



Inteligibilidade e convencionalidade em textos de divulgação da área médica em português brasileiro

Readability and conventionality in expository texts in Brazilian Portuguese

Yuli Souza Carvalho

Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul / Brasil

yuli_@live.ca

<https://orcid.org/0000-0002-7914-8459>

Rozane Rodrigues Rebechi

Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul / Brasil

rozane.rebechi@ufrgs.br

<https://orcid.org/0000-0002-1878-7548>

Resumo: O objetivo desta pesquisa é cotejar dados indicativos de inteligibilidade e convencionalidade em textos de divulgação da área médica em português para verificar sua adequação ao público brasileiro. Para tanto, apoiamos-nos nos pressupostos da Linguística de *Corpus* para a compilação e o processamento de um *corpus* paralelo, formado por textos escritos originalmente em inglês e suas traduções para o português, e um *corpus* comparável, composto pelos textos traduzidos em português e por textos originalmente escritos nesse idioma. A metodologia do estudo combina análises quantitativas – para levantamento de inteligibilidade, chaticidade e colocação – e qualitativas – para análise das palavras em contexto. Em relação à inteligibilidade, as ferramentas apontaram que os textos escritos em português são ‘difíceis’ para o leitor médio brasileiro, com grau de instrução inferior ao Ensino Médio. Já os textos traduzidos foram considerados ‘razoavelmente difíceis’, de acordo com esse mesmo critério de avaliação, que classificou os originais em inglês como ‘razoavelmente fáceis’, considerando-se seu público alvo, ou seja, o leitor médio estadunidense. A análise qualitativa apontou que os textos traduzidos apresentam quebras de convencionalidade, demonstrando preferência por equivalentes *prima facie*, nem sempre condizentes com os

padrões observados nos textos de mesmo gênero escritos originalmente em português. Apesar de a ferramenta de acessibilidade textual indicar que tanto os textos escritos originalmente em português quanto aqueles traduzidos não se mostram totalmente adequados para o leitor-alvo brasileiro de textos de divulgação médica, acreditamos que a quebra da convencionalidade, identificada nos textos traduzidos, pode dificultar ainda mais a compreensão do leitor médio de resultados de pesquisas científicas da área da saúde.

Palavras-chave: textos de divulgação; tradução; convencionalidade; inteligibilidade.

Abstract: The aim of this research is to collate data from intelligibility and conventionality in health-related expository texts in Portuguese to investigate their appropriateness to Brazilians. To this end, we rely on Corpus Linguistics for the compilation and processing of a parallel corpus, comprising texts originally written in English and their translations into Portuguese, and a comparable corpus, composed of texts translated into Portuguese and texts originally written in that language. Our methodology combines quantitative analysis – to assess readability, keyness, and collocation – and qualitative analysis – to investigate words in context. Regarding readability, the tools pointed out that texts written in Portuguese are ‘difficult’ for the average Brazilian reader, with a level of education lower than High School. The translated texts were considered ‘fairly difficult’, according to this same evaluation criterion, which classified the originals in English as ‘fairly easy’, considering its target audience, that is, the average American reader. The qualitative analysis pointed out that the translated texts may compromise conventionality, revealing a preference for *prima facie* equivalents, not always consistent with the patterns observed in original Brazilian Portuguese counterparts. Although the accessibility evaluation tool indicates that both the texts originally written in Portuguese and those translated into Portuguese do not prove to be entirely suitable for the Brazilian target reader of medical expository texts, we believe that, by breaking conventionality, the translated texts may hinder even more the average reader’s comprehension of results of scientific research.

Keywords: expository texts; translation; conventionality; readability.

Recebido em 09 de outubro de 2020

Aceito em 23 de novembro de 2020

1 Introdução

Os textos de divulgação são de suma importância, já que visam ao compartilhamento, com o público geral, de resultados de pesquisas desenvolvidas por especialistas em diversas áreas do conhecimento.

Na área médica, eles desempenham papel ainda mais importante, pois instruem a população em relação a questões de saúde. O papel preponderante dos textos de divulgação fica ainda mais evidente em momentos como o que estamos vivendo, quando o mundo todo enfrenta uma questão de saúde pública sem precedentes: a pandemia da COVID-19. Empenhados em pesquisar sobre o controle, a transmissão e o tratamento da doença, ao mesmo tempo em que buscam desenvolver uma vacina que consiga combater o vírus transmissor, médicos e cientistas têm publicado uma infinidade de artigos acadêmicos que relatam suas descobertas.

Em geral, as publicações acadêmicas são escritas para a comunidade discursiva que possui conhecimento prévio do domínio especializado no âmbito em que a pesquisa se insere, mas não se mostram adequadas para o público leigo, que espera receber informações menos técnicas para aplicá-las no dia a dia. Sendo assim, é necessário que os dados apresentados nos textos de divulgação sejam fornecidos de forma clara e acessível. Afinal, assim como elevadores e rampas visam tornar as construções acessíveis ao público geral, os textos de divulgação devem ter características que possibilitem sua acessibilidade para leitores de diferentes faixas etárias, classes sociais e níveis de escolaridade. Textos muito complexos para grande parcela do público podem acabar não atingindo o objetivo final, qual seja, o de instruir a população geral acerca de temas especializados.

Atualmente, grande parte das publicações científicas é compartilhada em língua inglesa, a *lingua franca* da academia (JENKINS, 2009), e traduzida para os inúmeros vernáculos falados no mundo. De acordo com Rosselli (2016), cerca de 96% de todos os artigos indexados em 2015 na base de dados especializada em artigos biomédicos PubMed foram publicados em inglês. Assim, a tradução tem papel central no compartilhamento de informações. Ao considerarmos que os textos de divulgação devem atingir uma grande parcela da população, é necessário também investigar se o processo de tradução desses materiais resulta em informações acessíveis ao leitor, especialmente no que diz respeito a aspectos como inteligibilidade e convencionalidade.

Muito já se discutiu sobre ‘fidelidade’ da tradução, conceito superado – especialmente no que tange aos textos especializados – e que deu lugar à ‘lealdade’ (NORD, 2006), conceito que coloca em foco a função do texto traduzido para o público-alvo a que se destina. Com

isso em mente, podemos afirmar que um texto de divulgação, escrito em inglês para determinado público-alvo, só vai ‘funcionar’ bem para o leitor da tradução se estiver adequado às características leitoras desse público.

Assim, o objetivo da pesquisa¹ é investigar, a partir de dados levantados por ferramentas de análise textual, como se dá a relação entre inteligibilidade e convencionalidade em textos de divulgação da área médica, a fim de verificar sua adequação para o público brasileiro, quer tenham esses textos sido escritos originalmente em português, quer sejam traduções. Para tanto, foram compilados dois *corpora* de estudo, um *corpus* paralelo e um *corpus* comparável. O *corpus* paralelo é composto por textos de divulgação escritos originalmente em língua inglesa e suas respectivas traduções para o português, extraídos do portal MedlinePlus (U.S. NATIONAL LIBRARY OF MEDICINE, 2020); e o comparável, com essas traduções para o português e textos originalmente escritos em português brasileiro do mesmo gênero, publicados pelo Ministério da Saúde (BRASIL, 2018). Primeiramente, esses *corpora* foram analisados pelas ferramentas Coh-Metrix (GRAESSER *et al.*, 2017) e Coh-Metrix-Port (NILC, 2020), para verificação do Índice Flesch – que estima a inteligibilidade de um texto. Palavras-chave e colocações foram identificadas por meio do AntConc (ANTHONY, 2019), enquanto o AntPConc (ANTHONY, 2017) foi usado para alinhar trechos originais e respectivas traduções, a fim de que se investigassem as palavras em contexto.

Em resumo, nossa investigação visa responder a duas perguntas: (i) de acordo com as ferramentas de inteligibilidade, os textos escritos em português e os traduzidos para esse idioma se mostram acessíveis para o público-alvo? e (ii) tomando como premissa que a convencionalidade característica dos gêneros textuais facilita a compreensão, os textos traduzidos revelam padrões observados em textos de divulgação escritos originalmente em português brasileiro?

Este artigo está dividido em cinco seções. Após esta introdução, será apresentada a fundamentação teórica, onde serão abordados conceitos de Linguística de *Corpus*, tradução, acessibilidade textual e gênero, pertinentes a este estudo. A terceira seção delineará a metodologia da pesquisa, apresentando os *corpora* de estudo e as ferramentas utilizadas

¹ Este artigo apresenta resultados de pesquisa de Mestrado desenvolvida no âmbito do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

para o levantamento dos dados. A quarta seção apresentará e discutirá os resultados obtidos. Por fim, apresentaremos as considerações finais do estudo.

2 Abordagem teórica

Para o desenvolvimento desta pesquisa, apoiamos-nos nos pressupostos da inteligibilidade textual, da tradução e da Linguística de *Corpus* para analisarmos textos de divulgação da área médica em português. Abaixo relatamos os conceitos de cada área julgados relevantes para o estudo.

2.1 Acessibilidade textual

Assim como as rampas de acesso para cadeirantes ou os elevadores, que podem ser observados em prédios públicos, e o piso tátil para pessoas cegas, que é visto pelas ruas, a acessibilidade atua, no âmbito textual, como um facilitador para o entendimento da mensagem pelo leitor (FINATTO, 2020). Dessa forma, a acessibilidade textual pode ser entendida como uma condição desejada de qualidade de texto, evitando que ele imponha barreiras linguísticas ao público leitor, para que este tenha condições de compreender as informações compartilhadas. Para os fins deste trabalho, apresentamos a inteligibilidade e a convencionalidade como aspectos cruciais para analisar a acessibilidade dos textos.

A fim de estimar o quão acessíveis são os textos, foram desenvolvidas fórmulas de inteligibilidade, que demonstram, matematicamente, a adequação de um texto para determinado público.

O Índice Flesch é o mais aplicado para pesquisas que se baseiam em estimativas de inteligibilidade, que, é importante ressaltar, não deve ser confundida com legibilidade. A primeira está relacionada ao que é inteligível, ou seja, aquilo que é de fácil compreensão, enquanto a segunda se refere ao que é legível, ou seja, que está claro e nítido (DUBAY, 2004). A legibilidade tem relação com design e *layout* – como a diagramação, o tipo e o tamanho de fonte etc. –, relacionando-se à acessibilidade visual. Ambas estão ligadas a formas de compreensão, respectivamente, à compreensão de forma mental e à compreensão de forma visual. Neste artigo, trataremos apenas de inteligibilidade, já que nos ocupamos apenas dos aspectos linguísticos do texto.

Desenvolvido originalmente para o inglês, o Índice Flesch leva em consideração duas variáveis: o comprimento médio das frases, ou seja, o número de palavras do texto dividido pelo número total de sentenças; e a média de sílabas por palavras, resultado do número total de sílabas dividido pelo número de palavras do texto. Esse índice foi adaptado para a língua portuguesa por pesquisadores do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (MARTINS *et al.*, 1996), que tinham como objetivo adequá-lo à realidade das palavras e sílabas do português brasileiro, já que o comprimento das palavras e das frases em língua portuguesa difere bastante daquele em língua inglesa.

A fim de exemplificarmos a importância da inteligibilidade textual, mencionamos um estudo realizado no ano de 2003, nos Estados Unidos, e relatado por DuBay (2004), que apontou que acidentes de trânsito vinham causando cerca de 46% das mortes acidentais de crianças, apesar de se esperar que o uso correto do bebê conforto, da cadeirinha ou do assento para as crianças reduzisse os riscos de fatalidades em 71%. Na verdade, a maioria desses assentos vinha sendo instalada incorretamente. A partir dessa conclusão, foi conduzido um estudo para analisar a inteligibilidade dos manuais de instalação desses assentos. Descobriu-se, então, que os 107 manuais analisados foram escritos em uma linguagem considerada difícil para 80% dos adultos no país, privados, portanto, da total compreensão das instruções descritas no material (DUBAY, 2004). Diante do exposto, podemos concluir que, apesar de terem domínio da língua em que os textos estavam escritos – inglês –, os leitores não tinham total acesso à mensagem que esses textos pretendiam transmitir. Daí a importância de se produzirem textos inteligíveis, especialmente aqueles que tenham como objetivo instruir o leitor correta e adequadamente.

Para tirar o maior proveito da comunicação, o leitor espera reconhecer nos textos padrões já observados em outros de mesmo gênero, ou seja, os textos devem revelar convenções características. Segundo Tagnin (2013), a convencionalidade ocorre em três níveis linguísticos, a saber, (i) sintático – relacionado à forma como os elementos se combinam gramaticalmente; (ii) semântico – com relação ao significado dos elementos que constituem as frases, sentenças, expressões idiomáticas etc.; e (iii) pragmático – quanto à interação entre os participantes do ato comunicativo. É importante ressaltar que as convenções linguísticas são consagradas pelo uso, nem sempre ‘respeitando’ uma lógica conceitual – por exemplo, em português convencionou-se dizer ‘achados e perdidos’,

apesar de esses acontecimentos ocorrerem em ordem inversa, ordem, esta, respeitada na expressão em língua inglesa – ‘*lost and found*’. Portanto, a convencionalidade está relacionada ao domínio da fluência de determinada língua (TAGNIN, 2013), já que os falantes dispõem de unidades linguísticas previamente armazenadas na memória.

Assim, ao produzir um enunciado, repetem-se sintagmas já utilizados, causando a cristalização de padrões linguísticos. Ao lermos textos de divulgação, esperamos identificar os padrões típicos desse gênero, a fim de absorvermos conteúdos novos de forma clara e precisa. Portanto, com base no que foi exposto, acreditamos que, caso se constate quebra da convencionalidade nos textos traduzidos, quando comparados àqueles originalmente escritos em determinado idioma, poderia haver prejuízos à acessibilidade textual, aspecto imprescindível para a divulgação de informações, especialmente aquelas de utilidade pública.

2.2 Os estudos de tradução e a Linguística de *Corpus*

A abordagem funcionalista da tradução, proposta por Nord (2006) com base em textos especializados, coloca em foco o propósito do texto traduzido, sempre com vistas a atingir seu público-alvo. Em outras palavras, o texto traduzido deve ‘funcionar’ apropriadamente para aqueles que o recebem, assim como se espera que o texto de partida tenha funcionado para seu público leitor. Dessa forma, além de fazer escolhas terminológicas adequadas às diferentes áreas de especialidade, o tradutor deve avaliar as capacidades de compreensão e cooperação de sua audiência, antecipando os possíveis efeitos que determinadas escolhas textuais poderão ter sobre o leitor. Nesse sentido, *corpora* textuais não só auxiliam os tradutores a evidenciar características próprias dos textos de partida, como também os ajudam a encontrar soluções na língua de chegada, já que evidenciam os padrões indicativos de convencionalidade (STEWART, 2000).

Ao se aliarem à Linguística de *Corpus*, os estudos de tradução possibilitaram apontamentos importantes sobre textos traduzidos em comparação com textos originais. No que diz respeito às tendências tradutórias, Baker (1993) cita o exemplo da explicitação, que fornece ao leitor da tradução informações que no original estavam subentendidas. Nessa direção, a tradução apresenta também movimentos de desambiguação e facilitação, como, por exemplo, o uso de formas

pronominais mais precisas, possibilitando que o leitor identifique o referente sem muito esforço.

Frankenberg-Garcia (2006) apresenta resultados similares para a língua portuguesa. A partir das análises feitas por meio do COMPARA,² a pesquisadora atesta que “as traduções tendem a ser mais longas do que os textos-fonte, tanto na direção inglês-português como na direção português-inglês” (FRANKENBERG-GARCIA, 2006, p. 147),³ corroborando o que foi apresentado em Baker (1993).

Além das tendências tradutórias, Baker (1993) menciona estudos sobre a chamada “terceira língua”⁴ na tradução, que consiste no resultado do confronto entre língua-fonte e língua-alvo, imprimindo ao texto traduzido características que o distanciam tanto do texto-fonte quanto de textos originalmente produzidos na língua-alvo. Portanto, a forma como o texto-fonte influencia o texto-alvo pode resultar em perda de padrões característicos dos gêneros textuais, comprometendo a acessibilidade. Aqui, entendemos essa “terceira língua” como a quebra da convencionalidade, ou seja, dos padrões observados nos textos escritos originalmente nas respectivas línguas e culturas, a depender das características dos gêneros em questão.

2.3 Texto de divulgação

Bhatia (1993, p. 49) define gênero como “um evento comunicativo reconhecível, caracterizado por um conjunto de propósito(s) comunicativo(s) identificado(s) e mutuamente compreendidos pelos membros da comunidade profissional ou acadêmica em que geralmente ocorre.”⁵ Textos de divulgação, assim como outros gêneros textuais, possuem características específicas, sendo a mais marcante delas o fato de estarem situados entre o falar científico e o falar comum, apresentando vocabulário desses dois registros. Nesse sentido,

² *Corpus* paralelo bidirecional de português e inglês. Disponível em: <https://www.linguateca.pt/COMPARA/Bem-vindos.html>. Acesso em: 25 ago. 2020.

³ No original: “translations tended to be longer than source texts in both the English-Portuguese and the Portuguese-English directions.”

⁴ No original: “third code”.

⁵ No original: “a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs.”

seria redutor pensar a divulgação científica apenas como uma redução ou adaptação de textos científicos, elaborados para a leitura de pares dotados de uma mesma competência profissional. Ao contrário, a divulgação científica em seu amplo universo, ainda carente de descrições, afigura-se como uma categoria textual autêntica, com regras próprias de produção de significação e de recursos que visam a uma comunicação eficiente. (KRIEGER, 2009, p. 9-10).

Massarani e Moreira (2005) distinguem três categorias na comunicação científica: i) os discursos científicos primários, escritos por pesquisadores para pesquisadores; ii) os discursos didáticos, como os manuais científicos para ensino; e iii) os discursos de divulgação científica. Enquanto o destinatário do texto científico é um par com conhecimento especializado sobre o tema, o texto de divulgação é voltado para pessoas leigas, sem, necessariamente, conhecimento prévio construído sobre o que será abordado no texto (SANTIAGO, 2007). Andreetto (2013, p. 8) corrobora essa visão:

Apesar de os artigos acadêmicos também terem o objetivo de informar, a organização dessa informação segue padrões diferentes daqueles na área da divulgação, que embora também tenham como objetivo transmitir conhecimento, o fazem respeitando a convencionalidade esperada por seu público alvo, que normalmente é leigo em uma dada área de especialidade.

A partir de uma pesquisa empírica com profissionais brasileiros da área médica que necessitavam se preparar para um exame de proficiência linguística em inglês contendo textos de divulgação desse domínio especializado, Andreetto (2013) concluiu que esses especialistas tinham mais dificuldade para compreender textos de divulgação do que textos acadêmicos, já que estavam mais habituados aos padrões destes. Portanto, observamos que a adequada compreensão da informação não depende exclusivamente do nível de escolaridade do público leitor, mas prevê a familiaridade com os discursos a que esse público é exposto, ou seja, aos padrões reconhecidos por determinada comunidade discursiva (cf. BHATIA, 1993).

Dessa forma, ao escrever um texto de divulgação, o especialista deve transmitir a mensagem de maneira diferente do que faria se estivesse compartilhando informações com outros especialistas no assunto. Além da terminologia adotada, as estruturas linguísticas características desse

gênero textual também devem ser reconhecidas pelo leitor. Ou, segundo Zamboni (2001 *apud* ANDREETTO, 2013), para essa tarefa, deve-se transformar o discurso científico em discurso do “cotidiano”.

3 Metodologia

Conforme mencionado na Introdução deste artigo, para este estudo foram realizadas análises quantitativas, por meio dos índices de inteligibilidade e chavicidade, e também análises qualitativas, auxiliadas por ferramentas de Linguística de *Corpus*, pois, para esgotar o objetivo da investigação proposta, a pesquisa não poderia se manter apenas no âmbito estatístico, mas sim tomá-lo como base para uma investigação mais aprofundada. Afinal, de acordo com Biderman (1967), os “primeiros senões facilmente apreensíveis são constituídos pelos dois aspectos irreduzíveis da realidade linguística: o elemento qualitativo e o quantitativo”, reiterando a importância de se analisarem os textos por esses dois vieses. Nesta seção, explicaremos a construção dos *corpora* de estudo, bem como as ferramentas utilizadas nas análises desses *corpora*.

3.1 Os *corpora* de estudo

A fim de compararmos as características de textos de divulgação da área médica escritos originalmente em português com aqueles traduzidos, e verificarmos se há diferenças significativas no que tange à convencionalidade do gênero, compilamos um *corpus* paralelo – originais em inglês e traduções para o português – e um *corpus* comparável, composto pelos mesmos textos traduzidos, e textos escritos originalmente em português.

O *corpus* paralelo foi compilado a partir de textos de divulgação do site MedlinePlus, portal da U.S. National Library of Medicine (2020), que publica informações sobre saúde para pacientes, seus familiares e amigos em diversas línguas, abrangendo sintomas e tratamentos, compondo um material que, de acordo com o site, é “de fácil leitura”.⁶ Esse *corpus* é composto pelos 66 textos em inglês e suas traduções para o português disponíveis.

O *subcorpus* em português apresenta estruturas bastante similares ao da sua contraparte de traduções, pois foi compilado com base em critérios compartilhados específicos desse gênero. Por exemplo, a grande

⁶ No original: “Easy-to-Read Materials”.

maioria dos textos tem subtítulos dividindo os diferentes tópicos; há, também, listas de itens introduzidas por elementos gráficos, que tornam as informações visualmente mais claras (DUBAY, 2004). Esse *subcorpus* contempla os 191 textos da Biblioteca Virtual em Saúde, mantida pelo Ministério da Saúde.

Ressaltamos que nosso objetivo não era balancear os *corpora* em número de textos ou *tokens*, mas, sim, coletar a totalidade de textos disponíveis em ambos os portais. A Tabela 1 resume os números relativos aos *corpora* de estudo.

TABELA 1 – Números de *types* e *tokens* dos *subcorpora* da pesquisa

Corpus	MedlinePlus (EN)	MedlinePlus (PT)	Ministério da Saúde (PT)
Textos	66	66	191
Tokens	34.765	39.476	84.085
Types	3.088	4.554	9.666
TTR*	8,88%	11,53%	11,49%

**Type-Token Ratio*

Fonte: Elaborada pelas autoras.

É interessante notar que, comparando-se a extensão média dos textos que compõem os *corpora*, os textos originalmente escritos em inglês são mais longos (526 *tokens*, em média) do que aqueles escritos em português (440 *tokens*), corroborando as conclusões de Rebechi (2017) e Fuchs (2018) sobre as diferenças nos tamanhos dos textos de mesmo gênero escritos em português brasileiro e em inglês estadunidense. Já as traduções para o português ficaram mais longas (598 *tokens*) do que os textos de partida, contrariando os estudos citados, mas confirmando os achados de Baker (1993) e Frankenberg-Garcia (2006) sobre a tendência de explicitação do texto traduzido.

O cálculo de *type-token ratio* indica a porcentagem da riqueza lexical do corpus, sendo demonstrado pelo cálculo $types \div tokens \times 100$. De acordo com Berber Sardinha (2004, p. 94), “Quanto maior o seu valor, mais palavras diferentes o texto conterà. Em contraposição, um valor baixo indicará um número alto de repetições, o que pode indicar um texto menos rico ou variado, do ponto de vista de seu vocabulário”. Por estarmos lidando com línguas diferentes, o índice calculado para o inglês não pode ser comparado com o do português. O que podemos

constatar, a partir desse levantamento, é que os índices de riqueza lexical dos textos de divulgação em português são muito próximos, com uma diferença de apenas 0,04% a mais para as traduções.

3.2 As ferramentas de análise

Para a análise de inteligibilidade, as ferramentas utilizadas foram o Coh-Metrix (GRAESSER *et al.*, 2017) e o Coh-Metrix-Port (NILC, 2020), para os textos em inglês e em português, respectivamente. O índice utilizado para a análise foi o Índice Flesch, que considera duas métricas para estimar o nível de dificuldade do texto: o comprimento médio das frases e a média de sílabas por palavras. Como resultado do cálculo, obtém-se um índice que pode ir de 0 a 100, sendo que quanto mais próximo de 0, mais difícil seria o texto; e quanto mais perto de 100, mais fácil.

Vale ressaltar que, de acordo com Graesser *et al.* (2004), as fórmulas matemáticas de inteligibilidade, como o Índice Flesch, ignoram componentes linguísticos e discursivos que influenciam a compreensão textual. Dessa forma, apesar de os parâmetros de tamanho de sentenças e de palavras serem indicativos de legibilidade, eles não conseguem revelar com precisão e por si só a complexidade de um texto. Por exemplo, aspectos como número de palavras diferentes, frequência de uso das palavras e sua regularidade ou irregularidade seriam interessantes de serem estudados em um cálculo de inteligibilidade. Por esse motivo, se viu a necessidade de aliar análises de inteligibilidade a análises manuais, partindo-se, também, de dados quantitativos revelados por ferramentas computacionais.

A análise de padrões linguísticos foi feita por meio do AntConc (ANTHONY, 2019), um *software* gratuito, que pode ser facilmente baixado e utilizado off-line. Com base em cálculos estatísticos, ele possibilita que sejam feitos levantamentos de palavras características do *corpus* de estudo – palavras-chave –, de padrões linguísticos, como colocações e *clusters*, entre outras aplicações. A partir desses levantamentos, o *software* permite que sejam feitas análises quantitativas aliadas ao olhar atento do pesquisador.

O *software* AntPConc (ANTHONY, 2017) foi utilizado na investigação do *corpus* paralelo, depois que os dois *subcorpora* – originais em inglês e traduções em português – foram devidamente alinhados, por meio do LF Aligner (FARKAS, 2018).

4 Análises e discussão

Nesta seção, explicitaremos as análises quantitativas e qualitativas realizadas a partir dos *corpora* de estudo, e apresentaremos e discutiremos os resultados dos levantamentos, que nos guiaram na reflexão sobre a adequação dos textos de divulgação da área médica para o público-alvo, ou seja, a população de não especialistas.

4.1 Levantamento de inteligibilidade

Iniciamos a análise quantitativa dos *corpora* de estudo a partir do levantamento de inteligibilidade, possibilitado pelo cálculo automático do Índice Flesch. Para esse levantamento, primeiramente foi aplicado o Índice Flesch à íntegra dos textos que compõem o *corpus* paralelo, ou seja, dos 66 textos em língua inglesa do MedlinePlus e suas respectivas traduções em língua portuguesa. Já para o levantamento do Índice Flesch dos textos escritos originalmente em língua portuguesa, foi utilizada uma amostra de 66 dos 191 textos, a fim de equiparar ao número de textos do *corpus* paralelo.

Ressaltamos que a ferramenta adaptada para a língua portuguesa (SCARTON *et al.*, 2009) possui algumas limitações em comparação com a da língua inglesa. Por exemplo, o Coh-Metrix-Port processa textos com até mil palavras, ao passo que o Coh-Metrix limita a análise a 15 mil caracteres. Apesar da diferença entre os parâmetros utilizados, observamos que o Coh-Metrix aceita textos mais longos do que o Coh-Metrix-Port. Por isso, todos os textos em inglês foram processados pelo Coh-Metrix, enquanto o Coh-Metrix-Port processou somente 59 textos do MedlinePlus (PT), já que os outros sete ultrapassavam o limite de palavras processáveis pela ferramenta. Quanto aos textos do Ministério da Saúde, todos foram processados, já que não extrapolaram o limite.

Como mencionado anteriormente, os resultados do cálculo do Índice Flesch vão de 0 a 100, variando entre ‘muito difícil’ e ‘muito fácil’. Na Tabela 2, a seguir, é possível observar a escala de valores, além de uma estimativa do nível de escolaridade que compreende cada um desses graus de inteligibilidade.

TABELA 2 – Interpretação do Índice Flesch

Valor do Índice	Descrição de Inteligibilidade	Escolaridade Estimada (EUA)	Escolaridade Estimada (BR)
0 a 29	Muito difícil	<i>College graduate</i>	Universitários*
30 a 49	Difícil	<i>13th to 16th grade</i>	EM ou universitários
50 a 59	Razoavelmente difícil	<i>10th to 12th grade</i>	EM
60 a 69	Padrão	<i>8th to 9th grade</i>	Até 8ª série do EF
70 a 79	Razoavelmente fácil	<i>7th grade</i>	Até 8ª série do EF
80 a 89	Fácil	<i>6th grade</i>	Até 8ª série do EF
90 a 100	Muito fácil	<i>5th grade</i>	Até 4ª série do EF

* Apenas para áreas acadêmicas específicas.

EM Ensino Médio

EF Ensino Fundamental

Fonte: Elaborada com base em Flesch (1949 *apud* DUBAY, 2004) e Martins *et al.* (1996).

Vale ressaltar que a tabela apresenta uma comparação somente em relação ao nível de escolaridade entre a população brasileira e a estadunidense, sem considerar fatores como nível de proficiência em leitura ou classe social dos indivíduos.

Com base nos valores de Índice Flesch levantados para cada texto que compõe os *corpora*, foram feitos os cálculos estatísticos de média, mediana, variância e desvio padrão de cada *subcorpus*, utilizando-se a ferramenta de equações do programa Excel. A média e a mediana são medidas de tendência central: a média consiste na soma de todos os elementos em análise (os índices) divididos pelo número total de elementos (número de textos processados); e a mediana é o valor central da série. Já a variância e o desvio padrão são medidas de dispersão, servindo para apontar a variabilidade dos dados em torno da média. A variância mostra o quão distantes os valores estão da média, sendo calculada a partir da soma dos quadrados da diferença entre cada valor e a média, dividida pelo número total de elementos. O desvio padrão é a medida do grau de dispersão em relação à média. Para calculá-lo, basta extrair a raiz quadrada da variância (MORATO, 2011).

As medidas de dispersão dos textos em inglês do MedlinePlus (EN) foram calculadas aplicando-se a equação de população, pois todos os textos foram avaliados pelo Coh-Metrix. Já dos *subcorpora* em

português, foram calculadas as medidas de dispersão de uma amostra, pois não foram avaliados todos os textos que os compõem, conforme explicamos anteriormente. Os dados estatísticos podem ser observados na Tabela 3, a seguir.

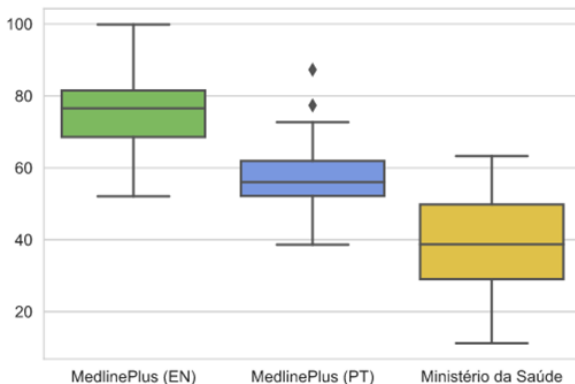
TABELA 3 – Resultado de cálculos estatísticos descritivos

	MedlinePlus (EN)	MedlinePlus (PT)	Ministério da Saúde
Média	74,845	57,659	39,115
Mediana	76,516	55,99	38,699
Variância	102,501	72,0444	170,942
Desvio padrão	10,124	8,488	13,074

Fonte: Elaborada pelas autoras.

A partir desses dados estatísticos levantados, foram construídos gráficos, a fim de facilitar a visualização e comparação entre os levantamentos dos diferentes *corpora*. O Gráfico 1, do tipo *boxplot*, apresenta visualmente os dados de mediana, que corresponde à linha horizontal no centro dos blocos, e de desvio padrão, que corresponde aos blocos coloridos. As linhas horizontais fora dos blocos representam o valor máximo e o valor mínimo observados nos textos de cada *subcorpus*. Por fim, os pequenos losangos representam valores discrepantes entre os textos pertencentes a cada amostra (*subcorpus*).

GRÁFICO 1 – *Boxplot* dos dados do Índice Flesch



Fonte: Elaborado pelas autoras.

A partir dos dados apresentados, é possível observar que as médias calculadas para o Índice Flesch apontam os textos do MedlinePlus (EN) com índices de inteligibilidade mais altos (média de 74,845), os do *subcorpus* traduzidos para o português, intermediários (57,659), e o *subcorpus* dos textos do Ministério da Saúde, originalmente escritos em português, com índices de inteligibilidade mais baixos (39,115). Além disso, os valores de desvio padrão mostram que os índices do *subcorpus* do Ministério da Saúde são os que contam com maior dispersão – de 13,074 –, seguidos pelos do MedlinePlus (EN) – de 10,124. Por último, vêm os do MedlinePlus (PT), com menor dispersão – de 8,488. O valor mínimo dos índices do MedlinePlus (EN) é de 52,05, o central (mediana) é de 76,516, e o máximo é de 99,821. No MedlinePlus (PT), o valor mínimo é de 38,654, o central é de 55,99, e o máximo, 72,367, com 77,367 e 87,276 como valores discrepantes. Para o *subcorpus* do Ministério da Saúde, observamos valor mínimo de 11,234, valor central de 38,699, e máximo de 63,272.

O acesso à informação é um direito assegurado na lei (BRASIL, 2015). Contudo, para que a população possa usufruir desse direito, devem-se garantir condições não só de acesso, mas também de compreensão desses materiais. Além disso, sendo a tradução imprescindível para a disseminação do conhecimento científico, que, muitas vezes, é transmitido em língua inglesa, ao traduzir textos com a finalidade de atingir o público geral, o tradutor deve levar em conta informações sobre o leitor a quem as traduções se destinam (NORD, 2006). Portanto, para analisar esses dados, é necessário utilizar como base informações sobre os leitores médios estadunidenses e brasileiros.

Segundo dados de 2017, 90% da população estadunidense possui, no mínimo, *High School* completo (U.S. CENSUS BUREAU, 2017). Já no Brasil, de acordo com o último censo, a parcela da população que tem grau equivalente de instrução, ou seja, Ensino Médio, não chega aos 50%. A discrepância fica ainda mais evidente se considerarmos que, enquanto 46,2% da população brasileira possui Ensino Médio ou graus mais altos de instrução, há uma considerável parcela de 33,7% que não completou ao menos o Ensino Fundamental (IBGE, 2019). Assim, ao produzir conteúdo escrito para o público geral, os órgãos competentes devem considerar esses dados, pois dentre os 53,8% da população brasileira que possui graus de instrução inferiores ao Ensino Médio, a maioria se concentra no Ensino Fundamental incompleto. Portanto, para

se mostrarem adequados ao leitor médio, os textos brasileiros deveriam se encaixar no nível ‘fácil’, ou seja, terem índices de inteligibilidade entre 70 e 100. Já os textos estadunidenses, para serem acessíveis à maior parcela da população, poderiam apresentar inteligibilidade ‘difícil’, com índices entre 30 e 60, mas nossos levantamentos mostraram que são ‘razoavelmente fáceis’, podendo atingir, inclusive, a pequena parcela da população com níveis inferiores de escolaridade. Assim, com base nos levantamentos apresentados e nos dados de escolaridade, podemos concluir que a população que mais necessitaria que as informações fossem fornecidas de maneira acessível seria, justamente, a população que a receberia com maior complexidade, ou seja, a brasileira.

A partir desses dados estatísticos de inteligibilidade textual, procederemos à análise de padrões linguísticos, utilizando os *softwares* AntConc e AntPConc, a fim de identificarmos o que ocorre em textos da área e verificarmos se tais padrões são mantidos nos textos traduzidos. Em outras palavras, pretendemos avaliar se os textos traduzidos mantêm a convencionalidade característica do gênero textual em estudo, que será analisada com base nos textos escritos originalmente em português. Nesse sentido, enfatizamos a necessidade de não apenas olhar para métricas superficiais de inteligibilidade, mas também para os padrões que são imprescindíveis para a caracterização dos gêneros e, portanto, para a sua compreensão pela comunidade discursiva a que se destina.

4.2 Levantamento de padrões linguísticos

A análise qualitativa manual dos textos de divulgação em língua portuguesa partiu das palavras-chave, ou seja, das palavras estatisticamente mais frequentes no *corpus* de estudo do que no *corpus* de referência. A decisão de olhar somente para palavras-chave das traduções se deve ao nosso interesse em buscar por padrões observados nos textos originalmente escritos em português e verificar se eles se mantêm nos textos traduzidos ou se estes indicariam quebras de convencionalidade, dificultando o entendimento do texto pelo público não especialista. Para esse contraste, compilamos textos também da área médica, contudo de um gênero distinto – artigos científicos –, a fim de que fossem destacadas as palavras estatisticamente relevantes relacionadas ao gênero dos textos, não necessariamente à área de especialidade.

Para a montagem do *corpus* de referência em português, foram selecionados 117 textos, compilados a partir da palavra de busca ‘medicina’ no Portal de Periódicos Capes, totalizando 450 mil *tokens*. O número de palavras do *corpus* de referência baseou-se nos tamanhos críticos apontados por Berber Sardinha (2004), sendo cinco vezes o tamanho do maior *subcorpus* de estudo em português.

Segundo Gabrielatos (2018), a chavicidade de um item deve ser estabelecida utilizando-se uma combinação de duas métricas que se complementam: tamanho de efeito e significância. A primeira é usada para determinar a diferença entre as frequências de determinada palavra nos dois *corpora* – quanto maior a diferença entre o *corpus* investigado e o *corpus* de referência, maior será a razão logarítmica –, enquanto a segunda aponta o tamanho dessa diferença. Nesta pesquisa, utilizamos, respectivamente, o *odds ratio* e o *log likelihood*, disponíveis no AntConc.

O *odds ratio* será levado em consideração para determinar o ponto de corte para as palavras serem ou não consideradas chave, estabelecido em 10. O *log-likelihood* (Keyness, no AntConc) foi estabelecido em $p < 0,0001$, indicando uma margem de erro de até 0,01% para as palavras-chave terem sido levantadas erroneamente. De acordo com Brezina (2018), para pesquisas na área de Ciências Humanas, é aceitável utilizar até $p < 0,05$. O *log-likelihood* foi o índice adotado para a organização das palavras-chave, apresentadas em ordem decrescente de chavicidade. De acordo com Pojanapunya e Todd (2018), o teste *log-likelihood* enfatiza palavras relativamente comuns, que servem ao propósito de pesquisas orientadas pelo gênero – interesse deste estudo –, ao passo que o *odds ratio* dá ênfase a palavras mais especializadas, que são mais adequadas para pesquisas terminológicas, por exemplo.

Por não se tratar de um *corpus* etiquetado morfossintaticamente e devido ao número de flexões possíveis na língua portuguesa (em substantivos, adjetivos, artigos e verbos), optamos por fazer a lematização manual das palavras-chave do *corpus* comparável. As palavras foram agrupadas sob aquela com o valor de *log-likelihood* mais alto. Por exemplo, os itens do *subcorpus* do MedlinePlus (PT) ‘precisa’ (com *log-likelihood* de 95,08) e ‘precisará’ (45,34) foram agrupadas sob ‘precisar’ (97,83).

Palavras homógrafas foram diferenciadas manualmente. Por exemplo, ‘sente’, que pode ser conjugação dos verbos ‘sentir’ e ‘sentar’, teve suas frequências mantidas separadamente. Após a distinção do número de ocorrências referentes a cada um dos verbos, foi recalculado

o *odds ratio* de cada uma das formas, utilizando-se uma calculadora de chavicidade e efeito,⁷ a fim de determinar se esses itens seguiriam ou não sendo palavras-chave, de acordo com os critérios estipulados.

Após a lematização manual, a lista de palavras-chave exclusivas do MedlinePlus (PT) mostrou 167 itens; a lista exclusiva do Ministério da Saúde, 115. Uma amostra da lista de palavras-chave do *corpus* comparável, organizada por chavicidade, pode ser observada na Tabela 4.

TABELA 4 – Amostra de palavras-chave exclusivas do *corpus* comparável

Posição	MedlinePlus (PT)		Ministério da Saúde	
	Frequência	Palavras-chave	Frequência	Palavras-chave
1	393	seu	153 [+25]	alimentos; alimento
2	251	médico	135	evitar
3	154	dor	48	camisinha
4	134	cirurgia	38	lixo
5	86 [+35]	tomar; tome	45	dentes
6	64	ligue	58	mulher
7	93	poderá	49	pé
8	100	fazer	35 [+16]	picada; picadas
9	116	pessoas	38	roupas
10	92	depois	43 [+21]	acidentes; acidente
11	52 [+20]	incisão; incisões	39	coluna
12	44	catapora	38	objetos
13	83	sinais	40	provocar
14	56	reação	25	manchas
15	35	vaginal	37	passo
16	55	use	24	hpv
17	35	alérgica	27	nariz
18	62	semanas	25	chão
19	48 [+13; +12; +36; +11; +12]	ajuda; ajudá[-lo/-la]; ajudam; ajudar; ajudará; ajude	36	veias
20	36	converse	22 [+18]	joelhos; joelho

Fonte: elaborada com base em AntConc (ANTHONY, 2019).

⁷ Disponível em: <http://ucrel.lancs.ac.uk/llwizard.html>. Acesso em: 15 maio 2020.

A partir da lista de palavras-chave, levantamos os itens que ocorreram estatisticamente com mais frequência nos textos traduzidos, sendo excluídos, contudo, aqueles que denominam doenças, como ‘catapora’ e ‘glaucoma’, sintomas de doenças, como ‘tontura’ e ‘erupção’, e partes do corpo, já que se referem a questões específicas, que podem não ter sido abordadas na mesma proporção nos dois *subcorpora*. Privilegiamos, portanto, as palavras-chave que, em tese, poderiam ocorrer em qualquer texto de divulgação da área médica, mas que não aparecem na lista de palavras-chave do *subcorpus* de textos originais em português.

Devido à limitação de espaço, neste artigo focaremos (i) o pronome adjetivo possessivo ‘seu’, palavra com maior chavicidade nos textos traduzidos; (ii) o substantivo plural ‘sinais’, na décima terceira posição, que é o primeiro substantivo exclusivo do *subcorpus* traduzido que se enquadra nos pré-requisitos da análise (o item ‘pessoas’ aparece na forma singular nas palavras-chave do *subcorpus* de originais em português); e (iii) o verbo no imperativo ‘use’, na décima sexta, já que, além de ser palavra-chave apenas dos textos traduzidos, é um entre os 28 verbos no modo imperativo dessa lista, enquanto a de palavras-chave dos textos originais compreende apenas dois – ‘utilize’ e ‘retire’.

4.2.1 Análise de ‘seu’

Como é possível observar na amostra de palavras-chave (TABELA 4), ‘seu’ é a primeira palavra estatisticamente relevante característica do *subcorpus* do MedlinePlus (PT), com 393 ocorrências (99,55 a cada 10.000). Para entender o seu papel nesse *corpus*, partimos para a análise dos colocados da palavra. Primeiro, buscamos os colocados em uma janela de 4 palavras à direita. A partir da lista de colocados, foi possível observar que as palavras que mais recorrem à direita de ‘seu’ são ‘médico’ (176 ocorrências), seguida de ‘bebê’ (48) e ‘filho’ (30). Já nos textos originais em português, há 193 ocorrências de ‘seu’, e, na mesma janela de até 4 palavras à direita, o substantivo ‘médico’ é colocado desse pronome adjetivo possessivo apenas 25 vezes. Nesse *subcorpus* (MS), o colocado mais frequente à esquerda de médico é ‘o’ (34 ocorrências). Esses índices parecem ainda mais discrepantes se considerarmos que o *subcorpus* do MS tem mais do que o dobro de palavras do MedlinePlus (PT).

A fim de confirmarmos se a alta recorrência da colocação ‘seu médico’ no MedlinePlus (PT) decorre da influência dos textos de partida, buscamos os colocados imediatamente à esquerda de ‘doctor’ – ou seja,

janela de 1 item. Os resultados corroboraram a hipótese levantada, mostrando que nos textos o colocado ‘*your*’ aparece com alta frequência com ‘*doctor*’ – 204 ocorrências. A Tabela 5 sintetiza esses dados.

TABELA 5 – Colocados imediatamente à esquerda de ‘médico’ e ‘*doctor*’ nos *corpora* da pesquisa

MedlinePlus (EN)			MedlinePlus (PT)			Ministério da Saúde		
Freq.	Colocado	Item de busca	Freq.	Colocado	Item de busca	Freq.	Colocado	Item de busca
204	<i>your</i>	doctor	168	seu	médico	27	o	médico
21	<i>the</i>		53	o		25	um	
8	<i>baby's</i>		8	ao		23	seu	
			6	um		17	pelo	
					7	ao		
					6	atendimento		
					6	do		

Fonte: elaborada com base em AntConc (ANTHONY, 2019).

A partir desse levantamento, foi possível observar que, apesar de ‘médico’ ser colocado de ‘seu’ no *subcorpus* do Ministério da Saúde, na verdade, ‘o médico’ e ‘um médico’ são mais utilizadas do que ‘seu médico’. Já no *subcorpus* de textos traduzidos, a sequência mais utilizada é ‘seu médico’, enquanto ‘o médico’ ocorre apenas 53 vezes, quase $\frac{1}{3}$ das ocorrências daquela. Essas duas sequências mais utilizadas vão ao encontro do que ocorre no *subcorpus* dos textos originais, sendo ‘*your doctor*’ a sequência mais utilizada, com 204 ocorrências, e ‘*the doctor*’, a segunda mais utilizada, com 21 ocorrências. Também é possível observar que, no MedlinePlus (PT), ‘o médico’ ocorre mais que o dobro de vezes de ‘*the doctor*’ no Medline (EN), sendo a escolha de tradução para ‘*your doctor*’ em 43 ocorrências.

4.2.2 Análise de ‘use’

A forma verbal ‘use’ tem 55 ocorrências (13,93 a cada 10.000 palavras) no *subcorpus* do MedlinePlus (PT), fazendo dela uma palavra-chave. A partir das linhas de concordância, foi possível observar construções de ‘use’ com diversos substantivos. Por exemplo, observamos

a utilização de sequências como ‘use absorventes’ (4 ocorrências), ‘use tampões’ (4 ocorrências) e ‘use preservativo(s)’ (2 ocorrências). Algumas linhas de concordância que exemplificam o tipo de orientações são apresentadas na Figura 1.

1 água morna. Sempre que for ao banheiro, use a garrafa de plástico para esguichar água
 2 no diário. 9 comparar o sódio nos alimentos Use a Informação Nutricional na embalagem
 3 e. Não use absorventes internos (tampões). Use absorventes externos. • Os seios ficarã
 4 com escuro e em seguida transparente. Não use absorventes internos (tampões). Use at
 5 vel. Use calcinha de algodão. • Absorventes Use absorventes se houver muita secreção.
 6 escuro e depois incolor. Não use tampões. Use absorventes íntimos. • Seus seios se en
 7 umidos em pequenas quantidades, mas não use adoçantes à base de sacarina (Sweet ‘N
 8 inchaço na pele ao redor dos testículos. 2. Use as duas mãos para tocar cada testículo
 9 ser tão simples quanto usar estas 10 dicas. Use as ideias nesta lista para equilibrar sua:
 10 elásticas para reduzir o inchaço. Neste caso, use as meias durante o dia e remova-
 11 não se mexa. (Se tiver uma contração, use as técnicas de respiração e relaxamentc
 12 . Não tome banho quente ou frio. Não use banheira de água quente, spa ou hidror
 13 . Mantenha a área tão seca quanto possível. Use calcinha de algodão. • Absorventes Use
 14 tocar carne crua. Cozinhe bem as carnes. • Use cintos de segurança abaixo da linha da
 15 zirá a rigidez da articulação. Dormir • Não use colchões de água enquanto o médico n

Fonte: AntConc (ANTHONY, 2019).

Como ‘use’ não aparece entre as palavras-chave do Ministério da Saúde, observamos a lista de palavras-chave a fim de identificar que outra possibilidade poderia ter sido a escolhida pelos autores. Identificamos, então, a palavra-chave ‘utilize’, que ocorre 17 vezes nesse *subcorpus* (MS) (2,02 a cada 10.000), enquanto ‘use’ ocorre 41 vezes (4,88 a cada 10.000) – mas pode ter ficado de fora das palavras-chave por ser também recorrente no *corpus* de referência. De fato, diversos dicionários, como, por exemplo, o Dicionário Online de Português⁸ e o Caldas Aulete Digital,⁹ apresentam ‘utilizar’ como sinônimo de ‘usar’. Supomos, assim, que a maior recorrência de ‘use’ nos textos traduzidos pudesse ser decorrente da forma cognata ‘use’ dos textos em inglês. A fim

⁸ Disponível em: <https://www.dicio.com.br/usar/>. Acesso em: 26 set. 2020.

⁹ Disponível em: <http://www.aulete.com.br/usar>. Acesso em: 26 set. 2020.

de confirmar ou refutar tal hipótese, procedemos com o alinhamento do *corpus* paralelo, por meio da ferramenta LF Aligner (FARKAS, 2018), e o investigamos usando o *software* AntPConc (ANTHONY, 2017). A Figura 2 apresenta uma amostra do alinhamento de ‘use’, partindo dos textos em inglês.

FIGURA 2 – Linhas de concordância de ‘use’ no *corpus* paralelo do MedlinePlus (PT-EN)

Line	KWIC
1	Use MyPlate to build your healthy eating style ar
2	Use a clean part of the washcloth and wash the
3	Use a nipple for your baby’s age.
4	• Use a piece of paper and a pen to mark
5	Use a pillow or folded blanket across your abdon
6	Use a pillow or folded blanket over your incision
7	• Use a sitz bath to relieve discomfort.
8	ess than 20 feet from any window, door, or vent. Use an extension cord that is more than 20 feet l
9	• Use an inflatable, donut-shaped, ring when sittin
Line	Reference
1	Fazer escolhas alimentares para uma vida saudável pode ser tão simples quanto usar estas 10 dicas.
2	Use uma parte limpa da toalha de mão para lavar o outro olho.
3	Use um bico compatível com a idade do bebê.
4	• Use uma folha de papel e uma caneta e anote os movimentos.
5	Use um travesseiro ou manta dobrada sobre o abdômen ou peito para proteger as incisões quando tossir.
6	Use um travesseiro ou manta dobrada sobre a incisão como apoio enquanto estiver respirando profundamente ou tossindo.
7	• Use um banho de assento para aliviar o desconforto.
8	Não instale geradores, lavadoras de alta pressão ou motores à gasolina a uma distância menor do que 6 metros de qualquer janela, porta ou saída de ar.
9	• Use um anel inflável como uma pequena bóia quando estiver sentada.

Fonte: AntPConc (ANTHONY, 2017).

Antes de procedermos para a análise das linhas em paralelo, é importante ressaltar que a forma ‘use’, em inglês, pode remeter ao substantivo e ao verbo em forma de infinitivo, além das conjugações em primeira e segunda pessoas do singular e das três pessoas do plural. A análise das 90 ocorrências de *use*, em inglês, revelou que a palavra foi traduzida como (i) diferentes formas do verbo ‘usar’, ‘utilizar’ e ‘consumir’ (62, 3 e 1 ocorrências, respectivamente), (ii) os substantivos ‘uso’ e ‘consumo’ (7 e 2 ocorrências, respectivamente), (iii) por paráfrases – por exemplo, a frase “*Each time you use the toilet*” foi traduzida por

“Todas as vezes que for ao banheiro” – (7 ocorrências). Para as restantes oito ocorrências de ‘use’, não foram identificados equivalentes.

A busca em direção contrária revelou que as 55 ocorrências de ‘use’ nas traduções são provenientes de ‘use’ (41 ocorrências), ‘wear’ (6 ocorrências), ‘take’ (1 ocorrência), ‘spend’ (1 ocorrência), ‘douche’ – traduzido por ‘use ducha vaginal’ – (1 ocorrência), e cinco das suas ocorrências não foram traduzidas. Já a forma ‘utilize’ foi a escolha tradutória para apenas uma frase – “*Never use a generator inside your home...*” [“Nunca utilize este gerador dentro de casa...”]. Portanto, podemos concluir que ‘use’ é a tradução mais recorrente do cognato ‘use’, enquanto ‘utilize’ e outras possibilidades são preteridas nas traduções.

Diferentemente do que foi observado no *subcorpus* traduzido, há recorrência tanto de ‘use’ quanto de ‘utilize’ nos textos originalmente escritos em português. A palavra ‘use’ ocorre 41 vezes (4,88 a cada 10.000) e ‘utilize’, 17 (2,02 a cada 10.000). A fim de verificar se existe diferença de uso entre esses dois verbos no *subcorpus* de textos originais em português, observamos seus colocados na janela de 4 itens, mais especificamente os substantivos à direita dos verbos. Esse levantamento mostrou que as ocorrências de ‘use’ no *subcorpus* do Ministério da Saúde estão bastante ligadas a objetos que são portados (‘chapéu’, ‘cintos’, ‘roupas’, ‘sapatos’ e ‘óculos escuros’), medicamentos e, como observado anteriormente, a forma verbal ‘use’ aparece próxima à palavra ‘camisinha’.

Já a palavra ‘utilize’, no *subcorpus* do Ministério da Saúde, ocorre em contextos semelhantes aos de ‘use’ no *subcorpus* do MedlinePlus (PT). Aparecem formulações como ‘não utilize’, ‘nunca utilize’ e ‘utilize sempre’, seguidos de alguns substantivos, porém sem recorrências, conforme as linhas de concordância a seguir (FIGURA 3).

FIGURA 3 – Linhas de concordância de ‘utilize’ no *subcorpus* do MedlinePlus (PT)

1	agens estragadas, sem rótulo ou bula; - não utilize a mesma receita médica mais de uma
2	quirir algo deficientemente projetado; - não utilize apoio de pulso durante a digitação, pois
3	armários que estão no alto; - no piso, utilize ceras que após a aplicação não deixem
4	s da barriga e das nádegas periodicamente; utilize esta técnica de relaxamento quando quiser aliviar
5	brindo as torneiras e dando descargas. Não utilize esta água para uso pessoal e outros
6	ocicleta, de forma segura. Se for necessário, utilize o transporte público (táxi ou ônibus); - não
7	use as escadas, ou então nem o utilize! Os 10 mandamentos do coração saudável: - evite
8	Salão de beleza: utilize sem prejudicar sua saúde O salão deve
9	para o uso! Cuidados durante a limpeza: - utilize sempre luvas no preparo da solução diluída
10	pegá-lo para brincar e danificá-lo; - utilize sempre pilhas adequadas para aparelhos auditivos.
11	nais objetos, como pneus velhos, lixo, etc.; - utilize telas em janelas e portas, use roupas
12	fique com os braços junto ao corpo. Utilize um suporte para que o texto fique
13	cia, com amortecimento. Para caminhadas, utilize um tênis adequado. Como sentar-se adequadament
14	o material não deslizante; - ao tomar banho, utilize uma cadeira de plástico firme com cerca
15	até a altura da cabeça. Se necessário, utilize uma escada, banco ou estrado. Também é
16	ou spray na direção do rosto; - não utilize xícaras, copos ou colheres de uso doméstico
17	para guardá-lo caso seja necessário; - nunca utilize álcool ou outras substâncias para limpá-lo;

Fonte: AntConc (ANTHONY, 2019).

Com exceção das linhas 3 e 6, nas quais ‘utilize’ está associado a itens de vestuário, as outras ocorrências se combinam a objetos que servirão para ajudar a cumprir determinada ação. Por exemplo, na linha 1, recomenda-se utilizar uma cadeira de plástico como auxílio para tomar banho. Contudo, devido ao tamanho reduzido dos *corpora* compilados para este estudo, julgamos inapropriado traçarmos generalizações sobre os contextos de ocorrência de ‘use’ e ‘utilize’. Por isso, recorreremos a um *corpus* de grandes proporções, mas de língua geral, o Corpus do Português (DAVIES, 2015). Esse *corpus*, que está disponível gratuitamente on-line, permite que seja feita a comparação entre os colocados de duas palavras por meio da ferramenta *Compare*.

Apesar de se tratar de um *corpus* geral de língua portuguesa, que abrange textos de diferentes países falantes da língua, o *subcorpus* Web/Dialects permite que a busca seja feita apenas em textos em português brasileiro, com cerca de 655 milhões de palavras. Para traçar um paralelo com os levantamentos do AntConc, optamos por levantar somente os substantivos que aparecem à direita dos verbos. A janela escolhida foi, também, de até 4 palavras.

A Tabela 6 apresenta os 25 primeiros substantivos que ocorrem à direita dos verbos ‘use’ e ‘utilize’, organizados pelo valor do índice (*score*) de relação entre o colocado e a palavra de busca.

TABELA 6 – Colocados de ‘use’ e ‘utilize’ no *subcorpus* Web/Dialects

Colocado de ‘use’	Freq.	Score	Colocado de ‘utilize’	Freq.	Score
curador	357	216,5	selo	101	333,1
bálsamo	211	128	créditos	136	224,3
camisinha	206	124,9	fins	575	99,8
comentários	88	53,4	login	106	87,4
autoridade	67	40,6	polegares	26	85,8
box	119	36,1	palhas	11	72,6
mente	44	26,7	pagamento	20	66
blusas	40	24,3	responsabilidade	114	62,7
maquiagem	66	20	fórum	26	42,9
vestido	33	20	x	141	42,3
mouse	237	18	email	123	14
perfume	29	17,6	sistemas	12	13,2
criatividade	342	17,3	formulário	139	10,2
kit	114	17,3	formas	33	8,4
instrumentos	104	15,8	rolagem	24	7,9
chicote	25	15,2	senha	118	7,6
calça	48	14,6	aparelhos	29	7,4
jeans	24	14,6	botão	321	6,9
saias	24	14,6	dispositivo	10	6,6
branco	23	13,9	seção	13	6,1
talento	22	13,3	forma	149	6,1
vestidos	22	13,3	letras	54	5,9
chapéu	21	12,7	código	117	5,8
salto	21	12,7	tag	12	5,7
seleção	21	12,7	aço	13	5,4

Fonte: Corpus do Português (DAVIES, 2015).

Corroborando os resultados observados no *subcorpus* de textos originais em português, no *subcorpus* de língua geral ‘use’ se associa

majoritariamente a objetos que são portados, especialmente peças de vestuário – ‘blusas’, ‘maquiagem’, ‘vestido’, ‘instrumentos’, ‘calça’, ‘jeans’, ‘saias’, ‘vestidos’, ‘chapéu’ e ‘salto’. Já a forma verbal ‘utilize’ é empregada com referência a palavras como ‘créditos’, ‘pagamento’, ‘responsabilidade’, ‘sistemas’, ‘aparelhos’, ‘dispositivo’, ‘forma’, ‘letras’, dentre outras. Esses resultados indicam que nem sempre o equivalente ‘use’ é a opção mais convencional para recuperar ‘use’ nos textos traduzidos, mas que outras possibilidades – ‘utilize’, além de paráfrases – deveriam ser consideradas a fim de se manter a convencionalidade observada em textos desse gênero escritos originalmente em português.

4.2.3 *Análise de ‘sinais’*

Diante da alta chavidade de ‘sinais’ nos textos traduzidos, diferentemente do que ocorre nos textos originalmente escritos em português, selecionamos mais essa palavra para fazermos uma análise quantitativa, a fim de identificar o porquê da discrepância.

A partir do alinhamento dos textos em inglês e suas traduções, buscamos identificar a(s) palavra(s) que tivesse(m) originado a tradução ‘sinais’. Observamos, então, que das 83 ocorrências de ‘sinais’ no *subcorpus* de traduções, apenas 11 não partiram de ‘signs’, mas foram resultado de explicitações, escolhas tradutórias de ‘changes’ e ‘cues’, ou simplesmente não tiveram o excerto de partida identificado.

Por meio desse levantamento, verificamos que, das 121 ocorrências (34,81 a cada 10.000) de ‘sign(s)’ como substantivo – observamos, também, duas ocorrências do verbo ‘sign’ [assinar] –, 83 foram traduzidas por ‘sinal’/‘sinais’, 31 por ‘sintomas’, e as restantes foram traduzidas por ‘efeitos’ (1 ocorrência), ou simplesmente não foram traduzidas (7 ocorrências).

Verificamos, posteriormente, que a palavra ‘sintomas’ ocorre nos textos traduzidos 52 vezes no total (13,17 a cada 10.000) – sendo 3 vezes na forma singular ‘sintoma’. Por observar que havia ainda 21 ocorrências da palavra com origens desconhecidas, alinhamos originais e traduções a partir da palavra de busca ‘sintoma(s)’. Das 21 ocorrências, 4 são traduções do cognato ‘symptoms’. Outras ocorrências têm origem em frases nas quais não se utiliza a palavra ‘signs’ nem ‘symptoms’, ou seja, Ø (10 ocorrências) – podendo ser observada na Figura 4. Por fim, há 7 ocorrências para as quais não foram localizadas correspondentes em inglês.

FIGURA 4 – Linhas de concordância de ‘sintomas’ no *corpus* paralelo do MedlinePlus (PT-EN)

Line	KWIC
38	spirar, batimentos cardíacos acelerados, tontura e fraqueza. Esses sintomas normalmente cc
39	spirar, batimentos cardíacos acelerados, tontura e fraqueza. Esses sintomas normalmente cc
40	• sintomas parecidos com c
41	es que sejam causadas por um vírus não afetado pela vacina ou • sintomas parecidos com ã
42	e um pronto-socorro ou entre em contato com seu médico se os sintomas piorarem ou se v
43	Contate seu médico caso estes sintomas piorem ou não c
44	Sarampo O vírus do sarampo causa sintomas que podem ind
Line	Reference
38	
39	
40	
41	
42	• Return to the Emergency Department or call your doctor if your signs get worse or you have a fever of more than 100.5 degrees F or 38 degrees C.
43	Call your doctor if this gets worse or does not go away in a few weeks.
44	MEASLES (M) can cause fever, cough, runny nose, and red, watery eyes, commonly followed by a rash that covers the whole body.

Fonte: AntPConc (ANTHONY, 2017).

Vale ressaltar que foram encontradas apenas 4 ocorrências (1,15 a cada 10.000) de ‘*symptoms(s)*’ nos textos originais em inglês, ao passo que seu cognato em português, ‘sintoma(s)’, ocorre 288 vezes (72,96 a cada 10.000) no Ministério da Saúde, o que parece indicar que os equivalentes *prima facie sign(s)* → sinal/sinais e *symptom(s)* → sintomas não são necessariamente utilizados em contextos semelhantes. A fim de tentar mapear se ocorre alguma distinção entre os contextos de ‘sinais’ e de ‘sintomas’ entre traduções e originais em português, foi feita a pesquisa por colocados. O levantamento de colocados de ‘sinais’ foi feito respeitando-se a janela de até 4 itens à direita e à esquerda, estabelecendo a frequência mínima de 6 ocorrências.

A partir das linhas de concordância mostradas pelo AntPConc e AntConc, é possível observar que os usos de ‘sinais’ e de ‘sintomas’ parecem estar indiscriminados no *subcorpus* traduzido. Ou seja, não é possível distinguir os contextos em que os substantivos aparecem. Assim como aparece próximo a nomes de doenças (como ‘catarata’, ‘asma’ e ‘ataque cardíaco’), ela também aparece perto de indícios de problemas de saúde (‘estresse’, ‘fome’ e ‘ruptura ou descolamento de retina’).

No *subcorpus* do Ministério da Saúde, enquanto ‘sintoma(s)’ ocorre 288 vezes (34,25 a cada 10.000), ‘sinal/sinais’ ocorre 54 vezes (6,42 a cada 10.000). Vale enfatizar que a palavra ‘sintomas’ aparece com frequência como um subtítulo dos textos, fazendo com que sua frequência aumente significativamente.

Em razão do número elevado de ocorrências, investigamos o motivo pelo qual a palavra não consta na lista de palavras-chave desse *corpus*. A partir de sua frequência, os índices da palavra foram determinados a partir da calculadora de chavidade e efeito. O *log-likelihood* da palavra é de 354,16 e o *odds ratio* é de 5,8163, que não alcança o ponto de corte para figurar na lista de palavras-chave. No *corpus* de referência de artigos científicos, a palavra ‘sintomas’ tem 280 ocorrências (6,19 a cada 10.000). Portanto, o *odds ratio* da palavra acabou sendo neutralizado, fazendo com que ela não fosse considerada chave.

Há ocorrências da palavra ‘sinais’ que se dão na sequência ‘sinais e sintomas’ (16 vezes). Essa sequência é, muitas vezes, utilizada como um subtítulo para organizar as partes do texto, como mencionado anteriormente sobre a palavra ‘sintomas’.

As linhas de concordância de ‘sinais’ no *subcorpus* do Ministério da Saúde parecem indicar aspectos em comum em seus contextos de uso (FIGURA 5).

FIGURA 5 – Linhas de concordância de ‘sinais’ no *subcorpus* do Ministério da Saúde

9	. Sintomas e sinais de alerta : Muitos sintomas são comuns aos
10	desidratação. Sinais de desidratação : - olhos fundos; - ausência
11	- observar os sinais de desidratação . Sinais de desidratação: - o
12	menstrual e sinais de desnutrição . Diagnóstico : A doença só p
13	tiverem dado sinais de erupção , é necessário procurar o dentista
14	oidamente os sinais de gravidade da doença, a tratar adequadarr
40	tistem outros sinais que indiquem que a fraqueza é ou
41	Os primeiros sinais são: fraqueza, transpiração , palidez,
42	bilização dos sinais vitais . Lembre-se : Não abra mão da

Fonte: AntConc (ANTHONY, 2019).

Podemos observar que, por exemplo, em ‘sinais de desidratação’, a primeira característica é ‘olhos fundos’. Em outro momento, é possível

ver a frase ‘Os primeiros sinais são: fraqueza, transpiração, palidez [...]’. Aqui, podemos observar que se usa ‘sinais’ para aspectos visíveis (‘fraqueza’, ‘transpiração’ e ‘palidez’).

Outro aspecto é que, em geral, a palavra ‘sinais’ não aparece associada diretamente às doenças, como no *subcorpus* do MedlinePlus (‘sinais de asma’; ‘sinais de doença arterial coronariana’; ‘sinais de derrame’; ‘sinais de glaucoma’; dentre outros). Pode-se observar que a palavra ‘sinais’ é mais associada a manifestações de problemas de saúde, por exemplo, ‘sinais de desidratação’, ‘sinais de desnutrição’, ‘sinais de erupção’, dentre outros.

Em um primeiro momento, apenas pelas linhas de concordância, é possível observar que, de fato, há uma distinção entre ‘sinais’ e ‘sintomas’. A fim de comprovar a existência dessa distinção, optamos por fazer este mesmo levantamento em um *corpus* de língua geral.

Utilizamos, novamente, os textos brasileiros do Corpus do Português, cujo papel, nesse caso, é de extrema relevância para comprovar os contextos convencionais de uso dessas palavras que são familiares ao público geral.

Assim como o levantamento de colocados feito nos *subcorpora* de estudo, mantivemos como padrão a janela de até 4 palavras. Esse levantamento foi feito apenas para colocados à direita das palavras ‘sinais’ e ‘sintomas’. Os primeiros 25 colocados de ‘sinais’ e de ‘sintomas’ que ocorrem nessa janela estão listados na Tabela 7. Pode-se observar suas frequências de co-ocorrência e o resultado estatístico da associação entre as palavras.

TABELA 7 – Colocados de ‘sinais’ e ‘sintomas’ no *subcorpus* Web/Dialects

Colocado de ‘sinais’	Freq.	Score	Colocado de ‘sintomas’	Freq.	Score
prodígios	491	983,9	TPM	159	317,4
tempos	354	709,4	psicóticos	157	313,4
maravilhas	277	555,1	hemolítico	112	223,6
vitais	243	486,9	urêmico	110	219,6
trânsito	233	466,9	sujeitos	93	185,6
elétricos	181	362,7	poliúria	84	167,7
libras	152	304,6	relatadas	76	151,7
rádio	149	298,6	artrite	150	149,7

distintivos	133	266,5	TDAH	75	149,7
pontuação	133	266,5	mosquito	62	123,8
emitidos	109	218,4	tratamentos	57	113,8
aparições	98	196,4	fibromialgia	51	101,8
gráficos	86	172,3	neuróticos	51	101,8
céu	81	162,3	queixas	87	86,8
seguirão	81	162,3	taquicardia	43	85,8
digitais	80	160,3	menopausa	167	83,3
sol	77	154,3	gastrite	39	77,8
vinda	74	148,3	cabeça	72	71,9
luminosos	147	147,3	ascensão	34	67,9
enviados	70	140,3	histéricos	34	67,9
sonoros	70	140,3	sinto	66	65,9
arrombamento	68	136,3	psiquiátricos	33	65,9
milagres	131	131,3	duram	32	63,9
espécie	112	112,2	alérgicos	31	61,9
terra	56	112,2	pré-menstrual	30	59,9

Fonte: Corpus do Português (DAVIES, 2015).

Por meio do levantamento, foi possível confirmar que há distinção entre as manifestações de ‘sinais’ e de ‘sintomas’ no *corpus* de língua geral. A palavra ‘sinais’ aparece associada a ‘trânsito’, ‘aparições’, ‘gráficos’ e ‘arrombamento’. Isso dá indícios de que a maneira como os ‘sinais’ se manifestam é de forma visual. Ou seja, são traços que você pode observar e enxergar. Por exemplo, ‘sinais de arrombamento’ em uma casa podem ser portas quebradas, com vestígios de terem sido forçadas por alguém, bagunça nos cômodos, demonstrando que alguém esteve por ali procurando algo. Assim, de acordo com os colocados observados, ‘sinais’ são traços que podem ser identificados por uma terceira pessoa observadora.

Além disso, podem ser observados colocados de ‘sinais’ mais relacionados à noção física de um conjunto que carrega informações ou dados. Aparecem colocados como ‘elétricos’, ‘rádio’, ‘emitidos’, ‘digitais’, ‘luminosos’, ‘enviados’ e ‘sonoros’, que indicam haver uma influência de contextos mais especializados.

Já associadas à palavra ‘sintomas’ podemos observar ocorrências como ‘psicótico’, ‘poliúria’, ‘neuróticos’, ‘taquicardia’, ‘histéricos’,

‘psiquiátricos’, ‘alérgicos’ e ‘pré-menstrual’. Ou seja, essas são manifestações que as pessoas sentem mais do que visualizam. Assim, em oposição às manifestações de sinais, que podem ser observadas por uma terceira pessoa, as manifestações de sintomas parecem ser mais facilmente identificadas pela própria pessoa. Até porque, no geral, antes de procurar atendimento que confirme a doença, é necessário que o paciente reconheça os sintomas, para então partir para a análise de uma terceira pessoa (médico), para quem serão relatadas as manifestações. Além dessas manifestações, há algumas doenças e distúrbios que estão na lista de colocados, como ‘[síndrome] hemolítico urêmica’, ‘artrite’, ‘TDAH’ (Transtorno do Déficit de Atenção com Hiperatividade), ‘fibromialgia’ e ‘gastrite’.

No *subcorpus* traduzido do MedlinePlus, a palavra ‘sinais’ parece ser utilizada de forma indiscriminada, aparecendo diversas vezes como um falso sinônimo de ‘sintomas’ e como equivalente de ‘*signs*’. Já a palavra ‘sintomas’, por se tratar de uma palavra motivada pelo uso de ‘*symptoms*’ ou por Ø, está sendo empregada seguindo um padrão semelhante aos aspectos aqui apresentados.

4.3 Discussão

Retomando as médias de Índice Flesch, para o *corpus* comparável, obtiveram-se índices de 57,659 para o *subcorpus* do MedlinePlus (PT) e 39,115 para o do Ministério da Saúde; para a língua inglesa, a média observada foi de 74,845 para o *subcorpus* do MedlinePlus (EN) – lembrando que índices mais próximos de 100 apontam para maior grau de facilidade, enquanto mais próximos de 0 demonstram maior grau de dificuldade. Com base nos níveis de escolaridade das populações estadunidense e brasileira, o índice de inteligibilidade dos textos em português não estão adequados para o seu público geral, ao passo que até mesmo o norte-americano com pouca escolaridade seria capaz de compreender os textos em inglês. O intervalo mais adequado dos índices para os textos em português seria de classificação ‘fácil’, entre 70 e 100. Já os textos em inglês, para serem acessíveis à maior parcela da população estadunidense, poderiam apresentar inteligibilidade ‘difícil’, com índices entre 30 e 60.

Enfatizamos que o levantamento quantitativo de inteligibilidade parte de noções superficiais do texto, quais sejam, o comprimento médio de palavras e de sentenças. Por esse motivo, o Índice Flesch

por si só não pode determinar com precisão o nível de dificuldade de um texto. Pode-se depreender, a partir dos resultados referentes a esse índice, que os textos originalmente escritos em português contam com estruturas, de forma geral – levando em consideração especificamente as palavras e as frases –, mais longas do que os originais em inglês e suas traduções. Entretanto, a investigação sobre a convencionalidade das traduções, por ser uma análise de padrões linguísticos desenvolvida em comparação com textos escritos originalmente na língua portuguesa, teve o papel imprescindível de indicar com mais precisão aspectos que podem dificultar a compreensão de um texto por leitores médios. Por isso, buscamos associar o levantamento puramente estatístico à análise qualitativa, já que consideramos indispensável que haja um olhar mais aprofundado sobre as questões linguísticas do texto para determinar se este conta ou não com barreiras que influenciam sua acessibilidade. Até o momento, não há ferramentas que permitam que a investigação sobre os padrões linguísticos seja desenvolvida com precisão a partir de análises exclusivamente quantitativas.

Nesse sentido, as ocorrências de ‘use’, no MedlinePlus (PT), e de ‘utilize’, no Ministério da Saúde, ocorrem em contextos similares, em que é dada alguma orientação para o leitor do que fazer quando se deparar com determinadas situações. Descobrimos que há, de fato, distinção entre aplicações de ‘use’ e de ‘utilize’, sendo que o primeiro se refere, principalmente, a objetos (com frequência, itens de vestuário), e o segundo é empregado com noções mais abstratas (como ‘créditos’, ‘pagamento’, ‘sistemas’ etc.). Essa descoberta foi feita a partir das linhas de concordância do *subcorpus* de textos originalmente escritos em português e, posteriormente, foi validada por meio do Corpus do Português (DAVIES, 2015).

A partir dos colocados de ‘seu’, foi possível observar nos textos traduzidos grande incidência da palavra ‘médico’. Isso chamou a atenção porque, nos textos do Ministério da Saúde, os principais colocados imediatamente à esquerda de ‘médico’ são os artigos ‘o’ e ‘um’. Vale ressaltar que das 251 ocorrências (63,58 a cada 10.000) de ‘médico’ no MedlinePlus (PT), 168 estão precedidas de ‘seu’; enquanto no Ministério da Saúde, das 130 ocorrências (15,46 a cada 10.000) da palavra, apenas 23 ocorrem ao lado de ‘seu’. Isso demonstra que ocorre uma interferência do texto em inglês sobre o texto traduzido, visto que no inglês, das 247

ocorrências (71,05 a cada 10.000) de ‘*doctor*’, 204 estão antecedidas por ‘*your*’.

Outra clara interferência dos textos-fonte é o uso de ‘sinais’ e ‘sintomas’ nos textos traduzidos. Enquanto há distinção entre os usos de ‘sinais’ e ‘sintomas’ no *subcorpus* dos textos escritos originalmente em português, nas traduções há indícios de que a motivação se dê puramente pelo uso de palavras cognatas no texto-fonte. Prova disso é que a grande maioria das ocorrências de ‘*signs*’ foram traduzidas como ‘sinais’ (83 de 117 ocorrências), enquanto todas as ocorrências de ‘*symptoms*’ foram traduzidas para ‘sintomas’. Adicionalmente, quando houve uso de Ø nos originais em inglês foram traduzidas como ‘sintomas’ (ou seja, não havia emprego nem da palavra ‘*symptoms*’ nem de ‘*signs*’), aparece o emprego de ‘sintomas’ nas traduções, fazendo com que a palavra seja empregada em contextos mais similares entre si. Isso pode ser explicado pela influência do texto original sobre as decisões do tradutor, sendo que, quando há Ø, o tradutor consegue se desprender do padrão de uso da língua inglesa. Por exemplo, pode-se observar que ‘sintomas’ aparece associado a ‘raiva’, ‘nervosismo’, ‘irritabilidade’ e ‘febre’, manifestações essas que podem ser sentidas pela pessoa doente.

Pôde ser estabelecida a diferença entre os contextos de uso de ‘sinais’ e ‘sintomas’ a partir do *subcorpus* do Ministério da Saúde. Posteriormente, buscamos colocados também no Corpus do Português (DAVIES, 2015), para fazer a comprovação dos resultados em um corpo de textos que abrange diversas facetas da língua. Depreendemos que as ocorrências de ‘sinais’ estão mais ligadas a demonstrações que podem ser vistas, enquanto ‘sintomas’ se refere a sensações.

5 Considerações finais

No que diz respeito aos resultados quantitativos, foi possível traçar algumas conclusões em relação aos levantamentos de palavras-chave e n-gramas. A partir dos levantamentos de colocados das palavras-chave, concluímos que os textos traduzidos apresentavam, em diversos momentos, quebras de convencionalidade (TAGNIN, 2013), distanciando-se dos padrões utilizados nos textos escritos originalmente em português. Essas quebras de convencionalidade ocorrem devido ao uso de palavras cognatas do inglês e de traduções *prima facie*, fugindo dos padrões esperados para o português.

Vale ressaltar que quebras de convencionalidade, causadas pela influência do texto-fonte sobre o texto-alvo, podem gerar dificuldades no entendimento do texto pelo leitor médio. Como exemplo, cita-se o uso indiscriminado de ‘sinais’ e ‘sintomas’ no *subcorpus* do MedlinePlus (PT). Ao deparar com um texto destinado ao público geral tratando de doenças, é comum que parte dele foque em abordar os sintomas. Entretanto, no âmbito dos textos de divulgação da área médica, o uso de ‘sinais’ pode causar uma quebra de expectativa no leitor, podendo fazer com que ele não entenda a que tipo de manifestações o texto se refere, por exemplo.

Além disso, por estar com a mente e o olhar no texto original, o tradutor pode acabar não associando a ocorrência de uma palavra a opções de tradução que vão além do seu cognato, como se observa na tradução de ‘*use*’ pelo cognato ‘*use*’. Esse fenômeno configura o que Baker (1993) denomina “terceira língua” na tradução, que é quando o texto traduzido fica com características do texto-fonte, distanciando-se, assim, de padrões de convencionalidade da língua-alvo.

Por essas razões, consideramos que a convencionalidade, sendo analisada a partir de dados quantitativos levantados a partir de critérios delimitados, pode auxiliar na avaliação da acessibilidade de um texto com maior precisão do que somente a aplicação de métricas que partem de noções superficiais do texto – como é o caso do Índice Flesch, que se baseia somente em medidas de comprimento de frases e de palavras. Nesse sentido, salienta-se a importância do uso de uma metodologia que associe a análise quantitativa à análise qualitativa. A associação dessas investigações, possibilitada por ferramentas que fazem cálculos estatísticos somadas ao olhar do pesquisador, permite o enriquecimento dos resultados da pesquisa.

Em relação a limitações do estudo, a maior delas foi o tamanho dos *corpora*. Ao utilizar *corpora* paralelos em análises, é comum que seu tamanho não seja tão expressivo quanto o de textos escritos originalmente na língua. Isso ocorre porque o número de textos que se enquadram nos critérios de seleção acaba se reduzindo, pois necessita-se que estejam disponíveis tanto o texto-fonte quanto o texto-alvo. Devido a essa barreira, foi necessário recorrer diretamente à análise das linhas de concordância em momentos que o levantamento de colocados era inconclusivo. Também foi necessário recorrer a um *corpus* maior, de

língua geral, a fim de confirmar os padrões de colocação de determinadas palavras-chave.

Além disso, há outras limitações impostas ao estudo pelos *softwares* utilizados para analisar os *corpora*. O Coh-Metrix-Port estabelece um número máximo de palavras por texto para a análise, o que fez com que parte dos textos do *subcorpus* do MedlinePlus (PT) não fossem analisados pela ferramenta. Ademais, vale ressaltar que o *software* AntConc não realiza etiquetagem ou lematização de *corpus*; portanto, a fim de reduzir o número de palavras-chave dos *corpora*, foi necessário partir para a lematização manual.

Por fim, julgamos que a quebra de convencionalidade em traduções, conforme resultados apresentados na seção 4.2, compromete mais a compreensão do texto do que o uso de palavras ou frases consideradas longas. Portanto, acreditamos que a inteligibilidade está diretamente relacionada à convencionalidade, ou seja, à manutenção de padrões reconhecíveis pelo público leitor dos diferentes gêneros textuais.

Declaração das contribuições de cada autora

As autoras Yuli Souza Carvalho e Rozane Rodrigues Rebechi produziram colaborativamente este artigo. A pesquisa de Mestrado relatada no texto foi desenvolvida por Yuli Souza Carvalho, sob a orientação de Rozane Rodrigues Rebechi, no Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul. Yuli Souza Carvalho escreveu uma primeira versão do texto, participando da escrita de todas as seções, principalmente das seções 2, 3 e 4, revisando as versões seguintes do manuscrito, assim como formatando a versão final. Rozane Rodrigues Rebechi participou da escrita de todas as seções, principalmente do Resumo e do Abstract, da Introdução e das Considerações Finais, bem como da revisão de todas as seções do manuscrito.

Referências

ANDREETTO, M. D. *Por que os textos de divulgação são mais difíceis para aprendizes de leitura com necessidades específicas do que textos científicos?* Um estudo direcionado pelo corpus. 2013. 172f. Dissertação (Mestrado em Estudos Linguísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2013.

ANTHONY, L. *AntConc*. Versão 3.5.8. Tokyo: Waseda University, 2019.

ANTHONY, L. *AntPConc*. Versão 1.2.1. Tokyo: Waseda University, 2017.

BAKER, M. Corpus Linguistics and Translation Studies: Implications and Applications. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (eds.). *Text and Technology: In Honour of John Sinclair*. Philadelphia: John Benjamins, 1993. p. 233-250. DOI: <https://doi.org/10.1075/z.64.15bak>

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

BHATIA, V. K. *Analysing Genre: Language Use in Professional Settings*. London; New York: Routledge, 1993.

BIDERMAN, M. T. C. Estatística linguística. *Alfa*, São Paulo, v. 11, p. 117-128, 1967.

BRASIL. *Lei nº 13.146, de 6 de julho de 2015*. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência). Brasília: Secretaria-Geral da Presidência da República, 2015.

BRASIL. Ministério da Saúde. *Biblioteca Virtual em Saúde*. Brasília: Ministério da Saúde, 2018. BREZINA, V. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press, 2018. DOI: <https://doi.org/10.1017/9781316410899>

DAVIES, M. *Corpus do Português*. Provo: Brigham Young University, 2015. Disponível em: <https://www.corpusdoportugues.org/>. Acesso em: 12 jun. 2020.

DUBAY, W. H. *The Principles of Readability*. California: Impact Information, 2004.

FARKAS, A. *LF Aligner*. Versão 4.2. [s. l.]: Source Forge, 2018.

FINATTO, M. J. B. Acessibilidade textual e terminológica: promovendo a tradução intralinguística. *Revista Estudos Linguísticos*, São José do Rio Preto, v. 49, n. 1, p. 72-96, 2020. DOI: <https://doi.org/10.21165/el.v49i1.2775>

FLESCHE, R. *The Art of Readable Writing*. Nova York: Harper, 1949.

FRANKENBERG-GARCIA, A. Using a Parallel Corpus in Translation Practice and Research. In: CONFERÊNCIA DE TRADUÇÃO PORTUGUESA, 1., 2006, Caparica, Portugal. *Actas da Contrapor*. Lisboa: [S.n.], 2006. p. 142-148.

FUCHS, S. N. *Orientações culturais e suas implicações para a tradução funcionalista: um estudo na área do turismo à luz da Linguística de Corpus*. 2018. 366f. Tese (Doutorado em Estudos da Tradução) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2018.

GABRIELATOS, C. Keyness Analysis: Nature, Metrics and Techniques. In: MARCHI, A.; TAYLOR, C. (ed.). *Corpus Approaches to Discourse: A Critical Review*. London: Routledge, 2018. p. 225-258. DOI: <https://doi.org/10.4324/9781315179346-11>

GRAESSER, A. C. *et al.* *Coh-Metrix*. Version 3.0. Tennessee: University of Memphis, 2017.

GRAESSER, A. C. *et al.* *Coh-Metrix: Analysis of Text on Cohesion and Language*. *Behavioral Research Methods*, [S.l.], v. 36, n. 2, p. 193-202, 2004. DOI: <https://doi.org/10.3758/BF03195564>

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. IBGE. *Pesquisa Nacional por Amostra de Domicílios Contínua*. Educação 2018. Rio de Janeiro: IBGE, 2019.

JENKINS, J. English as a lingua franca: Interpretations and Attitudes. *World Englishes*, [S.l.], n. 28, v. 2, p. 200-207, 2009. DOI: <https://doi.org/10.1111/j.1467-971X.2009.01582.x>

KRIEGER, M. G. Divulgação científica e terminologia. In: SIMPÓSIO INTERNACIONAL DE ESTUDOS DE GÊNEROS TEXTUAIS, 5., 2009, Caxias do Sul. *Anais [...]*. Caxias do Sul: UCS, 2009. p. 1-11

MARTINS, T. B. F. *et al.* Readability Formulas Applied to Textbooks in Brazilian Portuguese. *Notas do ICMSC*, São Paulo, n. 28, p. 1-11, 1996.

MASSARANI, L.; MOREIRA, I. C. A retórica e a ciência: dos artigos originais à divulgação científica. *MultiCiência*, Campinas, n. 4, p. 1-12, 2005.

MORATO, R. G. *Conceitos básicos de Estatística Descritiva*. São Paulo: Universidade de São Paulo, 2011.

NORD, C. Loyalty and Fidelity in Specialized Translation. *Confluências*, [S.l.], n. 4, p. 29-42, 2006.

NÚCLEO INTERINSTITUCIONAL DE LINGUÍSTICA COMPUTACIONAL (NILC). *Coh-Metrix-Port*. Versão 3.0. São Paulo: Universidade de São Paulo, NILC, 2020.

POJANAPUNYA, P.; TODD, R. W. Log-Likelihood and Odds Ratio: Keyness Statistics for Different Purposes of Keyword Analysis. *Corpus Linguistics and Linguistic Theory*, [S.l.], v. 14, n. 1, p. 133-167, 2018. DOI: <https://doi.org/10.1515/cllt-2015-0030>

REBECHI, R. R. Fraseologias bilíngues português-inglês da culinária brasileira: estudo direcionado pelo corpus. In: RIBEIRO, E. S.; TABOSA, L. M. A.; SILVA, N. R. B. (org.). *Tradução em três vertentes: teoria e prática, intersemiose e Linguística de Corpus*. Mossoró: Queima-Bucha, 2017. p. 201-220.

ROSSELLI, D. The Language of Biomedical Sciences. *The Lancet*, Londres, v. 387, n. 10029, p. 1720-1721, 2016. DOI: [https://doi.org/10.1016/S0140-6736\(16\)30259-8](https://doi.org/10.1016/S0140-6736(16)30259-8)

SANTIAGO, M. S. *Redes de palavras-chave para artigos de divulgação científica da Medicina: uma proposta à luz da Terminologia*. 2007. 151f. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2007.

SCARTON, C. E.; ALMEIDA, D. M.; ALUÍSIO, S. M. Análise da inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. In: BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 7., 2009, São Carlos. *Proceedings* [...]. São Carlos: WikiCFP, v. 1, 2009. p. 1-10.

STEWART, D. Conventionality, Creativity and Translated Text: The Implications of Electronic Corpora in Translation. In: OLOHAN, M. (org.). *Intercultural Faultlines*. Manchester, Northampton: St. Jerome Publishing, 2000. p. 73-91. DOI: <https://doi.org/10.4324/9781315759951-6>

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

U.S. CENSUS BUREAU. *Current Population Survey. Annual Social and Economic Supplement*. Suitland: Census Bureau, 2017.

U.S. NATIONAL LIBRARY OF MEDICINE. *MedlinePlus*. Bethesda, U.S.: Department of Health and Human Services, 2020.

ZAMBONI, L. M. S. *Cientistas, jornalistas e a divulgação científica: subjetividade e heterogeneidade no discurso de divulgação científica*. Campinas: Autores Associados, 2001.