

Do Sul para o Mundo

pensando a tradução no contexto pós-pandemia

Seleção de comunicações apresentadas no ENTRAD 2022

Cristiane Krause Kilian
Monique Pfau
Vinícius Martins Flores
Orgs.



Editora Fundação Fênix



**Cristiane Krause Kilian
Monique Pfau
Vinícius Martins Flores
Organização**

**Do Sul para o Mundo:
pensando a tradução no contexto pós-pandemia**

Seleção de comunicações apresentadas no ENTRAD 2022



Editora Fundação Fênix

Porto Alegre, 2024

Direção editorial: Agemir Bavaresco
Diagramação: Editora Fundação Fênix
Capa: Editora Fundação Fênix

O padrão ortográfico, o sistema de citações, as referências bibliográficas, o conteúdo e a revisão de cada capítulo são de inteira responsabilidade de seu respectivo autor.

Todas as obras publicadas pela Editora Fundação Fênix estão sob os direitos da Creative Commons 4.0 –
http://creativecommons.org/licenses/by/4.0/deed.pt_BR

Este livro foi editado com o apoio financeiro do Ministério das Relações Exteriores da República Federal da Alemanha através do Serviço Alemão de Intercâmbio Acadêmico (DAAD).



Série Humanidades e Interdisciplinaridade – 36

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Do Sul para o mundo [livro eletrônico] : pensando a tradução no contexto pós-pandemia : seleção de comunicações apresentadas no ENTRAD 2022 / organizadores Cristiane Krause Kilian, Monique Pfau, Vinicius Martins Flores. --
Porto Alegre, RS : Editora Fundação Fênix, 2024. -- (Série humanidades e interdisciplinaridade ; 36)
PDF

Vários autores.
Bibliografia.
ISBN 978-65-5460-172-6

1. Línguas e linguagem 2. Pandemia - Aspectos sociais 3. Tradução 4. Tradução e interpretação
I. Kilian, Cristiane Krause. II. Pfau, Monique.
III. Flores, Vinicius Martins. IV. Série.

24-225348

CDD-418.02

Índices para catálogo sistemático:

1. Tradução : Linguística 418.02

Eliane de Freitas Leite - Bibliotecária - CRB 8/8415

DOI – <https://doi.org/10.36592/9786554601726>

28. IMPACTOS SOCIAIS DA COVID-19: CONSTRUÇÃO DE UM GLOSSÁRIO BILÍNGUE¹



<https://doi.org/10.36592/9786554601726-28>

Ana Karina Borges Braun²

Cleci Regina Bevilacqua³

Patrícia Ramos Chitoni Reuillard⁴

1. Introdução

Este trabalho apresenta a construção de *corpora* em português-inglês para fins de elaboração de um glossário terminológico relativo à Covid-19 e a áreas inter-relacionadas das Ciências Humanas, Ciências Sociais e Letras, bem como os resultados iniciais da coleta de termos⁵.

O objetivo geral desta pesquisa é elaborar um glossário terminológico português-inglês relativo aos impactos sociais da Covid-19, do ponto de vista antropológico, a ser disponibilizado no site da Rede Covid-19 Humanidades MCTI⁶ para acesso livre e gratuito aos estudiosos das áreas envolvidas e demais interessados no tema. Seu objetivo específico, por sua vez, é a construção de *corpora* em português e inglês para extração dos termos que

¹ Este trabalho foi desenvolvido no âmbito do projeto “A Covid-19 no Brasil 2: análise e resposta aos impactos sociais da imunização, tratamento, práticas e ambientes de cuidado e recuperação de afetados” (Convênio FINEP/UFRGS 1212/21), da Rede Covid-19 Humanidades MCTI. A iniciativa integra o conjunto de ações da Rede Vírus MCTI, financiada pelo Ministério da Ciência, Tecnologia e Inovações (MCTI), Brasil.

² Pós-doutoranda em Letras pela Universidade Federal do Rio Grande do Sul (UFRGS), pesquisadora do projeto Covid-19 no Brasil – Fase 2. E-mail: anakarinabraun@gmail.com

³ Professora convidada do Departamento de Línguas Modernas e do PPG-Letras da UFRGS. Pesquisadora do projeto Covid-19 no Brasil – Fase 2 e do Grupo Termisul. E-mail: cleci.bevilacqua@ufrgs.br.

⁴ Professora do Departamento de Línguas Modernas e do PPG-Letras da UFRGS. Coordenadora do Instituto de Letras do projeto Covid-19 no Brasil – Fase 2 e do Grupo Termisul. E-mail: patricia.ramos@ufrgs.br.

⁵ Este capítulo é parte da pesquisa de pós-doutorado desenvolvida por Borges Braun (2022-2024), intitulada *A covid no Brasil divulgação científica: tradução e revisão em língua inglesa*. Esta pesquisa insere-se em um projeto maior denominado *Covid-19 no Brasil – fase 2 – Análise e resposta aos impactos sociais pandemia – imunização, tratamento e práticas de cuidado e de recuperação de afetados*.

⁶<https://www.ufrgs.br/redecovid/index.php/br/sobre-a-rede-covid-19-humanidades-mcti>. Acesso em: 28 mai. 2024.

comporão o glossário. Portanto, o projeto insere-se nos Estudos do Léxico, mais especificamente na área de Terminologia e na sua interface aplicada, a Terminografia, estabelecendo inter-relações com a Linguística de Corpus e a Tradução. No contexto do projeto, o recorte aqui apresentado diz respeito ao campo da Tradução, posto que prevê a construção de um glossário bilíngue português-inglês a partir da extração de termos de um *corpus* de textos publicados pelo grupo de pesquisadores responsável pela produção de pesquisas qualitativas que analisam os impactos da Covid-19 e conformam Rede Covid-19 Humanidades, Ministério da Ciência, Tecnologia e Inovações (MCTI).

O objetivo da criação do glossário é oferecer subsídios para a consulta terminológica em futuras traduções de textos produzidos pela rede em português para o inglês. Iniciamos nossa pesquisa com a construção do *corpus* em português e a extração inicial de alguns termos. Assim, nossa proposta no presente texto é apresentar um panorama geral sobre a construção do referido *corpus* e apresentar os primeiros resultados da extração dos termos.

Para dar conta desse objetivo, apresentamos inicialmente uma breve descrição do projeto geral. Em seguida, trazemos alguns aspectos teóricos relativos à Terminologia e à Terminografia (CABRÉ, 1999; KRIEGER, 2000 e 2017; KRIEGER e FINATTO, 2004) e descrevemos os procedimentos e as etapas metodológicas seguidas para a construção do *corpus*, bem como suas características, tomando como base principalmente Berber Sardinha (2000, 2004). Posteriormente, mostramos os resultados iniciais da extração dos termos utilizando o programa Word Sketch ⁷. Por fim, apresentamos algumas considerações relativas à continuidade da pesquisa. Esperamos, desse modo, ressaltar as contribuições das áreas da Terminologia, da Terminografia e da Linguística de *Corpus* para a Tradução e para o projeto Covid-19 como um todo, reforçando a importância da interdisciplinaridade e da inter-relação entre as diferentes áreas do conhecimento.

⁷ Disponível em: <https://www.sketchengine.eu/>. Acesso em: 28 mai. 2024.

2. Sobre a rede Covid-19 humanidades MCTI

O projeto Rede Covid-19 Humanidades MCTI tem o objetivo de estudar o impacto da pandemia de COVID-19 na sociedade brasileira a partir de pesquisa qualitativa e com foco em grupos distribuídos conforme sua classificação profissional (profissionais da saúde, da indústria frigorífica, da música e de entregas por aplicativo), racial (minorias indígenas e afrodescendentes), classe social (enfatizando os contrastes das condições de enfrentamento da pandemia entre a classe média predominantemente branca e a classe baixa das minorias afrodescendentes) e de gênero (mulheres)⁸.

Para dar conta desse objetivo, o projeto foi estruturado em duas fases. A primeira fase, denominada *A Covid-19 no Brasil: análise e resposta aos impactos sociais da pandemia entre profissionais de saúde e população em isolamento* teve vigência entre junho de 2020 a junho de 2022. Os resultados da primeira fase foram compartilhados em ampla produção acadêmica: 74 artigos; 24 livros/capítulos e coletâneas de artigos; teses e dissertações, relatórios técnicos e audiovisuais. Além disso, foram estabelecidas dez parcerias com projetos e organizações da sociedade civil sediados no Brasil, na Argentina, no Reino Unido e em Portugal. Quanto ao seu alcance, o site do projeto recebeu 6,5 mil usuários de dez países, cerca de mil usuários recorrentes; 501 postagens no site; teve 122 inserções na imprensa e está presente em quatro redes sociais: Twitter, Instagram, Facebook e YouTube.

A segunda fase, cuja vigência teve início em maio de 2022 e será concluída em maio de 2024, denomina-se *A Covid-19 no Brasil 2: análise e resposta aos impactos sociais da imunização, tratamento, práticas e ambientes de cuidado e recuperação de afetados*. Seu caráter interdisciplinar congrega profissionais das Ciências Humanas, Sociais e da Saúde, de várias regiões do Brasil e do exterior. É coordenado pelo professor Jean Segata do PPG em Antropologia Social da Universidade Federal do Rio Grande do Sul (PPGAS-UFRGS). Nesta fase, conta com a parceria da Fundação Oswaldo Cruz (Fiocruz), UFSC, UFRN, UNIDAVI, UNIPAMPA e University College of London. Também tem intercâmbio com a Associação de Vítimas e Familiares de

⁸ Conforme seu próprio nome indica foi e é financiado pelo Ministério da Ciência, Tecnologia e Inovações (MCTI).

Vítimas da Covid-19 (AVICO Brasil). Além dos pesquisadores, conta com a participação consultores, colaboradores e alunos egressos das instituições parceiras.

O público-alvo do projeto são pesquisadores, professores, estudantes, redatores, jornalistas e o público geral interessado nos impactos sociais da Covid-19. As pesquisas resultantes visam subsidiar ações voltadas a esse público, considerando de modo múltiplo e situado as implicações científicas, tecnológicas, sociais, políticas, históricas e culturais da pandemia.

3. Referencial teórico

Nesta seção, trazemos uma síntese dos aspectos teóricos que sustentam a pesquisa. São estudos relativos à Terminologia, à Terminografia e à Linguística de Corpus.

3.1 Aspectos da Terminologia e da Terminografia

Entre as diferentes perspectivas teóricas do âmbito, seguimos as propostas da Teoria Comunicativa da Terminologia (TCT – CABRÉ, 1999) e da Terminologia Linguístico-Textual. (KRIEGER, 2000 e 2017; KRIEGER e FINATTO, 2004)

No âmbito da TCT, Cabré (1999) defende que a Terminologia é interdisciplinar, isto é, baseia-se na inter-relação das perspectivas linguística, semiótica, cognitiva e comunicativa, posto que só assim é possível descrever a complexidade dos termos. A perspectiva linguística permite reconhecer, analisar e descrever os termos no âmbito da linguagem natural em relação à sua conformação, significado, variação etc. A perspectiva semiótica permite tratar de outros sistemas simbólicos (símbolos, fórmulas) que também representam conhecimentos especializados e fazem parte dos discursos das diferentes áreas de saber, como a Matemática, Física, Medicina etc. A perspectiva cognitiva ajuda a explicar a forma como a realidade é conceitualizada e como os conceitos se organizam e se correlacionam. A perspectiva comunicativa busca compreender as diferentes situações comunicativas e os contextos de uso em que se utilizam os termos, ou seja, quem participa das diferentes

situações de comunicação especializada (especialistas, aprendizes, leigos etc.), que nível de conhecimento sobre determinada área ou tema eles têm, a perspectiva a partir da qual o tema é abordado (perspectiva didática, divulgação entre os pares ou para leigos etc.).

Além dessa perspectiva interdisciplinar, a autora afirma que a Terminologia é transdisciplinar visto que recolhe e descreve os termos de diferentes áreas do saber (Antropologia, Linguística, Química etc.). Em outras palavras, toda área de estudo constituiu-se e se expressa pelo conjunto de termos que representam os conceitos atinentes a ela.

Nessa linha, a definição de termo proposta pela autora indica que são

“[...] unidades léxicas, que se descrevem como unidades denominativo-conceituais, dotadas de capacidade de referência, que podem exercer diferentes funções (referencial, expressiva, conativa) e que, integradas no discurso, constituem a quase totalidade dos argumentos das orações.” (CABRÉ, 1999, p. 132)

Ela esclarece ainda que os termos adquirem valor especializado por seu uso em situações comunicativas específicas; portanto, é uma ativação pragmática que considera aspectos como o âmbito de conhecimento, a temática, o gênero textual, os interlocutores e a situação comunicativa mais ou menos especializada. Sua categoria básica é a nominal, podendo ser simples (*vírus, pandemia*⁹) ou sintagmáticos (*rotas de transmissão, isolamento social e elite econômica*).

Em relação ao trabalho terminográfico, Cabré (1999) defende que, para atender ao princípio da adequação, é necessário adaptar as etapas metodológicas às peculiaridades de cada pesquisa, considerando a função e os usuários previstos na elaboração do produto terminográfico almejado.

Por sua vez, Krieger afirma que a Terminologia, além de organizar e representar os termos, “é um elemento constitutivo da produção do saber” (KRIEGER, 2000, p. 211). Nesse sentido, os textos especializados são considerados o *habitat* natural dos termos – tal como para Cabré – e são concebidos como um todo de sentido, caracterizando-se como uma espécie de moldura cognitiva dessas

⁹ Exemplos retirados no nosso *corpus* de estudo.

unidades. Por essa razão, a descrição de suas características estruturais e discursivas – por exemplo, as condições de produção, como finalidade, meio de divulgação, tema e nível de especialização; os sujeitos envolvidos na comunicação especializada; os diferentes gêneros em que são utilizados; a divisão e distribuição de suas partes e suas características linguísticas específicas – é fundamental, pois determina o valor especializado dos termos. Nas palavras de Krieger é preciso “considerar o funcionamento da linguagem em distintos contextos de uso para identificar os modos de comportamento dos termos e, conseqüentemente, as formas sinonímicas e as variantes dos termos pesquisados”. (KRIEGER, 2017)

Nessa perspectiva, os termos são definidos como signos linguísticos que têm conteúdo especializado ativado pelo seu uso e conformação em determinada comunidade de saber e remetem aos conceitos de uma área. São de natureza designativo-denominativa, tendo a função de representar e transmitir conhecimentos específicos das áreas especializadas, sejam elas científicas, técnicas ou tecnológicas ou ainda relacionadas a ofícios. (KRIEGER e FINATTO, 2004)

No que tange à Terminografia, ela é considerada a face aplicada da Terminologia e estabelece princípios e procedimentos metodológicos que orientam a construção de produtos terminográficos que registram os termos de uma área específica. Assim, para a identificação dos termos, é preciso estabelecer alguns critérios (por exemplo, sua frequência de uso, seu valor especializado considerando as características dos textos em que ocorrem, os usuários e a função da obra). Do mesmo modo, deve-se pensar em parâmetros para a construção de um glossário. Para tanto, é preciso definir, conforme já apontamos, quem são seus usuários (no nosso caso, pesquisadores e profissionais das áreas da saúde, da Antropologia, da Sociologia etc.), sua função (auxiliar na compreensão dos termos ou servir de apoio à tradução), sua macroestrutura (o conjunto de termos que o glossário terá) e sua microestrutura (as informações que oferecerá para cada entrada: definição, contextos de uso, equivalentes e seus contextos de uso, notas etc.).

3.2 Linguística de *corpus*

Considerando que as perspectivas anteriores têm os textos como o *habitat* natural dos termos e que é neles que os termos adquirem seu significado especializado e uso específico, justifica-se o uso de um *corpus* textual baseado na Linguística de Corpus, dado que ela oferece princípios para sua construção e ferramentas que auxiliam na extração dos termos. Um desses princípios defendidos por autores como Biber, Conrad e Reppen, (1998), Stubbs, (1996; 2001) e Berber Sardinha (2000, 2004), é que a configuração do significado se dá no contexto real da comunicação.

Segundo Berber Sardinha (2000),

A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador (BERBER SARDINHA, 2000, p. 325).

Entre os critérios para a coleta de textos mencionados por Berber Sardinha (2000), estão: a origem dos textos (preferentemente textos autênticos); a finalidade (destinar-se a determinado estudo linguístico); a composição (escolha criteriosa de seu conteúdo); a formatação (devem poder ser processados por ferramentas informáticas); a representatividade (o *corpus* deve ser suficientemente representativo para dar conta do objeto de estudo da pesquisa) e a extensão (o *corpus* deve ter um tamanho suficiente para ser representativo).

Além do *corpus* de estudo, é possível utilizar um *corpus* de referência para a extração de termos e de outras informações linguísticas. Esse *corpus* é composto, em geral, de textos da língua geral (mais comumente do gênero jornalístico) e deve ser, no mínimo, cinco vezes maior que o *corpus* de estudo. Nesta pesquisa, utilizaremos o *corpus* Portuguese web 2018 (pt tenten18), disponibilizado pelo Sketch Engine.

Apresentada a síntese dos fundamentos teóricos que orientam a pesquisa, passamos a explicitar os procedimentos e etapas metodológicas para a construção do *corpus* Covid-19 em português.

4. A construção do *corpus* COVID-19 em português

As etapas metodológicas para a construção do *corpus* de estudo compreendem: *download* e armazenamento dos textos provenientes do *site* do Projeto Rede Covid-19 Humanidades, Ministério da Ciência, Tecnologia e Inovações (MCTI); organização dos textos por língua e por gêneros textuais; conversão dos textos do formato PDF para Word e, posteriormente, para TXT para serem processados mais facilmente por ferramentas de extração de dados linguísticos; limpeza dos textos em TXT; e registro das informações bibliográficas dos textos em um catálogo para futura disponibilização para os usuários do glossário.

Para iniciar a construção do *corpus*, partimos das publicações reunidas no *site* do grupo. Essas publicações são de autoria dos pesquisadores da Rede e abarcam gêneros acadêmicos (artigos, livros e capítulos de livros, relatórios técnicos, trabalhos de conclusão de curso, dissertações e teses) e de divulgação (*blogs*, histórias de vida *blogs*, dossiês, brevíários, entrevistas e relatos de experiência). Há publicações em português, inglês, espanhol e alemão. Centramo-nos aqui nos textos elaborados em português, embora também já estejamos construindo o *corpus* em inglês. A fim de demonstrar a diversidade de gêneros textuais que compõem o *corpus*, coletamos todos os textos disponíveis entre julho e outubro de 2022.

A seguir, apresentamos nos quadros 1 e 2 a distribuição dos diferentes gêneros que totalizam, em conjunto, 100 textos.

Quadro 1: Número de textos acadêmicos por gênero

Gênero Textual	Artigo	Livro ou capítulo de livro	TCC	Dissertação	Produção técnica
Quantidade	19	25	2	4	22

Fonte: As autoras

Quadro 2: Número de textos de divulgação por gênero

Gênero Textual	Relato de experiência	Blog	Dossiê	Breviário	Entrevista
Quantidade	19	1	6	1	1

Fonte: As autoras

Retomando os critérios propostos por Berber Sardinha (2000), o *corpus* de estudo, em formato digital, é composto por textos autênticos em português, escritos e contemporâneos, cuja finalidade é a identificação de termos relativos à Covid-19 na perspectiva antropológica e seu conteúdo refere-se à Covid-19 com enfoque em diferentes grupos (profissionais da saúde, da indústria frigorífica, da música e de entregas por aplicativo; minorias indígenas e afrodescendentes; mulheres). Projeta-se que terá a extensão suficiente para ser representativo¹⁰ da temática tratada e para o objetivo principal da pesquisa.

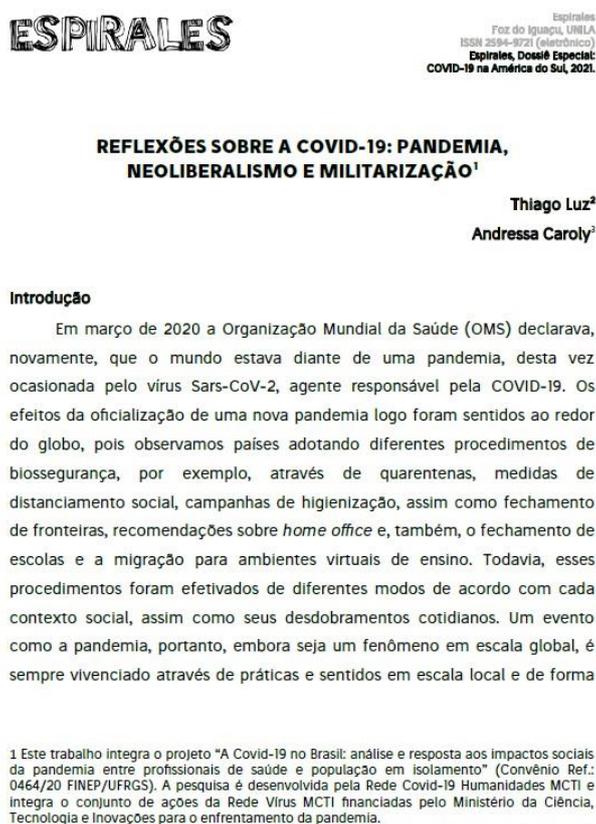
Após fazer o *download* dos textos, passamos para a etapa de conversão dos textos do formato pdf para word. Na sequência, procedemos sua limpeza que consiste na remoção de elementos extratextuais ou informações que não são

¹⁰ Ainda estamos em fase de limpeza e conversão dos textos, razão pela qual ainda não é possível ter a noção de seu tamanho final.

relevantes para a coleta: folha de rosto, agradecimentos, epígrafe, resumo e palavras-chave em português e nas línguas estrangeiras, listas de tabelas, siglas, ilustrações, símbolos e abreviaturas, figuras, notas de rodapé, gráficos e similares, juntamente com seus títulos, legendas e fontes. Após a limpeza no formato word os textos foram convertidos para txt (texto sem formatação) para serem processados de forma mais ágil pelas ferramentas de extração de informação linguística como o Sketch Engine.

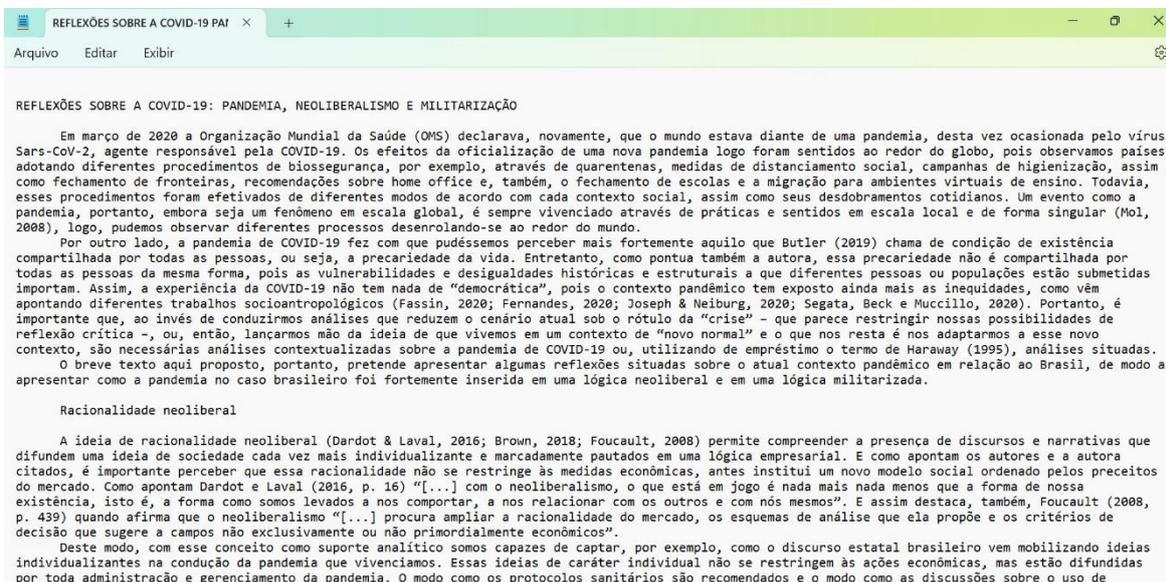
A título de comparação, trazemos um texto original (fig. 1) e um texto limpo (fig. 2).

Figura 1: Texto original



Fonte: Luz e Caroly (2021)

Figura 2: Texto limpo



Fonte: Luz e Caroly (2021)

Após a conversão e limpeza, os textos foram catalogados para o registro das referências bibliográficas de cada um deles. A catalogação se faz necessária para que se tenha todas as informações referentes ao texto em caso de necessidade: autor, data, fonte, data de coleta. Assim, os textos em português estão catalogados com os seguintes códigos que indicam o gênero textual, o número do texto e a língua: ART0001pt (artigo), BLO0001pt (blog), DOS0001pt (dossiê), BRE0001pt (breviário), ENT0001pt (entrevista) LIV/CAP0001pt (livro/capítulo de livro), DIS0001pt (dissertação), TCC0001pt (trabalho de conclusão de curso), RE0001pt (Relatos de experiência), PT0001pt (produções técnicas).

5. Levantamento terminológico inicial

Considerando as propostas teóricas da Terminologia e da Terminografia, estabelecemos critérios para a coleta inicial dos termos: pertinência temática, valor especializado e frequência. Os termos devem ocorrer nos textos da área da Covid-19, ou seja, no *corpus* compilado para fins da presente pesquisa; devem adquirir significado específico nesse *corpus* e ter frequência menor ou igual a 5. Para fins

deste trabalho, fizemos uma extração inicial com 11 textos de artigos científicos, com cerca de 40 mil palavras.

Para a extração dos termos, utilizamos as ferramentas *wordlist* (lista de palavras presentes no *corpus*) e *keywords* (palavras mais frequentes no *corpus* comparadas com as palavras de um *corpus* de referência) do programa Sketch Engine¹¹. Para a extração da *wordlist*, aplicamos uma lista de exclusão previamente construída, que permite retirar palavras gramaticais (artigos, preposições, conjunções, numerais etc.) e palavras lexicais que não interessam aos propósitos da pesquisa (nomes próprios, topônimos etc.).

A título de exemplificação, a figura 1, ilustra os dados obtidos. Nela já podemos visualizar candidatos a termos como *saúde*, *pandemia*, *vírus*, *transmissão*, *zika*, *covid-19*, *riscos*, *mulheres* etc. pois atendem aos critérios referidos anteriormente, isto é, referem-se às temáticas relacionadas à Covid-19, têm valor especializado e frequência igual ou superior a 5.

Figura 3: Resultados da extração pela *wordlist*

Word	Frequency ? ↓	Word	Frequency ? ↓	Word	Frequency ? ↓
1 saúde	148 ***	18 mulheres	51 ***	35 têm	37 ***
2 pandemia	122 ***	19 diferentes	50 ***	36 mundo	36 ***
3 vírus	119 ***	20 exemplo	49 ***	37 condições	36 ***
4 transmissão	105 ***	21 casos	48 ***	38 seja	35 ***
5 zika	89 ***	22 vida	48 ***	39 mosquitos	35 ***
6 dados	88 ***	23 social	48 ***	40 estado	34 ***
7 covid-19	80 ***	24 eu	47 ***	41 acesso	34 ***
8 sociais	75 ***	25 país	46 ***	42 formas	33 ***
9 pesquisa	74 ***	26 risco	46 ***	43 parte	33 ***
10 riscos	68 ***	27 foram	46 ***	44 meio	33 ***
11 brasil	64 ***	28 pessoas	45 ***	45 era	33 ***
12 tempo	63 ***	29 epidemia	43 ***	46 trabalhadores	33 ***

Fonte: *Sketch Engine*

¹¹ É possível utilizá-la de forma gratuita durante um mês para um *corpus* de até 1 milhão de palavras.

Para a extração das *keywords*, utilizamos a busca avançada, que permite selecionar um *corpus* de referência disponível na própria plataforma. Escolhemos o *corpus Portuguese web 2018* (pt tenten18) e definimos a frequência maior ou igual a 5. O resultado pode ser visto na figura 4.

Figura 4: Resultados da extração por *Keywords*

1	covid-19	...	14	vetorial	...	27	etnográfico	...
2	pandemia	...	15	multiespécie	...	28	epidemia	...
3	segata	...	16	beck	...	29	perambulação	...
4	coronavírus	...	17	latour	...	30	arboviroses	...
5	principalmente	...	18	placentário	...	31	antropológico	...
6	pandêmico	...	19	recentemente	...	32	antropólogo	...
7	zika	...	20	jauréguiberry	...	33	militarização	...
8	sars-cov-2	...	21	consequentemente	...	34	picada	...
9	keck	...	22	wuhan	...	35	mosquito	...
10	atualmente	...	23	patógeno	...	36	etnografia	...
11	wildavsky	...	24	reis-castro	...	37	vírus	...

Fonte: *Sketch Engine*

Nessa lista, há termos como *covid-19*, *pandemia*, *coronavírus*, *zika*, *sars-cov-2*, *multiespécie* etc, em sua grande maioria relacionados ao tema central do *corpus*, Covid-19 e, consequentemente, à área da saúde. Mas há também outras palavras que precisam ser retiradas, pois são irrelevantes, como nomes próprios (*Segata*, *Keck* etc.) e advérbios (*principalmente*, *atualmente* etc.). Essas palavras podem ser excluídas em etapa posterior, seguindo os critérios estabelecidos para a seleção dos termos. Observamos ainda que todas elas são simples, ou seja, formadas por uma palavra apenas, mas é possível gerar também termos sintagmáticos (*multi-word terms*), tal como se vê na figura 5.

Figura 5: Resultados da *multi-word terms*

SINGLE-WORDS ✓ MULTI-WORD TERMS ✓

reference corpus: Portuguese Web 2018 (ptTenTen18) (items: 603)

Term	Term	Term
1 rotas de transmissão ...	13 elites econômicas ...	25 utilização de dados ...
2 transmissão sexual ...	14 isolamento social ...	26 desigualdades sociais ...
3 epidemia de zika ...	15 mercados úmidos ...	27 risco de infecção ...
4 pandemia de covid-19 ...	16 racionalidade neoliberal ...	28 dados pessoais ...
5 transmissão vertical ...	17 uso de dados pessoais ...	29 instituições de pesquisa ...
6 elite econômica ...	18 condição de classe ...	30 número de casos ...
7 dimensão de classe ...	19 pesquisa participativa ...	31 pesquisa de campo ...
8 transmissão vetorial ...	20 uso de dados ...	32 mulheres negras ...
9 contexto pandêmico ...	21 proteção de dados ...	33 sistema de saúde ...
10 zona de alerta ...	22 controle vetorial ...	34 febre amarela ...
11 picadas de mosquitos ...	23 emergência de saúde ...	35 et al ...
12 barreira placentária ...	24 vírus zika ...	36 organização mundial ...

Fonte: *Sketch Engine*

Ao observar os resultados, constata-se que há termos relacionados à área a saúde como *rota de transmissão*, *transmissão sexual*, *epidemia de zika*, *transmissão vetorial* etc. Há também termos referentes a outras áreas (Sociologia, Antropologia e Economia), como *dimensão de classe*, *elite econômica*, *racionalidade neoliberal*, *pesquisa participativa*, *uso de dados* e *desigualdades sociais*.

Tanto para as *keywords* quanto para os *multi-word terms*, ao clicar-se nos três pontos ao lado do termo, é possível acessar seus contextos de uso. É o que ilustra a figura 6 para o termo *rota de transmissão*.

Figura 6: Contextos de rota de transmissão

The screenshot shows the CONCORDANCE interface with the search term "rotas de transmissão" and 14 results. The KWIC view displays the following results:

Doc#	Left context	KWIC	Right context
1	doc#10 análise que examina os complexos enredamentos das múltiplas	rotas de transmissão	do vírus Zika – transmissão vetorial (por meio c
2	doc#10 extensa literatura antropológica examinando de que maneira as	rotas de transmissão	são influenciadas pelas estruturas sociais e, do
3	doc#10 le da infecção.</s><s>Assim, por meio da análise das múltiplas	rotas de transmissão	, delineamos as condições que levaram Zika a
4	doc#10 partir desses dados, examinamos a epidemia em suas múltiplas	rotas de transmissão	.</s><s>Nossa proposta se inspira em esforços
5	doc#10 a doença ou epidemia, examina os estudos científicos sobre as	rotas de transmissão	, as políticas públicas elaboradas para conter tr
6	doc#10 iardian a complexa relação e implicações do vírus e suas várias	rotas de transmissão	: Se eu fosse uma cineasta oferecendo um cen
7	doc#10 J et al., 2018).</s><s>Em outras palavras, Zika possui múltiplas	rotas de transmissão	; porém, enquanto as transmissões vetorial e si
8	doc#10 um lugar como o Brasil", identificar a relevância das diferentes	rotas de transmissão	que, ao espalharem o vírus, causam a epidemi
9	doc#10 e também caracterizam o país.</s><s>Desigualdades Sociais e	Rotas de Transmissão	</s><s>Um estudo, publicado durante a epidem

Fonte: *Sketch Engine*

A análise dos contextos oferece subsídios na etapa de seleção dos termos, já que permite encontrar definições e entender seu funcionamento no texto, bem como selecionar o contexto que será apresentado no glossário.

Conforme indicamos, embora esses dados resultem de uma primeira extração, já é possível observar um conjunto importante de termos. Isso indica também que os critérios de construção do *corpus* e de seleção dos termos parecem ser adequados para os fins da pesquisa.

6. Considerações finais

Em primeiro lugar, finalizamos este artigo ressaltando a relevância do projeto de pós-doutorado *A covid no Brasil divulgação científica: tradução e revisão em língua inglesa* para os estudos das áreas das Ciências Sociais e Ciências da Saúde como para a área de Letras e, mais especificamente, da Tradução. Uma pesquisa que se debruça sobre os impactos sofridos pelos trabalhadores mais precarizados no período da crise sanitária pode ser útil para nos ajudar a entender como funciona a nossa sociedade em períodos de crise e quais os efeitos disso sobre a população vulnerável. Os resultados poderão fornecer dados úteis para a criação de políticas

públicas na área da saúde, além de permitir reflexões teóricas no âmbito das Ciências Sociais.

No que diz respeito à área da Tradução, acreditamos que a elaboração do glossário terminológico bilingue português-ínglês contribuirá para uma produção mais consistente e para a divulgação mais ampla dos diversos gêneros textuais publicados no *site* nas versões em português e/ou inglês, bem como para o fornecimento de subsídios para a consulta terminológica em futuras traduções.

Em relação à constituição do *corpus* de estudo, salientamos a necessidade de buscar ainda textos em inglês em outras fontes confiáveis (institucionais, acadêmicas) para a construção de um *corpus* comparável mais equilibrado. No que se refere à extração dos termos, é preciso processar o *corpus* total e, a partir dos resultados, aplicar os critérios estabelecidos para chegar à lista final dos termos que comporão o glossário. É necessário pensar ainda na estrutura do glossário (seus campos) para atender melhor sua função e seus usuários.

Finalmente, esperamos que a disponibilização do glossário terminológico bilíngue relativo à Covid-19 no *site* contribua para estudos futuros nas áreas de conhecimento contempladas neste projeto.

Referências

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri, São Paulo: Manole, 2004.

BERBER SARDINHA, T. Linguística de Corpus: Histórico e Problemática. **DELTA**, v. 16, n. 2, 2000.

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

CABRÉ, M. T. **Terminología**: representación y comunicación. Una teoría de base comunicativa y otros artículos. Barcelona: Universitat Pompeu Fabra, Instituto Universitario de Lingüística Aplicada, 1999.

KRIEGER, M. G. **Terminologia lingüístico textual (TLT)**: princípios teóricos e metodológicos. Primer Congreso Internacional de Lexicología, Lexicografía y Terminología. Córdoba, Argentina, julho de 2017. (Apresentação Oral)

KRIEGER, M. G. Terminologia revisitada. **DELTA**, v. 16, n. 2, p. 209-228, 2000.

KRIEGER, M. G.; FINATTO, M. J. B. **Introdução à Terminologia**: Teoria & Prática. São Paulo: Contexto, 2004.

STUBBS, M. **Words and phrases**; corpus studies of lexical semantics. Oxford: Blackwell, 2001.

STUBBS, M. **Text and Corpus Analysis**. Oxford: Blackwell Publisher Ltd., 1996.