

Processamento da Linguagem Natural e Complexidade Sintática

Bruna R. da Silva¹

¹Programa de Pós-Graduação em Letras – Universidade Federal do Rio Grande do Sul (UFRGS)

Caixa Postal 91540-000 – Porto Alegre – RS – Brasil

thu_du@hotmail.com

Resumo: Este trabalho se insere nos estudos de Acessibilidade Textual e Terminológica. O *corpus* de estudo é a publicação digital *Aprendendo sobre vírus e vacinas*. Os *corpora* de contraste são: (1) conjunto de textos sobre cuidados paliativos, e (2) conjunto de sentenças dos Juizados Especiais Cíveis. O objetivo é mapear as características sintáticas desses *corpora*, a fim de verificar qual o *corpus* mais simples. Tais *corpora*, de características distintas, foram etiquetados pela gramática UDPipe e explorados conforme alguns critérios pré-estabelecidos de análise sintática. Os contrastes indicam que o *corpus* de estudo tem sintaxe mais simples, embora alguns dos pontos explorados mereçam especial atenção. Sendo assim, tais comparações vão enriquecer a análise e a discussão sobre a acessibilidade desse material, além de servir de base para o estudo como um todo.

Abstract: This work is part of the studies of Textual and terminological accessibility. The *corpus* of study is the digital publication *Learning about viruses and vaccines*. The contrasting *corpora* are: (1) set of texts about palliative care, and (2) set of judgments of the Special Civil Courts. The goal is to map the syntactic features of these *corpora*, in order to verify which *corpus* is the simplest. Such *corpora*, of distinct features, were tagged by the UDPipe Grammar and explored according to some pre-established syntactic analysis criteria. The contrasts indicate that the *corpus* of study has simpler syntax, although some of the points explored deserve special attention. Thus, such comparisons will enrich the analysis and discussion about the accessibility of this material, as well as serving as foundation to the study as a whole.

Introdução

Os estudos de Acessibilidade Textual e Terminológica (ATT) tratam dos padrões que tornam um texto mais ou menos complexo, sempre de acordo com o público para o qual esse texto é direcionado e produzido. Nesse sentido, a pesquisa de Doutorado no qual esse estudo está inserido busca a união da experiência docente com a pesquisa acadêmica, por meio da investigação da leitura e da compreensão de materiais, em tese, adaptados para um público com doze anos ou mais, por jovens e adolescentes do Ensino Fundamental II de escola pública de Porto Alegre-RS.

O objetivo principal da pesquisa é descrever e analisar se um livro da área da saúde, disponível *on-line*, adaptado para um público leitor jovem, é compreendido por esse público e de que forma. Inicialmente, o foco será a publicação digital *Aprendendo sobre vírus e vacinas*, da Editora da UFCSPA. Essa editora lançou várias publicações, todas na área da saúde, adaptadas para diferentes públicos. A única dessas obras direcionada para público jovem, com doze anos ou mais, foi escolhida para análise porque essa é a faixa etária que corresponde aos alunos da pesquisadora responsável, com os

quais será possível dar continuidade à pesquisa, num próximo momento, por meio de testes de compreensão leitora.

O recorte que se apresenta neste resumo faz parte de estudo piloto sobre o *corpus* de estudo, em que serão realizados contrastes do *corpus* de estudo com outros *corpora*. O objetivo é mapear as características dos *corpora*, a fim de verificar se de fato o *corpus* de estudo está acessível para o público ao qual se destina. Os *corpora* escolhidos para contraste foram dois: (a) conjunto de textos de divulgação científica, da área da saúde, sobre cuidados paliativos; e (b) conjunto de textos de sentenças dos tribunais de pequenas causas, da área jurídica.

Ao lermos os três materiais, a hipótese principal é que o *corpus* de estudo seja o mais simples. A segunda hipótese é que, ainda que o *corpus* de estudo seja o mais simples entre os três, precisa de mais adaptação e simplificação para de fato se tornar acessível e compreensível para o público de alunos do Ensino Fundamental II.

A fim de verificar nossas hipóteses, todos os *corpora* foram submetidos ao mesmo procedimento, com as mesmas ferramentas computacionais: anotação pela UDPipe e buscas pelo Interrogatório. Ambas as ferramentas serão contextualizadas e explicadas na seção de Métodos.

Os resultados validaram a aposta inicial, de que o *corpus* de estudo seria o mais simples, de acordo com os critérios pesquisados. Porém, a segunda hipótese não foi confirmada, pois não houve critério em que o *corpus* de estudo fosse considerado o mais complexo, necessitando de atenção e simplificação.

Referencial Teórico

De maneira geral, a palavra “acessibilidade” corresponde a “ser acessível”. Então, acostumamo-nos a relacionar essa palavra com a inclusão, com os espaços, com os recursos. Mas, tal acessibilidade também pode ser analisada do ponto de vista textual, das palavras, das frases, dos parágrafos, dos textos, bem como da informação, da compreensão. É dessa forma que enxergamos a acessibilidade neste trabalho, alinhada ao que mencionam Finatto e Motta (2019, p. 317):

Hoje, porém, se entende que a noção de *acessibilidade* não é algo apenas associado a espaços físicos. Afinal, para as pessoas que buscam alguma inserção em diferentes contextos, sejam esses físicos, sejam abstratos, pode haver barreiras de vários tipos. É o caso, por exemplo, de muitos materiais escritos que, por várias razões, se tornam “ambientes” totalmente inacessíveis a um grande número de pessoas, portadoras de necessidades especiais ou não.

As autoras Fulgêncio e Liberato (1992) já discutiam, nos anos 1990, a importância da adequação da leitura ao público-alvo ao qual se destina. Segundo Perini (1988), que acreditava que só se aprende a ler, lendo, “[...] a leitura funcional nascerá do convívio com o material escrito adequado, e somente dele” (PERINI, 1988, p. 81). Dessa forma, o autor propõe que textos, quanto ao nível de dificuldade de leitura, devem ser graduados a diferentes públicos, “[...] de modo que um texto da terceira série fosse significativamente mais simples do que um de oitava série, ou de nível universitário” (PERINI, 1988, p. 82).

Baseados nessas referências, o interesse deste estudo está no texto do livro *Aprendendo sobre Vírus e Vacinas*, que, conforme os autores, foi adaptado para jovens com idade entre doze anos ou mais, alunos de Ensino Fundamental II. Mas, como a

publicação não foi testada de fato, na prática, não há como saber se o público a compreende. Em função disso, o estudo pretende explorar esse material, contrastando-os com outros *corpora* e realizando testes de compreensão leitora com esse público.

Assim, nesse momento, analisaremos a sintaxe do texto, em comparação com outros dois *corpora* também escritos para pessoas leigas. Como auxílio para tal análise, utilizamos o NILC-Metrix, um sistema computacional composto por 200 métricas propostas em estudos sobre discurso, psicolinguística, linguística cognitiva e computacional, para avaliar a complexidade textual do português brasileiro (PB). Esse sistema é a adaptação mais moderna da ferramenta Coh-Metrix para o inglês.

Desenvolvida por pesquisadores da Universidade de Memphis, a ferramenta Coh-Metrix é uma ferramenta de análise de textos que “[...] calcula índices que avaliam a coesão, a coerência e a dificuldade de compreensão de um texto (em inglês) [...]” (SCARTON, ALUÍSIO, 2010, p. 47). No Brasil, a ferramenta foi adaptada para o português brasileiro por pesquisadores do Núcleo Interinstitucional de Linguística Computacional da USP e da UFSCar (NILC-USP), ao longo dos últimos treze anos.

No índice do NILC-Metrix, é possível selecionar métricas específicas de complexidade sintática. Entre essas métricas estão três da utilizadas neste estudo: quantidade de conjunções subordinativas em relação ao total de conjunções, quantidade de orações subordinadas em relação ao total de orações, e quantidade de orações com voz passiva em relação ao total de orações.

Outro estudo que também mediu a complexidade sintática de textos literários para público neoleitor (adultos que estão sendo alfabetizados tardiamente) foi o de Rodrigues, Freitas e Quental (2013). As autoras investigaram a Coleção Literatura para Todos 1, de 2006. Em função do público pretendido, imaginava-se que a escrita dessas obras fosse mais acessível linguisticamente. Foi em função disso que as autoras investigaram a complexidade sintática dessas obras de maneira objetiva e quantitativa, por meio de ferramentas computacionais. As ferramentas utilizadas na época foram o analisador morfossintático PALAVRAS (BICK, 2000) e o programa Coh-Metrix Port (ALMEIDA e ALUÍSIO, 2009). As buscas utilizadas pelas autoras naquele momento, que foram reutilizadas em nossa pesquisa atual, foram: quantidade de verbos por sentença, quantidade de vírgulas por sentença, quantidade de elementos intercalados por sentença e quantidade de orações reduzidas de gerúndio em relação ao total de orações.

Neste momento, em que novas e diferentes ferramentas de Linguística Computacional estão a nossa disposição, utilizamos a UDPipe (STARAKA *et al.*, 2016) para etiquetar os *corpora* e o Interrogatório (de SOUZA e FREITAS, 2019) para realizar as buscas nesses *corpora* etiquetados.

Materiais

Os três *corpora* utilizados nesse estudo foram: (a) *corpus* de estudo: livro Vírus e Vacinas, (b) *corpus* de contraste I: conjunto de textos de divulgação científica sobre cuidados paliativos, e (c) *corpus* de contraste II: conjunto de sentenças dos tribunais de pequenas causas.

Um *corpus* de estudo é a base de uma pesquisa, é o material, o compilado de textos que pretendemos descrever e avaliar. Já o *corpus* de contraste é aquele compilado de

textos utilizado na comparação, a fim de aproximar ou distanciar os materiais quanto aos mais diferentes aspectos: semânticos, sintáticos, léxicos, estruturais, quantitativos, etc.

É importante ressaltar que esses materiais foram escolhidos porque se destinam à população em geral. Dessa forma, todos deveriam ser de leitura e entendimento acessíveis. Enquanto o *corpus* de estudo tem como público alunos do Ensino Fundamental II, o *corpus* de contraste I é formado por textos sobre cuidados paliativos para pessoas leigas. Pessoas leigas são aquelas que não têm conhecimento sobre o assunto, geralmente cuidadores e familiares de pessoas que necessitam de (ou fazem escolha por) cuidados paliativos, então, os textos precisam ser compreendidos por elas. Da mesma forma, o *corpus* de contraste II é formado por sentenças de tribunais de pequenas causas, ou seja, englobam processos de menor complexidade, que não ultrapassem 40 salários mínimos e qualquer pessoa pode utilizar sem a necessidade de um advogado, sendo assim, tais sentenças também precisam ser acessíveis à população.

O *corpus* de estudo é o livro *Aprendendo sobre Vírus e Vacinas*, da Editora da UFCSPA, publicado no ano de 2020. O livro tem 72 páginas e está disponível para *download* gratuito no site da Universidade. O projeto de extensão de mesmo nome, *Aprendendo sobre Vírus e Vacinas*, objetiva auxiliar o aprendizado dos alunos dos anos finais do Ensino Fundamental no aprendizado sobre vírus e vacinas utilizadas na sua prevenção (RODRIGUES, 2020, p. 65). Os autores são estudantes de Biomedicina na UFCSPA que, juntamente com o professor Luiz Carlos Rodrigues Jr., diante da pandemia do novo Coronavírus, criaram histórias, baseadas nas suas pesquisas e estudos, adaptando-as para diferentes públicos. O único direcionado para a idade dos alunos (12+) com os quais será possível dar continuidade à pesquisa num segundo momento, por meio de testes de compreensão leitora, é o livro *Aprendendo sobre Vírus e Vacinas*, que, conforme os autores:

“Apresenta um panorama geral das principais epidemias virais, incluindo a do SARS-CoV-2, aspectos da estrutura, da infecção e da replicação viral no hospedeiro, a resposta imunológica formada e o mecanismo imunogênico das vacinas. Cada parte é ilustrada com imagens simplificadas e lúdicas dos sistemas e dos mecanismos envolvidos.” (RODRIGUES, 2020, p. 65).

Os *corpora* de contraste foram escolhidos porque também lidam com um público específico: a população em geral.

O *corpus* de contraste I também é da área da saúde, foi organizado durante a Dissertação de Esteves (2023), e é composto por conjuntos de textos de divulgação científica sobre cuidados paliativos. A autora explica que, a partir de busca no Google, foram selecionados textos com as seguintes palavras-chave: a) “cuidados paliativos” e b) “cuidados paliativos” + “Ministério da Saúde”. Essa mesma pesquisa foi realizada em três datas diferentes: 1) 12 de maio de 2021, com a seleção de 10 textos; 2) 15 de junho de 2021, com a seleção de 5 textos; e 3) 19 de setembro de 2021, com a seleção de 15 textos. Ao selecionar os textos, foram usados alguns critérios de inclusão: a) apenas em português brasileiro; b) de acesso livre e gratuito; c) de fontes idôneas, como instituições do Governo, portais de notícias reconhecidos e entidades bem estabelecidas; d) preferência a textos escritos originalmente em português; e) preferência a textos jornalísticos ou de divulgação científica. Em função da pandemia, foram compilados outros 15 textos, com resultados para buscas Google com as palavras-chave: “cuidados paliativos” + “CPI da pandemia”. Essa última seleção ocorreu em dois momentos: 1) 12

de novembro de 2021 e 2) 4 de fevereiro de 2022. Foram utilizados os mesmos critérios de inclusão e exclusão das compilações anteriores, totalizando um corpus de 45 textos sobre cuidados paliativos. (ESTEVES, 2023)

O *corpus* de contraste II é da área jurídica, foi organizado na Tese de Motta (2022), e compreende um total de 440 Sentenças Judiciais dos Juizados Especiais Cíveis (JECs) do Poder Judiciário do Rio Grande do Sul (PJRS). Essas Sentenças representam as decisões tomadas pelos juízes em relação aos pedidos das pessoas que procuram os JECs para resolver algum conflito. O Juizado Especial Cível tem como atribuição a conciliação, o processo e o julgamento de ações cíveis de menor complexidade, cujo valor de causa não ultrapasse 40 salários mínimos. As Sentenças foram coletadas do *site* do Tribunal de Justiça do Rio Grande do Sul, entre os anos de 2018 e 2019. As temáticas das Sentenças são: água, cadastro de inadimplentes, energia elétrica, plano de saúde, prestação de serviços e telefonia (de forma que o *corpus* ficasse variado). O *corpus* organizado compreende um total de 440 sentenças. (MOTTA, 2022)

Métodos

Primeiramente, os três *corpora* foram formatados em arquivo de texto, seguindo o padrão UTF8, necessário para anotação em UDPipe. A ferramenta UDPipe, de código aberto, “[...] realiza sequencialmente as etapas de tokenização (segmentação do texto em unidades básicas, como palavras e sinais de pontuação), anotação gramatical, lematização e análise de dependências em qualquer *corpus* que esteja no formato CoNLL-U.” (de SOUZA e FREITAS, 2021, p.1043). Utilizamos esse modelo de anotação por que:

O modelo fornecido para o português (versão 2.5) tem índices de acerto (F1) de 96.4%, 95%, 87.2% e 83.1% para os níveis de classes gramaticais (POS), características morfológicas (feats), dependência sintática (unlabeled attachment score (UAS) e relação de dependência sintática (labeled attachment score (LAS)), respectivamente. .” (de SOUZA e FREITAS, 2021, p.1043)

Esses *corpora* anotados foram processados no Interrogatório, uma ferramenta para busca e revisão de *corpora* anotados (de SOUZA e FREITAS, 2019). Essa ferramenta realiza buscas sintáticas em arquivos anotados em formato CoNLL-U.

É importante mencionar que o *corpus* de sentenças judiciais, por ser muito extenso, teve de ser dividido em duas partes, pois as ferramentas não conseguiram realizar seu processamento de uma única vez. Então, as buscas e resultados para esse *corpus* em específico foram dobradas. Todos os resultados foram somados para obtermos um resultado final do todo do *corpus*.

A partir disso e de estudos anteriores já mencionados nas referências, foi verificada a quantidade de ocorrências dos seguintes indícios de complexidade textual:

- Quantidade de verbos por período;
- Quantidade de vírgulas por período;
- Quantidade de elementos explicativos intercalados por período;
- Orações subordinadas em relação às orações;
- Voz passiva em relação às orações;
- Orações reduzidas de gerúndio em relação às orações;
- Conjunções subordinativas em relação a todas as conjunções.

Essas buscas específicas aconteceram por terem sido delimitadas como indícios de complexidade textual, conforme os estudos de Rodrigues, Freitas e Quental (2013) e conforme apontamentos do NILC-Metrix (LEAL *et al*, 2022).

Algo que notamos desde o início é que as buscas a partir dessa anotação nem sempre são óbvias e/ou diretas. Por exemplo, para analisarmos o primeiro item, a quantidade de verbos por período, não somente buscamos os verbos, pois há verbos anotados de outra maneira, como auxiliares, por exemplo, ainda que sejam verbos principais, dependendo da frase. Alguns exemplos estão na Tabela 1 a seguir:

Tabela 1: Exemplos de verbos auxiliares como principais

<i>Corpus</i>	Exemplo
Vírus e Vacinas	Os neutrófilos são os primeiros a chegar.
Cuidados Paliativos	O que leva um médico a atuar com quem está na reta final da vida?
Sentenças Jurídicas	A parte Demandada é prestadora de serviço de energia elétrica, logo fornecedora .

Então, para encontramos a quantidade total de verbos, foram necessárias duas buscas separadas, nas quais a quantidade de verbos encontrada em cada *corpus* e em cada busca foi somada. Primeiro, buscamos os verbos. Em seguida, os verbos auxiliares com função de verbo principal, que buscou verbos como os da Tabela 1.

Para encontrarmos as vírgulas, buscamos pela pontuação do texto, todas que eram iguais à vírgula. Mas, para buscarmos os elementos intercalados também não foi possível uma única busca, pois não há como a ferramenta extrair todos de uma única vez. Assim, para encontrarmos esses elementos intercalados, foram realizadas cinco buscas diferentes para cada *corpus*: aposto intercalado, adjunto adverbial anteposto ao verbo, oração adverbial anteposta ao verbo, oração adjetiva explicativa e oração adjetiva reduzida. Na Tabela 2 estão alguns exemplos de elementos intercalados:

Tabela 2: Exemplos de elementos intercalados

<i>Corpus</i>	Exemplo
Vírus e Vacinas	PARA ESCREVER ESTE LIVRO , NÓS TAMBÉM ESTUDAMOS MUITO! (oração adverbial)
Cuidados Paliativos	E, na Bélgica , ocorre, atualmente, a mesma situação jurídica. (adjunto adverbial)
Sentenças Jurídicas	A parte autora alega perda de carne bovina , entre outros alimentos , que se encontravam congelados para consumo próprio, em dois frízers, no montante de R\$1.690,00 (um mil seiscentos e noventa reais). (oração adjetiva explicativa)

Ao final das buscas, todos os números encontrados em cada *corpus* foram somados e relacionados com a quantidade de períodos. Dessa forma, encontramos a média de elementos intercalados por período e por *corpus*.

Para buscarmos as orações subordinadas, foi preciso realizar várias buscas diferentes para cada *corpus*. Orações subordinadas substantivas subjetivas e objetivas diretas, além de casos de coordenação entre elas; orações subordinadas adjetivas e adverbiais, além de casos de coordenação entre elas. Todos os números foram somados em cada *corpus*.

Para encontrarmos a voz passiva também foi mais tranquilo, pois foi possível utilizar apenas um comando por *corpus*. Entretanto, para encontrarmos as orações reduzidas de gerúndio, também foram necessárias algumas buscas para cada *corpus*: orações subordinadas adjetivas reduzidas de gerúndio, orações subordinadas adverbiais reduzidas de gerúndio, orações subordinadas substantivas que podem ser reduzidas de gerúndio.

E, por fim, para encontrarmos as conjunções subordinativas, foi possível somente uma busca, porém, para relacionarmos com a quantidade total de conjunções, foram necessárias duas buscas (conjunções subordinativas + conjunções coordenativas).

Resultados

Primeiramente, os dados gerais dos *corpora* são apresentados na Tabela 3 a seguir.

Tabela 3: Informações sobre os *corpora*

	<i>Corpus</i> de estudo	<i>Corpus</i> de Contraste (1) - SAÚDE	<i>Corpus</i> de Contraste (2) - JUSTIÇA
<i>Tokens</i>	4.465	33.0731	702.511
Sentenças	342	13.883	30.604
Quantidade de documentos	1 publicação digital	45 textos do Google	440 Sentenças Judiciais

É importante ressaltar que, embora os *corpora* tenham extensões diferentes, o que é visível pelos dados da Tabela 3, isso não compromete os resultados e as comparações, uma vez que todas as análises foram relacionadas de acordo com os dados de cada *corpus*. Por exemplo, a quantidade de verbos, de vírgulas e de elementos intercalados do *corpus* de estudo foi relacionado com a quantidade de sentenças desse *corpus*, 342. Já a quantidade de verbos, de vírgulas e de elementos intercalados do *corpus* de contraste (2) foi relacionada com a quantidade de sentenças do *corpus* da justiça: 30.604. Dessa forma, os resultados foram sempre relativos ao próprio *corpus*, sua extensão, quantidade de sentenças e características, o que permitiu as comparações a seguir.

A Tabela 4 a seguir mostra os resultados finais, de cada soma de critérios, obtidos com todas as buscas, separadas por *corpus* e por critério pesquisado:

Tabela 4: Resultados Gerais

	<i>Corpus</i> de estudo	<i>Corpus</i> de Contraste (1) - SAÚDE	<i>Corpus</i> de Contraste (2) - JUSTIÇA
Verbos/Período	1,38	2,47	2,24
Vírgulas/Período	0,37	1,3	1,5
Elementos intercalados/Período	0,23	0,54	0,66
Subordinadas/Orações	0,39	0,49	0,49
Voz Passiva/Orações	0,04	0,06	0,05
Reduzidas de Gerúndio/Orações	0,01	0,03	0,05
Conjunções Subordinativas/Conjunções	0,39	0,03	0,58

Os números grifados em negrito apontam para maior resultado em cinco critérios para o *corpus* de sentenças jurídicas. Em relação às orações subordinadas, houve empate de resultado entre os dois *corpora* de contraste. Ainda assim, o *corpus* de cuidados paliativos é o segundo mais complexo, com resultados maiores em três critérios.

Conforme a nossa hipótese inicial, o *corpus* de estudo, sobre vírus e vacinas, é o menos complexo em relação aos outros dois, pois não apresenta resultados maiores em nenhum dos critérios pesquisados. Apesar disso, nestes critérios pesquisados, a segunda hipótese, de que o *corpus* de estudo ainda poderia ser adaptado e simplificado para o público ao qual se destina, não se confirmou, visto que não houve resultado maior em nenhum quesito, então, não há como afirmar a segunda hipótese.

Considerações

O objetivo principal do estudo foi mapear as características sintáticas dos *corpora* e verificar qual deles seria o mais simples, sintaticamente. O objetivo foi realizado e o resultado principal comprovou a hipótese de que o *corpus* de estudo é o mais simples, sintaticamente, entre os *corpora* analisados e de acordo com os critérios estabelecidos, baseados em estudos anteriores.

Porém, algumas dificuldades surgiram ao longo do caminho. A primeira delas foi a familiarização com os termos e a anotação realizada. Não é tão óbvio encontrar as estruturas da gramática tradicional, pois a anotação na gramática UDPipe nem sempre segue a mesma lógica, então, a busca precisa ser ajustada (e compreender a anotação exige abstração, foco e experimentação).

No passo seguinte, as buscas no Interrogatório, foi preciso também experimentação nas buscas, pois nem sempre o que pensamos de acordo com a gramática tradicional já internalizada em nosso raciocínio, é o que a busca realizada nos traz como resultado. Às vezes, os caminhos precisam ser refinados, acrescentando mais de uma condição, e/ou filtros. Então, em vários dos critérios foi preciso mais de uma busca, sobretudo nos elementos intercalados, critério mais complicado, em nosso entendimento e experiência.

Além disso, o *corpus* de sentenças precisou ser dividido em duas partes, pois a ferramenta não funcionou com ele por completo. Isso tornou as buscas nesse *corpus* especificamente mais trabalhosas, pois foram o dobro de pesquisas e de somas.

Ainda assim, acreditamos que o estudo contribuiu muito para a familiarização com os termos, os critérios e os funcionamentos desses sistemas computacionais, os quais não conhecíamos, nem nunca havíamos experimentado. Além disso, é somente na prática que aprendemos a lidar com algo que não é do nosso uso frequente e familiar. Adaptar nosso pensamento e lógica tão acostumados com a gramática tradicional também foi um exercício importante.

Por fim, de forma geral, acreditamos que o estudo foi útil não só para o aprendizado relacionado às ferramentas e à análise linguística, como também para as discussões sobre complexidade textual de diversos *corpora* que podem ser explorados dessa mesma maneira. Além disso, acreditamos que tais contrastes também vão enriquecer a análise e a discussão sobre a acessibilidade do *corpus* de estudo para esse público, servindo de base para as etapas subsequentes da pesquisa e da Tese.

Referências

- de SOUZA, E.; FREITAS, C. ET: uma Estação de Trabalho para revisão, edição e avaliação de corpora anotados morfosintaticamente. In: **WORKSHOP DE INICIAÇÃO CIENTÍFICA EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA** (TILic 2019), VI., 2019. Salvador. Proceedings [...]. Salvador: SBC, 2019. p. 15-18.
- de SOUZA, E.; FREITAS, C. Sujeito oculto às claras: uma abordagem descritivo-computacional. In: **Revista Estudos da Linguagem**. Belo Horizonte, v. 29, n. 2, p. 1033-1058, 2021.
- ESTEVES, F. F. **DEFINIÇÕES ACESSÍVEIS: POR UMA LINGUAGEM SIMPLES EM CUIDADOS PALIATIVOS**. 157f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2023.
- FINATTO, M. J. B. ; MOTTA, E. Terminologia e Acessibilidade: novas demandas e frentes de pesquisa. **REVISTA GTLEX**, v. 2, p. 316-356, 2019.
- FULGÊNCIO, L.; LIBERATO, Y. **Como Facilitar a Leitura**. São Paulo: Contexto, 1992.
- LEAL, Sidney Evaldo, DURAN, Magali Sanches, SCARTON, Carolina Evaristo, HARTMANN, Nathan Siegle, ALUÍSIO, Sandra Maria. **NILC-Metrix: avaliando a complexidade da linguagem escrita e falada no português brasileiro**. CoRR abs/2201.03445 (2022). Disponível em: <https://arxiv.org/abs/2201.03445> Acesso em: 11 jun. 2023.
- MOTTA, E. **SENTENÇAS JUDICIAIS E LINGUAGEM SIMPLES: um encontro possível e necessário**. 411f. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2022.
- PERINI, M. A. A Leitura Funcional e a Dupla Função do Texto Didático. In: ZILBERMAN, Regina; SILVA, Ezequiel Theodoro (org.). **Leitura: Perspectivas Interdisciplinares**. São Paulo: Ática, 1988.
- RODRIGUES, E. C.; FREITAS, C.; QUENTAL, V. Análise de inteligibilidade textual por meio de ferramentas de processamento automático do português: avaliação da Coleção Literatura para Todos. In: **Letras de Hoje**, Porto Alegre, v.48, n. 1, p.91-99, jan./mar. 2013.
- RODRIGUES, L. C. J. **Aprendendo sobre vírus e vacinas** [recurso eletrônico]. Porto Alegre: Editora da UFCSPA, 2020. ISBN 978-65-87950-28-0. Disponível em: <https://www.ufcspa.edu.br/vida-no-campus/editora-da-ufcspa/obras-publicadas>. Acesso em: 30 set. 2021.
- SCARTON, C.; ALUISIO, S. M. Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o português. **LinguaMática 2** (2010): p. 45-62.
- STRAKA, M.; HAJIC, J.; STRAKOVÁ, J. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In: **TENTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION** (LREC'16), 10th., Portorož, Proceedings [...]. Portorož: ELRA, 2016. p. 4290-4297.