

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS  
DEPARTAMENTO DE LÍNGUAS MODERNAS  
CURSO DE BACHARELADO EM LETRAS - TRAD. PORTUGUÊS-INGLÊS**

**GABRIELE VASCONCELOS PAPARELLI**

**Complexidade Textual em textos de divulgação sobre a Doença de Parkinson e seu impacto  
para a tradução: um estudo baseado em *corpus***

**PORTO ALEGRE  
2018**

Gabriele Vasconcelos Paparelli

**Complexidade Textual em textos de divulgação sobre a Doença de Parkinson e seu impacto para a tradução: um estudo baseado em *corpus***

Monografia apresentada ao Curso de Letras da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção de grau de Bacharel em Letras - Trad. Português-Inglês.

Orientadora: Prof.<sup>a</sup> Dra. Rozane Rodrigues Rebechi

PORTO ALEGRE  
2018

Gabriele Vasconcelos Paparelli

**COMPLEXIDADE TEXTUAL EM TEXTOS DE DIVULGAÇÃO SOBRE A DOENÇA DE PARKINSON E SEU IMPACTO PARA A TRADUÇÃO: UM ESTUDO BASEADO EM *CORPUS***

Trabalho de conclusão de curso de graduação apresentado à Comissão de Graduação do curso de Bacharelado em Letras – Tradutor Português-Inglês – do Instituto de Letras da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel(a) em Letras.

Aprovado em: \_\_\_\_ de dezembro de 2018

**BANCA EXAMINADORA**

---

Prof<sup>a</sup>. Dr<sup>a</sup> Rozane Rodrigues Rebechi

---

Prof<sup>a</sup>. Dr Márcia Moura da Silva (UFRGS)

---

Prof. Mestre Asafe Davi Cortina Silva (Colégio Israelita Brasileiro)

## RESUMO

A doença de Parkinson (DP) é um distúrbio progressivo do sistema nervoso central que afeta o movimento e, muitas vezes, provoca tremores. Além de não ter cura, pesquisadores preveem que no futuro o número de pessoas com a doença irá aumentar, devido ao aumento da expectativa de vida. Como, em geral, cabe aos familiares sem preparo assumir a função de cuidadores, a internet acaba desempenhando o importante papel de servir como fonte de instrução para essas pessoas. Além disso, essa rede pode despertar no próprio paciente o desejo de saber mais sobre a sua doença, os tratamentos disponíveis e os estudos que estão sendo realizados. Por esse motivo, um número significativo de websites especializados, criados por estudiosos, doutores e especialistas, possuem o intuito de oferecer informações sobre essa doença. Apesar do aumento do uso da internet como forma de busca de informações sobre doenças, basta uma simples pesquisa em dissertações e artigos acadêmicos para notarmos como há poucas pesquisas sobre a complexidade textual desses websites. Sendo assim, nosso objetivo, nesta monografia, é analisar a complexidade textual de textos de divulgação da fundação Michael J. Fox e da Parkinson's UK, suas traduções para o português e cinco websites de associações de Parkinson escritos originalmente em Português. Utilizando a metodologia da Linguística de *Corpus*, nosso intuito é verificar se os textos originais em português e em inglês (estadunidense e britânico) são adequados para o seu público alvo, e se houve alteração na complexidade textual das traduções para o português. Para isso, concentramo-nos na voz passiva analítica – considerada critério de complexidade textual –, presente no *corpus* de estudo, relacionando seu uso com os resultados do índice Flesch. Apoiamo-nos na teoria funcionalista da tradução (NORD 2006, 2007 e 2012) para analisarmos as traduções dos textos para o português, pois essa teoria tem como um dos focos o público-alvo da tradução. Por fim, apresentamos o número de ocorrências de voz passiva dos textos originais e suas traduções. Quando comparadas com os textos originais, as traduções dos textos dos websites das fundações Michael J.Fox e Parkinson's UK tiveram um aumento de sentenças na voz passiva, indicando, portanto, um aumento da complexidade textual, ao menos no que tange a esse quesito.

**Palavras-Chave:** Complexidade Textual. Tradução. Doença de Parkinson. Website.

## ABSTRACT

Parkinson's disease (PD) is a progressive nervous system disorder that affects movement and causes tremors. PD does not have a cure and researchers predict that the number of people with Parkinson will grow more in the future, due to the fact that people are living longer. Often, it is a relative without training that becomes a caregiver. Thus, the information available on the internet is used as a kind of guide. Also, the patient might want to learn more about his or her disease, treatments and studies being developed it. For this reason, an expressive number of specialized websites, created by researchers, doctor and specialists, are being developed to provide information about PD. However, although the number of people who seeks online information about the disease is increasing exponentially, few are the dissertations and thesis that focus on the subject of textual complexity on these websites. Therefore, my aim is to investigate the textual complexity of the informative texts of Michael J. Fox foundation and Parkinson's UK websites, their translations to Portuguese, and five websites of Parkinson associations, originally written in Portuguese. My objective is to verify if the original texts in Portuguese and English (American and British) are suitable for the target public. Also, I want to determine if there was any change in terms of textual complexity in the translated texts. In order to do that, I analyzed sentences in the passive voice in analytical position in the study *corpora* – since it is considered a criterion of textual complexity –, relating them with the scores of the Flesch index score. This study draws on the functional theory (NORD 2006, 2007 e 2012) to analyze the Portuguese translations, because this theory has the target audience of a translation text as focus. Lastly, we present the number of occurrences of the passive voice in both the original and the translated texts. When the translations are confronted with the originals texts, I could observe an increase in the use of passive voice in both the translations of the websites Michael J.Fox and Parkinson's UK texts, which leads to higher complexity, at least in what regards this criterion.

**Keywords:** Textual Complexity. Translation. Parkinson's disease. Website.

## LISTA DE FIGURAS E QUADROS

Figura 01 — Tela principal do Coh-Metrix 3.0 .....	30
Figura 02 — Tela principal do Coh-Metrix-Port .....	31
Figura 03 — Fórmula Índice Flesch .....	32
Figura 04 — Fórmula Índice Flesch versão Português .....	32
Figura 05— Linhas de concordância dos verbos <i>ser</i> no <i>corpus</i> comparável em pt .....	35
Figura 06 — Distribuição da escolaridade dos brasileiros com 25 anos ou mais .....	37
Figura 07 — Visualização de voz passiva no AntPconc 1 .....	48
Figura 08 — Visualização de voz passiva no AntPconc 2 .....	48
Quadro 01 — <i>Corpus</i> comparável .....	26
Quadro 02 — <i>Corpus</i> Paralelo .....	27
Quadro 03 — Nomeação dos <i>corpora</i> .....	27
Quadro 04 — O uso da voz passiva analítica no <i>corpus</i> comparável .....	41
Quadro 05 — O uso da voz passiva analítica no <i>corpus</i> paralelo .....	43
Quadro 06 — Índice Flesch do <i>subcorpus</i> paralelo estadunidense e suas traduções .....	44
Quadro 07— Índice Flesch do subcorpus paralelo em inglês britânico e suas traduções .....	45
Quadro 08 — Índice Flesch do <i>corpus</i> comparável em português .....	46
Quadro 09 — Simplificação Textual – voz passiva para ativa .....	49

## **LISTA DE ABREVIATURAS**

DP	Doença de Parkinson
ABP	Associação Brasil Parkinson
APARS	Associação Parkinson do Rio Grande do Sul
APC	Associação Parkinson Carioca
APPP	Associação Paranaense dos Portadores de Parkinsonismo APPP
CT	Complexidade Textual
SEU	Survey of English Usage
NIFL	National Institute for Literacy

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	9
<b>2 ESCOPO TEÓRICO</b> .....	12
<b>2.1 Breve panorama sobre os estudos de complexidade textual</b> .....	12
<b>2.2 Linguística de <i>Corpus</i></b> .....	17
<b>2.3 Pensando no leitor durante o processo de tradução: a teoria funcionalista de Nord</b> .....	20
<b>3 METODOLOGIA</b> .....	24
<b>3.1 <i>Corpus</i> de estudo</b> .....	24
3.1.1 <i>Corpus</i> Comparável .....	24
3.1.2 <i>Corpus</i> Paralelo.....	26
<b>3.2 Limpeza e armazenamento dos textos</b> .....	27
<b>3.3 Ferramentas computacionais utilizadas</b> .....	28
3.3.1 Coh-Metrix .....	28
3.3.2 Coh-Metrix-Port.....	30
3.3.3 Índice Flesch .....	32
3.3.4 Sketch Engine .....	33
3.3.5 LF Aligner e AntPconc.....	35
<b>3.4 Escolaridade dos públicos alvos dos textos de divulgação sobre a DP</b> .....	37
<b>5 ANÁLISE</b> .....	42
<b>6 CONSIDERAÇÕES FINAIS</b> .....	50
<b>REFERÊNCIAS</b> .....	53
<b>ANEXO</b> .....	57



## 1 INTRODUÇÃO

Logo nos primeiros semestres de minha graduação em Bacharelado Letras Português/Inglês, matriculei-me em uma disciplina obrigatória do curso chamada Introdução à Terminologia. Nessa disciplina, fomos apresentados aos conceitos básicos da Linguística de *Corpus* e o quão útil essa metodologia poderia ser para nossos estudos. Com um simples clique de mouse podíamos processar uma grande quantidade de textos, possibilitando diferentes análises. Além disso, Introdução à Terminologia foi uma das poucas disciplinas do curso em que foi possível interagir diretamente com os diferentes programas de computador que processam esses textos. Sendo assim, meu primeiro trabalho dentro da faculdade foi como monitora dessa disciplina. Nessa monitoria, meu trabalho era auxiliar os alunos a compilarem um *corpus* de bulas de remédios para Parkinson.

Foi devido ao meu envolvimento com essa monitoria que entendi a importância de estudar sobre a doença de Parkinson. De acordo com a Organização Mundial da Saúde (World Health Organization), 1% da população mundial com mais de 65 anos tem a Doença de Parkinson (DP). Nos Estados Unidos, estima-se que mais de 60.000 cidadãos sejam diagnosticados com essa doença a cada ano. No Brasil, mais de 200 mil pessoas sofrem dessa doença. A DP não tem cura e pesquisadores preveem que no futuro o número de pessoas com essa doença irá aumentar, devido ao aumento da expectativa de vida.

Karsch (2003) aponta que geralmente as mulheres que residem no mesmo domicílio se tornam as cuidadoras de seus maridos, pais e até mesmo filhos. A autora também comenta que não é apenas no Brasil que as mulheres desempenham esse papel, já que diversos dados indicam que, por razões culturais, é a mulher quem acaba se tornando a “grande cuidadora”. Além disso, por motivos financeiros, os pacientes acabam sendo tratados em suas casas sob os cuidados de sua família. Quando não há disponibilidade de algum membro da família para cuidar desse paciente, as famílias com menor valor aquisitivo acabam recorrendo a pessoas que não possuem formação e nem preparo para lidar com o Parkinson.

Sendo assim, acaba sendo uma situação normal esses cuidadores ou familiares sem preparo recorrerem à internet como forma de se instruírem. Além disso, quando em estágio inicial, o próprio paciente pode sentir o desejo de saber mais sobre a sua doença, os tratamentos e os estudos que estão sendo feitos. Por esse motivo, há um significativo no número de websites especializados,

criados por estudiosos, doutores e especialistas, com o intuito de oferecer informações sobre essa doença. No Brasil, as associações de Parkinson ajudam a disseminar informações sobre a doença em seus websites, bem como sobre os tratamentos, além de ajudar pacientes, cuidadores e familiares de pacientes a conviver com o Parkinson.

Apesar do aumento do uso da internet como forma de busca de informações sobre doenças, basta uma simples pesquisa em dissertações, teses e artigos acadêmicos para notarmos como há poucos trabalhos que lidam com a complexidade textual desses websites. Sendo assim, nosso objetivo nesta monografia é analisar a complexidade textual de textos de divulgação<sup>1</sup> sobre a doença de Parkinson.

Para nossa análise, foram compilados dois *corpora* de estudo: um *corpus* paralelo e um *corpus* comparável, ambos no par linguístico português-inglês, a partir de textos de divulgação presentes em websites de associações e fundações que lidam com a DP. As definições de *corpus* paralelo e *corpus* comparável são baseadas em Bowker e Person (2002), que os definem, respectivamente, como “[...] um conjunto de textos escritos originalmente em um idioma e traduzidos para o outro(s) e um conjunto de textos originais nos idiomas envolvidos” (Bowker e Person, 2002, apud Rebechi, 2011, p.60).

O *corpus* paralelo foi compilado a partir de textos do site da fundação Michael J. Fox<sup>2</sup> (doravante MJF) e da Parkinson's UK<sup>3</sup> (doravante PUK), e suas traduções para o português, realizadas por alunos da disciplina obrigatória “Estágio Supervisionado de Tradução do Inglês”, que é requisito da graduação em Bacharelado em letras português/inglês da Universidade Federal do Rio Grande do Sul. Essa disciplina foi orientada pela Profa. Dra. Rozane Rebechi.

O *corpus* comparável em inglês foi construído a partir dos mesmos textos em língua inglesa descritos acima, além de cinco websites de associações de Parkinson escritos originalmente em Português. Nosso intuito é verificar se os textos originais em português e em inglês (estadunidense e britânico) são adequados para o seu público alvo, e se houve alteração na complexidade textual das traduções para o português. Esperamos que este trabalho ajude a motivar os autores e tradutores a escreverem de forma acessível, para que pessoas com diferentes níveis de escolaridade possam ter acesso às informações que procuram.

---

<sup>1</sup> Textos de divulgação são textos que possuem o objetivo de difundir o conhecimento científico.

<sup>2</sup> <https://www.michaeljfox.org/> (acesso 4 de novembro de 2018).

<sup>3</sup> <https://www.parkinsons.org.uk/> (acesso 4 de novembro de 2018).

Nesta pesquisa, analisaremos as sentenças em voz passiva analítica presentes em ambos os *corpora* de estudo, relacionando-as com os resultados do índice Flesch, a fim de verificarmos se esses textos estão compatíveis com a escolaridade de seus públicos-alvo. Esses *corpora* foram processados por dois programas de levantamento de índice de complexidade textual - Coh-Metrix e Coh-Metrix-Port. Já que os websites da fundação MJF e a PUK possuem um papel importante na disseminação de informações sobre a DP, espera-se que esses textos estejam de acordo com a escolaridade do adulto, respectivamente, estadunidense e britânico. Nossa hipótese é de que os resultados apontem que nas traduções dos textos de divulgação dos dois websites para o português e textos originais das associações brasileiras haja um índice de dificuldade menor, levando-se em conta a escolaridade média do leitor brasileiro. Além disso, espera-se que esta monografia também possa contribuir para incentivar as discussões sobre complexidade textual de sites especializados e o uso da Linguística de *Corpus* como metodologia para a realização de tais pesquisas.

## **2 ESCOPO TEÓRICO**

De acordo com Finatto (2011), as discussões sobre Complexidade Textual (CT), tanto em português como em inglês, abarcam diversos tipos de pesquisas dentro dos estudos da linguística aplicada. Esses estudos, na maioria das vezes, versam sobre estratégias de leitura, compreensão de texto, tipos de leitores e elementos linguísticos associados ao aumento da dificuldade de leitura. Porém, a autora acredita que, mesmo sendo pioneiros e tendo criado importantes contribuições para a área de CT, esses estudos ainda ignoravam o uso da LC, especialmente no Brasil. Essa relutância em utilizar a LC pode ser atribuída a certa “dificuldade operacional” na utilização de ferramentas computacionais e software. Apesar das possíveis dificuldades, a LC permite investigar, com a ajuda de programas de computador, uma quantidade ilimitada de textos e aspectos que seriam difíceis – ou até mesmo impossíveis – de serem observados por meio de métodos tradicionais.

Esta seção do trabalho será dividida em três subseções. A primeira subseção (2.1) visa traçar um panorama sobre a CT, ao mesmo tempo em que discute conceitos essenciais da área. Já a segunda seção (2.2) é dedicada à discussão sobre a importância da LC, especialmente neste tipo de estudo. Por fim, na última seção (2.3) é discutida a teoria funcionalista da tradução de Christiane Nord, que privilegia o público alvo específico de cada tradução.

### **2.1 Breve panorama sobre os estudos de complexidade textual**

Apesar de parecer que a Complexidade Textual é uma área nova de estudo, pesquisas sobre esse tema existem, fora do Brasil, desde os anos 20 (Davison e Green, 1988, apud Finatto, 2011). Esses estudos lidavam com diversos tópicos, mas o processo de compreensão da leitura era o mais estudado. Smith (1971), por exemplo, foi um dos pioneiros em discutir o processo de leitura a partir de uma abordagem psicológica. Para o autor, o processo de leitura é algo extremamente natural nas crianças, pois esse é apenas mais um aspecto do mundo que as crianças devem compreender e ao qual se adaptar. Ademais, o pesquisador acredita que, desde que nascemos, usamos a interpretação para ler diversos tipos de situação, como o clima, os sentimentos das pessoas, gestos, emoções e muito mais. Nesse caso, o processo de leitura se torna “a atividade mais natural do mundo” (Smith,

1971, p.2)<sup>4</sup>, já que não é limitada apenas a textos escritos. Portanto, o autor acredita que o processo de leitura deveria ser como qualquer outro “[...] aspecto compreensível da existência” (Smith, 1971, p.5)<sup>5</sup>. Por esse viés, o ato de leitura deve ser atribuído às pessoas, e não somente aos cérebros. Portanto, cabe aos pesquisadores não focar apenas o aspecto cognitivo da leitura e ignorar o aspecto humano envolvido.

Outro aspecto humano envolvido no processo de leitura é a compreensão. Compreensão é o simples ato de entender, de perceber o significado de algo que lhe é apresentado. Para Smith (1971), temos que entender o significado, a natureza e o sentido em diversas situações diárias, como diferentes conversas e leituras. O autor também acredita que a compreensão tem o poder de unir aspectos do mundo com o que já sabemos, como nossas intenções, expectativas e conhecimentos. Por fim, Smith (1971) acredita que os atos de compreensão e aprendizado estão intrinsecamente ligados, já que “[...] nós aprendemos a ler, e aprendemos pela leitura, elaborando o que já sabemos” (p.13)<sup>6</sup>.

Com base nessa concepção, não é surpreendente quando o leitor se depara com problemas de compreensão ao ler um texto de um assunto sobre o qual não possui um grande conhecimento prévio. Em textos científicos ou técnicos, como, respectivamente, artigos e manuais de instrução, isso é mais provável de acontecer, já que os autores tendem a usar jargões próprios da área e construções frasais mais elaboradas. Ademais, esses autores tendem a esperar que os leitores preencham suas lacunas no texto, fazendo inferências durante a leitura. Best et al (2005) acreditam que essas inferências que os leitores devem fazer quando leem um texto científico ou técnico é indício de baixa coesão do texto. Para os autores, “[...] coesão textual está ligada às propriedades do texto que determinam o grau em que os leitores precisam gerar inferências para construir representações mentais coerentes das informações explicitadas no texto” (Best et al, 2005,p.3)<sup>7</sup>. Sendo assim, o sucesso no ato de compreensão de um texto está ligado a diversos fatores, como, por exemplo, a qualidade do texto, o conhecimento prévio do leitor e as estratégias de leitura.

---

<sup>4</sup> “[...]the most natural activity in the world” (SMITH, 1971, p. 02). Todas as traduções nesta monografia são de nossa autoria, salvo quando especificado.

<sup>5</sup> “[...]comprehensible aspect of existence” (SMITH, 1971, p.05).

<sup>6</sup> “we learn to read, and we learn through reading, by elaborating what we know already” (SMITH, 1971, p.13).

<sup>7</sup> “[...]text cohesion refers to properties of the text that determine the degree to which readers need to generate inferences to construct a coherent mental representation from the information explicitly stated in the text.” (BEST et al, 2005,p.3).

A discussão sobre a CT de textos técnicos ou científicos é um tópico muito importante. DuBay (2004) relata que, em 1988, nos Estados Unidos, acidentes de trânsito causaram 46% da morte das crianças de 1 a 14 anos. O autor menciona que um estudo feito por Johnston et al. (1994) revelou que o motivo principal das mortes era o uso impróprio da cadeira infantil para automóveis. DuBay (2004) menciona que um estudo realizado por Dr. Mark Wegner e Deborah Girasek (2003) indicou que a falta de compreensão do manual de instrução foi o que contribuiu para esse problema. Mesmo com tamanha importância, estudos que lidavam com a CT de textos técnicos ou científicos eram esquecidos no começo das discussões desse assunto.

Finatto (2011) menciona que alguns autores, como, por exemplo, Davison e Green (1988) foram uns dos primeiros a discutirem problemas de compreensão e acessibilidade de textos técnicos ou científicos. A autora explica que o motivo da falta de discussões sobre esse assunto possa se justificar pelo atraso do reconhecimento dos textos científicos como gênero textual. Apenas a partir dos anos 1930, quando ocorre a primeira reunião internacional de editores de textos científicos, o texto científico passa a ser considerado como tal. No Brasil, de acordo com Fulgêncio e Liberato (2004), Mário Perini, no final dos anos 70, foi um dos primeiros autores a tratar de leitura funcional e da capacidade de um texto ser mais ou menos legível. Porém, o trabalho de Perini só foi publicado em 1988, quando havia um maior interesse pelo processo de leitura. O desejo de Perini era simplificar livros didáticos para as crianças de escolas mais carentes, de forma que tivessem um material especial e acessível, pensado para elas. Ele acredita que a simplificação poderia ser alcançada com discussões sobre o assunto e mudanças nesses livros didáticos (Fulgêncio; Liberato, 2004).

Nos anos 80 e 90, tivemos uma forte tendência de trabalhos focados em discutir o processo de leitura e compreensão. Neis (1982), Kato (1982), Averbuck, Appel e Hessel (1983), Kleiman (1987, 1989, 1993, 1997) e Leffa (1996) contribuíram para estabelecer um vasto território sobre esses assuntos. Kleiman (1989), por exemplo, acreditava que a compressão de um texto envolvesse diversos processos cognitivos, atividades e estratégias mentais. Essa é a razão pela qual muitos autores referirem-se à leitura como um processo.

Também nesse período houve um grande investimento nas fórmulas de inteligibilidade. De acordo com DuBay (2004), no final dos anos 80, já existia mais de 200 fórmulas e inúmeros trabalhos publicados que atestavam a eficácia dessas fórmulas. O autor também menciona que as fórmulas criadas para se calcular a inteligibilidade de um texto sempre estiveram no centro de

controvérsias. Alguns críticos questionavam as limitações dessas fórmulas, enquanto outros propunham testes de funcionalidade. DuBay (2004) argumenta que apesar de os críticos terem razão, eles ainda não conseguiam propor um instrumento objetivo que calculasse a complexidade de um texto, como as fórmulas de inteligibilidade. Vale a pena lembrar que inteligibilidade é a característica que faz com que alguns textos sejam mais fáceis de serem entendidos do que outros, enquanto legibilidade se associa com a qualidade tipográfica dos textos (DuBay, 2004). Nesta monografia, usaremos essas mesmas definições de inteligibilidade e legibilidade.

Uma das mais famosas fórmulas de inteligibilidade é, ainda nos dias atuais, a fórmula Flesch. A fórmula Flesch foi criada em 1948 por Rudolf Flesch e considera apenas duas variáveis: o número de sílabas e o número de sentenças em uma amostra de cem palavras. Essa fórmula ficou conhecida, especialmente no Brasil, graças a sua inclusão no programa *Microsoft Word*. De acordo com Scarton e Aluísio (2010), os valores do índice Flesch variam entre 100-75 (muito fácil), 75-50 (fácil), 50-25 (difícil) e 25-0 (muito difícil). Há também uma versão chamada *Flesch-Kincaid*, que foi financiada pela Marinha dos Estados Unidos com o objetivo de medir a dificuldade de manuais técnicos usados em treinamento. A fórmula Flesch foi aprimorada para medir o nível escolar dos textos, baseando-se no sistema de notas escolar estadunidense.

Apesar de essas fórmulas terem sido por muito tempo o único instrumento disponível para calcular a complexidade textual, elas ainda deixam de lado inúmeros fatores que não podem ser calculados. Para Kleiman (1989), os conhecimentos linguístico, textual e enciclopédico do leitor são essenciais para a compreensão de um texto. A autora acredita que a falta de um bom conhecimento linguístico pode afetar a compreensão de um texto, já que ele se refere a palavras e conceitos que o leitor não entende ou não conhecia antes da leitura. Já o conhecimento textual consiste de uma série de noções preconcebidas sobre a estrutura do texto. Esse tipo de conhecimento pode ser aumentado durante a vida do leitor, após fazer a leitura de diversos tipos textuais e narrativos. Esse conhecimento também ajuda a criar a expectativa do leitor sobre o texto. Por fim, o conhecimento enciclopédico é essencial para a compreensão de um texto, pois é a habilidade do leitor em fazer as inferências necessárias para tornar o texto como um todo coerente. Esse conhecimento está ligado ao conhecimento que o leitor tem do mundo. Por ser um processo inconsciente, pode levar à falta de compreensão de um texto quando o leitor não possui as informações prévias necessárias sobre o assunto. Kleiman (1989) acredita que esses três tipos de conhecimentos precisam ser ativados durante a leitura para que haja a compreensão de um texto.

Outro aspecto importante na compreensão de um texto é a organização sintática. DuBay (2004) menciona que por muito tempo os professores ensinaram processos cognitivos e estruturais, como organização e coerência, para a redação de um texto.

Cunha (2000) ressalta que a voz passiva, por exemplo, é vista como um dificultador da compreensão de um texto, pois ela se “desvia da estratégia mais comum de apresentação desses constituintes: Sujeito – Verbo – Objeto...” (CUNHA, 2000, p.2). Para Cunha (2000), a ocorrência das orações em voz passiva em textos escritos são menos frequentes do que as orações em voz ativa, e que aparece mais em textos não ficcionais, como os textos científicos.

A complexidade sintática da voz passiva também está ligada a uma espécie de surpresa comunicativa, já que o leitor espera uma oração ativa no lugar da passiva (Givón, 1979, p.87 apud Cunha, 2000, p.4). Quanto mais surpreendente for uma construção ou uma palavra, mais difícil será sua compreensão. Aluísio et al (2014) confirmam que essa dificuldade também está presente na língua inglesa. Para os pesquisadores, a voz passiva torna as sentenças em inglês mais longas e difíceis de serem compreendidas. Para Aluísio et al (2014), as grandes revistas científicas estadunidenses já esperam que o autor faça uso da voz ativa em sua escrita. Por exemplo, as diretrizes para autores no website da associação *IEEE Computer Society* dizem que:

Os leitores de hoje em dia precisam entender a informação oferecida de forma rápida. Longos artigos escritos em voz passiva não conseguem atender a essa necessidade. Os autores devem se empenhar em usar verbos na voz ativa e sentenças simples e declarativas, fazendo todo o tipo de esforço possível para que os leitores possam entender os conceitos apresentados. (2018)<sup>8</sup>

Kleiman (1989) também comenta sobre a dificuldade de um texto, dizendo que:

O processamento [da leitura] é essencialmente de carácter cognitivo, mas quanto mais complexo for o texto, mais se faz necessário o controle ativo desse processo mediante as estratégias metacognitivas de manutenção de objetivos e monitorização e desautomatização do processo de compreensão (p.62)

---

<sup>8</sup> “Today's readers need to grasp information quickly. Extremely long articles presented in a passive writing style don't meet this need. Authors should strive for active verbs and straightforward declarative sentences, making every effort to help readers understand the concepts presented.” (IEEE COMPUTER SOCIETY, 2018).



A autora também menciona que um parágrafo pode ser mais difícil de ser entendido se o seu tema ou tópico não aparecer logo no início. Enfatiza, também, que a compreensão de crianças com problemas de leitura é mais afetada por um parágrafo mal organizado. Porém, parágrafos mal organizados, construções sintáticas inesperadas, jargões e termos desconhecidos podem afetar a compreensão de leitores de qualquer idade. Mesmo o leitor mais experiente, quando depara com algo inusitado, irá recorrer a algum tipo de estratégia para que possa entender determinado trecho. Há diversos tipos de estratégias, como, por exemplo, ler diversas vezes o mesmo trecho, ler em voz alta e sublinhar ou reescrever o trecho. Nesse caso, o leitor precisa usar passos conscientes durante a leitura para que possa entender o parágrafo.

## 2.2 Linguística de *Corpus*

De acordo com Berber Sardinha (2004), a Linguística de *Corpus* se ocupa da coleta e exploração de *corpora* para investigar certa língua ou variedade linguística. Segundo Berber Sardinha (2000), a definição de LC que mais engloba as principais características de um *corpus* computadorizado é a de Sanchez (2000), que diz que a LC é:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (Sanchez, 1995 p. 8-9 apud Berber Sardinha, 2000, p. 338)

Para o autor, essa definição reúne os mais importantes pontos a serem seguidos na construção de um *corpus*:

- (a) Origem: os dados devem ser autênticos.
- (b) Propósito: o *corpus* deve ter a finalidade de ser um objeto de estudo linguístico.
- (c) Composição: o conteúdo do *corpus* deve ser criteriosamente escolhido.
- (d) Formação: os dados do *corpus* devem ser legíveis por computador.
- (e) Representatividade: o *corpus* deve ser representativo de uma língua ou variedade.
- (f) Extensão: o *corpus* deve ser vasto para ser representativo. (Berber Sardinha, 2004, p.18-

19).

Foi justamente a criação dos computadores que facilitou e tornou mais confiável a criação dos *corpora*. De acordo com Berber Sardinha (2004), em 1960 os computadores mainframe já estavam presentes nos centros de pesquisas universitários. O primeiro *corpus* computadorizado foi criado por Henry Kučera e W. Nelson Francis na Universidade Brown em 1964. Esse *corpus* recebeu o nome de *Brown University Standard Corpus of Present-Day American English* (mais conhecido como *Corpus Brown*). O *Corpus Brown* foi considerado revolucionário pelo seu um milhão de palavras. É preciso lembrar que naquele tempo era muito trabalhoso compilar um *corpus*, pois os textos precisavam ser transferidos para o computador por cartões que tinham que ser perfurados um de cada vez (Berber Sardinha, p.2-3). Atualmente, com o avanço dos computadores, é possível ter acesso a *corpora* com centenas de milhões de palavras, como, por exemplo, o *Bank of English*. Mesmo assim, alguns dos grandes *corpora* só são acessíveis para pesquisadores ou pagantes. Esses *corpora* disponíveis gratuitamente podem motivar um pesquisador a usar a LC em seu estudo, já que um dos processos mais trabalhosos na criação de um *corpus* é a coleta e a limpeza dos textos.

Berber Sardinha (2004) afirma que a LC atua dentro de um sistema empirista, que percebe a linguagem como um sistema de probabilidade. Na visão racionalista da linguagem, o estudo da língua é feito por meio da introspecção, como “[...] forma de verificar modelos de funcionamento estrutural e processamento cognitivo da linguagem.” (Berber Sardinha, 2004, p.30). Diferentemente, a LC se baseia nas observações da linguagem proveniente de um *corpus*. Na visão da linguagem como probabilidade, a frequência se torna um elemento essencial, já que diversos traços podem ser possíveis, mas não necessariamente ocorrem com a mesma frequência.

Nesta monografia, é reconhecida a importância da LC para os estudos linguísticos e o valor de se adotar uma abordagem baseada em *corpus* em estudos que lidam com textos autênticos. Uma abordagem baseada em *corpus* permite uma análise empírica de padrões reais de uso da língua, e com o apoio das ferramentas computacionais, permite análises sistemáticas, que não seriam possíveis – ou seriam menos confiáveis –, de outra maneira. Uma abordagem baseada em *corpus* permite também que o pesquisador compartilhe o *corpus* com outros pesquisadores para que outros aspectos de uma mesma coleção de textos possam ser investigados.

Apesar de poucos, existem trabalhos relevantes que tratam da complexidade textual de textos médicos por meio da metodologia da LC. Elhadad (2006) identificou termos de uma revista médica que um leitor leigo provavelmente não entenderia, e propôs um método simples para

fornecer definições retiradas da internet para esses termos. O autor compilou um *corpus* de textos escritos para leigos e selecionou os termos de menor frequência do *corpus*. Elhadad (2006) acredita que esses termos com menor frequência possuiriam menos chances de serem entendidos pelos leitores. Severance e Cohen (2015), por outro lado, examinaram a CT de resumos submetidos a revistas médicas, mas não utilizaram a LC em sua abordagem. Os autores usaram o *Coleman-Liau Index* para calcular a complexidade dos resumos e descobriram que tem havido um aumento na dificuldade dos resumos submetidos às revistas médicas com o passar do tempo

Porém, quando se trata da análise da complexidade textual de websites, ainda parece haver uma escassez de estudos que investiguem esse aspecto, sobretudo por meio de uma abordagem baseada em *corpus*. No Brasil, Barboza e Nunes (2007) exploraram a inteligibilidade dos textos de websites das agências governamentais federais brasileiras. Os autores utilizaram a Fórmula de *Lecturabilidad* de Fernández-Huerta (1959), uma adaptação da fórmula Flesch para a língua espanhola. Barboza e Nunes (2007) optaram por utilizar o índice de Fernández-Huerta,<sup>9</sup> devido à mesma base latina dos idiomas espanhol e português. Os autores concluíram que apesar de o governo ter a obrigação de escrever de uma forma que todos possam facilmente entender, na prática isso não acontece. O índice de Fernández-Huerta apontou que 92% dos websites de ministérios investigados apresentaram índices de inteligibilidade que exigiriam do usuário de cinco a seis anos de escolaridade (5º e 6º série completos) para que pudesse compreendê-los inteiramente.

O conceito de website é algo extremamente novo. Em 1969, a ARPANET (considerada a mãe da internet) tinha o objetivo de interligar as bases militares e os departamentos de pesquisa do governo americano. Hoje em dia, ela abriga websites que são complexos repositórios de informações com diferentes tipos de conteúdo, espalhados ao redor do mundo. Como os websites especializados que iremos analisar servem como repositório de informações sobre doenças, acreditamos que os textos dentro deles se encaixem no gênero de divulgação científica, pois são textos expositivos com o propósito de disseminar o conhecimento científico de uma maneira mais simples. É nos websites que esse tipo textual consegue abranger um maior número de pessoas.

O website internacional *Internet Usage and World Population Statistic*<sup>10</sup> fornece estatísticas sobre o uso da internet de mais de 243 países e regiões. De acordo com sua última pesquisa divulgada em julho de 2018, atualmente, 95% da população norte americana usa a

---

<sup>9</sup> Vale a pena lembrar que, durante a pesquisa de Barboza e Nunes (2007), o Coh-Matrix-Port ainda não havia sido lançado e nem a sua adaptação do *Flesch Reading Ease*.

<sup>10</sup> <https://www.internetworldstats.com/stats.htm> (acesso 4 de novembro de 2018)

internet. Na América do Sul, esse número cai para 71,8%. Esses usuários possuem acesso a um número ilimitado de informações, podendo pesquisar diversos assuntos em ferramentas gratuitas de busca, como *Google*, *Yahoo Search*, *Bing* e etc. Os assuntos pesquisados variam muito: fornecem desde informações sobre doenças, seus sintomas, tratamentos e remédios, até sobre curiosidades da vida da duquesa Meghan Markle.<sup>11</sup>

Uma rápida pesquisa no *Google Trends*<sup>12</sup> mostra que as buscas por *mal de Parkinson sintomas iniciais*, *Parkinson mata?*, *acinesia*, *epilepsia*, *mal de Parkinson tratamento*, *mal de Parkinson causa* e *o que é doença de Parkinson* estão entre as 15 pesquisas mais recorrentes com o termo Parkinson. Em 3º lugar aparece o medicamento *memantina*, que vem sendo utilizado como uma nova opção terapêutica no tratamento da DP. Apenas com os 15 primeiros resultados já é possível perceber um número expressivo de dúvidas relacionadas a medicamentos, tratamentos, informações sobre a doença e sintomas. Às vezes, pacientes, familiares ou cuidadores podem se sentir envergonhados ou com medo de conversar com o médico sobre suas dúvidas e inseguranças relacionadas à doença. Cada vez mais esses usuários se apoiam nas ferramentas de buscas da internet para encontrar uma maneira de se manterem informados. Sendo assim, essas pessoas têm o direito de poder entender adequadamente o material textual que está sendo oferecido para elas, de forma a sanar suas dúvidas. Isso demonstra o quão importante é discutir a complexidade textual de websites que servem como reservatórios de informações.

A analogia com a rampa de acesso físico para cadeirantes é bastante utilizada para explicar a importância da acessibilidade textual. Todas as pessoas podem usufruir da rampa de entrada para cadeirantes, porém o cadeirante depende mais do direito de usar a rampa do que outras pessoas. Uma situação parecida pode ser pensada sobre o texto. Um texto simples demais pode entediar um leitor experiente, porém, aumentando-se a dificuldade desse mesmo texto, pode-se dificultar a leitura de pessoas com menor índice de escolaridade. Acessibilidade é um direito de todos, só que para alguns ela é mais necessária do que para outros.

---

<sup>11</sup> De fato, *Meghan Markle* aparece no *Google Trends* de 2017 em primeiro lugar como atriz mais pesquisada, devido ao seu casamento com o príncipe Harry.

<sup>12</sup> Uma ferramenta gratuita criada pelo *Google* que mostra a frequência que palavras chaves, assuntos, frases, pessoas e etc foram pesquisadas na ferramenta de busca *Google*.

### 2.3 Pensando no leitor durante o processo de tradução: a teoria funcionalista de Nord

Gentzler (2009) afirma que a teoria da tradução é e ao mesmo tempo não é uma área nova. Para o autor, mesmo que a primeira aparição da teoria da tradução tenha ocorrido em 1983 no *Modern Language Association International Bibliography*, ela é tão antiga quanto a história da Torre de Babel. O mito da Torre de Babel, segundo Gênesis 11:1-9, é usado para explicar as diferentes línguas usadas ao redor do mundo. Segundo essa lenda, a humanidade era uniforme e compartilhava um único idioma. Até que, na terra de Sinar, os homens combinaram de construir uma cidade e uma torre tão alta que alcançasse os céus. Quando Deus soube dessa construção ambiciosa, ele castigou os humanos, misturando suas vozes para que eles não pudessem entender uns aos outros e dar continuidade à construção da torre. Gentzler (2009) ressalta que alguns estudiosos literários não reconhecem a teoria da tradução como uma disciplina e que alguns tradutores alegam que não precisam dela. Porém, para o autor a tradução é “[...] inerente a todas as línguas por meio de suas relações com outros sistemas significativos, tanto do passado quanto da atualidade” (Gentzler, 2009, p.21).

Nord (2012) ressalta que os estudos da tradução começaram como uma ramificação da linguística aplicada em 1960, influenciada pelo crescente interesse nas traduções automáticas. A autora relata que foi a necessidade da Alemanha e de outros países Europeus, durante a década de 1970, em treinar seus tradutores e intérpretes que impulsionou uma tendência de teorias pensadas para o treino de tradutores, como, por exemplo, a teoria funcionalista. Existem diversas teorias contemporâneas da tradução, algumas bastante divergentes entre si. A dicotomia mais antiga em torno da tradução é a tradução literal e a livre. Geralmente, a tradução literal é associada com a ideia de tradução palavra por palavra, neutralidade por parte do tradutor, e busca por uma tradução fiel ao texto de partida. Já a tradução livre está associada à ideia de tradução infiel, onde o tradutor coloca sua subjetividade no texto na hora de traduzir. Nord (2012) afirma que essa dicotomia da tradução já vem de bastante tempo. Tradutores como Cícero (1º. século A. C.), São Jerônimo (384 AD), e Martinho Lutero (século XVI) já se confrontaram com o dilema básico da tradução: traduzir literalmente, palavra por palavra, e gerar um texto de chegada confuso e difícil para seus leitores, ou traduzir o sentido do texto, ao invés das palavras. Os três escolheram a tradução livre. Naquela época, a tradução livre de um livro sagrado poderia ser interpretada como um ato de infidelidade a Deus, e seu tradutor poderia ser perseguido pela Inquisição. Mesmo assim, os tradutores advogaram

em favor da tradução do sentido do texto e não das palavras. Porém, Nord (2012) lamenta que, com o avanço dos estudos de tradução baseados em linguística, esse enfoque no sentido tenha sido substituído por uma perspectiva linguística mais rigorosa, que buscava relações de equivalências entre as línguas. Para a autora, tal modelo dominou a primeira fase dos Estudos de Tradução.

No começo da década de 1970 a linguista e tradutora alemã Katharina Reiss elaborou uma teoria de tradução com o propósito de servir de guia para tradutores. Nord (2012) afirma que a teoria de Reiss postulava que o público-alvo, do texto traduzido, é quem “determina” as estratégias de tradução. Contudo, com base em sua experiência tradutória, Reiss deixou claro que havia certas limitações na hora de encontrar equivalências e que a funcionalidade do texto era muito importante na hora da tradução. Apesar de Reiss ainda se basear na equivalência, foi a questão da funcionalidade no seu modelo que deu início à teoria funcionalista da tradução.

Foi um aluno de Katharina Reiss, Hans J. Vermeer, quem apresentou a teoria que, de acordo com Nord (2012), “[...] rompeu com as teorias linguísticas e virou de cabeça para baixo o modelo de Reiss”<sup>13</sup> (p.2). Na teoria de *Skopos* (Skopostheorie), palavra de origem grega que significa propósito, objetivo, Hans J. Vermeer propõe que se pense na função dos textos de partida e chegada. Nessa teoria, equivalência é uma das várias possibilidades de propósito comunicativo, e não a única. Para Nord (2006), o princípio básico da teoria do *Skopos* é que o propósito do texto de partida determine os métodos e as estratégias da tradução.

O modelo funcionalista de Christiane Nord é uma sistematização das ideias de Hans J. Vermeer e Katharina Reiss, porém dedicada à formação de tradutores e na prática tradutória em si. Esse propósito do texto de chegada que determina os métodos e estratégias da tradução, é o que Nord (2007) prefere chamar de "princípio funcional". Para a autora, a experiência tradutória mostra que há sempre mais de uma maneira de traduzir um texto. Sendo assim, Nord (2007) afirma que a tradução é um processo de decisão que parece ser guiada por algum tipo de critério intersubjetivo ou estratégia. Para a teoria de *Skopos*, esse critério deve ser a função comunicativa para qual o texto de chegada é necessário. Por isso, Nord (2007) afirma que muitos críticos condenam o funcionalismo por priorizar as expectativas e necessidades do público alvo, e não do texto de partida.

Como as associações e fundações de Parkinson ajudam a disseminar informações sobre a doença em seus websites, assim como oferecem diversos tratamentos e ajuda para os pacientes,

---

<sup>13</sup> “[...] who broke with linguistic translation theory and turned Reiss’ model upside down” (NORD, 2012, p.2).

cuidadores e familiares de pacientes com Parkinson, elas precisam que os seus leitores compreendam o máximo possível das informações que estão sendo apresentadas. Assim, cabe ao tradutor levar em consideração diversos aspectos de seu público alvo, como, por exemplo, a sua escolaridade e o nível de inteligibilidade de seus textos. É a teoria funcionalista da tradução que permite que finalmente o leitor se torne juiz da funcionalidade de um texto. Como Nord (2006) ressalta, a funcionalidade não é uma característica inerente de um texto, mas sim “[...] atribuída ao texto pelo receptor, no momento da recepção” (p.31)<sup>14</sup>. Dessa forma, é o leitor quem decide se uma determinada tradução funciona em um determinado espaço de tempo e cultura.

---

<sup>14</sup> “[...] is a quality attributed to the text by the receiver, in the moment of reception” (NORD, 2006, p.31).

### 3 METODOLOGIA

A proposta, nesta monografia, é focar na análise das sentenças em voz passiva analítica, que são consideradas dificultadores da compreensão de um texto. Para isso, usaremos os pressupostos da Linguística de *Corpus* (LC).

A seguir, apresentaremos maiores detalhes sobre a compilação do *corpus* de estudo e sobre as ferramentas utilizadas para o seu processamento.

#### 3.1 *Corpus* de estudo

Na década de 1990, acreditava-se que um *corpus* representativo teria em torno de um milhão de palavras. Atualmente, existem *corpora* de bilhões de palavras, em geral utilizados para estudos lexicográficos. Berber Sardinha (2004) defende que um *corpus* “[...] deve ser adequado aos interesses do pesquisador, que deve ter uma questão a investigar para a qual necessite de um *corpus* específico” (p. 29), enfatizando a impossibilidade de tentar incluir toda uma língua em um *corpus*.

Contudo, um *corpus* especializado não precisa ser muito extenso para que revele resultados interessantes, considerando-se os objetivos da pesquisa. Para nossa análise, foram compilados dois *corpora* de estudo: um *corpus* paralelo e um *corpus* comparável, ambos no par linguístico inglês-português, a partir de textos de divulgação presentes em websites de associações e fundações que lidam com a DP.

##### 3.1.1 *Corpus* Comparável

O *corpus* comparável em inglês é formado de dois *subcorpora* de textos de divulgação: um em inglês estadunidense, extraído do website da associação MJF, e o outro em inglês britânico, do website PUK. A fundação MJF se dedica a encontrar uma cura para a DP e aprimorar terapias já existentes através de pesquisas financiadas. O seu website é bem sofisticado e age como um reservatório de informações para pacientes, cuidadores, familiares e estudantes universitários. A Parkinson's UK é a sigla para a Sociedade de Parkinson do Reino Unido<sup>15</sup> e foi fundada em 1969

---

<sup>15</sup> Do inglês, *Parkinson's Disease Society of the United Kingdom*.



por Mali Jenkins. Durante as últimas cinco décadas, a PUK vem se dedicando a financiar pesquisas e ajudar milhares de pessoas com Parkinson no Reino Unido. Seu site também possui bastante material informativo, mas ainda é menor quando comparado com o site da fundação MJF.

Para descobrirmos o que é característico dos websites com textos de divulgação da DP no Brasil, e não apenas de um único website, buscamos textos de divulgação em diferentes websites de associações brasileiras de Parkinson para formar o nosso *corpus* comparável em português. Escolhemos trabalhar com os textos de quatro associações de diferentes regiões e um projeto social, sendo eles: Associação Brasil Parkinson (ABP), Associação Parkinson do Rio Grande do Sul (APARS), Associação Parkinson Carioca (APC), Associação Paranaense dos Portadores de Parkinsonismo (APPP) e Projeto Vibrar Parkinson.

A ABP<sup>16</sup> é uma organização filantrópica pioneira na América Latina, criada em 1985 por Marylandes Grossmann, uma paciente de Parkinson. O seu website é bastante rico em informações sobre a doença, além de ser simples e fácil de utilizar, com links que fornecem informações sobre a DP e as atividades que a associação promove. O website da APARS<sup>17</sup> possui pouco material informativo sobre a DP. Porém, no menu *viver com Parkinson* há uma mensagem que avisa que o site está em construção e promete falar sobre atividades físicas e fisioterapias, psicologia e terapias, alimentos e nutrição. O website da APC<sup>18</sup> possui pouco material informativo sobre a doença. De fato, somente o material sobre a *DP e a Fisioterapia* e *O que é a Doença de Parkinson?* foram usados. A partir de 13 de abril de 2018, o site da APC incluiu um menu *Últimas Notícias*, com textos curtos de divulgação sobre medicamentos, entre outros. Porém, como nossa pesquisa foi conduzida antes dessa data, infelizmente, esse material, ainda que pouco, não foi utilizado em nosso *corpus*. O website da APPP<sup>19</sup> possui textos mais longos, com destaque para o menu *DIREITOS ASSEGURADOS AOS PARKINSONIANOS*, que possui um texto informativo sobre os direitos e benefícios dos portadores da doença de Parkinson. Por fim, o website do Projeto Vibrar Parkinson<sup>20</sup> disponibiliza bastante material informativo, uma vez que o website se caracteriza por:

Site desenvolvido para divulgar informações sobre a doença de Parkinson, sobre o Vibrar com Parkinson, dicas úteis para auxiliar no gerenciamento do tratamento da doença, entre outros. Sob coordenação da cientista Dra.

<sup>16</sup> <http://www.parkinson.org.br/> (acesso em 1 de dezembro de 2018).

<sup>17</sup> <https://www.apars.org.br/> (acesso em 1 de dezembro de 2018).

<sup>18</sup> <http://parkinsoncarioca.com.br/> (acesso em 1 de dezembro de 2018).

<sup>19</sup> <http://www.appp.com.br/site/> (acesso em 1 de dezembro de 2018).

<sup>20</sup> <http://vibrarcomparkinson.com.br/> (acesso em 1 de dezembro de 2018).

Danielle Lanzer, o Vibrar com Parkinson é um projeto de extensão da Universidade Federal de Goiás, que atua em todo o país. (PROJETO VIBRAR PARKINSON, 2016)

No quadro 1, a seguir, apresentamos o número de *tokens* (palavras) do *corpus* comparável. Salientamos que, para a LC, uma sequência de letras precedida e seguida de espaço constitui uma palavra:

Quadro 1. *Corpus* comparável

<b>Português</b>	Nº tokens	<b>Inglês</b>	Nº tokens
	9928	estadunidense	16474
		Britânico	10815
Total	9928		27.289

Fonte: Elaboração Autoral

Conforme apontado anteriormente, o *corpus* comparável não está balanceado pelo número de palavras, uma vez que há pouca disponibilidade de material disponível em português brasileiro. Apesar de nossos *corpora* serem considerados “pequenos” (Berber Sardinha, 2004), acreditamos que sejam representativos. Afinal, a representatividade do nosso *corpus* é assegurada pelo fato de os textos que os compõem pertencerem a um mesmo gênero especializado, ou seja, sobre a DP.

### 3.1.2 *Corpus* Paralelo

Para a compilação do *corpus* paralelo, foram selecionados os textos do *corpus* comparável em inglês e suas respectivas traduções para o português, realizadas por estagiárias do curso de bacharelado em tradução da UFRGS. A escolha pela tradução desses textos teve como propósito contribuir com o website da associação APARS. Quatro estagiárias<sup>21</sup> assumiram a tarefa de traduzir os textos cedidos pelas fundações britânica e estadunidense. É importante mencionar que, para garantir a uniformidade na tradução de termos e fraseologias durante a tarefa coletiva, as alunas desenvolveram um glossário de termos e fraseologias da área da doença de Parkinson compartilhado via *Google Sheets*, com metodologia baseada na Linguística de *Corpus*.

<sup>21</sup> Fernanda Cerveira, Gislaíne Flor Alves, Kelly Carrion e Laís Callegaro Fritzen.

O quadro 2 apresenta o número de palavras do *corpus* paralelo:

Quadro 2. *Corpus* Paralelo

Inglês	Nº tokens	Português	Nº tokens
MJF	16474	MJF	16562
PUK	10815	PUK	9928
Total	27.289		26.490

Fonte: Elaboração Autoral

### 3.2 Limpeza e armazenamento dos textos

Primeiramente foram selecionados os websites que seriam utilizados e foi feita a coleta dos textos de divulgação sobre a DP. Em seguida, foi feita uma limpeza para remoção dos elementos extratextuais, como tabelas, gráficos, imagens, agradecimentos e links externos. Cada texto foi precedido de cabeçalho contendo o *link* para o *website* de onde o texto foi retirado e a data de acesso. Depois, esse material foi salvo em formato de texto sem formatação (TXT), na codificação UTF-8, que é a mais apropriada para línguas que usam diacríticos, como o português. Nas traduções feitas pelos alunos da UFRGS, foi acrescentado um cabeçalho contendo o nome dos autores das traduções. Os arquivos dos *corpora* foram nomeados da seguinte maneira: abreviatura da língua (pt para português e en para Inglês), a abreviatura do nome da associação ou fundação (MJF, PUK, ABP, APARS, APC, APPP ou VP), seguido pelo número do texto e a letra que indica a qual *corpus* se refere (C para comparável e P para Paralelo). Os *subcorpora* em inglês também foram acrescentados das abreviaturas NA (estadunidense) e BR (britânico) para diferenciar a variante de inglês utilizado.

No quadro 3, é possível ver alguns exemplos da nomeação dos arquivos.

Quadro 3. Nomeação dos *corpora*

ptABP001C	Exemplo de nomeação de texto do <i>corpus</i> comparável em português.
enNAMjF001C	Exemplo de nomeação de texto do <i>subcorpus</i> comparável em inglês

	estadunidense
enBRPUK001C	Exemplo de nomeação de texto do <i>subcorpus</i> comparável em inglês britânico.

Fonte: Elaboração Autoral

Foi encontrado um número expressivo de textos de divulgação que relatavam os objetivos, valores e serviços prestados pelas associações brasileiras e fundações estadunidenses. O motivo disso é porque seus websites, além de informarem sobre a DP, também servem para divulgar suas pesquisas, trabalhos, projetos, etc. Esses textos ficaram de fora dos nossos *corpora*, pois buscamos textos que tratassem diretamente sobre a doença de Parkinson. O anexo A apresenta o tema de cada um dos textos que compõem o *corpus* de estudo.

Em seguida, apresentaremos as ferramentas computacionais utilizadas nesta pesquisa.

### 3.3 Ferramentas computacionais utilizadas

#### 3.3.1 Coh-Metrix

Coh-Metrix foi o primeiro programa gratuito que permitiu que pesquisadores analisassem textos sob 200 medidas de coesão, língua e legibilidade. De acordo com Graesser et al (2003), a coesão ajuda o leitor a unir mentalmente as ideias de um texto. Já a coerência de um texto são as características que contribuem para construir essas ideias mentais. Para esses autores, as antigas fórmulas de legibilidade só levavam em conta o número de palavras e o tamanho das sentenças. Graesser et al (2003) acreditam que o programa Coh-Metrix se difere das outras fórmulas, pois é “[...] sensível a uma ampla quantidade idiomas e características coesivas.” (p.2).<sup>22</sup> Essa ferramenta foi criada por Graesser e outros pesquisadores na Universidade de Memphis, Estados Unidos. De acordo com Pasqualini (2012), o nome da ferramenta significa métricas de coesão, pois seu objetivo é calcular índices de coesão e coerência de um texto em um “[...] amplo espectro de medidas lexicais, sintáticas, semânticas e referenciais, a fim de indicar a adequação de um texto a seu

<sup>22</sup> “[...] is sensitive to a broad profile of language and cohesion characteristics.” (GRAESSER et al, 2003, p.2)

público-alvo (a demanda cognitiva e a legibilidade do texto) e de apontar problemas textuais de ordem estrutural” (p.73).

Na versão gratuita 3.0, 108 métricas estão disponíveis. Para Pasqualini (2012), essas métricas variam desde as mais simples, como, por exemplo, contagem de palavras, até as mais complexas, como o algoritmo de resolução de anáfora. No website do programa, há um documento que explica sobre essas métricas, que se dividem em 11 categorias diferentes<sup>23</sup>. Essas categorias são:

1. *Descriptive*: fornece índices que descrevem o texto em seus aspectos gramaticais, como, por exemplo, o número de parágrafos, sentenças e palavras.
2. *Text Easability Principal Component Scores*: essa categoria fornece uma visão mais completa da dificuldade (ou facilidade) das características linguísticas do texto. Ela foi baseada em estudos feitos por Graesser, McNamara, & Kulikowich (2011), e calcula alguns índices, como, por exemplo, palavras concretas, coesão verbal, etc.
3. *Referential Cohesion*: essa métrica analisa a coesão referencial local e global.
4. *Latent Semantic Analysis*: essa métrica calcula as sobreposições semânticas entre as sentenças ou parágrafos.
5. *Lexical Diversity*: essa categoria analisa a variedade de palavras no texto em relação ao seu número total.
6. *Connectives*: essa categoria calcula todos os tipos de conectivos que aparecem no texto. Ela também diferencia cinco classes de conectivos: causal, lógico, adversativo/contrastivo, temporal e aditivo.
7. *Situation Model*: essa categoria fornece a taxa de verbos, partículas ou ações intencionais, causais e temporais para analisar o nível de representação mental de um texto.
8. *Syntactic Complexity*: essa categoria calcula a sintaxe de um texto, fornecendo métricas simples, como, por exemplo, o número de palavras que aparece antes de um verbo principal. Porém, ela também fornece métricas mais complexas, como o número de intersecções entre frases adjacentes.
9. *Syntactic Pattern Density*: essa categoria calcula a incidência de frases verbais, nominais, adverbiais e preposições.

---

<sup>23</sup> É possível acessar a versão integral dessas métricas do Coh-Metrix 3.0 no site: <http://cohmetrix.com/>

10. *Word Information*: essa categoria se relaciona com a ideia que cada palavra pertence a um tipo de categoria gramatical. O Coh-Metrix associa cada palavra do texto a uma única categoria gramatical.

11. *Readability*: essa última categoria calcula a inteligibilidade de um texto, baseando-se na Fórmula Flesch e no *Flesch Kincaid Grade Level*.

A figura 1 (abaixo) apresenta a tela principal do Coh-Metrix

Figura 1. Tela principal do Coh-Metrix 3.0

Created: September 1, 2012 **Coh-Metrix 3.0** Last updated: Aug. 16, 2017

Save Data

Enter your input

1) Prestar atenção em cada um dos passos na realização de todos os movimentos, por exemplo: para comer, o paciente deve prestar atenção para pegar com o garfo o alimento, depois se concentrar em levantar o braço e por último em levar o garfo até a boca.  
 2) Andar com as costas bem retas, olhando para a frente, dar passos largos apoiando primeiro o calcanhar no chão.  
 3) Procurar orientar os passos através de pistas visuais, por exemplo, um quadriculado de um ladrilho, colocando cada pé em um quadrado.  
 4) Sempre que errar na realização de um movimento, parar, respirar fundo e iniciar novamente desde o início.  
 5) Evitar ficar muito tempo na mesma posição durante o dia.

Type text in the image

Submit

Number	Label	Label V2.x	Text	Full description
Descriptive				
1	DESPC	READNP	5	Paragraph count, number of paragraphs
2	DESSC	READNS	12	Sentence count, number of sentences
3	DESWC	READNW	132	Word count, number of words
4	DESPL	READAPL	2.400	Paragraph length, number of sentences in a paragraph, mean
5	DESPLd	n/a	0.894	Paragraph length, number of sentences in a paragraph, standard deviation
6	DESSL	READASL	11	Sentence length, number of words, mean
7	DESSLd	n/a	12.413	Sentence length, number of words, standard deviation
8	DESWLsy	READASW	1.735	Word length, number of syllables, mean
9	DESWLsyd	n/a	0.940	Word length, number of syllables, standard deviation
10	DESWLit	n/a	4	Word length, number of letters, mean
11	DESWLitd	n/a	2.630	Word length, number of letters, standard deviation
Text Easability Principle Component Scores				
12	PCNARz	n/a	-1.645	Text Easability PC Narrativity, z score
13	PCNARp	n/a	5.050	Text Easability PC Narrativity, percentile
14	PCSYNz	n/a	0.288	Text Easability PC Syntactic simplicity, z score
15	PCSYNp	n/a	61.030	Text Easability PC Syntactic simplicity, percentile
16	PCCNCz	n/a	1.014	Text Easability PC Word concreteness, z score
17	PCCNCp	n/a	84.380	Text Easability PC Word concreteness, percentile
18	PCREFz	n/a	-2.407	Text Easability PC Referential cohesion, z score
19	PCREFp	n/a	0.820	Text Easability PC Referential cohesion, percentile
20	PCDCz	n/a	-2.788	Text Easability PC Deep cohesion, z score
21	PCDCp	n/a	0.270	Text Easability PC Deep cohesion, percentile
22	PCVERBz	n/a	-3.781	Text Easability PC Verb cohesion, z score
23	PCVERBp	n/a	0	Text Easability PC Verb cohesion, percentile
24	PCCONNz	n/a	0.627	Text Easability PC Connectivity, z score
25	PCCONNp	n/a	73.240	Text Easability PC Connectivity, percentile
26	PCTEMPz	n/a	1.615	Text Easability PC Temporality, z score
27	PCTEMPp	n/a	94.630	Text Easability PC Temporality, percentile
Referential Cohesion				
28	CRFNO1	CRFBN1um	0.091	Noun overlap, adjacent sentences, binary, mean

Fonte: Print Screen da tela principal Coh-Metrix 3.0

### 3.3.2 Coh-Metrix-Port

Coh-Metrix-Port<sup>24</sup> é uma adaptação da ferramenta Coh-Metrix para o português brasileiro.. Esse software foi criado por Scarton e Aluísio, pois eles sentiram que apenas o Índice Flesch não

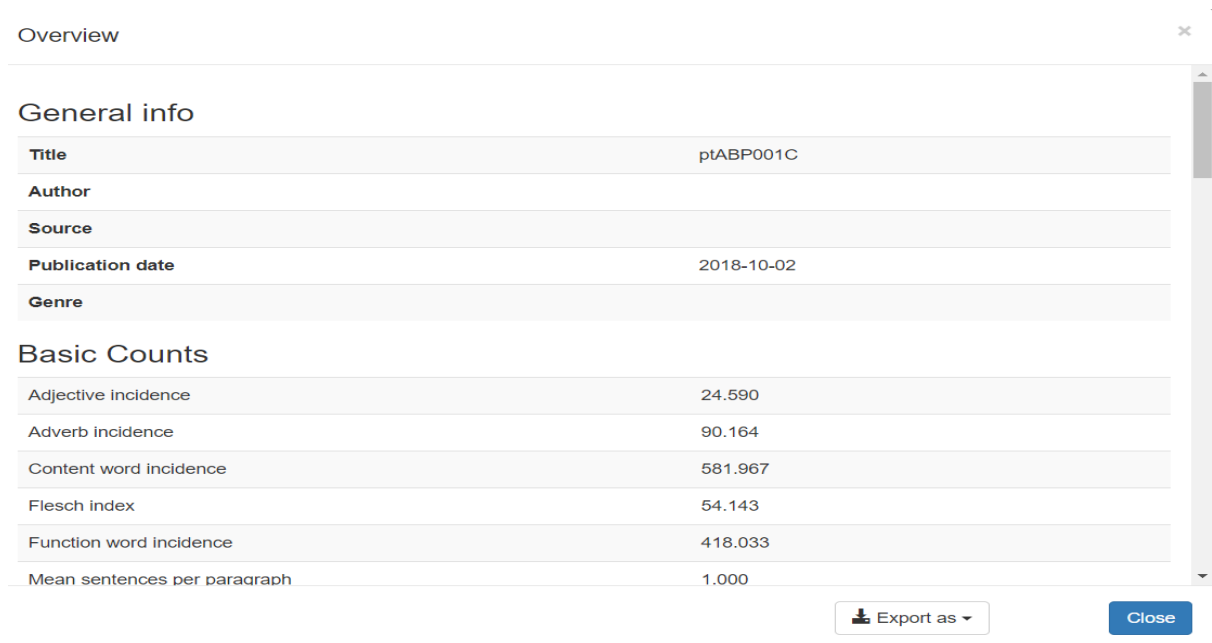
<sup>24</sup> <http://143.107.183.175:22680/>

era suficiente para medir a complexidade textual de um texto, já que ele se baseia apenas no número de palavras e no número de sílabas (Scarton e Aluísio, 2010). Assim como o Coh-Metrix, o Coh-Metrix-Port também é gratuito, porém seu website é muito mais instável, e pode travar diversas vezes, quando sobrecarregado, e ficar fora do ar por horas.

O Coh-Metrix-Port possui 48 métricas para medir coesão, coerência e complexidade de um texto. Porém, de acordo com Pasqualini (2012), apenas oito dessas métricas podem ser comparadas diretamente com a versão inglesa: contagens básicas, Índice Flesch, constituintes, conectivos, operadores lógicos, pronomes, *types* e *tokens*, correferenciais e anáforas.

A figura 2 (abaixo) apresenta a tela principal do Coh-Metrix-Port. Como é possível observar na figura, os campos *author*, *source* e *genre* estão vazios, pois, depois de diversas tentativas, descobrimos que o website da ferramenta funciona melhor quando deixamos esses campos em branco.

Figura 2. Tela principal do Coh-Metrix-Port



The screenshot shows the main interface of Coh-Metrix-Port. It has a title bar 'Overview' with a close button. Below is a 'General info' section with a table of metadata. The 'Author', 'Source', and 'Genre' fields are empty. Below that is a 'Basic Counts' section with a table of linguistic metrics. At the bottom right, there is an 'Export as' button and a 'Close' button.

General info	
Title	ptABP001C
Author	
Source	
Publication date	2018-10-02
Genre	

Basic Counts	
Adjective incidence	24.590
Adverb incidence	90.164
Content word incidence	581.967
Flesch index	54.143
Function word incidence	418.033
Mean sentences per paragraph	1.000

Fonte: Print Screen da tela principal Coh-Metrix-Port

### 3.3.3 Índice Flesch

O Índice Flesch, também conhecido como Índices de Legibilidade de Flesch ou Índice Facilidade de Leitura Flesch, foi criado pelo australiano Rudolf Franz Flesch em 1948. O índice Flesch corresponde a uma escala que varia de 0 a 100, onde índices altos indicam uma leitura mais fácil, enquanto índices baixos indicam uma mais difícil. A sua fórmula é:

Figura 3. Fórmula Índice Flesch

$$206.835 - 1.015 \left( \frac{\text{total de palavras}}{\text{total de sentenças}} \right) - 84.6 \left( \frac{\text{total de sílabas}}{\text{total de palavras}} \right)$$

Fonte: Scarton e Aluísio (2010)

De acordo com Scarton e Aluísio (2010), a adaptação do Índice Flesch para o português resultou na fórmula:

Figura 4. Fórmula Índice Flesch versão Português

$$248.835 - 1.015 \left( \frac{\text{total de palavras}}{\text{total de sentenças}} \right) - 84.6 \left( \frac{\text{total de sílabas}}{\text{total de palavras}} \right)$$

Fonte: Scarton e Aluísio (2010)

Essa fórmula é a mesma fórmula original do Índice Flesch somada com o número 42 que, de acordo com Martins et al. (1996), corresponde ao número que diferencia textos em inglês de textos em português. De acordo com Scarton e Aluísio (2010), os valores do índice Flesch variam entre 100-75 (muito fácil), 75-50 (fácil), 50-25 (difícil) e 25-0 (muito difícil).

Apesar de o Índice Flesch ser considerado por alguns críticos como um teste superficial, é importante reconhecer que, ainda hoje, a fórmula é um elemento relevante para medir a complexidade de um texto. Sendo assim, o Índice Flesch será usado como meio para calcular a dificuldade dos nossos *corpora* nesta monografia, pois, de acordo com Pasqualini (2012) o índice Flesch, quando contextualizado e enriquecido pelo acréscimo de outros elementos textuais, pode ser tomado como um indicador bastante confiável.



### 3.3.4 Sketch Engine

O *Sketch Engine* é uma ferramenta bastante sofisticada de análise textual. Esse software teve seu desenvolvimento iniciado pela *Lexical Computing Limited* em 2003. Suas principais funções são<sup>25</sup>: (i) *Word Sketch*: função que mostra o comportamento gramatical de uma palavra escolhida e suas relações com outras palavras; (ii) *Word Sketch difference*: diferentemente do *Word Sketch*, essa função permite que se analisem e comparem duas palavras para observar os seus usos. (iv) *Thesaurus*: essa função mostra o grupo de palavras próximas de uma palavra escolhida; (v) *Keyword and term extraction*: essa função combina estatística com critérios linguísticos para extrair possíveis termos e palavras-chave do *corpus*; (vi) *Concordance*: essa função pode ser usada pelo usuário para procurar exemplos de uma palavra, lema, frase, *tag* ou uma estrutura gramatical ou lexical complexa em seu contexto de uso; (vii) *Frequency lists*: essa função permite gerar listas de frequência de vários tipos, como, por exemplo, palavras, lemas, substantivos, verbos etc.

Além de o *Sketch Engine* ser usado para contagens básicas, como *types*, *tokens* e listas de frequências, nesta monografia ele foi usado para outra função. Para textos escritos em língua inglesa, há diversas ferramentas disponíveis gratuitamente que calculam a quantidade de sentenças na voz passiva, de forma automática. Inclusive, o *Microsoft Word* possui uma ferramenta que calcula a porcentagem de frases na voz passiva de um texto em inglês. De fato, utilizamos o próprio *Microsoft Word* para realizar os cálculos com os *corpora* em inglês. Porém, não foi encontrada nenhuma ferramenta que faça esse cálculo de forma automática para os textos em português. Como dito anteriormente, o uso da voz passiva é considerado um dificultador da complexidade de um texto. Portanto, calcular a quantidade de sentenças em voz passiva usadas em nossos *corpora* e comparar com o Índice Flesch era de vital importância para esta pesquisa. Sendo assim, encontramos um meio de calcular a quantidade de frases na voz passiva em português usando o *Sketch Engine*.

Em português, a voz passiva pode ser de duas formas: sintética ou analítica. A voz passiva sintética é formada por um verbo transitivo na terceira pessoa mais o pronome apassivador *se* (exemplo: Invadiu-se a festa). Já a voz passiva analítica é formada pelo verbo auxiliar *ser* ou *estar* mais o particípio de um verbo transitivo (Ex.: A casa será pintada). Como o *Sketch Engine* não

---

<sup>25</sup> É possível acessar a versão integral do tutorial das funções do *Sketch Engine* em seu site: <https://www.sketchengine.eu/quick-start-guide/>

distingue entre as diferentes formas da palavra *se* (como, por exemplo, índice de indeterminação do sujeito, partícula apassivadora, pronome etc.), por limitação de tempo e espaço, nesta pesquisa teremos como foco apenas os casos de passiva analítica. Sendo assim, nossa análise foi conduzida de modo semiautomático no *Sketch Engine* para encontrarmos as sentenças em voz passiva analítica em português. Primeiramente, utilizamos a ferramenta *concordance* para gerar as listas com todas as concordâncias e contextos do verbo *ser* e do verbo *estar*. Na figura 5, é possível ver algumas das linhas de concordância do verbo *ser*, em diferentes conjugações:

Figura 5. Linhas de concordância do verbo *ser* no *corpus* comparável em pt.

13/11/2018 ✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓ Sketch Engine

CONCORDANCE

simple ser **470** (21,091 per million)

Details	Left context	KWIC	Right context
1 ⓘ #188	ais próximos, como se toda a família adoecesse junto. O apoio psicológico	é	fundamental na superação dos problemas emocionais. Pacientes e familia
2 ⓘ #339	ões. O cuidador sente-se apoiado e compreendido, não se sente mais só,	é	incentivado ao convívio social, a reestruturar seu estilo de vida, escala de v
3 ⓘ #395	para falar de suas maiores dificuldades ao lidar com a doença. O encontro	é	na segunda segunda-feira do mês, das 14:00 s 15:30, na sede da ABP. Dur
4 ⓘ #427	cam experiências e buscam orientações. O contato com outros familiares	é	importante porque o cuidador descobre que é possível fazer muito pelo pac
5 ⓘ #434	ntato com outros familiares é importante porque o cuidador descobre que	é	possível fazer muito pelo paciente e que não é só de medicação que ele pre
6 ⓘ #443	je o cuidador descobre que é possível fazer muito pelo paciente e que não	é	só de medicação que ele precisa, mas também de amor, paciência e solida
7 ⓘ #714	oficina.html;acess 26/08/2018 <header> Oficina de Artes A Arte, além de	ser	uma expressão da criatividade, funciona como um recurso terapêutico, ind
8 ⓘ #938	os de seus colegas. Integração e socialização. Os Trabalhos Executados	são	Especialmente: Pinturas com guache sobre papel canson. Desenhos com
9 ⓘ #1026	ia familiar. Sua mãe, portadora da doença de Parkinson há 25 anos, vinha sendo	sendo	por ela incentivada a pintar como distração. A pintura proporcionou-lhe um
10 ⓘ #1298	entre 55 e 70 anos de idade, e alguns até acima dessa faixa. Os trabalhos	são	iniciados com música de fundo, que auxilia o aquecimento dos participante
11 ⓘ #1336	ráfica, a fim de soltar os movimentos e concentrar a atenção. Os trabalhos	são	iniciados com música de fundo, que auxilia o aquecimento dos participante
12 ⓘ #1376	soltar os movimentos e concentrar a atenção. Terminado o aquecimento,	é	dado a cada participante um trabalho de acordo com suas preferência e co
13 ⓘ #1562	sultados e metas alcançadas Os trabalhos executados na Oficina de Artes	são	vendidos nos bazares montados nas festas que a Associação promove ou
14 ⓘ #1591	n de angariar recursos para a manutenção da própria Oficina. Desde 2001	são	realizadas exposições dos quadros pintados na Oficina, evento que pode s
15 ⓘ #1603	realizadas exposições dos quadros pintados na Oficina, evento que pode	ser	realizado na entidade ou em espaços públicos, como já aconteceu no Espa
16 ⓘ #1695	z xícaras, canecas, copos e objetos de decoração. As aulas de artesanato	são	ministradas no mesmo horário de funcionamento da Oficina de Artes, toda
17 ⓘ #1735	da voluntária Elli Ahlemeyer, também diretora da ABP. Origami O Origami	é	uma técnica japonesa de dobraduras, transforma papel em diversas figura
18 ⓘ #1757	uras representativas da natureza. Acredita-se que a origem dessa prática	é	tão antiga quanto a do papel, que surgiu na China em 105 a.C. O contato d
19 ⓘ #1841	v.parkinson.org.br/firefox/oquee.html;acess 26/08/2018 <header> O que	é	Parkinson? A doença de Parkinson é uma enfermidade que foi descrita pel

Fonte: *Print Screen* da tela do *Sketch Engine*

Em seguida, procuramos por construções na voz passiva de modo manual em cada uma das listas, seguindo as regras gramaticais das construções em voz passiva em português – verbos no participio. Por fim, o *Sketch Engine* possui uma ferramenta que calcula de forma automática o número total de sentenças do *corpus* escolhido. Então, depois de encontrarmos as sentenças em

voz passiva, defrontamos a quantidade encontrada com a quantidade total de sentenças de cada um dos nossos *corpora* em português, para obtermos a porcentagem total de sentenças em voz passiva analítica.

Como foi possível observar na figura 5, a ferramenta *concordance* do *Sketch Engine* mostra todos os contextos do verbo escolhido, que nesse caso foi o verbo *ser*. Portanto, a distinção das sentenças em voz passiva de outras sentenças, como, por exemplo, as frases interrogativas (exemplo: O que é Parkinson?) e frases na voz ativa (exemplo: O origami é uma técnica japonesa de dobraduras...) tiveram que ser feita de forma manual, pois o *Sketch Engine* não consegue diferenciar sentenças em voz passiva de forma automática. O mesmo procedimento foi feito para o verbo *estar*. Assim como o verbo *ser*, a ferramenta *concordance* mostrou todos os contextos do verbo *estar*, incluindo sentenças em voz ativa (exemplo: O tremor não está presente em todos os pacientes com DP) e a forma verbal gerúndio (exemplo: No entanto, as alterações não são restritas à substância negra e nem restritas à dopamina, estando presentes também em outros núcleos cerebrais).

### 3.3.5 LF Aligner e AntPconc

Para fazermos o alinhamento de nosso *corpus* paralelo, usamos a ferramenta chamada *LF aligner*, criada por um tradutor húngaro chamado Andras Farkas. O *LF aligner* se baseia na ferramenta de código aberto *hunalign* para fazer, de forma automática, o alinhamento de textos paralelos em nível sentencial. O *hunalign* é um código aberto, desenvolvido por Dániel Varga do Departamento de Sociologia e Comunicações da Universidade de Tecnologia e Economia de Budapeste. Em nosso *corpus*, mais de 90% das frases foram alinhadas perfeitamente pelo *LF aligner*. Quando o programa não consegue alinhar alguma frase ou alinha de maneira errada, é possível corrigir o alinhamento.

Depois de os textos serem alinhados e salvos em formato arquivo de texto (\*.txt), com codificação UTF-8, eles foram processados em uma ferramenta chamada *AntPconc*<sup>26</sup>. O *AntPconc* é um programa gratuito que permite a análise de *corpus* paralelo alinhado. Ele foi criado pela equipe de Laurence Anthony, o responsável por criar um dos softwares gratuitos de análise de *corpus* mais populares, o AntConc. Apesar do programa não permitir que se faça nele o

<sup>26</sup> <http://www.laurenceanthony.net/software/antpconc/>. Acesso em 3 de dezembro de 2018.

alinhamento dos textos, ele permite que sejam carregados os textos já alinhados e que se busquem por palavras ou termos específicos.

### **3.4 Escolaridade dos públicos alvos dos textos de divulgação sobre a DP**

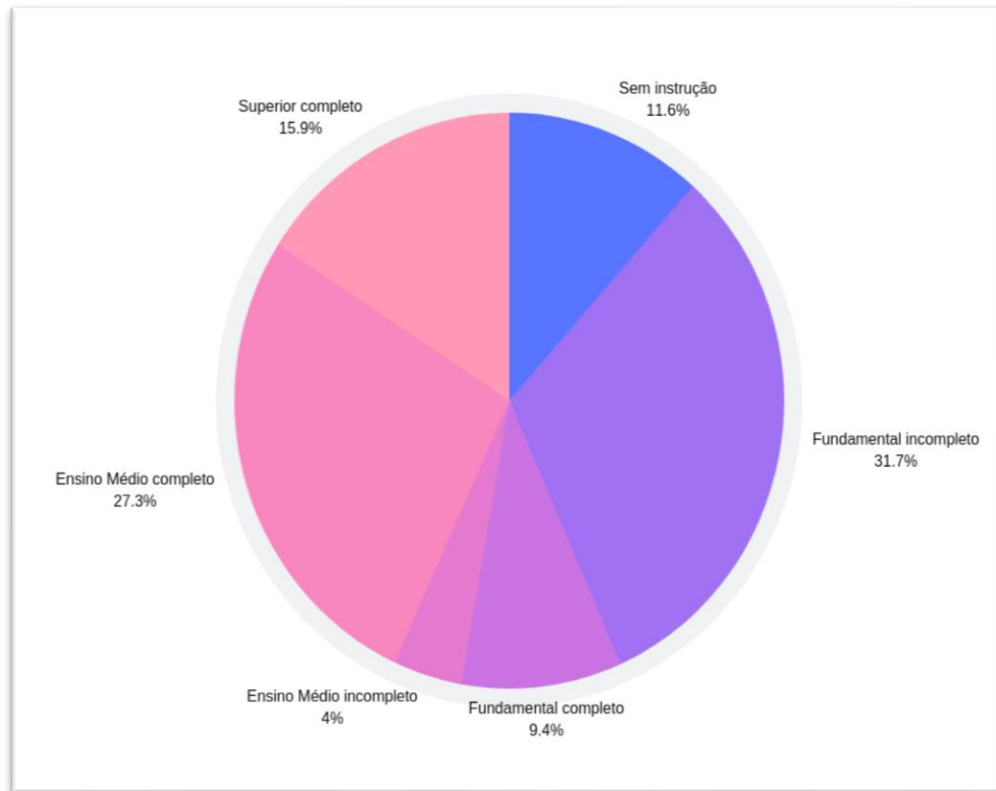
A habilidade de leitura e uso da internet permite que se abra um leque de possibilidades com um simples clique. Como dito anteriormente, é muito comum que o paciente, familiares ou cuidadores, ao receberem a notícia do diagnóstico de Parkinson, sintam-se atordoados, confusos e com medo e/ou vergonha de conversar com o médico sobre suas dúvidas e inseguranças. Eles recorrem à internet como forma de se instruírem e se informarem sobre tratamentos experimentais, sintomas, etc. Dessa forma, o público alvos dos textos de divulgação da DP acaba sendo os cuidadores, familiares até mesmo os pacientes.

Os usuários têm o direito de acessar um material textual mais simples, para que possam compreender esses materiais em sua totalidade ou a maior parte possível das informações. Sabemos que há muitas variáveis que influem na compreensão de um texto, e que não há como prever que um material será compreendido por todos os seus leitores. Porém, cabe ao autor ter empatia para com o público alvo e procurar tornar esse texto o mais acessível possível.

Segundo a pesquisa do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação, entre 2012 e 2016 a porcentagem de brasileiros com mais de 60 anos que usam a internet passou de 8% para 19%. O uso da internet aumenta a independência dos idosos, além de diminuir as chances de depressão e ansiedade e manter a mente ativa, com a possibilidade de aprender algo novo a cada dia. Apesar de sabermos que a doença de Parkinson pode ocorrer em jovens, ela ainda é mais frequente em idosos com mais de 65 anos. Cerca de uma em cada 250 pessoas com 40 anos de idade ou mais pode ser afetada pela doença, enquanto uma em 100 pessoas com 65 anos de idade ou mais pode ser afetada.

No Brasil, de acordo com o PNAD 2016, a taxa de analfabetismo das pessoas de 15 anos ou mais foi estimada em 7,2%, o que representa 11,8 milhões de analfabetos. Dentro dessa porcentagem, 20,4% possuem mais de 60 anos de idade. Na figura 6, é possível observar a distribuição da escolaridade dos brasileiros com 25 anos ou mais, segundo o PNAD 2016

Figura 6. Distribuição da escolaridade dos brasileiros com 25 anos ou mais



Fonte: Elaboração autoral

Em 2016, 11,2% da população brasileira com 25 anos ou mais não possuía nenhum tipo de instrução, 30,6% tinha fundamental incompleto, 9,1% tinha fundamental completo, 3,9% tinha ensino médio incompleto, 26,3% tinha ensino médio completo e 15,3%, superior completo. Sendo assim, podemos concluir que mais da metade da população de 25 anos ou mais no Brasil possui apenas até o ensino fundamental completo. Dentre os homens que fazem parte dos 11,2% da população de 25 anos ou mais que não possuía nenhum tipo de instrução, a razão mais frequente para a baixa instrução está relacionada ao trabalho (50,5%). Além disso, 24,1% disse não ter interesse e 8,2%, que já tinha concluído o nível de estudo que desejava.

Entre as mulheres, o motivo principal também foi o trabalho (30,5%), depois a obrigação de cuidarem da casa ou de crianças, adolescentes, idosos ou pessoas com necessidades especiais (26,1%) e por último não ter interesse nos estudos (14,9%). Vale a pena mencionar que, segundo o PNAD 2016, a proporção de mulheres jovens que não estudavam para realizar tarefas da casa ou cuidar de pessoas era 32,6 vezes superior a dos homens envolvidos nessas atividades. Isso, mais

uma vez, corrobora as conclusões de Karsch (2003), que afirma que geralmente as mulheres que residem no mesmo domicílio ficam com a tarefa de se tornarem as cuidadoras de seus filhos, pais e maridos acometidos pela doença de Parkinson, e que isso não é uma questão demográfica, e sim cultural. Esse *fenômeno* não ocorre apenas no Brasil, pois em diversos países as mulheres acabam assumindo o papel de *grandes cuidadoras*.

Morley et al (2012) conduziram um estudo para determinar as principais influências na qualidade de vida dos cuidadores de pacientes com Parkinson. Para isso, os autores pediram que a organização PUK divulgasse entre suas filiais e seus voluntários o convite para que pacientes de DP e seus cuidadores participassem da pesquisa. Os autores concluíram que no Reino Unido também há uma predominância de cuidadoras femininas. Das entrevistadas, 92% se identificaram como esposas ou parceiras de um paciente com Parkinson, e metade revelaram sofrer de algum tipo de doença ou doença crônica.

Parece importante enfatizar que a organização PUK é a principal organização de Parkinson no Reino Unido. Ela oferece diversos tipos de ajuda e suporte que um paciente de Parkinson pode precisar, além de recomendar grupos de suporte que se encontram perto do paciente. A população residente na Inglaterra e no País de Gales em 2011, de acordo com *Key Statistics for England and Wales*, era de 56,1 milhões de pessoas, onde 1/6 da população é composta de indivíduos com 65 anos ou mais. De acordo com uma pesquisa feita pela PUK em 2018, a incidência estimada de DP para pessoas com 45 anos ou mais no Reino Unido é de 18 mil. A cada hora, duas pessoas no Reino Unido são diagnosticadas com DP, o que equivale a dizer que uma em 37 pessoas será diagnosticada com DP em algum estágio da vida. É estimado que entre 2018 a 2025 a incidência aumente em 18%, já que a população do Reino Unido está crescendo e atingindo uma expectativa de vida maior.

A educação no Reino Unido é descentralizada, o que significa que cada país possui um sistema diferente de educação, gerenciado pelo seu próprio governo. Na Inglaterra, por exemplo, a alfabetização adulta é tratada por “níveis”. As estatísticas governamentais mostram que 14,9% (1 em 7) dos adultos ingleses possuem nível de alfabetização igual ou menor do que nível 3, que é equivalente ao nível de uma criança de 9 a 11 anos. De acordo com o website *National Literacy Trust*<sup>27</sup>, as últimas estatísticas sobre alfabetização adulta feita na Inglaterra (2011), Escócia (2009),

---

<sup>27</sup> <https://literacytrust.org.uk/>

País de Gales (2010) e Irlanda do Norte (2016) mostram que 1 em cada 7 adultos ingleses (14.9% / 5 milhões de pessoas) não possui habilidades básicas de alfabetização, 1 em 4 (26.7% / 931 mil pessoas) adultos escoceses sofre com problemas devido à falta de alfabetização, 1 em 8 (12% / 216 mil pessoas) adultos do país de Gales carece de habilidades básicas de alfabetização e 1 em 5 (17,9% / 550 mil pessoas) adultos na Irlanda do norte carece de habilidades básicas de alfabetização. Mesmo assim, a estimativa da taxa de analfabetos no Reino Unido é bastante baixa, o que corresponda a cerca de 0,5% da população. Portanto, mesmo que essas pessoas sejam alfabetizadas, ainda assim elas não possuem uma habilidade satisfatória de leitura e compreensão.

Apesar de não podermos comparar os níveis de alfabetização de cada país do Reino Unido, é possível perceber que a falta de alfabetização também é um problema na região. De fato, a pesquisa da Organização para a Cooperação e Desenvolvimento Econômico (OECD) de 2012 concluiu que o grupo de pessoas que carecem de alfabetização em língua inglesa é bastante variado, sendo composto por um grande número de imigrantes (que são alfabetizados em suas respectivas línguas nativas).

O índice de escolaridade nos EUA é bastante diferente do Brasil e Reino Unido. Apesar de apenas 1% da população ser analfabeta, de acordo com um estudo feito em abril de 2017 pelo Departamento de Educação dos Estados Unidos e pelo National Institute for Literacy (NIFL), 32 milhões de adultos americanos não conseguem ler um livro destinado para um leitor da oitava série. Isso equivale a 14% da população. A pesquisa ainda revelou que 21% dos adultos americanos possuem a habilidade de leitura inferior à de uma criança na quinta série, e 19% dos americanos com ensino médio completo possuem dificuldades para compreender um texto. De acordo com Sheida White, pesquisadora do Departamento de Educação dos Estados Unidos, eles possuem dificuldades em compreender parágrafos e sentenças que possuem ligações entre si. Em 2008, o *National Commission on Adult Literacy* publicou um relatório buscando discutir a crise de alfabetização que os EUA estão enfrentando. De acordo com a pesquisa, 1 em cada 3 americanos abandonou o ensino médio e 1 em cada 4 famílias americanas possui baixa renda, sendo composta por pais que possuem pouca ou nenhuma alfabetização. Isso representa um ciclo de baixa escolaridade, que poderá afetar as próximas gerações.

Como foi possível observar, as diferenças dos níveis médios de escolaridade e de alfabetismo do Brasil, Reino Unido e Estados Unidos são grandes, e apontam um sério problema na habilidade de leitura de seus cidadãos. Apesar disso, são poucas as ferramentas computacionais

que levam em conta os índices de escolaridade de cada região na hora de calcular a complexidade de um texto. De fato, o *Flesch-Kincaid Grade Level*, uma ferramenta americana, é um dos poucos que relaciona o resultado do Índice Flesch com uma determinada série escolar dos Estados Unidos. Essa fórmula é frequentemente usada por diversos países, sendo ignorada a diferença dos níveis médios de escolaridade e particularidades de cada país.



## 5 ANÁLISE

Como dito anteriormente, a voz passiva é um dos fatores que podem aumentar a dificuldade de um texto, tanto em português, quanto inglês. Mesmo assim, em português não há uma ferramenta que calcule a dificuldade de um texto que leve em conta a voz passiva. Em nosso *subcorpus* comparável dos textos em inglês britânico da PUK, foi encontrada uma porcentagem menor de sentenças na voz passiva analítica quando contrastado com o *corpus* comparável em português e o *subcorpus* comparável dos textos em inglês estadunidense da fundação MJF. Das 577 sentenças dos textos em inglês britânico, apenas 44 estão na voz passiva, o que corresponde a 8% do total de sentenças. O *corpus* em português obteve segundo lugar em menor presença de sentenças na voz passiva. Das 1101 sentenças do *corpus* comparável em português, compostos pelos textos de divulgação dos websites de associações de Parkinson, 100 eram em voz passivas, correspondendo a 9% do total de sentenças. O maior percentual de sentenças em voz passiva foi dos textos da fundação MJF. Das 845 sentenças, 125 eram em voz passiva, o que corresponde a 14,79% do total de sentenças.

No quadro 4 é possível observar a porcentagem das sentenças em voz passiva analítica do *corpus* comparável:

Quadro 4. O uso da voz passiva analítica no *corpus* comparável

Corpus	Quantidade de sentenças	Porcentagem das sentenças em voz passiva analítica
<i>Subcorpus</i> comparável fundação MJF	845	14,79%
<i>Subcorpus</i> comparável PUK	599	8%
<i>Corpus</i> comparável português	1101	9%

Fonte: Elaboração Autoral

Como vimos, as taxas de analfabetismo dos Estados Unidos e do Reino Unido são baixíssimas, especialmente quando comparadas com as do Brasil. Porém, a crise na alfabetização

parece ser um problema que também está presente nos países desenvolvidos, já que as pesquisas mostraram que, apesar de essas pessoas serem alfabetizadas, elas possuem uma habilidade fraca em ler e escrever. De acordo com DuBay (2004), materiais de leitura que possuem o propósito de instruir, utilizados em sala de aula com a supervisão de um professor, podem e devem ser mais difíceis do que o nível do leitor. Porém, materiais destinados para o público geral, como textos médicos, declarações de taxas etc. devem ser o mais fácil possível. Dessa forma, o uso da voz passiva seria um dificultador extra para essas pessoas com pouca habilidade de leitura. É importante ressaltar que um leitor com habilidades medianas ou acima da média não se importaria com tal presença. Porém para um leitor com pouca habilidade, a voz passiva pode ser o “empurrão” que faltava para ele abandonar o texto.

Já que estamos tratando de textos naturalmente produzidos que são representativos dos idiomas em questão, tais achados nos levam a intuir que tanto as associações brasileiras, quanto a PUK tiveram um cuidado maior na hora de usar a voz passiva analítica em suas sentenças. A associação MJF, por outro lado, foi a que fez maior uso da voz passiva. Podemos relacionar essa alta incidência da voz passiva nos textos da associação da MJF com a hipótese de Aluísio et al (2014), que diz que a voz passiva foi utilizada por muito tempo pelos americanos em artigos acadêmicos. Porém, a tendência vem mudando e as revistas que publicam artigos esperam que os autores deem preferência pela voz ativa em detrimento da passiva. Mesmo assim, como esse fenômeno é novo, podemos também intuir que o uso da voz passiva ainda é mais aceito pelos estadunidenses do que os britânicos e brasileiros.

A quantidade de uso da voz passiva também foi bastante surpreendente nas traduções que compõem nosso *corpus* paralelo em português. Quando as traduções são comparadas com os textos originais, é possível perceber que as traduções de ambos os textos da associação MJF e do PUK tiveram um aumento de sentenças na voz passiva. No caso dos textos do PUK, que possuía a menor porcentagem das sentenças em voz passiva, aumentou-se em 2% o uso da voz passiva nas traduções. A incidência de voz passiva nos textos originais em inglês estadunidense da associação MJF, que já era alto, teve um aumento de 1,3% na sua tradução.

No quadro 5, é possível observar a quantidade de sentenças no *corpus* paralelo e sua porcentagem em voz passiva.

Quadro 5. O uso da voz passiva analítica no *corpus* paralelo

<i>Corpus</i>	Quantidade de sentenças	Porcentagem das sentenças em voz passiva analítica
<i>Subcorpus</i> paralelo estadunidense fundação M.J. Fox	845	14,7%
<i>Subcorpus</i> paralelo português fundação M.J. Fox	845	16%
<i>Subcorpus</i> paralelo britânico PUK	599	8%
<i>Subcorpus</i> paralelo português PUK	577	10%

Fonte: Elaboração Autoral

Apesar de as alunas relatarem que o intuito eram traduções funcionais, mantendo-se o foco no leitor, houve um aumento no número de sentenças em voz passiva em relação aos originais. Isso tornou a incidência de voz passiva do *corpus* paralelo em português maior que as do *corpus* comparável em português, composto de textos escritos originalmente em português. Como foi dito anteriormente, o Brasil, em comparação aos EUA e ao Reino Unido, é o país com maior número de analfabetos (11,2% da população), enquanto a maioria dos brasileiros (30,6%) possui o ensino fundamental incompleto. Portanto, era esperado que os textos de divulgação da DP tivessem um menor número de sentenças em voz passiva, do que os textos em inglês estadunidense e britânico, pois o uso da voz passiva é considerado um dificultador da compreensão.

De fato, os resultados do índice Flesch confirmam que houve um aumento no potencial de dificuldade dos textos traduzidos para o português. Como já mencionado, os valores do índice Flesch variam entre 100-75 (muito fácil), 75-50 (fácil), 50-25 (difícil) e 25-0 (muito difícil). É possível observar no quadro 6 o índice Flesch de cada um dos textos do *subcorpus* paralelo da associação Michael J. Fox.

Quadro 6. Índice Flesch do *subcorpus* paralelo estadunidense e suas traduções

Arquivo	Índice Flesch	Nº de type/tokens	Arquivo	Índice Flesch	Nº de type/tokens
enNAMjF001P	41.289	357/833	ptMjF001P	39.623	336/780

enNAMjF002P	41.169	233/468	ptMjF002P	32.756	207/423
enNAMjF003P	48.251	415/829	ptMjF003P	34.352	426/893
enNAMjF004P	44.660	103/154	ptjF004P	50.510	119/181
enNAMjF005P	65.471	173/307	ptMjF005P	58.933	189/346
enNAMjF006P	59.608	1756/7776	ptMjF006P	50.961	1833/7323
enNAMjF007P	66.063	67/91	ptMjF007P	22.162	59/74
enNAMjF008P	27.603	505/1131	ptMjF008P	22.073	534/1316
enNAMjF009P	40.181	467/1302	ptMjF009P	41.361	493/1431
enNAMjF010P	29.542	987/2312	ptMjF010P	27.732	983/2561
enNAMjF011P	44.163	468/1271	ptMjF011P	39.926	475/1234

Fonte: Elaboração Autoral

A mediana do índice Flesch dos textos originais em inglês estadunidense da associação MJF foi 44,163, o que encaixa na categoria *difícil*, com um desvio padrão de 12,9. A predominância foi de textos na categoria *difícil*, com apenas 3 se encaixando na categoria *fácil*. Já a mediana das traduções desses mesmos textos para o português foi de 39, com um desvio padrão de 11,9, o que ainda continua na categoria *difícil*. Dos textos traduzidos da MJF para o português, foi mantida a predominância de textos *difíceis*, sendo que um texto passou da categoria *difícil* para a *fácil* e dois textos passaram das categorias *fácil* e *difícil* para *muito difícil*. Apenas na tradução a categoria *muito difícil* foi encontrada.

Nas traduções dos textos da PUK, também houve um aumento na dificuldade, de acordo com o índice Flesch. Nas medidas a seguir usamos os valores de mediana e desvio padrão. Mediana representa o valor que se encontra no meio de todas as medidas registradas. Já o desvio Padrão demonstra a dispersão das medidas registradas em torno da mediana, representando a uniformidade entre as medidas. A mediana do índice Flesch dos textos originais em inglês britânico da PUK foi de 63, que se encaixa na categoria *fácil*, com um desvio padrão de 17. A predominância foi de

textos na categoria *fácil*, com 1 texto na categoria  *muito fácil* e apenas 1 texto na categoria *difícil*. Já a mediana das traduções desses mesmos textos para o português caiu para 51, com desvio padrão de 8. As traduções continuaram na categoria *fácil*. Também, 3 textos passaram para a categoria *muito difícil*, o que não existia nos originais, e nenhum texto se encaixou na categoria de *muito fácil*.

Abaixo, é possível observar o score Flesch de cada um dos textos do *corpus* paralelo – inglês e português – da Parkinson’s UK.

Quadro 7. Índice Flesch no *subcorpus* paralelo em inglês britânico e suas traduções

Arquivo	Índice Flesch	Nº de type/tokens	Arquivo	Índice Flesch	Nº de type/tokens
enBRPUK001P	72.918	441/1608	ptBRPUK001P	61.134	441/1608
enBRPUK002P	68.266	606/2050	ptBRPUK002P	51.422	606/2050
enBRPUK003P	66.786	163/392	ptBRPUK003P	52.104	163/392
enBRPUK004P	56.730	733/3017	ptBRPUK004P	43.738	733/3017
enBRPUK005P	63.415	208/515	ptBRPUK005P	57.917	208/515
enBRPUK006P	58.168	134/266	ptBRPUK006P	62.197	134/266
enBRPUK007P	58.767	448/1476	ptBRPUK007P	42.615	448/1476
enBRPUK008P	66.063	67/91	ptBRPUK008P	22.162	67/91
enBRPUK009P	64.323	175/405	ptBRPUK009P	56.528	175/405
enBRPUK010P	46.691	153/296	ptBRPUK010P	17.564	153/296
enBRPUK011P	61.020	117/199	ptBRPUK011P	37.920	117/199
enBRPUK012P	56.251	134/223	ptBRPUK012P	22.417	134/223
enBRPUK013P	80.614	135/277	ptBRPUK013P	71.949	135/277

Fonte: Elaboração Autoral

No *corpus* comparável em português, a mediana obtida foi 30, com desvio padrão 11. Essa mediana baixíssima se enquadra na categoria Flesch *difícil*. A predominância foi de textos na categoria *difícil*, com apenas 2 na categoria *fácil* e 7 na categoria muito *difícil*. Embora esse *corpus* tenha obtido uma baixa porcentagem de sentenças na voz passiva, ele apresentou um baixo índice Flesch. Como DuBay (2004) afirma, não adianta apenas reduzir o tamanho das sentenças e usar palavras menores para diminuir a complexidade de um texto. É preciso observar diversos outros fatores que podem afetar a leitura de um texto. No quadro 8 é possível observar o score Flesch de cada um dos textos do *corpus* comparável em português, formado pelas associações de Parkinson

Quadro 8. Índice Flesch no *corpus* comparável em português.

Arquivo	Score Índice Flesch	Nº de type/tokens
ptABP001C	54.143	91/128
ptABP002C	38.406	165/271
ptABP003C	37.694	118/185
ptABP004C	30.157	470/1018
ptABP005C	43.704	629/1543
ptABP006C	28.916	395/726
ptABP007C	31.868	333/731
ptABP008C	37.430	281/497
ptABP009C	38.284	214/389
ptAPARS001C	26.969	538/1324
ptAPARS002C	12.532	232/433
ptAPC001C	43.954	273/552
ptAPC002C	30.789	295/571
ptAPPP001C	46.760	774/1767
ptAPPP002C	12.532	569/1299
ptAPPP003C	24.569	687/1771
ptVP001C	22.223	253/548
ptVP002C	35.254	461/1147
ptVP003C	26.249	215/430
ptVP004C	14.890	99/173
ptVP005C	32.826	349/823
ptVP006C	14.021	489/1129
ptVP007C	30.206	570/1376
ptVP008	54.294	214/432
ptVP009C	21.284	135/226

Fonte: Elaboração Autoral

O Índice Flesch leva em consideração apenas o tamanho das sentenças e a quantidade de sílabas, ignorando outros elementos dificultadores, como a voz passiva, construções inesperadas,

termos e jargões pouco utilizados. Seria necessária uma análise mais profunda para podermos de fato afirmar que os textos das associações de Parkinson são mais difíceis do que as traduções dos textos originais em inglês, ou afirmar que as traduções são mais difíceis. Ademais, é necessário ressaltar que os resultados encontrados podem ser alterados com a presença das sentenças em voz passiva sintética, que, conforme mencionado anteriormente, não foram possíveis de ser detectadas nesta monografia. Contudo, a fórmula Flesch e a porcentagem de sentenças em voz passiva analítica podem ser tomadas como um ponto de partida para uma análise mais profunda de um texto, buscando por maneiras de torná-lo mais fácil e compreensível.

A simplificação textual é um conjunto de ações que busca, depois de uma análise, tornar um texto mais acessível para um determinado grupo de pessoas. De acordo com Finnato (2016):

[...] para alcançar a acessibilidade de informação textual escrita, existem propostas que vão desde a simplificação do texto, com trocas de estruturas linguísticas e de vocabulário, até a apresentação das informações em diferentes formatos – como a inclusão de imagens e outros recursos audiovisuais, como os infográficos. (Finnato, 2016, p.3)

Uma das principais técnicas utilizadas na hora de simplificar sintaticamente um texto, além de encurtar frases muito longas, é justamente transformar a voz passiva em ativa. Com o intuito de propor uma simplificação, buscamos pelas sentenças em voz passiva nos textos em português que tiveram menor Índice Flesch: ptMjF007P, ptMjF008P, ptBRPUK008P, ptBRPUK010P, ptBRPUK012P. Os textos selecionados possuem um total de 67 sentenças, sendo 5 em voz passiva. Sendo que dois arquivos, ptBRPUK010C e ptBRPUK012C, não apresentaram nenhuma sentença em voz passiva. Nas figuras 7 e 8 é possível ver uma foto da tela do *AntPconc*, que mostra as sentenças em voz passiva encontradas em português e a sentença original em inglês abaixo. Nas figuras 7 e 8, é possível perceber também que as traduções optaram por manter as sentenças em voz passiva. Contudo, conforme Alúcio et al (2014), a passiva no inglês foi utilizada por muito tempo pelos americanos em artigos acadêmicos. Diferentemente do português, onde a voz passiva pode soar mais formal e menos habitual.

Figura 7. Visualização passivas no AntPconc 1

AntPconc

File Help

Search Term  Words  Case Show Nth Entry 1 Search Corpus  pt  eng

\*ido Start Context Size 50 words

Sort 1: CEN Sort 2: 1R Sort 3: 2R Sort Page Size All entries

Target: pt Hits: 2 100% Font Size

Line	KWIC
1	A postura distônica pode fazer com que o braço seja mantido próximo ao corpo, dobrado no cotovelo, enquanto o
2	Como ocorre com a DP, o diagnóstico é clínico, obtido com base em histórico médico e avaliação física do pac

Line	Reference
1	The dystonic posture may cause the arm to be held close to the body and bent at the elbow and the wrist and fingers to be flexed toward the palm.
2	(As with PD, the diagnosis is based on a person's medical history and physical examination.)

Fonte: *Print Screen* da tela do *AntPconc*

Figura 8. Visualização de voz passiva no AntPconc 2



AntPConc

File Help

Search Term  Words  Case Show Nth Entry 1 Search Corpus  pt  eng

são \*adas Start Context Size 50 words

Sort 1: CEN Sort 2: 1R Sort 3: 2R Sort Page Size All entries

Target: pt Hits: 3 100% Font Size

Line	KWIC
1	Geralmente essas doenças são chamadas de "Parkinson plus" (no inglês plus significa "mai
2	Terapias diferentes são recomendadas para pessoas diferentes; experimente e veja
3	Terapias diferentes são recomendadas para pessoas diferentes; experimente e veja

Line	Reference
1	These conditions are often referred to as "Parkinson's plus" because they mimic PD but have extra associated symptoms (the "plus").
2	Different therapies will appeal to different people, so try things out and see what works for you.
3	Different therapies will appeal to different people, so try things out and see what works for you.

Fonte: *Print Screen* da tela do *AntPConc*

Convertendo a voz passiva em voz ativa, acreditamos que o leitor conseguiria saber de maneira mais fácil quem é o autor da ação ou quem está sofrendo a ação, além de evitar que o leitor não entenda e/ou entenda errado quem está realizando ou sofrendo a ação na sentença. Sendo assim, no quadro 9, apresentamos uma sugestão de conversão das sentenças em voz passiva para voz ativa:

Quadro 9. Simplificação Textual – voz passiva para ativa

Voz passiva	Voz Ativa
Ex. 1: Geralmente essas doenças são chamadas de “Parkinson plus” (no inglês plus significa “mais”), já que se assemelham à DP, mas apresentam outros sintomas também. (ptMjF008P)	Ex. 1: Geralmente, as doenças chamadas “Parkinson plus” (no inglês plus significa “mais”) são doenças que se assemelham à DP, mas apresentam outros sintomas também. (ptMjF008P)
Ex. 2: Terapias diferentes são recomendadas para pessoas diferentes; experimente e veja o que funciona para você. (ptMjF007P e ptBRPUK008P)	Ex. 2: Pessoas diferentes devem receber terapias diferentes; experimente e veja o que funciona para você. (Essa frase apareceu 2x)

	em dois textos diferentes). (ptMjF007P e ptBRPUK008P)
Ex. 3: A postura distônica pode fazer com que o braço seja mantido próximo ao corpo, dobrado no cotovelo, enquanto o pulso e os dedos ficam flexionados em direção à palma da mão. (ptMjF008P)	Ex. 3: Na postura distônica, o braço fica próximo ao corpo, dobrado no cotovelo, enquanto o pulso e os dedos ficam flexionados em direção à palma da mão. (ptMjF008P)
Ex. 4: Como ocorre com a DP, o diagnóstico é clínico, obtido com base em histórico médico e avaliação física do paciente. (ptMjF008P)	Ex. 4: Como ocorre com a DP, o histórico médico e a avaliação física do paciente servem de base para o diagnóstico clínico. (ptMjF008P)

Fonte: Elaboração Autoral

Depois de passarmos as sentenças em voz passiva para a voz ativa, utilizamos o índice Flesch, disponível no website do Coh-Metrix-Port, para calcularmos se houve, uma diminuição do potencial de complexidade dos textos. Apenas no exemplo 4 o número de palavras da sentença na voz passiva se manteve igual ao número de palavras da sentença em voz ativa, em todas as outras houve uma redução do número de palavras. Em dois textos, ptMjF007P e ptBRPUK008P, o score do índice Flesch da categoria  *muito difícil*  passou para  *difícil* , já que os dois subiram de, respectivamente, 22,162 para 25,203 e 22,162 para 25,203. No arquivo ptMjF008P, pelo contrário, houve uma diminuição de 22,073 para 21,883, mesmo com a quantidade de palavras diminuindo de 15 para 13.

Ainda não existe uma fórmula ideal que calcule a complexidade de um texto. Existem ferramentas que auxiliam os autores na busca pela simplificação de um texto. O  *PorSimples*  é uma dessas ferramentas, desenvolvida para o português brasileiro pelo NILC (Núcleo Interinstitucional de Linguística Computacional) da Universidade de São Paulo. De acordo com o website da Associação Brasileira de Editores de Livros Escolares<sup>28</sup>, essa ferramenta busca simplificar a linguagem de textos em português, facilitando o entendimento deles por crianças e adultos em processo de alfabetização, pessoas com problemas de linguagem ou cognitivos, além dos considerados analfabetos funcionais. O programa identifica tanto estruturas sintáticas complexas

<sup>28</sup> <http://www.abrelivros.org.br/home/index.php/noticias/2589-porsimples>. Acesso em 22 de novembro de 2018.

e as divide em orações mais curtas ou altera a sua estrutura, quanto simplifica lexicalmente ao destacar palavras complexas e oferecer sinônimos para elas. Porém, o site é bastante instável, e durante esta pesquisa (último acesso em novembro) ele se encontrava fora do ar. O *CorPop* é outra ferramenta *on-line* criada com o intuito de ajudar na simplificação de um texto. Ele foi criado a partir da pesquisa de doutorado de Bianca Pasqualini (2018) com o intuito de servir como um *corpus* de referência do português popular. O acervo do *CorPop* foi desenvolvido através da análise de dados sobre o nível de letramento dos leitores brasileiros, portanto seu acervo foi desenvolvido para esses leitores. O *CorPop* pode ser utilizado como referência para a escolha de termos ou palavras acessíveis, já que ele possui ferramentas como lista de palavras, concordanciador e N-Gram.

## 5 CONSIDERAÇÕES FINAIS

Existem poucos trabalhos acadêmicos que tratem da complexidade textual de websites, principalmente de websites especializados, criados por estudiosos, doutores e especialistas e que servem como repositórios de informações sobre diversas doenças. Os tratamentos e sintomas da doença de Parkinson encontram nos textos de divulgação das associações e fundações um lugar em que podem informar e, ao mesmo tempo, conscientizar outras pessoas. Como a DP não tem cura e pesquisadores preveem que no futuro o número de pessoas com essa doença irá aumentar, devido, principalmente, à expectativa do aumento de vida, é de fundamental importância que se discuta sobre essa doença, com o foco nos pacientes, familiares e cuidadores. Por isso, esperamos que esta monografia contribua para incentivar mais pesquisas sobre esse assunto, principalmente utilizando a Linguística de *Corpus*.

Assim, além de outras vantagens de usar a Linguística de *Corpus* em sua pesquisa, o pesquisador encontra nesse campo a possibilidade de compartilhar o seu trabalho (o *corpus*) com outros estudiosos, para que possam explorar diversas outras características dos textos compilados.

Além disso, a LC não se baseia apenas na intuição do pesquisador, e sim em textos autênticos, que servem como uma representação de determinado idioma.

Entendemos que alguns pesquisadores podem se sentir relutantes em aderir à LC em suas pesquisas, devido ao medo e preconceito em usar novas tecnologias. Mesmo assim, esperamos motivar os autores a usarem ferramentas que possam auxiliar na simplificação de seus textos, com o intuito de não só ajudar os pacientes de Parkinson, seus familiares e cuidadores a entenderem sobre a doença, mas também qualquer pessoa que busca por algum tipo de informação, mas que não possui domínio da área ou nível satisfatório de alfabetização.

Como apontado na análise e embasado pelo referencial teórico, o índice Flesch sozinho não representa a complexidade de um texto. Porém, quando complementado com outros elementos, considerados dificultadores, ele passa a ser um relevante índice para medir a complexidade de um texto. O índice Flesch também serve como um ponto de partida para explorar outros aspectos na escrita de um texto. Concordamos com DuBay (2012) quando afirma que as pesquisas sobre alfabetização tornaram as pessoas conscientes das limitações nas habilidades de leitura de outras pessoas, enquanto as pesquisas em complexidade textual e inteligibilidade nos tornaram conscientes dos diversos fatores que afetam a leitura. Por fim, DuBay (2012) acredita que as fórmulas de inteligibilidade, quando usadas de maneira apropriada, ajudam a aumentar o sucesso da compreensão da leitura.

Elementos sintáticos, como o uso da passiva, possui um papel bastante importante na compreensão de um texto. Segundo pesquisas, a voz passiva pode confundir o leitor, já que alterna a ordem canônica, surpreendendo o leitor com uma construção menos esperada. Nossos resultados mostraram que as traduções feitas pelas alunas tiveram um aumento no número de sentenças em voz passiva em relação aos originais. Isso tornou a incidência de voz passiva do *corpus* paralelo em português maior que as do *corpus* comparável em português, composto de textos escritos originalmente em português. Levando em consideração que o leitor médio brasileiro possui apenas o ensino fundamental completo e que ainda há uma grande taxa de analfabetos, esperava-se menor incidência de voz passiva nas traduções.

Durante esta monografia, pudemos conhecer algumas das ferramentas que auxiliam os autores na busca pela simplificação de um texto. Porém, é preciso lembrar que as ferramentas de análise de complexidade e de simplificação ainda são muito incipientes, mas geram métricas

importantes para discutir a complexidade de um texto, como o *Coh-Metrix* e o *Coh-Metrix-Port*, além de outras que ajudam a simplificar esses textos, como o *PorSimples* e o *CorPop*.

Por fim, precisamos ressaltar, mais uma vez, que a compreensão de um texto envolve diversos fatores. Segundo Kleiman (1989), três aspectos são essenciais e precisam ser levados em consideração: o conhecimento linguístico, textual e enciclopédico do leitor. Também, há diversos estudos, especialmente médicos, que focam na função do cérebro durante o processo de compreensão. Nesta pesquisa foi possível lidar apenas com um aspecto sintático que afeta a compreensão. No futuro, uma pesquisa com participantes de diferentes idades e níveis de alfabetização poderia agregar mais confiabilidade aos resultados. Para isso, poderia ser relacionado o uso da voz passiva com os resultados do índice Flesch em contextos reais de leitura. Ainda, por falta de espaço e tempo não foi possível observar outras formas da voz passiva em português. Em pesquisas futuras, pretendemos analisar essas outras formas e propor a criação de um software que faça essa identificação de forma automatizada. A criação desse software pouparia o pesquisador de realizar esses cálculos de maneira manual, garantindo maior confiabilidade, e seria um indicador para o autor que deseja tornar seu texto mais acessível.

## REFERÊNCIAS

ALUÍSIO, Sandra M. et al. Models for Scientific Writing. In: SCHUSTER, Ethel et al (Ed.). *Writing Scientific Papers in English Successfully: Your Complete Roadmap*. São Carlos: Compacta, 2014. Cap. 2. p. 31-56.

BERBER SARDINHA, Tony. Linguística de Corpus: histórico e problemática. *DELTA*, São Paulo, v. 16, n. 2, p. 323-367, 2000. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S010244502000000200005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010244502000000200005&lng=en&nrm=iso)>. Acesso em: 25 nov. 2018.

\_\_\_\_\_. *Linguística de corpus*. Barueri: Manole, 2004.

BEST, Rachel M. et al. Deep-Level Comprehension of Science Texts. *Topics In Language Disorders*, [S.l.], v. 25, n. 1, p.65-83, jan. 2005. Ovid Technologies (Wolters Kluwer Health). <http://dx.doi.org/10.1097/00011363-200501000-00007>.

BARBOZA, Elza Maria Ferraz; NUNES, Eny Marcelino de Almeida. A inteligibilidade dos websites governamentais brasileiros e o acesso para usuários com baixo nível de escolaridade. *Inclusão Social*, Brasília, v. 2, n. 2, p.19-33, 2007. Disponível em: <<http://revista.ibict.br/inclusao/article/view/1599/1805>>. Acesso em: 02 dez. 2018.

DAVISON, Alice; GREEN, Georgia M. *Linguistic Complexity and Text Comprehension: Readability Issues Reconsidered: Psychology of Reading and Reading Instruction Series*. [S. L.]: Routledge, 1988.

DUBAY, William H. *The Principles of Readability*. Califórnia: Impact Information, 2004. Disponível em: <<http://www.impact-information.com/impactinfo/readability02.pdf>>. Acesso em: 2 dez. 2018.

ELHADAD, Noemie. Comprehending technical texts: Predicting and defining unfamiliar terms. In: *Proceedings of the American Medical Informatics Association Symposium*. Washington, DC: [s.n],p. 239-243, 2006.

FINATTO, Maria José Bocorny. Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. *Organon*, Porto Alegre, v. 25, n. 50, p. 30-45, 2011. Disponível em: <<https://seer.ufrgs.br/organon/article/view/28340/16989>>. Acesso em: 26 junho.2018.

FULGÊNCIO, Lúcia.; LIBERATO, Yara. *Como Facilitar a Leitura*. São Paulo: Contexto, 2004.

CUNHA, Maria Angélica Furtado da. A complexidade da passiva e as implicações pedagógicas do seu uso. *Linguagem & Ensino*, [s. L.], v. 3, n. 1, p.107-116, 2000. Disponível em: <<http://www.rle.ucpel.tche.br/index.php/rle/article/view/285/251>>. Acesso em: 2 dez. 2018.

GENTZLER, Edwin. *Teorias contemporâneas da tradução*. São Paulo: Madras, 2009.

GRAESER, A.; MCNAMARA.; LOUWERSE, M. What readers need to learn in order to process coherence relations in narrative and expository text. In: SWEET, A.; SNOW, C. (Org.). *Rethinking Reading Comprehension*. The Guilford Press, 2003. p. 82-98.

KARSCH, Ursula M.. Idosos dependentes: famílias e cuidadores. *Cadernos de Saúde Pública*, [s.l.], v. 19, n. 3, p.861-866, jun. 2003. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0102-311x2003000300019>.

KLEIMAN, Angela. *Texto e leitor: aspectos cognitivos da leitura*. Campinas: Pontes, 1889. 1.ed.

LEFFA, Vilson. J. O conceito de leitura. In: LEFFA, V. J. *Aspectos da leitura*. Porto Alegre: Sagra-Luzzato, 1996. p. 9-24.

NORD, Christiane. Loyalty and fidelity in specialized translation. *Confluências: Revista de Tradução Científica e Técnica*, Germany, v. 4, p.29-41, 2006.

\_\_\_\_\_.Function plus Loyalty: Ethics in Professional Translation. *Génesis*, [s. L.], p.7-17, 2007.

\_\_\_\_\_. Functional approaches to translation. In: CHAPELLE, Carol A. (ed.) *The Encyclopedia of Applied Linguistics*. Hoboken: Blackwell Publishing, 2012. p. 2223-2228.

PASQUALINI, Bianca Franco. *Leitura, tradução e medidas de complexidade textual em contos da literatura para leitores com letramento básico*. 2012. 155 f. Monografia (Especialização) - Curso de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012. Disponível em: <<https://www.lume.ufrgs.br/handle/10183/61188>>. Acesso em: 2 nov. 2018.

REBECHI, Rozane Rodrigues. *A Tradução da Culinária Típica Brasileira para o Inglês: um Estudo sob o Enfoque da Linguística de Corpus*. 2015. 355 f. Tese (Doutorado) - Curso de Estudos Linguísticos e Literários em Inglês, Universidade de São Paulo, São Paulo, 2015. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8147/tde-29092015-162829/pt-br.php>>. Acesso em: 25 nov. 2018.

SCARTON, Carolina; ALUÍSIO, Sandra. Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. *Linguamática*, v. 2, n. 1, p. 45-61, 7 abr. 2010.

SEVERANCE, Samuel; COHEN, Kevin Bretonnel. Measuring the readability of medical research journal abstracts. In: WORKSHOP ON BIOMEDICAL NATURAL LANGUAGE PROCESSING (BIONLP 2015), 1., 2015, Beijing. *Proceedings*. China: Association For Computer Linguistics, 2015. p. 127 - 133.

SMITH, Frank. *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*. 6. ed. New Jersey: Lawrence Erlbaum, 1971.

## SITES CONSULTADOS

AntPconc: Disponível em: <http://www.laurenceanthony.net/software/antpconc/>. Acesso em 3 de dez. de 2018.

Associação Brasil Parkinson: Disponível em: <http://www.parkinson.org.br/>. Acesso em: 25 nov. 2018.

Associação Brasileira de Editores de Livros Escolares. Disponível em: <http://www.abrelivros.org.br/home/index.php/noticias/2589-porsimples>. Acesso em: 25 nov. 2018.

Associação Paranaense dos Portadores de Parkinsonismo Disponível em: <http://www.appp.com.br/site/>. Acesso em: 25 nov. 2018.

Associação Parkinson Carioca. Disponível em: <http://parkinsoncarioca.com.br/> Acesso em: 25 nov. 2018.

Associação Parkinson do Rio Grande do Sul. Disponível em: <https://www.apars.org.br/> Acesso em: 25 nov. 2018.

Internet World Stats. Disponível em: <https://www.internetworldstats.com/stats.htm>. Acesso em: 25 nov. 2018.

Michael J.Fox Foundation: Disponível em: <https://www.michaeljfox.org/foundation/index.html?navid=foundation-home>. Acesso em: 25 nov. 2018.

Parkinson's UK: Disponível em: <https://www.parkinsons.org.uk/>. Acesso em: 25 nov. 2018

Sketch Engine. Disponível em: <https://www.sketchengine.eu/quick-start-guide/>. Acesso em: 25 nov. 2018.



UCB Canadá. Disponível em: <https://www.ucb-canada.ca/en/Patients/Conditions/Parkinson-s-Disease>. Acesso em: 25 nov. 2018.

Vibrar com Parkinson. Disponível em: <http://vibrarcomparkinson.com.br/>. Acesso em: 25 nov. 2018.

#### **ANEXO A – Tema dos textos que compõem os nossos *corpora***

Arquivo	Tema
ptABP001C	Orientações para Atividades Diárias
ptABP002C	Psicologia
ptABP003C	Orientações Gerais para os Exercícios
ptABP004C	Oficina de Artes
ptABP005C	O que é Parkinson?
ptABP006C	Nutrição
ptABP007C	Fonoaudiologia

ptABP008C	Fisioterapia
ptABP009C	Coral da ABP
ptAPARS001C	Doença de Parkinson
ptAPARS002C	Atividades físicas
ptAPC001C	O que é a Doença de Parkinson?
ptAPC002C	Doença de Parkinson
ptAPPP001C	A doença O que é a doença de parkinson?
ptAPPP002C	Benefícios A quem se destina: (sobre tratamentos)
ptAPPP003C	Direitos assegurados aos parkinsonianos
ptVP001C	Doença de Parkinson O que é?
ptVP002C	Sinais e Sintomas
ptVP003C	Diagnóstico
ptVP004C	Classificação da doença de Parkinson (sub-tipos) de acordo com a causa
ptVP005C	Escalas de Avaliação da Doença de Parkinson
ptVP006C	Tratamentos
ptVP007C	Aspectos psicológicos – O Impacto Psicológico
ptVP008C	Recebeu o diagnóstico de Parkinson. E agora?
ptVP009C	Direitos e Benefícios
enNAMjF001C	Anxiety and Depression with Parkinson's Disease
enNAMjF002C	Exercise and Parkinson's Disease
enNAMjF003C	Diet and Parkinson's Disease
enNAMjF004C	Driving with Parkinson's Disease
enNAMjF005C	Dexterity and Parkinson's Disease

enNAMjF006C	Guide for the newly diagnosed
enNAMjF007C	Therapies
enNAMjF008C	Atypical parkinsonism
enNAMjF009C	Ask the MD: faqs on Lewy Body Dementia
enNAMjF010C	Dementia with lewy bodies: in the news and in the clinic
enNAMjF011C	Commonly Used Drugs Improve Depression in People with PD
enBRPUK001C	Talking to children and teenagers
enBRPUK002C	Being in a relationship
enBRPUK003C	Forming new relationships
enBRPUK004C	How can parkinson's affect sex?
enBRPUK005C	Newly diagnosed with Parkinson's
enBRPUK006C	How does Parkinson's progress?
enBRPUK007C	Drug treatments
enBRPUK008C	Therapies
enBRPUK009C	Newly diagnosed with Parkinson's
enBRPUK010C	Money, grants and benefits
enBRPUK011C	Driving
enBRPUK012C	Exercise with mild symptoms
enBRPUK013C	Questions to ask my doctor or nurse