

Acessibilidade textual e terminológica para o português brasileiro: pesquisa, estratégias e orientações de [re]escrita simplificada

Accessible scientific writing in Brazilian Portuguese language: research, strategies and guidelines for simplified [re]writing

Asafe Davi CORTINA SILVA*

Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

Heloísa Orsi Koch DELGADO**

Mais H Consultoria Linguística Internacional

Maria José Bocorny FINATTO***

Universidade Federal do Rio Grande do Sul (UFRGS)

RESUMO: Este artigo retoma e discute a pesquisa de mestrado intitulada “Textos de divulgação para leigos sobre o Transtorno do Estresse Pós-Traumático em português: alternativas para a Acessibilidade Textual e Terminológica”. De natureza quali-quantitativa, a pesquisa associou preceitos para a promover a Acessibilidade Textual e Terminológica, a Linguística Textual, o Processamento da Linguagem Natural e a Linguística de *Corpus*. Esses campos teórico-aplicados, conforme dispostos pelo trabalho, permitiram mensurar a complexidade da linguagem de textos informativos sobre o Transtorno do Estresse Pós-Traumático voltados para leitores leigos com escolaridade limitada. Como resultados, o estudo demonstrou o potencial de eficiência de diferentes estratégias para simplificar textos complexos e uma série de instruções para a redação simplificada dos textos analisados. Esses resultados são aqui ponderados quanto à sua abrangência e aplicabilidade em diferentes situações da divulgação científica para leigos.

PALAVRAS-CHAVE: Acessibilidade textual. Processamento da Linguagem Natural. Redação Simplificada. Linguagem simples. *Coh-Matrix Dementia*.

ABSTRACT: The article partially describes and analyses the research of the thesis entitled “*Textos de divulgação para leigos sobre o Transtorno do Estresse Pós-Traumático em português: alternativas para a acessibilidade textual e terminológica*”. This research integrated both qualitative and quantitative methods, and established the relationship between the fields of Accessible scientific writing, Text Linguistics, Multi-dimensional Approach, and the Natural Language Processing combined with *Corpus* linguistics. The study analyzed the complexity of the language used in informative texts about the Post Traumatic Stress Disorder, written in Brazilian Portuguese for lay people with low education level, and indicated – through software metrics and multi-dimensional analysis - that they are potentially complex for the intended

* Doutorando em Linguística (Pontifícia Universidade Católica do Rio Grande do Sul, PPG-Letras, Porto Alegre, RS). E-mail: asafedcs@gmail.com

** Doutora em Letras (Universidade Federal do Rio Grande do Sul, PPG-Letras, Porto Alegre, RS). E-mail: heloisa@maish.com.br

*** Doutora em Letras (Universidade Federal do Rio Grande do Sul, PPG-Letras, Porto Alegre, RS), pesquisadora do CNPq. E-mail: mariafinatto@gmail.com

readers. The analysis also proved the efficiency of a range of techniques to make scientific texts more accessible for lay people, which were used as the basis for the recommended simplification guidelines presented in the text.

KEY WORDS: Accessible scientific writing. Natural Language Processing. Simplified writing. Plain Language. Coh-Metrix Dementia.

Introdução

Ainda que atualmente o acesso à informação sobre temas de ciências seja facilitado por meio da internet, as informações encontradas tendem a ser, via de regra, textual e terminologicamente pouco acessíveis (SAGGION, 2017). Isto é, embora possamos encontrar uma grande quantidade de textos *on-line* sobre inúmeros assuntos, veiculados em linguagem geral ou especializada, sua compreensão pode ser difícil para uma grande parte da população, principalmente para pessoas com escolaridade limitada, com experiências de leitura limitadas ou com pouco letramento científico. Portanto, pode-se dizer que o atual "acesso ilimitado" e imediato a informações não representa, necessariamente, a ampliação de conhecimento adequado e pertinente, principalmente se considerarmos fontes que fornecem informações de divulgação científica de qualidade duvidosa, sem autoria confiável.

Considerando, dessa forma, que o acesso ao conhecimento democraticamente disponível na internet não perpassa de forma igualitária todas as camadas da sociedade brasileira, retomamos aqui uma pesquisa que se ocupou, como tema central, da acessibilidade de textos (em língua portuguesa) que tratam sobre um transtorno estudado pela Psicologia, tendo em mente um leitor leigo adulto com escolaridade limitada. Essa pesquisa intentou, com base em um conjunto de textos produzidos por profissionais de Saúde sobre o tema do Transtorno do Estresse Pós-Traumático (TEPT¹) e oferecidos na internet para o público leigo, descrever os padrões de sua organização textual (CORTINA SILVA, 2018). A partir dessa descrição, buscou orientar as suas [re]escritas, apresentando e testando estratégias de simplificação textual e terminológica da linguagem veiculada.

Assim, na sequência deste artigo, revisitamos as contribuições da pesquisa de Cortina Silva (2018) intitulada "Textos de divulgação para leigos sobre o Transtorno do

¹ Esse assunto foi escolhido devido ao aumento da incidência desse transtorno psicológico entre os brasileiros estar em um crescimento exponencial nos últimos anos (Medeiros, 2017).

Estresse Pós-Traumático em português: alternativas para a Acessibilidade Textual e Terminológica”. A pesquisa (de natureza quali-quantitativa) associou preceitos para promover a Acessibilidade Textual e Terminológica, aproveitando experiências da Linguística Textual, do Processamento da Linguagem Natural e da Linguística de *Corpus*. Esses campos teórico-aplicados, conforme dispostos pelo trabalho, permitiram mensurar a provável complexidade da linguagem de textos informativos voltados para leitores leigos. Como resultados, o estudo demonstrou o potencial de eficiência de diferentes estratégias para simplificar textos complexos e uma série de instruções para a redação simplificada dos textos analisados. Esses resultados são aqui ponderados quanto à sua abrangência e aplicabilidade em diferentes situações da divulgação científica para leigos.

Isso posto, sintetizamos, a seguir, os principais objetivos e questões norteadoras do trabalho em foco e como a pesquisa foi teoricamente sustentada, apresentando o tema da Acessibilidade Textual e Terminológica (ATT) em textos sobre o TEPT. Seus resultados e indicações de alternativas para a produção de textos simplificados serão discutidos ao longo deste artigo.

1 Objetivos e questões da pesquisa

Os principais objetivos da pesquisa de Cortina Silva (2018) foram: a) analisar métricas textuais que indicassem altos potenciais de complexidade de textos informativos sobre o TEPT escritos para os públicos leigos; b) verificar a eficiência de diferentes estratégias de simplificação na tentativa de tornar esses textos mais acessíveis para o perfil de leitor estipulado. A partir dos resultados observados, o autor descreveu uma série de orientações que devem ser consideradas para a redação de textos simplificados.

Os objetivos gerais do estudo foram os seguintes:

- I. Verificar a possível complexidade de textos em língua portuguesa brasileira sobre o Transtorno do Estresse Pós-Traumático para públicos leigos de escolaridade limitada; II. Observar o impacto de diferentes estratégias de simplificação nas métricas indicativas de complexidade dos textos estudados a partir do sistema Coh-Metrix Dementia; III. Ponderar a possível adequação de textos simplificados para o público determinado; IV. Indicar quais estratégias de reescrita testadas poderiam apresentar maior

potencial de eficiência para acessibilidade textual em termos de medidas/métricas. (CORTINA SILVA, 2018, p. 32)

Já as questões norteadoras gerais e específicas são resumidas a seguir: a) o que torna uma formulação textual mais potencialmente acessível para um leitor receptor como o nosso? b) quais estratégias de reescrita poderiam ser mais eficientes, considerando as indicações de uma ferramenta computacional que aponta métricas de provável complexidade em diferentes níveis da tessitura do texto? c) os textos sobre TEPT da amostra reunida, conforme se apresentam originalmente, tendem a favorecer a ATT ou contribuem para uma potencial complexidade?; d) quais particularidades podem perfazer o perfil de uma linguagem acessível em português brasileiro, no cenário dos textos e da comunicação sob exame?

As próximas seções apresentam um resumo dos pressupostos teóricos que sustentaram a pesquisa, seguido pela sua configuração metodológica e pelos resultados levantados a respeito do potencial de complexidade dos textos sobre o TEPT (o *corpus* ou amostra de estudo) e de suas versões simplificadas. Por fim, retomamos um sumário das orientações sugeridas por Cortina Silva (2018) para a redação adaptada de textos.

2 Orientações teóricas adotadas: síntese

É importante destacar que o trabalho de Cortina Silva (2018) se tratou de uma pesquisa transdisciplinar, situando-se nas áreas da Terminologia, da Acessibilidade Textual e Terminológica (ATT), da Linguística Textual e Sistêmico-Funcional (LT/LST), da Análise Multidimensional (AMD), e do Processamento da Linguagem Natural (PLN) aliado à Linguística de *Corpus* (LC).

Áreas relacionadas à Psicologia e à Linguística internacionais há muito vêm buscando teorias, fórmulas e modelos que possam nos ajudar a entender e prever quais constituintes ou características de formulação linguística tornariam um texto potencialmente mais complexo. Essas pesquisas pretendem embasar reescritas que tornem o acesso mais claro por leitores de diferentes níveis (cf. PASQUALINI, SCARTON & FINATTO, 2010). Vale reiterar, conforme apontou Finatto (2011), que os estudos a respeito da complexidade em nível textual global são recentes no Brasil. Há estudos mais antigos sobre a microestrutura dos componentes textuais (palavras, frases ou expressões

sintagmáticas), mas não sobre o objeto texto como um todo. Por outro lado, ao compararmos com a literatura estrangeira, podemos observar que existem registros de estudos sobre legibilidade e complexidade desde 1920, em trabalhos que lidavam com a compreensão desde o nível micro até o nível macro do texto. Finatto, a propósito, ainda observa que esses estudos eram voltados, principalmente, para a compreensão de leitura de textos literários e que referências sobre textos científicos e técnicos são escassas.

Ciapuscio (2003) já afirmava que redatores (ou editores) de textos simplificados não têm muitos recursos que possam colaborar para a compreensão linguística dos perfis dos leitores aos quais os textos serão direcionados, de forma a poder identificar quais elementos textuais tornam a leitura e a compreensão mais complexa. Por essa razão, o desenvolvimento de estudos a respeito de complexidade textual (e, por conseguinte, de ferramentas que possam trabalhar com a área) parece ser importante para o entendimento a respeito dos graus de legibilidade de textos potencialmente complexos.

Cortina Silva (2018) também utilizou, em sua pesquisa, o PLN, uma área da Ciência da Computação que desenvolve e utiliza recursos informatizados para o tratamento de fenômenos das línguas naturais. No PLN, tratam-se e modelam-se ocorrências das línguas naturais de modo que isso dê suporte a diferentes tarefas computacionais que envolvem informação linguística (MANNING & SCHÜTZE, 1999), como a tradução automática, a sumarização automática de textos, a indexação linguística de motores de busca ou a recuperação de informações.

Várias ferramentas computacionais de PLN possibilitam análise de dados partindo-se de textos individuais, como também são capazes de lidar com centenas de milhares de textos ao mesmo tempo. Programas como o *Coh-Matrix Dementia*² analisam um texto por vez e dispõem informações (métricas textuais) a respeito desse texto específico. Conforme apontam Finatto et al. (2018), para determinados fins de pesquisas linguísticas, é necessário e fundamental o tratamento individual de textos, ainda que seja importante destacar que, ao dizer isso, não se reduz ou elimina o tratamento de textos “agrupados” em grandes conjuntos, como em pesquisas que se baseiam na LC. Ao contrário, a validade e os trabalhos da LC, lidando com dados em larga escala, são essenciais para a Linguística e permitem o desenvolvimento de diversas pesquisas. Por

² Disponível em <http://143.107.183.175:22380/> - nova versão (2021) disponível em : <http://fw.nilc.icmc.usp.br:23380/nilcmatrix>

outro lado, o *modus operandi* de ferramentas do PLN “aplicado” a unidades textuais que se contrastam entre si permite dar suporte, por exemplo, a estudos que precisem descrever e ponderar elementos textuais e discursivos particulares.

Embora o desenvolvimento da pesquisa de Cortina Silva (2018) tenha lidado, primeiramente, com os textos de forma individualizada (com o apoio de uma ferramenta de PLN), os princípios probabilísticos e estatísticos da LC contribuíram para a averiguação dos seus padrões linguísticos. Esses padrões permitiram observar colocações mais corriqueiras em textos com linguagens mais "rebuscadas" e colocações mais comuns em textos de linguagem mais "simples". Essa averiguação, feita por meio de *corpus*, auxiliou no entendimento a respeito de características de uma linguagem simplificada.

Esse entendimento conjugando LC e PLN permitiu o desenvolvimento de subsídios para um guia de simplificação textual para a língua portuguesa, desenvolvido a partir dos achados do trabalho de Cortina Silva (2018) - e de outros - a respeito do assunto. Em suma, o autor ressaltou que os trabalhos de descrição e de análise de complexidade textual podem ser sustentados pela soma das perspectivas das duas áreas.

Ferramentas de PLN podem ser usadas para “levantar” uma quantidade de dados a respeito de textos em particular, mas, naturalmente, não os interpretam. Assim, Cortina Silva (2018) concluiu que uma interpretação humana era essencial para a compreensão das informações apresentadas pelo *Coh-Matrix Dementia* e para a validação da sua hipótese de que os textos estudados eram complexos para o público da pesquisa. Para isso, conduziu uma interpretação dos dados assumindo a sua posição de linguista e de leitor proficiente.

A sua análise holística de complexidade foi sustentada por princípios canônicos da LT, principalmente aqueles associados à percepção da coerência e da coesão textuais, conforme apresentadas por Koch (2015). Os fatores de textualidade apresentados por essa autora foram utilizados, como referência, com a finalidade de averiguar a possível complexidade de cada um dos textos do conjunto de textos sob exame. Assim, os fatores analisados foram: a informatividade (quantidade de informações novas apresentadas no texto, apoiadas ou não por informações de “senso comum”); a intencionalidade (propósitos do emissor e maneiras como ele utiliza os recursos textuais para cumprir seus objetivos); a aceitabilidade (“esforço” feito pelo

leitor para adequar a leitura à sua realidade) e a situacionalidade (elementos de um texto para que ele possa ser corretamente situado a determinados contextos).

Na próxima seção, apresentaremos a configuração metodológica da pesquisa.

3 Corpus, materiais e métodos

O *corpus* da pesquisa (doravante *corpus* de estudo) de Cortina Silva (2018) foi composto pelos dez primeiros textos³ dispostos pela busca no Google sobre o TEPT. Em termos de estudos de LC, não se poderia considerar que se trate, *stricto sensu*, de um *corpus*, mas, sim, de uma pequena amostra textual que será profundamente explorada. Os textos selecionados para estudo foram escritos em português-brasileiro e encontrados em páginas que tivessem por objetivo central informar leitores leigos. Nenhum dos textos selecionados foi escrito para a comunidade científica específica da Psiquiatria ou áreas afins.

Esse *corpus* amostral é comparado com seis textos, tomados como *corpus* de referência (ou um *corpus* de contraste), considerados simples: cinco narrativas para crianças retiradas de páginas da internet e um texto previamente simplificado.

Após a coleta e a preparação dos textos componentes do *corpus* de estudo, cada um, individualmente, foi submetido à ferramenta *Coh-Matrix Dementia* (CUNHA, 2015) para que os índices previamente selecionados pudessem ser coletados e analisados pelo autor. Vale lembrar que o *Coh-Matrix Dementia* é uma ferramenta gratuita de PLN que auxilia na análise automatizada de métricas textuais (setenta e três ao total) que, de acordo com as orientações dadas por Cunha (2015), na aba “sobre” da ferramenta, “capturam complexidade sintática, densidade de ideias, coerência textual e níveis de disfluência.”

No que diz respeito à AMD, utilizada na pesquisa, lembramos que é um tipo de estudo muito reconhecido no âmbito da pesquisa mais tradicional da LC. Embora a LC não tenha sido estritamente operacionalizada em seu estudo, Cortina Silva (2018) julgou válido referir à AMD para apontar, ao leitor interessado, outras metodologias e

³ Cortina Silva (2018) optou por utilizar os dez primeiros resultados dispostos pelo Google para que os textos encontrados refletissem, com verossimilhança, os primeiros resultados que usuários que procuram informações sobre o TEPT encontrariam.

alternativas por meio da verificação de traços de potencial complexidade em textos para além do uso de medidas do sistema *Coh-Matrix Dementia*.

Tal como o nome sugere, a AMD é o tipo de prática que leva em consideração mais de uma dimensão, parâmetro ou métrica ao analisar um determinado objeto de estudo. No âmbito da análise textual, a AMD foi proposta por Biber (1988/1995) com o objetivo de delimitar uma combinação dos níveis macro e micro para a análise de um *corpus*. Segundo o estudioso, nenhum traço isolado é suficiente para que uma descrição adequada seja encontrada e que um estudo baseado somente em características situacionais ou linguísticas não é igualmente satisfatório.

Berber Sardinha (2000, p.101) acrescenta que “a AMD foi criada com o objetivo de permitir uma descrição rica e complexa de *corpora* inteiros de textos por meio estatísticos bem como a extração precisa de características textuais em comum entre *corpora*”. Dessa forma, o mais apropriado, na visão do autor, “seria combinar-se as duas perspectivas, isto é, a descrição firmada em características situacionais com a descrição baseada em traços linguísticos” (BERBER SARDINHA, 2000, p.100). Esse tipo de análise permitiu um detalhamento rico e complexo de *corpora* por intermédio de dados estatísticos e da extração acurada de traços textuais comuns entre textos, reconhecendo que a análise deve apoiar-se em uma quantidade grande de parâmetros para possibilitar uma comparação mais compreensiva.

Feita essa breve caracterização da AMD, vale salientar que, ao trabalhar com a análise da complexidade em uma amostra de textos, o autor abordou, do seu próprio jeito, distintas dimensões de um texto. Tratou não apenas do léxico, mas também da sintaxe, da terminologia, e assim por diante. Uma vez que a sua pesquisa levou em conta diferentes dimensões textuais para a análise da CT e das estratégias de acessibilidade, preceitos da AMD serviram como inspiração para o desenvolvimento do trabalho de Cortina Silva (2018). Entretanto, vale frisar, o autor não empreendeu uma pesquisa de LC.

Ao fim da interpretação dos índices coletados e da verificação do alto potencial de complexidade dos textos analisados, foram aplicadas oito estratégias de simplificação, em cada um dos textos, provenientes da literatura sobre o assunto. Essas estratégias foram aplicadas de forma isolada (ou seja, uma estratégia por vez) para que seus potenciais de simplificação pudessem ser observados sem interferência entre elas.

Embora existam diferentes estratégias de simplificação, Cortina Silva (2018) optou pelas que lidassem diretamente com o léxico, com aspectos semânticos e com a estruturação dos textos. As estratégias adotadas serão brevemente explicadas na seção “Estratégias aplicadas”.

Após a aplicação dessas estratégias, foi gerado um novo *corpus* de aproximadamente 80 textos que são transformações dos textos originais coletados. Cada um desses textos (que são versões adaptadas pelo autor dos textos do *corpus* de estudo) foi submetido novamente ao *Coh-Matrix Dementia* para que os seus índices de complexidade pudessem ser comparados aos índices originais e os impactos de cada estratégia pudessem ser observados.

Desse modo, diante dos resultados das diferentes estratégias de simplificação que empregou, Cortina Silva (2018) identificou quais delas contribuem para uma maior acessibilidade e quais não possibilitam um potencial de simplificação tão significativo (embora o autor tenha concluído que todas as estratégias aplicadas tenham validade em determinados momentos). A partir da leitura desses dados, uma série de orientações de redação e simplificação foram indicadas, as quais serão explicadas mais adiante. Nas subseções abaixo, apresentamos a análise dos índices e seus resultados.

3.1 Índices da ferramenta *Coh-Matrix* aproveitados no trabalho

Embora a ferramenta *Coh-Matrix Dementia* ofereça setenta e três métricas textuais, Cortina Silva (2018) optou pelas que trabalhassem com o léxico, com a semântica e com a estrutura dos textos de uma forma mais abrangente. Os índices analisados foram:

1. Índice Flesch (IF): estima, por meio da extensão textual, o possível grau de complexidade e a quantidade de anos de educação formal necessários para a compreensão de um texto.

2. Análise Semântica Latente (ASL): calcula a similaridade semântica entre as frases do texto. Quanto maior semelhança semântica houver entre as frases de um texto, mais fácil será compreendê-lo.

3. Relação *Type-Token* (TT): utiliza as palavras de conteúdo (substantivos, verbos, entre outras) para calcular a variação lexical. Cortina Silva et al. (2018), ao

relacionar a variação lexical com o grau de coesão de um texto, explicam que quanto maior a variação do léxico, maior será a quantidade de informações que o leitor precisa compreender e, conseqüentemente, maior será o potencial de complexidade de um texto.

4. Densidade Semântica (DS): divide a quantidade de palavras de conteúdo pela quantidade de palavras funcionais (pronomes, artigos, etc.) e calcula o quão semanticamente complexo um texto aparenta ser. Ao que tudo indica, quanto maior for o valor desse índice, mais potencialmente complexo o texto se torna (CORTINA SILVA, 2018).

5. Incidência de palavras: mostra a quantidade de substantivos, verbos, adjetivos, advérbios e pronomes presentes em um texto. Esses índices foram indicados para verificar se a maior ou menor presença de determinadas classes de palavras pode contribuir para a maior ou menor complexidade textual verificada pelas métricas supracitadas (IF, ASL, TT e DS).

Na próxima seção, discutiremos os valores dos índices encontrados nos textos originais e suas interpretações a respeito de complexidade.

4 Resultados obtidos

Uma série de análises interpretativas com as métricas geradas pelo *Coh-Matrix Dementia* para o *corpus* de estudo foram conduzidas com o objetivo de descrever os textos, verificando a hipótese de que seriam potencialmente complexos para o público leigo com escolaridade limitada.

Esse conjunto de análises se baseou nas especificidades de cada uma das métricas escolhidas e na comparação dessas métricas entre os textos do *corpus* de estudo (dez textos originais encontrados na internet escritos para leigos sobre o TEPT) e seis textos (*corpus* de referência/contraste) considerados simples (cinco narrativas para crianças retiradas de páginas da internet e um texto previamente simplificado que apresentou bons resultados pela análise métrica). A finalidade de analisar as métricas dos seis textos considerados “simples” era possibilitar uma comparação dos valores numéricos do *corpus* de estudo com o *corpus* de referência. Esses textos foram escolhidos não só pela sua simplicidade, mas também para que pudesse existir um *tertio*

comparationis, uma vez que, embora o *Coh-Metrix Dementia* apresente explicações sobre as métricas, valores numéricos de referência não são apresentados.

Para cada um dos índices, Cortina Silva (2018) utilizou quadros classificatórios para demonstrar os valores numéricos e interpretar, qualitativamente, esses números em “níveis de complexidade”, utilizando as diretrizes de cada métrica. Sendo assim, Cortina Silva (2018) utilizou sistemas classificatórios de sua autoria (de acordo com as especificidades de cada métrica) para classificar e interpretar os valores de cada índice. Além de apresentar os números, Cortina Silva (2018) apresentou os dados com as seguintes “etiquetas de classificação”⁴: [complexidade] baixa, muito baixa, média, alta e muito alta. Os resultados de complexidade das métricas dos textos originais estão dispostos no quadro abaixo (Quadro 1). Nele podemos observar que as quatro métricas analisadas indicaram potenciais que variavam entre graus de complexidade "médio" e "muito alto". Esses indicativos nos permitiram concluir que os textos podem ser considerados complexos para o público estudado. Cortina Silva (2018, p. 168) concluiu que “Essa classificação nos indica que 17,5% das métricas totais indicam textos de complexidade muito alta, 57,5% apontam que os textos têm complexidade alta e 25% nos mostram que os textos têm complexidade média”. Conforme o autor, um texto acessível para esse perfil de leitor deveria apresentar métricas classificadas entre "muito baixo" e "baixo" em potencial de complexidade.

Quadro 1: Interpretação do potencial de complexidade das métricas do *corpus* de estudo

	Classificação de complexidade (IF)	Classificação de complexidade (ASL)	Classificação de complexidade (TT)	Classificação de complexidade (DS)
PORT ⁵ 1	Alta	Média	Média	Média
PORT 2	Muito alta	Alta	Alta	Muito alta
PORT 3	Muito alta	Alta	Alta	Média
PORT 4	Muito alta	Alta	Alta	Média
PORT 5	Alta	Alta	Alta	Alta
PORT 6	Muito alta	Alta	Média	Alta

⁴ Por limitações de espaço, optamos por não apresentar os quadros neste texto. Contudo, eles podem ser observados na seção *Apresentação e Análise de Dados* (p. 131-178) da dissertação de Cortina Silva (2018). Disponível em: <https://lume.ufrgs.br/handle/10183/189547>

⁵ PORT (abreviação de Português) foi a etiqueta utilizada para se referir aos dez textos sobre o TEPT componentes do corpus de estudo.

PORT 7	Alta	Alta	Alta	Média
PORT 8	Muito alta	Alta	Alta	Alta
PORT 9	Muito alta	Média	Média	Alta
PORT 10	Alta	Alta	Média	Alta

Fonte: Cortina Silva (2018, p. 168)

O *corpus* de referência, por outro lado, apresentou resultados diferentes do *corpus* de estudo. Cortina Silva (2018, p. 168) afirma que esses resultados apontam que “37,5% das métricas são consideradas de alta complexidade, 29,1% são classificadas como médias, 29,1%, como baixas e 4,1% como muito baixas.” Entretanto, o autor também observou que as métricas que demonstraram potencial de complexidade alta são, em grande maioria, a Análise Semântica Latente (ASL) e a Relação *Type-Token* (TT). Segundo ele, isso ocorre porque cinco dos seis textos comparativos eram narrativas para crianças, e narrativas apresentam desenvolvimentos e manutenções tema-remas⁶ que configuram uma diferença na estrutura semântica das frases ao decorrer do texto, fator que impacta no valor de ASL. Além disso, narrativas tendem a descrever diferentes partes da história e, dessa forma, a variação lexical tende a ser maior e, conseqüentemente, o valor de TT também se altera.

Além da interpretação dos índices, Cortina Silva (2018) também realizou uma análise holística dos textos e, com base em preceitos da LT já referidos, concluiu que a maior parte dos textos pode ser considerada complexa para o público destinado por uma série de motivos, os quais resumimos a seguir: a) grau de informatividade alto; b) linguagem complexa; c) grande presença de termos específicos da Psicologia/Psiquiatria; d) baixos elementos de situacionalidade que permitissem que o leitor construísse a interpretação com base na sua posição; e) intertextualidade baixa; f) intencionalidade pouco clara; g) organização complexa e remissividade; h) extensão grande do texto como todo, de parágrafos e de frases; i) repetição de informações; j) falta de informações.

⁶ Manutenção tema-remas é um fenômeno de “desdobramento” textual. Ocorre quando um subtema se forma partindo de um tema original. Essas manifestações podem ocorrer de diferentes formas: um novo tema pode surgir (ou não) de um tema anterior. Normalmente, o Tema expressa a informação dada, a qual já é conhecida pelo nosso ouvinte ou que é recuperável no contexto. O Remas, por sua vez, expressaria a informação nova: aquela que nosso ouvinte desconhece, e que corresponde, efetivamente, ao conteúdo que queremos que ele passe a conhecer (SIPPERT, 2018).

Portanto, diante da interpretação das métricas e de uma análise qualitativa holística, Cortina Silva (2018) corroborou uma das hipóteses mencionadas na seção Introdução: a) há uma tendência para que textos disponíveis gratuitamente na internet voltados para o público leigo sobre o TEPT sejam potencialmente complexos conforme as medidas do sistema *Coh-Matrix-Dementia*. Portanto, constatou que os textos encontrados na internet sobre o TEPT para leitores leigos são potencialmente complexos para um leitor adulto de escolaridade limitada e que estratégias de simplificação deveriam ser adotadas para permitir maior acessibilidade.

5 Estratégias aplicadas

Com a finalidade de observar técnicas de simplificação que possam ser eficientes para colaborar para a acessibilidade de textos científicos, oito estratégias de simplificação foram adotadas para uma análise comparativa entre as métricas dos textos originais e os valores dessas mesmas métricas nos textos “simplificados”. As estratégias adotadas por Cortina Silva (2018) foram:

1. Simplificação por edição de pronomes: a maior parte possível dos pronomes foram substituídos por seus referenciais, ainda que isso indicasse repetição de palavras.

2. Simplificação por quebra de sentenças/parágrafos em estruturas menores: essa estratégia constituiu na redução do tamanho de sentenças e de parágrafos. Quando orações eram compostas por mais de um verbo, essas orações foram transformadas em duas ou mais. Quando parágrafos foram considerados longos, esses parágrafos foram divididos em dois ou mais, mesmo que isso indicasse repetição dos sujeitos e/ou verbos ou necessidade de preenchimento (repetição) de informações.

3. Simplificação por redução de informação: foram retirados trechos considerados como “desnecessários” (como componentes de medicamentos e explicações de procedimentos técnicos), redundantes ou repetitivos, tendo em mente leitores leigos buscando informações básicas sobre o TEPT. Sobre essa estratégia, Cortina Silva (2018) comenta que as informações eliminadas seguiram uma série de critérios pré-determinados. Entretanto, se os textos editados fossem ser divulgados para públicos reais, seria necessário que um profissional da Psicologia ou Psiquiatria validasse a escolha das informações excluídas.

4. Simplificação por ampliação de informação: por meio de um movimento “contrário” à estratégia anterior, foram inseridas informações complementares sobre termos ou partes do texto julgadas complexas para um leitor leigo com o objetivo de evitar que ele buscasse informações em outras fontes e interrompesse o fluxo da leitura.

5. Simplificação lexical: essa estratégia consiste em substituir palavras consideradas complexas por sinônimos mais simples. Para isso, o autor buscou, em *corpus* da língua portuguesa, sinônimos que fossem mais frequentes do que a palavra original como, por exemplo, substituir a palavra “eloquente” por “expressivo”, uma vez que “eloquente” ocorre 1.116 vezes no *corpus* utilizado e “expressivo”, 8.292.

6. Simplificação por eliminação da voz passiva: substituição de todas as instâncias da voz passiva analítica pela voz ativa.

7. Simplificação por redução de adjetivos: eliminação de adjetivos considerados desnecessários para a compreensão do texto (como adjetivos que refletissem uma opinião pessoal).

8. Simplificação por redução de advérbios: exclusão de advérbios considerados desnecessários para a compreensão do texto.

Na próxima seção, discutiremos os resultados do potencial de simplificação encontrados por meio da aplicação das estratégias supracitadas.

5.1 Resultados das estratégias

Ao verificar o potencial de complexidade dos textos analisados, Cortina Silva (2018) aplicou as estratégias supracitadas de forma isolada em cada um dos textos e submeteu esses textos editados ao *Coh-Matrix Dementia*. Isso foi feito para analisar o possível impacto que cada estratégia teria nas métricas.

Para a verificação da validade das estratégias aplicadas, o autor interpretou os dados de duas principais formas. Primeiramente, comparou os valores das métricas dos textos originais com os valores das métricas após a aplicação de cada estratégia. Ou seja, o autor comparou os valores de IF, DS, TT e ASL dos textos originais com os valores dessas métricas após a edição de pronomes, depois comparou os valores dos índices dos textos originais com os resultados provenientes da simplificação lexical e assim por

diante. Essa comparação, portanto, foi feita após a aplicação de cada uma das estratégias para a validação de cada uma delas de forma isolada.

Em um segundo momento, Cortina Silva (2018) isolou cada uma das métricas e comparou (lado a lado) os valores encontrados nos textos originais e nos textos simplificados (comparando todos os valores de IF, por exemplo, entre os textos do *corpus* de estudo e de cada versão após a aplicação das estratégias). Sendo assim, o autor pode verificar quais estratégias tiveram maiores impactos nos valores numéricos de cada índice⁷.

Após um estudo comparativo de todos os dados, o autor desenvolveu um sistema classificatório de pontos⁸ que ordenou as estratégias adotadas de acordo com a eficiência apresentada por elas. O quadro abaixo (Quadro 2) apresenta as pontuações recebidas por cada estratégia e sua classificação de eficiência:

Quadro 2: Classificação das estratégias de simplificação adotadas

CLASSIFICAÇÃO	ESTRATÉGIA	PONTUAÇÃO
1	Simplificação lexical	66
2	Redução de adjetivos	44
3	Redução de informação	36
4	Edição de pronomes	25
5	Quebra de parágrafos e frases	24
6	Ampliação de informação	17
7	Redução de advérbios	11
8	Eliminação da voz passiva	4

Fonte: Cortina Silva (2018, p. 200)

Com base na interpretação dos resultados, Cortina Silva (2018) pôde corroborar a hipótese de pesquisa previamente mencionada: b) dentre as diferentes estratégias de reescrita que favoreçam a acessibilidade textual e terminológica de textos escritos sobre TEPT, algumas se mostram mais eficazes do que outras apontadas pelo mesmo sistema⁹.

⁷ Para uma leitura mais detalhada dos valores das métricas, consultar a seção *Resultados dos Testes de Reescrita com vistas à Simplificação* (p. 178-202) da dissertação de Cortina Silva (2018).

⁸ Para uma leitura mais detalhada sobre o sistema desenvolvido por Cortina Silva (2018), consultar a seção *Sistema classificatório para interpretação das estratégias com melhores resultados de simplificação* da dissertação (p. 199-202).

⁹ É importante frisar que, embora a pesquisa tenha sido realizada com textos da área da Psiquiatria, os resultados e orientações de [re]escrita simplificada não se limitam apenas a textos da área da Saúde, mas

Ainda sobre os resultados observados, Cortina Silva (2018) salientou que, embora algumas estratégias tenham demonstrado maior eficiência que outras, todas puderam ser validadas.

Cortina Silva (2018) observa que a eliminação da voz passiva, por exemplo, embora seja indicada na literatura sobre acessibilidade textual, não apresentou grande eficiência de acordo com as métricas estudadas. Ele observou que as estratégias com melhores resultados foram a simplificação lexical, a redução de adjetivos e a redução de informações. Salientou, igualmente, que alguns resultados podem ser diferentes e ter sua eficiência comprovada quando textos editados puderem ser testados com leitores reais. Entretanto, as interpretações dos índices dos textos simplificados ao serem comparados com os valores dos textos originais permitiram levantar conclusões a respeito de práticas de composição que colaboraram para a acessibilidade textual.

Baseando-se nos resultados deste estudo e em trabalhos dos grupos de pesquisa associados ao seu mestrado, Cortina Silva (2018) elaborou orientações para a composição de um futuro guia de simplificação para o português brasileiro. Trataremos dessas orientações na próxima seção.

6. Orientações para simplificação textual com base nos resultados

Tendo aproveitado as produções de um grupo de pesquisa que se dedica especialmente a estudos sobre complexidade e facilitação de textos técnico-científicos para pessoas leigas¹⁰, Cortina Silva (2018) elaborou uma série de orientações para instruir a composição ou edição de textos acessíveis. A seguir, abordaremos as instruções levantadas pelo autor.

1. Delimitar o perfil de leitor: uma vez que o potencial de complexidade de um texto só pode ser analisado diante de um leitor (ou perfil de leitor), é importante entender para quem o texto será escrito e/ou simplificado. Uma simplificação para um estudante de Medicina deve ser diferente de uma simplificação para um leitor com baixa escolaridade, por exemplo.

podem ser utilizados para a redação/edição simplificada de materiais de diferentes tipologias e provenientes de distintas áreas do conhecimento.

¹⁰ <http://www.ufrgs.br/textecc/acessibilidade/>

2. **Ponderar sobre a necessidade de adicionar informações extras ao texto:** é importante se posicionar no papel do leitor e se certificar que sejam adicionadas informações “extras” somente para palavras/termos essenciais para a compreensão de algum assunto (e que não possam ser substituídos por referenciais mais simples).

3. **Escrever o que é necessário:** textos em língua portuguesa do Brasil tendem a ser mais detalhados em comparação a textos em inglês, por exemplo. Entretanto, quando se compõe um texto com objetivo de ser simplificado, a regra “menos é mais” pode ser válida, desde que esse “menos” seja entendido como “as informações que sejam necessárias”. Ser direto em um texto simplificado cumpre melhor os objetivos do que detalhar com informações que são desnecessárias e exigem mais da cognição do leitor.

4. **Utilizar palavras simples:** para um texto simplificado, palavras rebuscadas são desnecessárias e tornam o texto complexo. A educação formal do brasileiro tem a tendência a ensinar a utilizar a maior quantidade possível de sinônimos em um texto para evitar repetições. Entretanto, isso não se aplica a um texto simplificado. É preferível que palavras sejam repetidas do que a utilização de palavras que possam ser complexas, principalmente para o leitor que não tem domínio da área sobre o qual o texto se refere. Deve-se manter em mente que os objetivos de um texto simplificado não devem se preocupar com a “estética” do texto, mas com seu propósito informativo.

5. **Evitar substituir substantivos por pronomes:** sempre que possível, evitar utilizar um pronome no lugar do seu referencial, principalmente quando esse referencial não está próximo do pronome. Embora essa técnica nos “force” a repetir palavras, o estudo comprovou (bem como outros estudos anteriores a esse, como apontado pelos Guias de Simplificação do Governo Norte Americano¹¹) que os pronomes podem ser fatores complicadores na compreensão do texto, uma vez que exige um raciocínio inferencial que não se pode presumir que exista no leitor.

6. **Evitar frases e parágrafos longos:** além da extensão do texto como um todo, é importante manter frases e parágrafos curtos. É preferível a repetição de sujeitos e

¹¹ Os *guidelines* de simplificação são disponibilizados pelo governo dos Estados Unidos no projeto *Plain Language*. Esse manual orienta, com detalhes, alternativas de como compor um texto de forma simplificada. Esse manual está disponível em <https://www.plainlanguage.gov/guidelines/> e foi traduzido e adaptado para a língua portuguesa por Cortina Silva (2016) e uma versão resumida pode ser acessada em <http://www.ufrgs.br/textecc/acessibilidade/page/simplificar/>.

verbos do que frases longas. Orações com mais de um verbo devem ser divididos em duas ou mais.

7. **Evitar um texto longo:** um texto longo pode ter um impacto psicológico negativo no leitor. Uma extensão grande, além de exigir cognitivamente mais dos leitores, geralmente tende a ser abandonada ou apenas “escaneada”, o que pode gerar má interpretação por parte de quem o lê.

8. **Evitar adjetivos e locuções adjetivas:** é importante lembrar que um texto informativo se difere de um texto com opiniões pessoais. Tente inserir somente adjetivos que sejam essenciais para a compreensão do texto (por exemplo: o remédio não é indicado para quem tem pressão *alta*).

9. **Evitar advérbios:** advérbios devem ser evitados. Essa classe de palavras é apontada pelas diretrizes do Governo Americano e pela dissertação de Cortina Silva (2018) como difícil de ser compreendida. Evite, principalmente, advérbios longos. Ao inserir um advérbio, é importante observar sua posição na frase. Além da complexidade das palavras por si só, um advérbio deslocado pode interromper o fluxo de pensamento.

10. **Deixar o texto limpo:** evite textos que sejam “sujos” em formatação, marcadores visuais, gráficos e imagens. Insira gráficos e imagens somente quando forem necessários. Além de exigir outro tipo de interpretação (a não-textual), a inserção desses elementos pode interromper o fluxo de pensamento do texto quando não forem bem posicionados ou quando a interpretação for complexa.

11. **Organize suas informações por temas/tópicos:** novamente, fluxo de leitura é importante para a compreensão de um texto. Evite textos que sejam remissivos e que “forcem” o leitor a ir e voltar. Se falar a respeito de uma doença, por exemplo, distribua os tópicos de forma clara: origem da doença, causas, tratamentos, medicamentos, e evite abordar conteúdos sobre tópicos diferentes e que não sejam específicos sobre o assunto. Se for preciso, adicione uma nota informando que determinado assunto será tratado mais adiante no texto. Ao adicionar notas, entretanto, insira no corpo do texto. Evite notas de rodapé ou outras formas que exijam que o leitor interrompa o fluxo da leitura.

12. **Evitar fazer com que o leitor saia do texto:** tente organizar seu texto de forma que o leitor não tenha a necessidade de buscar informações em dicionários ou páginas na internet, por exemplo. Essa estratégia deve ser bem alinhada às demais.

Deve-se ponderar sobre os conteúdos necessários ao texto e adicionar as informações complementares de forma sucinta e clara.

13. Optar pela voz ativa em vez da voz passiva: a voz passiva pode ser considerada muito complexa, mesmo para leitores proficientes. Esse tipo de voz pode confundir o leitor sobre o agente/sujeito e a omissão desse componente pode causar confusão. Embora a pesquisa de Cortina Silva (2018) tenha apontado que a substituição da voz passiva foi a estratégia com menores resultados, a literatura a respeito de acessibilidade textual enfatiza sobre a importância de evitar esse tipo de voz verbal. Importante ressaltar que testes com leitores reais podem comprovar (ou refutar) a complexidade da voz passiva.

14. Manter, sempre que possível, a ordem canônica de formação de frases: procure apresentar as frases na ordem sujeito, verbo, objeto. O deslocamento de algum componente das frases pode desconectar os referenciais e, por esse motivo, tornar a leitura complexa.

15. Lembrar que seus conhecimentos são diferentes dos conhecimentos do leitor: é preciso se posicionar como leitor do texto e lembrar que sua formação é diferente da formação do leitor. Não pressuponha que uma informação está escrita de forma clara porque você a compreendeu; revise-a e, se possível, solicite a revisão por um colega, para validá-la.

As orientações descritas por Cortina Silva (2018), como anteriormente mencionado, puderam ser concluídas com base na sua pesquisa com um conjunto de dados textuais e nos seus estudos sobre CT. Entretanto, é válido lembrar que apenas essas orientações não são suficientes para desenharmos um perfil de linguagem simples, acessível ou simplificada por completo. Sendo assim, prosseguir pesquisas a respeito do tema é fundamental para que possamos, cada vez mais, compreender características que perfazem uma linguagem acessível para diferentes públicos, considerando suas especificidades, saberes, culturas e necessidades. Também é preciso destacar que a pesquisa de Cortina Silva (2018) dedicou-se a antecipar as necessidades de apenas um tipo de um público leigo, conformado por pessoas adultas de escolaridade limitada que, em tese, buscam informações escritas na internet sobre um dado tema de Saúde. Como sabemos, diferentes perfis de leitores “exigem” diferentes estratégias de acessibilidade.

Isso ainda se precisa melhor reconhecer e sistematizar no âmbito dos estudos linguísticos.

Considerações finais

O trabalho aqui tomado como tópico chama a atenção para a questão da complexidade textual e terminológica de textos científicos para pessoas com escolaridade limitada. Destaca, principalmente, a relevância do tema da Acessibilidade Textual e Terminológica (ATT) por meio da descrição e da análise de textos escritos sobre o TEPT disponíveis na internet. A partir disso, oferece alternativas de facilitação de textos potencialmente complexos. Suas indicações baseiam-se em uma análise minuciosa de métricas de complexidade, geradas por uma ferramenta computacional, aliadas a uma avaliação multidimensional do funcionamento dos elementos textuais, lexicais e gramaticais. Ao final, é apresentado um guia para redatores que precisam produzir textos simplificados. Essa é uma fonte opcional de recomendações, a qual deve ser adaptada conforme o perfil do leitor a ser contemplado, o tipo de texto a ser facilitado e a língua na qual o texto é veiculado.

Um ponto limitante da pesquisa, reconhecido no próprio trabalho, é a falta de uma testagem dos textos simplificados pelo pesquisador com os leitores reais cuja compreensão se pretende auxiliar. Essa lacuna, sem ser demérito, abre espaço para uma continuidade da investigação. E, justamente nesse ponto, vale que diferentes pesquisadores dos Estudos da Linguagem e de áreas afins possam conhecer o que foi feito, em termos de alternativas de simplificação, reescrita e comparações diversas, reunindo ideias e subsídios para possíveis continuidades. A promoção da acessibilidade da informação sobre temas de Utilidade Pública e o tema da linguagem simples são, além de tópicos de pesquisa acadêmica na área de Letras e Linguística, fatores de promoção da cidadania e do Letramento em Saúde no Brasil.

Agradecimentos

Os autores agradecem aos apoios do PPG-LETRAS-UFRGS, da PUCRS, do CNPq e da FAPERGS, instituições que estiveram em sinergia para a concretização da pesquisa aqui relatada e para a produção deste artigo.

REFERÊNCIAS

- BERBER SARDINHA, T. **Análise multidimensional**. São Paulo: Delta, 2000.
- BIBER, D. **Variation across Speech and Writing**. Cambridge: Cambridge University Press, 1988.
- BIBER, D. **Dimensions of Register Variation – A Cross-Linguistic Comparison**. Cambridge: Cambridge University Press, 1995.
- CIAPUSCIO, G. E. **Textos especializados y terminología. Panorama de las tipologías**. Barcelona: IULA, 2003.
- CORTINA SILVA, A.D.C., et al. **Fundamentos da Língua Portuguesa**. Porto Alegre: Editora Sagah, 2018.
- CORTINA SILVA, A.D.C. **Textos de divulgação para leigos sobre o Transtorno do Estresse Pós-Traumático em português: alternativas para a acessibilidade textual e terminológica**. Dissertação (Mestrado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.
- CUNHA, A. L. V. **Coh-Metrix-Dementia: análise automática de distúrbios de linguagem nas demências utilizando Processamento de Línguas Naturais**. Tese (Doutorado em Ciência da Computação e Matemática Computacional) – Universidade de São Paulo, São Paulo, 2015.
- FINATTO, M.J.B. **Acessibilidade Textual e Terminológica: um Novo Tópico de Pesquisas em Terminologia no Brasil 1** - Estudos Geossociolinguísticos do Português Brasileiro. Porto Alegre: Editora Pontes, 2018.
- FINATTO, M.J.B. **Complexidade Textual em Artigos Científicos: contribuições para o estudo do texto científico em português**. Porto Alegre: Organon (UFRGS), v. 50, 2011.
- KOCH, I.V. **Introdução à Linguística Textual**. São Paulo Editora Contexto, 2015.
- INEP, **Letramento científico**. 2010. Disponível em http://download.inep.gov.br/download/internacional/pisa/2010/letramento_cientifico.pdf. Acessado em 20 de junho de 2020.
- MANNING, C.D., SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Massachusetts: MIT Press, 1999.
- MEDEIROS, R. **Crescem casos de estresse pós-traumático no Brasil**. *Psique Ciência & Vida*. no. 137. 2017. Disponível em: <http://psiquecienciaevida.com.br/cresce-casos-de-estresse-pos-traumatico-no-brasil>. Acessado em 11 de abril de 2018.

PASQUALINI, B.F., SCARTON, C., FINATTO, M.J.B. **Comparando Avaliações De Inteligibilidade Textual Entre Originais E Traduções De Textos Literários.** In: VIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, 2011, Cuiabá: Anais do STIL 2011. Cuiabá: Sociedade Brasileira de Computação, 2011.

SIPPERT, L. **Análise da estrutura temática em resenhas de alunos do Ensino Superior numa perspectiva da Linguística Sistêmico-Funcional.** São Paulo: DELTA, v.31, n.1, 2018.

SAGGION, H. **Automatic Text Simplification: Synthesis Lectures on Human Language Technologies.** Barcelona: Morgan & Claypool, v. 10, n.1, 2017.