: https://www.sketchengine.eu/my_keywords/word-sketch-grammar/. Acesso em: 18 abr. 2021.

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

LIU, X. **New Practical Chinese Reader**. Pequim: Beijing Language and Culture University

Press, v. 2, 2. ed., 2017.

LIU, Z.; WANG, H.; WU, H. et al. Collocation extraction using monolingual word alignment method. In: **Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing**. 2009. p. 487-495.

LÜ, Y.; ZHOU, M. Collocation translation acquisition using monolingual corpora. In: **Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)**. 2004. p. 167-174.

MARTELLI, A. **Lexical Collocations in Learner English: a Corpus-Based Approach**. Alessandria: Edizioni dell'Orso, 2007.

MCENERY, T.; HARDIE, A. **Corpus Linguistics**. Cambridge: Cambridge University Press, 2012.

MCENERY, T.; OAKES, M. Sentence and word alignment in the Crater project: methods and assessment. In: WARWICK-ARMSTRONG, S. (ed.) **Proceedings of the Association for Computational Linguistics Workshop SIG-DAT Workshop**. Dublin: ACL, 1995. p. 104–116.

MCENERY, T.; TANAKA, I.; BOTLEY, S. P. 1997. Corpus annotation and reference resolution. In: MITKOV, R.; BOGURAEV, B. (eds.) **Proceedings of the Association for Computational Linguistics Workshop on Anaphora Resolution for Unrestricted Texts**. Madri, 1997. p. 67–74.

MCENERY, T.; XIAO, R. Z.; MO, L. Aspect marking in English and Chinese: using the Lancaster Corpus of Mandarin Chinese for contrastive language study. In: **Literary and Linguistic Computing**. 2003. p. 361–78.

MCKEOWN, K. R.; RADEV, D. R. Collocations. In: **Handbook of Natural Language Processing**. Marcel Dekker, 2000. p. 1-23.

MENDES, A. ANTUNES, S.; GONÇALVES, A.; JANSSEN, M. The COPLE2 corpus: a learner corpus for Portuguese. In: **Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)**. 2016. p. 3207-3214.

MENDES, A.; ANTUNES, S.; GONÇALVES, A.; JANSSEN, M.; ALEXANDRE, N.; AVELAR, A.; CASTELO, A.; DUARTE, I.; FREITAS, M. J.; PASCOAL, J.; PINTO, J. Corpus de Português Língua Estrangeira/Língua Segunda–COPLE2. In: **Livro de Resumos–XXX Encontro Nacional da Associação Portuguesa de Linguística**, 22-24 de outubro, 2014.

MICHAELIS moderno dicionário da língua portuguesa. São Paulo: Melhoramentos. Disponível em: <http://michaelis.uol.com.br/moderno/portugues/index.php>. Acesso em: 15 out. 2021.

OLIVEIRA, J. T. **Um corpus de aprendizes de tradução: compilação, tratamento e aplicação ao ensino por meio de planilha eletrônica**. São Paulo, 2019. 259f. Tese (Doutorado em Estudos Linguísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

ORENHA-OTTAIANO, A. The creation of an Online English Collocations Platform to help develop collocational competence. In: **PHRASIS. RIVISTA DI STUDI FRASEOLOGICI E PAREMIOLOGICI**, v. 1, p. 59-81, 2020.

ORENHA-OTTAIANO, A. Escollas colocacionais a partir dun corpus de estudantes de tradución e a importancia do desenvolvemento da competencia colocacional. **Cadernos de Fraseoloxía Galega**, v. 21, p. 35-64, 2019.

ORENHA-OTTAIANO, A. The compilation of an online *Corpus*-Based bilingual Collocations Dictionary: motivations, obstacles and achievements. In: **Proceedings of eLex 2017–Electronic lexicography in the 21st century: Lexicography from Scratch**. 2017. p. 458-473.

ORENHA-OTTAIANO, A. Collocations workbook: a corpus-based online pedagogical support material for the teaching of English collocations. In: **Revista de Estudos da Linguagem**. Belo Horizonte: Universidade Federal Minas Gerais, Faculdade de Letras, v. 23, n. 3, p. 833-881, 2015. Disponível em: <http://hdl.handle.net/11449/158743>.

ORENHA-OTTAIANO, A. English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. In: **Acta Scientiarum language and culture**, 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. **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) – IBILCE, Universidade Estadual Paulista, São José do Rio Preto, 2009.

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

ORENHA-OTTAIANO, A. O uso de *corpora* e o ensino de ESP. **Estudos linguísticos**, v. 33, p. 1036-1041, 2004b.

ORENHA-OTTAIANO, A.; GARCIA-GONZALEZ, M.; SILVA, M. E. O. O. ; L'HOMME, M. ; ALONSO RAMOS, M. ; VALENCIO, C. R. ; TENORIO, W. Corpus-based Methodology for an Online Multilingual Collocations Dictionary: First Steps. In: Elex 2021 - Electronic lexicography in the 21st century, 2021, University of Ljubljana. **Proceedings of Elex2021**, 2021.

PACHECO, A. **A Aquisição de Morfemas em Inglês como L2: Uma análise dos padrões evolutivos através do BELC (Brazilian English Learner Corpus)**. Orientador: Sérgio de Moura Menuzzi. 188 f. 2010. Tese (Doutorado em Letras) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

PEARSON, J. Using parallel texts in the translator training environment. In: ZANETTIN, F.; BERNARDINI, S.; STEWART, D. **Corpora in Translator Education**. Manchester: St Jerome Publishing, 2003. p.15-24.

PENG, W. Confucius Institutes in Brazil and BRICS Education Cooperation. The Diplomat, 9 jan. 2020. Disponível em: <https://thediplomat.com/2020/01/confucius-institutes-in-brazil-and-brics-education-cooperation/>. Acesso em: 03 jul. 2020.

PROJETO. São Paulo: *Corpus Multilíngue para Ensino e Tradução – COMET*. Disponível em: <http://comet.fflch.usp.br/projeto>. Acesso em: 07 jul. 2020.

PUTAOYAYU zhuan ye jieshao. China: Gaokao wang: Zhuanye sousuo, 2020. Disponível em: <http://college.gaokao.com/speciality/653/>. Acesso em: 03 jul. 2020.

RACINE, J. P. Lexical Approach. In: LIONTAS, J. L.; DELLICARPINI, M. (eds.) **The TESOL Encyclopedia of English Language Teaching**. Hoboken: Wiley-Blackwell, v. 2, 2018.

ROCHA, J. M. P. **Tradução de fraseologismos metafóricos do português para o inglês: um estudo de corpus de aprendizes brasileiros**. 2020. 258 f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2020.

ROCHA, J. M. P.; VIANA, V.; ORENHA-OTTAIANO, A. Tradução de fraseologismos metafóricos: contribuições teórico-metodológicas da linguística de corpus. In: **Revista da ABRALIN**, v. 19, n. 1, p. 1-26, 16 nov. 2020. Disponível em: doi 10.25189/rabralin.v19i1.1697. Acesso em: 03 fev. 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.

SALDANHA, C.T.; LAIÑO, M.J. Ensino de língua espanhola: prática de tradução sob o viés funcionalista. In: **E por falar em tradução**. Departamento de Linguística Aplicada da Universidade Estadual de Campinas, 7-11 dez. 2020.

SINCLAIR, J.M. *Corpus and Text – Basic and Principles*. In: WYNNE, M. (ed.). **Developing linguistic corpora: A guide to good practice**. Oxford: Oxbow Books, 2005. p. 1-16.

SINCLAIR, J.M. **Corpus Concordance Collocation**. Oxford University Press, 1991.

SINCLAIR, J. M. Collocations: a progress report. In: STEELE, R.; THREADGOLD, T. (eds.) **Language Topics: Essays in Honour of Michael Halliday**. Amsterdam: John Benjamins, p. 319-332, v. 2, 1987.

SOUZA, G. A. *Gay Language* e as colocações no seriado *Queer as Folk* à luz da linguística de *corpus* e da fraseologia. Separata de: **Miguilim - Revista Eletrônica do Netlli**, Crato. v. 7, n.1, p.05-25, jan.-abr. 2018.

TAGNIN, S. E. O. A multilingual learner corpus in Brazil. In: Corpus Linguistics 2003, 2003, Lancaster. **Proceedings of the Corpus Linguistics 2003 UCREL Technical Papers**, Special Issue. Lancaster: Lancaster University, 2003. v. 1. p. 940-945. <https://comet.fflch.usp.br/sites/comet.fflch.usp.br/files/u30/A%20multilingual%20learner%20corpus%20in%20Brazil.pdf>. Acesso em: 29 mai. 2020a.

TAGNIN, S. E. O. A multilingual learner *corpus* in Brazil. In: **Language and Computers**. 2006. p. 195-202. Disponível em: <https://www.researchgate.net/publication/279744152>. Acesso em: 24 abr. 2020b.

TAGNIN, S. E. O. E a Linguística de Corpus vai desbravando novos horizontes...In: FINATTO, M. J. B. *et al. Linguística de corpus: perspectivas*. Porto Alegre: Instituto de Letras – UF, 2018. p. 11-18.

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

TAGNIN, S. E. O.; TEIXEIRA, E.D. Lingüística de *Corpus* e Tradução Técnica - relato da montagem de um *corpus* multivarietal de culinária. **Tradterm**, v. 10, p. 313-358, 2004a.

TAGNIN, S. E. O. Um *corpus* multilíngüe para ensino e tradução-o COMET: da construção à exploração. **Tradterm**, v. 10, p. 117-141, 2004b.

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

TAGNIN, S. E. O. **Convencionalidade e produção de texto: um dicionário de colocações verbais inglês/português; português/inglês**. São Paulo, 1998. Tese (Livre-docência), FFLCH-USP.

TOGNINI-BONELLI, E. **Corpus linguistics at work**. Amsterdam/Philadelphia: John Benjamins, 2001.

VASQUEZ-AYORA, G. **Introducción a la traductología: curso básico de traducción**. Washington Georgetown University Press, 1977.

VIANA, V.; TAGNIN, S. E. O. (Org.). **Corpora na tradução**. São Paulo: Hub Editorial, 2015.

XU, J. *Corpus-based Chinese Studies: A historical review from the 1920s to the present*. In: **Chinese Language and Discourse**. v. 6, 2ª edição. 2015. p. 218-244.

XU, J; CUI, N.; MU, Y. **Developing Chinese (2nd Edition)**: Intermediate Comprehensive Course I. 2. ed. Pequim: Beijing Language and Culture University Press, 2011.

XU, R.; LU, Q.; LI, S. The Design and Construction of a Chinese Collocation Bank. In: **LREC**. 2006. p. 1880-1885.

YUQI, S. **A produção de Hedges por falantes brasileiros de português e aprendizes chineses de PLA**. Orientadora: Cristina Becker Lopes Perna. 155 f. 2011. Dissertação (Mestrado em Letras) - Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2011.

ZANETTIN, F. Translation and corpus design. **SYNAPS – A Journal of professional communication**. 2011. p. 14-23.

ZANETTIN, F. Swimming in Words. In: ASTON, Guy.(ed.) **Learning with Corpora**. Houston: Athelstan, 2001. p. 177-197

ZHANG, J. Construção de um Corpus de Português Constituído por Produções Escritas de Aprendentes Chineses. In: **Fórum de Ensino e Estudos Portugueses**, Guangdong: Departamento de Português do Instituto Xinhua da Universidade Sun Yat-Sem, 2 -4 Set. 2019.

ZHONGGUO tigong 21 ge putaoyayu benke xuewei. Pingtai, 26 mai. 2014. Disponível em: <https://www.plataformamedia.com/2014/05/26/中国提供 21 个葡萄牙语本科学位/?lang=zh-hans>. Acesso em: 03 jul. 2020.

APÊNDICES

Apêndice A— Perguntas e resultados do questionário final sobre o perfil dos alunos na plataforma

腾讯调查问卷

第 1 页							
1. Você é aluno(a) de qual universidade?							
选项	百分比%	小计					
Universidade de Hubei	64.70%	11					
Instituto de Comunicações de Hebei	35.30%	6					
有效填写量		17					
2. Você é aluno(a) de qual semestre do curso de graduação?							
选项	百分比%	小计					
Primeiro semestre do terceiro ano	0.00%	0					
Segundo semestre do terceiro	76.50%	13					
Primeiro semestre do quarto ano	0.00%	0					
Segundo semestre do quarto ano	5.90%	1					
Já me formei	17.60%	3					
有效填写量		17					
3. Quantos anos você tem?							
选项	百分比%	小计					
Menos de 20	0.00%	0					
21	58.80%	10					
22	29.40%	5					
23 ou mais	11.80%	2					
Em caso de menos de 20 ou 23 ou mais	especifique:	0.00%	0				
有效填写量		17					
4. Você se considera como pertencente ao sexo							
选项	百分比%	小计					
Masculino	23.50%	4					
Feminino	76.50%	13					
Não-binário	0.00%	0					
Prefiro não responder	0.00%	0					
有效填写量		17					
第 2 页							

5. Qual o seu nível de proficiência em português?							
选项	百分比%	小计					
Fluente/nativo	0.00%	0					
Avançado	5.90%	1					
Intermediário avançado	58.80%	10					
Intermediário	23.50%	4					
Básico a intermediário	11.80%	2					
Básico	0.00%	0					
有效填写量		17					
6. Qual o seu nível de proficiência em chinês - mandarim?							
选项	百分比%	小计					
Fluente/ nativo	94.10%	16					
Avançado	5.90%	1					
Intermediário avançado	0.00%	0					
Intermediário	0.00%	0					
Básico a intermediário	0.00%	0					
Básico	0.00%	0					
有效填写量		17					
7. Qual a sua língua nativa?							
选项	百分比%	小计					
Chinês - mandarim	94.10%	16					
Cantonês	5.90%	1					
Outra língua e/ou dialeto da China_____	0.00%	0					
有效填写量		17					
8. Você já morou em algum país de língua portuguesa?							
选项	百分比%	小计					
Sim	76.50%	13					
Não	23.50%	4					
有效填写量		17					
9. Caso a resposta anterior seja "sim"		por quanto tempo?					
选项	百分比%	小计					
Período inferior a um semestre (seis meses)	21.40%	3					
De seis meses a um ano	64.30%	9					
De um a três anos	14.30%	2					
De três a cinco anos	0.00%	0					
Mais de cinco anos	0.00%	0					

有效填写量		14					
10. Caso a sua resposta para a pergunta 8 seja "sim"	em qual(is) país(es)?						
选项	百分比%	小计					
Brasil	64.30%	9					
Angola	0.00%	0					
Moçambique	0.00%	0					
Portugal	35.70%	5					
Guiné-Bissau	0.00%	0					
Timor-Leste	0.00%	0					
Cabo Verde	0.00%	0					
São Tomé e Príncipe	0.00%	0					
Territórios de outras nações (Macau	Goa)	14.30 %	2				
有效填写量		14					
11. Caso a sua resposta para a pergunta 8 seja "sim"	qual foi o propósito da sua viagem?						
选项	百分比%	小计					
Turismo	7.10%	1					
Intercâmbio/ estudos	92.90%	13					
Visita a parentes	0.00%	0					
Negócios	0.00%	0					
Outro ____	0.00%	0					
有效填写量		14					
12. Além do português e do chinês - mandarim	você fala alguma outra língua?						
选项	百分比%	小计					
Sim	94.10%	16					
Não	5.90%	1					
有效填写量		17					
13. Caso a sua resposta para a pergunta anterior seja	qual(is) língua(s)?						
选项	百分比%	小计					
Inglês	100.00%	16					
Francês	6.20%	1					
Espanhol	6.20%	1					
Árabe	0.00%	0					
Russo	0.00%	0					
Italiano	6.20%	1					
Coreano	0.00%	0					

Japonês	6.20%	1					
Alemão	0.00%	0					
Línguas e/ou dialetos chineses	18.80%	3					
Outra(s)_____	0.00%	0					
有效填写量		16					
第 3 页							
14. Você possui alguma experiência profissional?							
选项	百分比%	小计					
Sim	23.50%	4					
Não	76.50%	13					
有效填写量		17					
15. Caso sua resposta anterior seja "sim"			especifique em que área você atuou:				
选项	百分比%	小计					
Tradução e/ou revisão	75.00%	3					
Ensino de idiomas	0.00%	0					
Consultoria linguística	0.00%	0					
Escrita criativa e/ou copywriting	0.00%	0					
Dublagem	legendagem e/ou transcrição linguística	25.00 %	1				
Outra_____	0.00%	0					
有效填写量		4					
16. Caso a sua resposta para a questão 15 seja "sim"			por quanto tempo?				
选项	百分比%	小计					
Até um semestre (seis meses)	50.00%	2					
De seis meses a um ano	25.00%	1					
De um a três anos	25.00%	1					
De três a cinco anos	0.00%	0					
Mais de cinco anos	0.00%	0					
有效填写量		4					
第 4 页							
17. Durante o andamento do projeto			de onde você realizou suas traduções?				
选项	百分比%	小计					
China ou um de seus territórios	52.90%	9					
Um país de língua portuguesa	47.10%	8					
De um país estrangeiro cuja língua	0.00%	0					

oficial não é o português ou chinês							
有效填写量		17					
18. Durante as traduções	você usou alguma ferramenta de memórias de tradução?						
选项	百分比%	小计					
Sim	52.90%	9					
Não	47.10%	8					
有效填写量		17					
19. Caso a sua resposta anterior seja "sim"	especifique qual(is)						
选项	百分比%	小计					
WordFast	11.10%	1					
MemoQ	0.00%	0					
Trados	22.20%	2					
Outro____	66.70%	6					
有效填写量		9					
20. Durante as traduções	você utilizou algum tradutor automático?						
选项	百分比%	小计					
Sim	88.20%	15					
Não	11.80%	2					
有效填写量		17					
21. Caso a sua resposta anterior seja "sim"	qual(is)?						
选项	百分比%	小计					
Google Translate	93.30%	14					
Baidu Translate	13.30%	2					
Bing Microsoft Translator	6.70%	1					
Outro____	0.00%	0					
有效填写量		15					
22. Quais outras ferramentas você utilizou para traduzir?							
选项	百分比%	小计					
Buscadores online (Google	Baidu	Sogou	entre outros)	70.60 %	12		
Dicionários eletrônicos	70.60%	12					
Dicionários físicos (de papel)	11.80%	2					
Corpora linguísticos	5.90%	1					
Outro____	5.90%	1					
有效填写量		17					

23. Em algum momento		você usou uma língua como intermediária para realizar as suas traduções?					
选项	百分比%	小计					
Sim	52.90%	9					
Não	47.10%	8					
有效填写量		17					
24. Caso a sua resposta anterior seja "sim"		qual(is)?					
编号	文本答案	提交时间					
17	Inglês						
16	inglês						
13	Inglês						
10	Inglês						
9	Inglês						
8	inglês						
4	inglês						
2	inglês						
1	Inglês						
备注：更多详情请导出原始数据查看							
第 5 页							
25. Você já teve aula sobre teorias da tradução?							
选项	百分比%	小计					
Sim	41.20%	7					
Não	58.80%	10					
有效填写量		17					
26. Você já teve aula sobre ferramentas de tradução?							
选项	百分比%	小计					
Sim	23.50%	4					
Não	76.50%	13					
有效填写量		17					
27. Você já ouviu falar sobre linguística de corpus?							
选项	百分比%	小计					
Sim	41.20%	7					
Não	58.80%	10					
有效填写量		17					

28. Você já ouviu falar sobre colocações e sua importância?							
选项	百分比%	小计					
Sim	29.40%	5					
Não	70.60%	12					
有效填写量		17					

ANEXOS

Anexo A– Termo de Consentimento Livre e Esclarecido



Termo de Consentimento Livre e Esclarecido - TCLE

(Conselho Nacional de Saúde, Resolução 466/2012/Resolução 510/2016)

Você está sendo convidado a participar como voluntário dos projetos de pesquisa *CAT (Corpus de Aprendizes de Tradução)* e *Multilingual Student Translation Corpus (MUST)*, sob responsabilidade da pesquisadora Adriane Orenha-Ottaiano, da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto, 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", também sob coordenação da Profa. Dra. Adriane Orenha-Ottaiano. O estudo será realizado com um grupo experimental de alunos do curso de Letras – Português da Universidade de Hubei e do Curso de Letras – Português do Instituto de Comunicações de Hebei e terá a colaboração do mestrando Hernandes dos Santos Miguel, membro discente do Projeto de Pesquisa acima mencionado. O objetivo é analisar o comportamento linguístico das colocações e outras unidades fraseológicas 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: _____

ID: () Passaporte () Cartão de identificação chinês (身份证) _____

Endereço: _____

_____, Contato: _____

_____, ____ de _____ de 2020

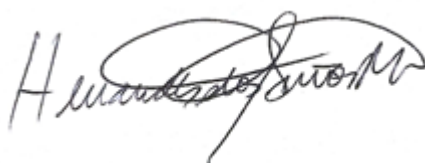
OBS.: Termo apresenta duas vias, uma destinada ao participante e a outra ao pesquisador.

Nome Pesquisador(a): Adriane Orenha Ottaiano	Cargo/Função: Professora Assistente Doutora
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) 32212518 ou 981064849	
Projeto submetido ao Comitê de Ética em Pesquisa do IBILCE/UNES Rua Cristóvão Colombo, 2265. Bairro: Jardim Nazareth. São José do Rio Preto/SP – Fone 17-3221.2480 e 3221.2545	

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, 10 / 04 / 2022



Assinatura do autor