

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ÁREA DE CONCENTRAÇÃO: ESTUDOS DA LINGUAGEM
LINHA DE PESQUISA: LINGUÍSTICA APLICADA**

ELLEN YURIKA NAGASAWA

**O CONTEÚDO DE INSUMO EM TAREFAS QUE INTEGRAM LEITURA E
ESCRITA NO CELPE-BRAS: UMA ABORDAGEM INFORMADA POR
*CORPUS***

PORTO ALEGRE

2023

ELLEN YURIKA NAGASAWA

**O CONTEÚDO DE INSUMO EM TAREFAS QUE INTEGRAM LEITURA E
ESCRITA NO CELPE-BRAS: UMA ABORDAGEM INFORMADA POR
*CORPUS***

Tese em Linguística Aplicada apresentada como requisito parcial para a obtenção do título de Doutora em Letras pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Orientadora: Profa. Dra. Margarete Schlatter
Coorientadora: Profa. Dra. Adriana Picoral

PORTO ALEGRE

2023

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

REITOR

Carlos André Bulhões Mendes

VICE-REITORA

Patricia Pranke

DIRETORA DO INSTITUTO DE LETRAS

Carmem Luci da Costa Silva

VICE-DIRETORA DO INSTITUTO DE LETRAS

Márcia Montenegro Velho

DIRETORA DO PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

Simone Sarmento

VICE-DIRETOR DO PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

Antônio Marcos Vieira Sanseverino

CIP - Catalogação na Publicação

Nagasawa, Ellen Yurika

O conteúdo de insumo em tarefas que integram
leitura e escrita no Celpe-Bras: uma abordagem
informada por corpus / Ellen Yurika Nagasawa. -- 2023.
302 f.

Orientadora: Margarete Schlatter.

Coorientadora: Adriana Picoral.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Letras, Programa de
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2023.

1. Celpe-Bras. 2. Avaliação de leitura. 3.
Avaliação de linguas informada por corpus. 4.
Complexidade textual. 5. Argumentos de validade. I.
Schlatter, Margarete, orient. II. Picoral, Adriana,
coorient. III. Título.

ELLEN YURIKA NAGASAWA

**O CONTEÚDO DE INSUMO EM TAREFAS QUE INTEGRAM LEITURA E
ESCRITA NO CELPE-BRAS: UMA ABORDAGEM INFORMADA POR
*CORPUS***

Tese em Linguística Aplicada apresentada como requisito parcial para a obtenção do título de Doutora em Letras pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Porto Alegre, 16 de outubro de 2023.

Resultado: Aprovada.

BANCA EXAMINADORA:

Prof. Dra. Margarete Schlatter (Orientadora)
Universidade Federal do Rio Grande do Sul

Prof. Dra. Adriana Picoral (Coorientadora)
University of Arizona

Prof. Dra. Juliana Roquele Schoffen
Universidade Federal do Rio Grande do Sul

Prof. Dra. Larissa Goulart da Silva
Montclair State University

Prof. Dr. Leandro Rodrigues Alves Diniz
Universidade Federal de Minas Gerais

Ao Cícero e à Maria Cláudia.

AGRADECIMENTOS

Ao longo do processo de doutoramento, enfrentei não apenas desafios teórico-metodológicos, mas também vivenciei um período historicamente trágico, durante o qual lutamos, literalmente, para sobreviver a uma pandemia mundial, e, figurativamente, a quatro anos de uma gestão governamental marcada pela desvalorização da docência e da pesquisa no Brasil.

No âmbito pessoal, atravessei uma série de fases críticas que, em muitos momentos, pareciam obstáculos intransponíveis ao avanço desta pesquisa. Diante de tantas incertezas, a conclusão do doutorado parecia realmente impossível. No entanto, aqui estou, delineando meus agradecimentos. Tal conquista se deve, em grande parte, ao fato de eu não ter percorrido esta jornada sozinha; estive constantemente amparada por pessoas que me ergueram, queda após queda. Registro aqui meus agradecimentos a elas, embora saiba que essas palavras não capturam a profundidade de meu sentimento de gratidão.

Primeiramente, agradeço ao Cícero e à Maria Cláudia, que me apoiaram na escolha de iniciar o doutorado, me incentivaram durante o processo e, sobretudo, me aturaram no período final.

Aos meus colegas do grupo de pesquisa Avalia, que continuam me instigando a refletir e a rir, uma combinação que certamente representa a nossa maior força. Agradeço em particular aos colegas que contribuíram de alguma maneira para a concretização deste estudo: Fernanda Xavier, Gabrielle Rodrigues Sirianni, Giovana Lazzaretti Segat, Isadora Dahmer Hanauer, Kaiane Mendel, Mariana Alves e Matheus de Aguilar Garcia.

À minha colega, companheira e amiga, Gabrielle Rodrigues Sirianni (e família), que esteve ao meu lado desde o início desta jornada. Foi um prazer imenso trilhar esse caminho ao seu lado.

Aos meus colegas do corpo docente de Letras Inglês da Universidade do Vale do Rio dos Sinos, por compreenderem os meus momentos de distanciamento e incentivarem a conclusão da minha pesquisa. Em especial, agradeço à Adila Beatriz Naud de Moura e Marcia de Oliveira Del Corona, pela confiança depositada, pela valorização do meu trabalho e pela forma, simultaneamente afetuosa e profissional, com a qual me guiaram dentro da universidade. Foi uma honra aprender a ser professora universitária sob a orientação de ambas.

Ao Sidney Evaldo Leal, que atendeu prontamente nosso pedido, fornecendo o código da ferramenta de análise e gentilmente gerou para nós os dados que necessitávamos.

Às gurias do meu grupo “Shut up and Write”, Marine Laísa Matte, Kaiane Mendel e Gabrielle Rodrigues Sirianni, carinhosamente apelidadas de “Sussa”. Agradeço por me apoiarem, entenderem e inspirarem. Agradeço pelas risadas compartilhadas durante as partidas de Gartic no período de isolamento social. Acima de tudo, sou grata pela constante presença e disponibilidade, pois estão sempre a uma simples mensagem de distância.

Às gurias do meu grupo de leitura “Reading Out Loud”, pelas oportunidades de desacelerar os pensamentos e dar espaço à fruição.

À Letícia Fossati, pela escuta atenta e pela atenção dedicada para a minha saúde mental.

Nos ombros de gigantes

Toda a minha formação acadêmica foi guiada por mulheres extraordinárias, um fato que eu não poderia deixar de registrar.

Esta tese não teria se concretizado sem as inúmeras conversas com minha orientadora, que sempre enfatizou a importância desta pesquisa e a certeza de que eu era a pessoa certa para conduzi-la. Foram incontáveis “me dá um minuto, vou pegar um chá e já abro o Zoom”. Quando eu parecia não ter mais forças e esperança, do seu jeito, sempre me persuadiu a acreditar que tudo estava sob controle, mesmo quando o meu mundo estava um caos. Agradeço à Margarete Schlatter, pela melhor orientação que eu poderia ter tido nesse período tumultuado da vida e pela amizade que anseio cultivar.

À Adriana Picoral, coorientadora deste trabalho, professora e pesquisadora fenomenal, pela disponibilidade e atenção, por todo o aprendizado e paciência com minhas dúvidas de aprendiz. Sem sua orientação, esta tese não teria se materializado da maneira que se concretizou. Não tenho palavras para expressar suficientemente a minha gratidão pela dedicação das tantas horas de trabalho voluntário para me apresentar um mundo novo; foram dois anos de encontros virtuais semanais, inclusive durante suas férias, feriados ou viagens a trabalho. Mesmo estando a nove mil quilômetros de distância, me acompanhou muito de perto.

À Luciene Simões, por aceitar ser nossa interlocutora desde o projeto da tese, incluindo o artigo publicado até a banca de qualificação. Pela leitura cuidadosa e interessada, pelo diálogo e direcionamentos que mudaram o rumo da minha pesquisa.

À Larissa Goulart, pela interlocução durante o exame de qualificação, cujos questionamentos e apontamentos contribuíram efetivamente para aprimorar este trabalho. Pelo aceite em também fazer parte da banca examinadora.

À Juliana Roquele Schoffen, sem a qual eu não teria me tornado uma linguista. Por me enxergar, por acreditar no meu potencial, por me acolher e por possibilitar que eu encontrasse

um espaço dentro da universidade. Pelos cinco anos de orientação durante a graduação e o mestrado, pelos ensinamentos, discussões, risadas e amizade. Agradeço o aceite para compor a banca examinadora deste trabalho.

Ao Leandro Diniz, por prontamente aceitar participar dessa interlocução e a qualificar este trabalho, compondo a banca examinadora desta tese.

Finalmente, expresso minha profunda gratidão à Universidade Federal do Rio Grande do Sul por fornecer uma formação de excelência, tanto na graduação quanto na pós-graduação, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pelo apoio financeiro e pelo incentivo da pesquisa no Brasil.

A leitura da palavra não é apenas precedida pela leitura do mundo, mas por uma certa forma de “escrevê-lo” ou de “reescrevê-lo”, quer dizer, de transformá-lo através de nossa prática consciente.

Paulo Freire

RESUMO

Os testes de línguas de larga escala e alto impacto social têm sido amplamente investigados devido a dimensão das consequências sociais do uso de seus resultados como ferramentas para classificação e exclusão de indivíduos na sociedade. Essa preocupação está ligada à busca científica por argumentos que validem as inferências derivadas desses resultados. Um argumento plausível de validade é baseado em evidências que demonstram a representatividade do conteúdo do teste em relação ao domínio de uso da linguagem avaliado, conforme definido a partir dos usos pretendidos dos resultados e delineados nas especificações. A finalidade das especificações é possibilitar a reprodução de edições equivalentes, fornecendo detalhes suficientes para a recriação de versões com conteúdo e demandas similares, incluindo informações relativas à complexidade textual, que podem impactar no desempenho de leitura. Avaliar a representatividade do conteúdo e da complexidade textual dos textos de insumo oferece evidências para a validade do exame. O objetivo desta tese é fornecer evidências para verificar o argumento de validade de que o domínio de uso da linguagem que o Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras) visa avaliar é representado pelo conteúdo das tarefas que integram leitura e produção textual desse exame brasileiro de proficiência em Português como Língua Adicional. Adicionalmente, a pesquisa visa treinar e testar um classificador binário, utilizando 200 métricas automáticas da ferramenta NILC-Matrix, para verificar suas capacidades preditivas de complexidade textual dos textos de insumo dessas tarefas e o grau de adequação ao público-alvo. Para tanto, a pesquisa se apoiou na abordagem da Avaliação de línguas informada por corpus, com bases teórico-metodológicas das áreas de Avaliação de línguas e Linguística de *Corpus*. Por meio da linguagem de programação R, foram conduzidas análises de caráter descritivo e exploratório do conteúdo e características das tarefas e testes estatísticos de redução de multivariáveis com o algoritmo Boruta. O *corpus* de estudo, constituído por 90 tarefas aplicadas entre 1998 e 2020 no Celpe-Bras, dois *corpora* de referência, documentos oficiais e estudos prévios foram utilizados nas análises. Com as 33 métricas preditivas selecionadas pelo algoritmo, um modelo estatístico de complexidade foi criado com 80% dos dados dos *corpora* de complexidade. Os 20% restantes foram utilizados na testagem, que distinguiu textos de alta e baixa complexidade com 100% de acurácia. Os resultados indicaram que o conteúdo das tarefas não reflete as especificações, que carecem de detalhes sobre a seleção de texto de insumo e o processo de elaboração. Ademais, a maioria dos textos de insumo foram classificados como altamente complexos, correspondente ao grau mais alto de letramento brasileiro, de acordo com o Indicador de Alfabetismo Funcional. As evidências teóricas e empíricas levantadas neste estudo não permitiram confirmar o argumento de validade pretendido. Candidatos em nível intermediário podem ser potencialmente prejudicados pela inacessibilidade dos textos de insumo e pela falta de diversidade de repertórios comunicativos e culturais desses materiais. Os resultados desta pesquisa devem ser complementados por estudos dos outros instrumentos que compõem o exame para uma compreensão integral da representatividade do domínio de uso da linguagem no Celpe-Bras.

Palavras-chave: Celpe-Bras; avaliação de leitura; Avaliação de línguas informada por corpus; complexidade textual; argumentos de validade.

ABSTRACT

Large-scale and high-stakes language tests have been widely investigated due to the extent of the social consequences of the use of their results as tools for the classification and exclusion of individuals in society. This concern is linked to the scientific pursuit of arguments that validate the inferences derived from test results. A plausible argument for validity is based on evidence that shows the representativeness of the test content in relation to the language domain being assessed, as defined from the intended uses of the results and outlined in the specifications. The purpose of the specifications is to enable the replication of equivalent versions, providing sufficient detail for the creation of versions with similar content and demands, including aspects of textual complexity that can impact reading performance. Assessing the representativeness of the content and textual complexity of the input texts provides evidence for the validity of the test. The aim of this dissertation is to provide evidence to verify the validity argument that the language domain that the *Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras)* intends to assess is represented by the content of the integrated reading-to-write tasks in this Brazilian Portuguese as an Additional Language test. In addition, the research aims to train and test a binary classifier, using 200 automatic metrics from the NILC-Matrix tool, to verify its predictive potential of the textual complexity of the tasks' input texts and their suitability for the target audience. For this purpose, the study was based on the Corpus-informed language testing approach, with theoretical and methodological frameworks from the areas of Language Testing and Corpus Linguistics. Using the R programming language, descriptive and exploratory analyses were carried out on the content and characteristics of the tasks, as well as statistical multivariable reduction tests using the Boruta algorithm. The study corpus, consisting of 90 tasks administered between 1998 and 2020 in Celpe-Bras, two reference corpora, official documents and previous studies were used in the analysis. Using the 33 predictive metrics selected by the Boruta algorithm, a statistical model of complexity was created using 80% of the data from the reference corpora. The remaining 20% was used for testing, which differentiated between high and low complexity texts with 100% accuracy. The results showed that the content of the tasks does not correspond to the specifications, which lack details about the selection of input material and the development process. Furthermore, most of the input texts were classified as highly complex, corresponding to the highest level of Brazilian literacy, according to the National Index of Functional Literacy. The theoretical and empirical evidence provided by this study did not confirm the intended validity argument. Test-takers at Intermediate level may potentially be harmed by the inaccessibility of the input texts, as well as by the lack of diversity in the communicative and cultural repertoires of these materials. The results of this research should be complemented by studies of other instruments that make up the exam to fully understand the representativeness of the Celpe-Bras' language domain.

Keywords: Celpe-Bras; reading assessment; Corpus-informed language testing; text complexity; validity arguments.

LISTA DE FIGURAS

Figura 1 – Modelo de argumentação de Toulmin.....	44
Figura 2 – Possível estrutura de argumento sobre complexidade de tarefas	46
Figura 3 – Tarefa 3 da Parte Escrita de 2016-2	58
Figura 4 – Modelo de Compreensão Leitora RAND.....	59
Figura 5 – Modelo de Medida de Dificuldade Textual em Três Partes.....	60
Figura 6 – Descrição do modelo de Medida de Dificuldade Textual em Três Partes	61
Figura 7 – Relações entre leitor, texto e tarefa	62
Figura 8 – Três níveis de análise de complexidade	66
Figura 9 – Tarefa 4 da Parte Escrita de 2016-2	73
Figura 10 – Ocorrências de esferas de atuação nas tarefas 3 e 4 de 1998 a 2017	84
Figura 11 – Ocorrências de temáticas nas tarefas 3 e 4 de 1998 a 2017	85
Figura 12 – Enunciado da tarefa 3 da Parte Escrita de 2016-2.....	88
Figura 13 – Enunciado da tarefa 3 da Parte Escrita de 2014-2.....	89
Figura 14 – Ocorrências de esferas de atuação na PE de 1998 a 2017 em todas as tarefas	91
Figura 15 – Descritor da nota 2 do quadro de parâmetros de avaliação da Parte Escrita do Celpe-Bras	95
Figura 16 – Descritor do nível de proficiência Intermediário do Celpe-Bras	98
Figura 17 – Perfil dos consumidores de jornais por classe econômica	142
Figura 18 – Alfabetismo por níveis de escolaridade do Inaf.....	143
Figura 19 – Tarefas 3 e 4 da edição 2013-2 com 1 e 2 textos de insumo.....	161
Figura 20 – Edição 2014-2 com a capa do caderno de questões impressa colorida.....	162
Figura 21 – Etapas do algoritmo Boruta de seleção de variáveis.....	173
Figura 22 – Comparação de número de palavras em duas tarefas.....	179
Figura 23 – Tarefa 3 da edição 2012-1 em orientação retrato.....	181
Figura 24 – Recorrência de palavras na tarefa 4 da edição do Celpe-Bras 2016-2.....	188
Figura 25 – Gêneros de leitura por ano do corpus Celpe	195
Figura 26 – Fonte dos textos de leitura do corpus Celpe por ano	204
Figura 27 – Esfera de atuação da produção textual do corpus Celpe por ano.....	207
Figura 28 – Propósito da produção textual do corpus Celpe por ano.....	209
Figura 29 – Relação de interlocução da produção textual do corpus Celpe por ano.....	210
Figura 30 – Natureza da interlocução da produção textual do corpus Celpe por ano	211
Figura 31 – Gênero do discurso da produção textual do corpus Celpe por ano	212

Figura 32 – Suporte de produção textual do corpus Celpe por ano.....	213
Figura 33 – Tarefa 4 de 2013-1	217
Figura 34 – Tarefa 4 de 2014-1	219

LISTA DE GRÁFICOS

Gráfico 1 – Distribuição binomial de experimentos repetidos	174
Gráfico 2 – Distribuição de textos por tokens	177
Gráfico 3 – Média de tokens por corpora	183
Gráfico 4 – Distribuição de textos por tokens após exclusão de outliers	184
Gráfico 5 – Gráfico da média de TTR por corpora	186
Gráfico 6 – Número de edições em impressão colorida e monocromática	193
Gráfico 7 – Bootstrapping da distribuição de gêneros de leitura no corpus Celpe	196
Gráfico 8 – Número de palavras por gênero de leitura do corpus Celpe	198
Gráfico 9 – Número de suporte dos textos de leitura por ano do corpus Celpe	200
Gráfico 10 – Bootstrapping de distribuição de suportes de gênero de leitura do corpus Celpe	201
Gráfico 11 – Número de palavras por temática do corpus Celpe	208
Gráfico 12 – Efeito dos índices NILC-Metrix na probabilidade de ser PorSimples	228
Gráfico 13 – Número de tarefas do corpus de estudo por ano e nível de complexidade	215

LISTA DE QUADROS

Quadro 1 – Possíveis aplicações de corpora no aprimoramento de testes.....	35
Quadro 2 – Especificações para a elaboração de tarefas do Celpe-Bras - Parte Escrita	78
Quadro 3 – Resposta esperada para a tarefa “Shopping em casa”	81
Quadro 4 – Especificações dos textos de insumo da Parte Escrita do Celpe-Bras.....	90
Quadro 5 – Sub-habilidades de compreensão avaliadas na Parte Escrita do Celpe-Bras.....	93
Quadro 6 – Aplicações da Linguística de Corpus por área de estudos.....	101
Quadro 7 – Tipologia de corpora.....	101
Quadro 8 – Métricas de complexidade textual de trabalhos relacionados	117
Quadro 9 – Exemplo de transcrição do corpus Celpe	126
Quadro 10 – Exemplo de transcrição do corpus PorSimples	130
Quadro 11 – Exemplo de transcrição do corpus PorPopular.....	133
Quadro 12 – Escala de proficiência do Inaf	136
Quadro 13 – Texto de insumo da avaliação de leitura do Inaf	138
Quadro 14 – Headers do corpus Celpe	163
Quadro 15 – Categorias do grupo Avalia.	164
Quadro 16 – Métricas do NILC-Matrix excluídas das análises.....	167
Quadro 17 – Gêneros de leitura nas especificações e no corpus Celpe.....	197
Quadro 18 – Fontes dos textos de insumo nas especificações e no corpus Celpe.....	201
Quadro 19 – Métricas do NILC-Matrix aceitas pelo modelo Boruta	224
Quadro 20 – Descrição dos proxies de complexidade textual.....	226
Quadro 21 – Efeito dos índices NILC-Matrix na probabilidade de classificação	230
Quadro 22 – Níveis do Índice Flesch	233
Quadro 23 – Exemplo de análise de texto do corpus PorSimples.....	235
Quadro 24 – Texto de insumo da tarefa 4 de 2013-1	218
Quadro 25 – Texto de insumo da tarefa 4 de 2014-1	220

LISTA DE TABELAS

Tabela 1 – Composição inicial dos corpora.....	134
Tabela 2 – Distribuição da população por níveis de alfabetismo do Inaf 2018.....	139
Tabela 3 – Proposta de equiparação de estratos socioeconômicos entre ABEP e Inaf	140
Tabela 4 – Níveis de Alfabetismo Inaf por estrato socioeconômico	141
Tabela 5 – Composição inicial dos corpora.....	175
Tabela 6 – Medidas de tendência central e de dispersão de palavras dos três corpora	182
Tabela 7 – Resultados da regressão linear de número de tokens e corpora.....	183
Tabela 8 – Composição final dos corpora	185
Tabela 9 – Médias das notas finais das quatro tarefas das edições 2015-2, 2016-1, 2016-2 e 2017-1	190
Tabela 10 – Total de gêneros de leitura do corpus Celpe.....	194
Tabela 11 – Fonte dos textos de leitura do corpus Celpe	203
Tabela 12 – Tarefas do corpus de estudo classificadas por nível de complexidade.....	214

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

SUMÁRIO

Prefácio	17
1 INTRODUÇÃO	18
2 AVALIAÇÃO DE LÍNGUAS INFORMADA POR CORPUS	29
3 AVALIAÇÃO DE LÍNGUAS	40
3.1 AVALIAÇÃO DE LÍNGUAS E ARGUMENTOS DE VALIDADE	40
3.2 LEITURA E LETRAMENTO	52
3.3 LEITURA E COMPLEXIDADE TEXTUAL	59
4 O CELPE-BRAS	70
4.1 TAREFAS 3 E 4	83
4.2 AVALIAÇÃO DE LEITURA NA PARTE ESCRITA	92
4.3 NÍVEL DE PROFICIÊNCIA INTERMEDIÁRIO	97
5 LINGUÍSTICA DE CORPUS	100
5.1 COMPLEXIDADE TEXTUAL EM LÍNGUA INGLESA	105
5.2 COMPLEXIDADE TEXTUAL EM LÍNGUA PORTUGUESA	112
6 METODOLOGIA	123
6.1 OS CORPORA	124
6.2 INDICADOR DE ALFABETISMO FUNCIONAL	134
6.3 NILC-METRIX	144
6.3.1 Medidas Descritivas	147
6.3.2 Simplicidade Textual	148
6.3.3 Coesão Referencial	149
6.3.4 Coesão Semântica	150
6.3.5 Medidas Psicolinguísticas	151
6.3.6 Diversidade Lexical	152
6.3.7 Conectivos	153
6.3.8 Léxico Temporal	153
6.3.9 Complexidade Sintática	154
6.3.10 Densidade de Padrões Sintáticos	155
6.3.11 Informações Morfossintáticas de Palavras	155
6.3.12 Informações Semânticas de Palavras	156
6.3.13 Frequência de Palavras	157

6.3.14 Índices de Leiturabilidade	158
6.4 DESCRIÇÃO DAS TAREFAS	160
6.5 MEDIDAS DE COMPLEXIDADE TEXTUAL	165
7 RESULTADOS	175
7.1 DESCRIÇÃO DAS TAREFAS	176
7.1.1 Medidas de tendência central e dispersão	176
7.1.2 Composição das tarefas	191
7.1.3 Categorias do Grupo Avalia	206
7.2 PREDIÇÃO DE COMPLEXIDADE TEXTUAL	213
7.2.1 Métricas preditivas de complexidade textual	223
7.2.2 Complexidade textual no corpus de estudo	213
7.3 CONCLUSÕES ACERCA DO CONTEÚDO DAS TAREFAS E DA DIFICULDADE TEXTUAL	237
8 CONSIDERAÇÕES FINAIS	245
Posfácio	256
REFERÊNCIAS	257
ANEXO A – Especificações para a elaboração de tarefas do Exame Celpe-Bras - Parte Escrita	277
ANEXO B – Eixos de sub-habilidades das partes escrita e oral do Exame Celpe-Bras	283
ANEXO C – Quadro dos Parâmetros de avaliação da Parte Escrita	285
ANEXO D – Quadro de níveis de proficiência certificados pelo Exame Celpe-Bras	286
ANEXO E – As 200 métricas do NILC-Metrix	288

Prefácio

Era 1964. Ano de golpe de Estado no Brasil. No dia 15 de abril, após mais de 40 dias no mar, o navio Africa Maru atracou no porto de Santos. Dentre os mais de sessenta passageiros, um jovem japonês de 20 anos, Shigeru Nagasawa, que deixava para trás sua pátria ainda fragilizada pela guerra, chegava em novas terras em busca de um recomeço.

Ao desembarcar e ser questionado por soldados armados, Shigeru Nagasawa repetia as palavras anotadas em seu caderninho e que lhe serviam de escudo: “não sei português”. Nos primeiros meses, o trabalho na lavoura de tomate paulista lhe permitiu um tempo de aclimatação em ambiente acolhedor, ao lado de outros compatriotas. No entanto, a curiosidade e o desejo por novas experiências o impulsionaram a sair daquela zona de conforto. Com a mala pronta e confiante em seus instintos, seguiu rumo ao sul. Em Porto Alegre, capital mais meridional do país, conseguiu uma vaga temporária como repórter fotográfico em um novo jornal da cidade. Para garantir a posição na Zero Hora, era imprescindível naturalizar-se brasileiro.

Acompanhado de colegas jornalistas, compareceu à Polícia Federal para a obtenção de sua certidão de naturalização, apresentando apenas seus documentos de identificação e a promessa de admissão na empresa. A nacionalidade brasileira não apenas lhe oportunizou uma nova profissão, mas também lhe permitiu estabelecer raízes em Porto Alegre, um fato que geraria mudanças significativas em sua trajetória de vida. À medida que os anos passavam, a frase “não sei português” perdeu sua utilidade. Assim, agarrado em seu prolífero caderninho, Shigeru dedicou-se a aprender português e foi fazendo daquela língua alheia a sua própria língua. E ao fazer isso, fazia deste país estrangeiro o seu novo lar. Como repórter da Zero Hora, viajou por todo o Rio Grande do Sul fotografando, casou-se com uma gaúcha e tornou-se pai de três meninas: Tatiana, Mariana e Ellen.

Cinco décadas após a chegada de Shigeru, Ellen passou a compor um grupo de pesquisa do Instituto de Letras da Universidade Federal do Rio Grande do Sul, dedicando-se ao estudo do exame de proficiência de português para estrangeiros, uma ferramenta utilizada, dentre outros propósitos, como uma das possibilidades de se comprovar proficiência para fins de naturalização de imigrantes. Movida por uma conexão íntima com caderninhos rabiscados com palavras repletas de possibilidades e a compreensão de que a expressão “não sei português” pode encapsular uma multiplicidade de sentidos, Ellen se viu impulsionada a investigar esse instrumento de avaliação capaz de transformar vidas. Seus inúmeros questionamentos a conduziram à escrita de uma monografia de graduação, uma dissertação de mestrado e, finalmente, à elaboração desta tese de doutorado.

1 INTRODUÇÃO

O campo de estudos referido como *Corpus-informed language testing* é o resultado da interseção de duas especialidades da Linguística Aplicada – Avaliação de línguas (AL) e Linguística de *Corpus* (LC). Embora tenhamos procurado, não foram encontradas publicações brasileiras anteriores a esta tese que tenham nomeado esse ramo de pesquisas, portanto, é uma tradução nossa, o termo em língua portuguesa, “Avaliação de línguas informada por corpus”. Em língua inglesa, o termo estruturado dessa forma, ao que nos consta, parece ter sido registrado pela primeira vez no texto *Corpora and language assessment: trends and prospects*, de Fiona Barker, publicado no relatório *Research Notes* da comissão examinadora do *Cambridge Assessment English*, de 2006. O texto de Barker parece responder à pergunta feita uma década antes por Charles Alderson, no capítulo intitulado *Do corpora have a role in language assessment?*, publicado em 1996 no livro *festschrift, Using Corpora for Language Research. Studies in the Honour of Geoffrey Leech* (CUSHING, 2017).

Desde então, muitas pesquisas têm sido publicadas na área (BARKER, 2006, 2010; BARKER *et al.*, 2015; EGBERT, 2017; CUSHING, 2017; XI, 2017; BIJANKHAN; SHAYESTEFAR, 2023) e “parece evidente que a previsão de Alderson está agora confirmada; a Avaliação de línguas informada por corpus está aqui para ficar” (EGBERT, 2017, p. 556, tradução nossa)¹. As pesquisas nessa área são realizadas por linguistas especializados em AL interessados em trabalhar com métodos de análise possibilitados pela LC e por linguistas especializados em LC interessados em trabalhar com *corpora* relacionados a testes de línguas² (CUSHING, 2017). Egbert (2017), contudo, é bastante enfático em alertar para os riscos da falta de especialidade em ambas as áreas, indicando o perigo da falta de alto padrão de rigor metodológico, por um lado, e a interpretação de resultados que não se alinham com o construto teórico dos testes de língua, por outro. Esta tese de doutorado contou com a orientação de duas especialistas, uma em AL, Professora Doutora Margarete Schlatter, e outra em LC, Professora Doutora Adriana Picoral, possibilitando uma abordagem sinérgica³. Indubitavelmente que, sem

¹ Ao longo do texto, todas as traduções são de nossa autoria. No entanto, dado que a maioria do referencial teórico está em língua estrangeira, as citações não serão seguidas da expressão “tradução nossa”. Contudo, os textos originais serão incluídos em notas de rodapé. No original, lê-se: It seems evident that Alderson’s prediction is now confirmed; corpus-informed language testing is here to stay.

² Usaremos “instrumento de avaliação”, “teste” e “exame” como sinônimos, para nos referir a instrumentos utilizados para avaliar uma língua nomeada.

³ Devido ao seu caráter interdisciplinar, a metodologia empregada neste estudo se diferencia dos métodos convencionais da Linguística de *Corpus* no Brasil e pode-se ter a impressão de que se trata de um trabalho em Linguística Computacional. No entanto, é importante salientar que os algoritmos utilizados não foram desenvolvidos por nós, mas sim aplicados para atender às necessidades específicas de nossas questões de pesquisa.

o apoio das duas orientadoras, não haveríamos navegado o que Egbert assinala como as águas desconhecidas e desafiadoras dessa área de estudos, uma vez que não é tarefa simples “cruzar as fronteiras de dois campos bem situados, altamente técnicos e muito diferentes um do outro” (EGBERT, 2017, p. 563)⁴.

Os estudos conduzidos na área visam realizar análises computacionais de extensas coleções de textos, que abrangem tanto produções de usuários de uma determinada língua para quem ela foi a primeira língua de socialização quanto de produções para quem ela é uma língua adicional. Esses estudos são de grande auxílio para os desenvolvedores de testes de línguas, pois fornecem evidências que, aliadas às intuições e experiências dos elaboradores, são utilizadas para embasar decisões relacionadas ao construto teórico, elaboração de tarefas e grades de avaliação, balizamento e descrição de níveis de proficiência, além de contribuir para correções automáticas e semiautomáticas (BARKER, 2010; BARKER *et al.*, 2015; CUSHING, 2017; EGBERT, 2017; XI, 2017; GOTZ, 2022). Apesar de ainda pouco explorado por pesquisadores brasileiros, de acordo com Gotz (2022), o campo, de modo geral, tem experimentado progresso significativo nos últimos anos.

No campo da avaliação, cada vez mais, as pesquisas têm concentrado sua atenção na investigação das dimensões sociais dos testes, especialmente nas consequências resultantes do uso de seus resultados como ferramentas para classificação e exclusão de indivíduos na sociedade (MCNAMARA; ROEVER, 2006; SHOHAMY, 2017; YOUNG, 2021). Os testes são usados como filtros sociais que selecionam sujeitos para acessar (ou não) instituições educacionais, empresas e países. Além disso, esses testes são usados como mecanismos para a manutenção de línguas dominantes e perpetuação de ideologias de linguagem que legitimam línguas nomeadas⁵ como sistemas fixos relacionados a identidades e territórios, sendo considerados por Shohamy (2011, p. 426) como “um dos instrumentos mais poderosos nas sociedades modernas”.

⁴ No original, lê-se: crossing boundaries of two well-established fields that are highly technical and vastly different from one another.

⁵ O termo “língua” é utilizado neste trabalho para nomear línguas específicas, tais como a língua inglesa e a língua portuguesa. Já o termo “linguagem” será empregado em um sentido mais abrangente, relacionado às práticas sociais que se dão por meio das diversas formas de comunicação. Como será observado no decorrer da escrita, o Celpe-Bras, exame foco desta investigação, prioriza a segunda expressão, visto que “fundamenta-se em uma visão de uso da língua(gem) com propósitos sociais, construída social e localmente por seus participantes” (BRASIL, 2020, p. 28). Desse modo, ao abordarmos o exame e estudos correlatos, esta terminologia será utilizada. Quanto às traduções de nossa autoria, conforme o contexto, o termo “language” será traduzido ora como “língua”, quando se tratar de línguas nomeadas, ora como “linguagem”, quando refletir o uso mais amplo das práticas interacionais.

Essa preocupação social está estreitamente ligada à busca científica por argumentos que validem as inferências derivadas dos resultados dos testes, especialmente em contextos de testes de alto impacto. Fulcher e Owen explicam que

[e]m testes de alto impacto, a validade diz respeito a se e até que ponto nossas inferências sobre o significado da nota do teste são verdadeiras. Um argumento de validade apresenta evidências e base teórica que fundamentam a asserção que fazemos sobre o significado da nota. Os tipos de evidências que normalmente esperamos encontrar em um argumento podem incluir uma comparação do conteúdo do teste com os diferentes tipos de interações que poderíamos encontrar no mundo real (FULCHER; OWEN, 2016, p. 5).⁶

Dito de outra forma, a validade de um teste é um processo multifacetado, no qual várias evidências podem ser empregadas como argumentos para a validade dos usos pretendidos de seus resultados, considerando as implicações sociais que tais usos podem ter para os usuários. Após os elaboradores dos testes estabelecerem o argumento de interpretação e de uso que almejam para os resultados, as evidências de validade do exame são utilizadas para respaldar esse argumento (CHAPELLE; LEE, 2021).

Como ressaltado por Fulcher e Owen (2016), um argumento plausível de validade é baseado em evidências que demonstram a representatividade do conteúdo do teste em relação à língua-alvo que ele pretende avaliar. A elaboração de tarefas que abarque todas as áreas da língua-alvo é complexa e, por isso, “com muita frequência, o conteúdo dos testes é determinado pelo que é fácil testar e não pelo que é importante testar” (HUGHES; HUGHES, 2020, p. 30)⁷. Dada a complexidade da avaliação da representatividade do conteúdo dos testes, pesquisas que examinam o conteúdo das tarefas dos exames e o processo de sua elaboração são ainda escassas (SHIN, 2021; PLAKANS, 2021). De maneira semelhante, as discussões sobre o impacto da complexidade textual dos textos de insumo na avaliação de leitura são limitadas (SPENCER *et al.*, 2019). Plakans (2021) enfatiza o papel crítico da seleção de textos autênticos para compor tarefas que integram a avaliação da compreensão e produção textual, as quais devem ser pilotadas e testadas para aferir seu nível de complexidade textual. A autora defende o uso de ferramentas computacionais para uma análise mais aprofundada dos aspectos linguísticos, com vistas a fortalecer a validade do teste. Isso porque, por um lado, tarefas excessivamente fáceis

⁶ No original, lê-se: In high-stakes testing, validity is about whether and to what extent our inferences about the meaning of a test score are true. A validity argument sets out the evidence and theory to support the claim we make for score meaning. The kinds of evidence that we would typically expect to find in an argument may include a comparison of the test content with the kinds of communication we would find in the real world.

⁷ No original, lê-se: Too often the content of tests is determined by what is easy to test rather than what is important to test.

podem impossibilitar a distinção dos níveis de proficiência mais avançados entre os participantes (GOMES, 2009), o que indica uma sub-representação do construto do exame, ocorrendo quando as tarefas exigem um nível de proficiência inferior ao efetivamente necessário nas práticas de uso da língua-alvo externas ao exame (MCNAMARA; ROEVER, 2006; SCARAMUCCI, 2011). E, por outro lado, se o nível de complexidade textual ultrapassar o pretendido, os examinandos podem ser prejudicados. Sobre essa questão, Bachman já pontuava que

[...] se os testes de língua que usamos não fornecem informações confiáveis e se os usos que fazemos dos seus resultados não são corroborados por evidências plausíveis, então corremos o risco de tomar decisões incorretas e injustas que serão potencialmente prejudiciais para os indivíduos que precisamente desejamos beneficiar (BACHMAN, 2004, p. 3)⁸.

Análises computacionais de aspectos linguísticos presentes em uma grande quantidade de textos, como apontado por Plakans (2021), requerem um entendimento abrangente da coesão, coerência, análises lexicais, sintáticas, semânticas e psicolinguísticas da língua. Contudo, esse conhecimento deve estar associado à capacidade de fazer uso dos programas computacionais destinados a realizar tais análises. Isso envolve uma compreensão básica dos algoritmos utilizados e da interpretação das métricas, mas, mais crucialmente, exige a capacidade de fazer inferências qualitativas a partir dos resultados quantitativos obtidos. Trata-se, portanto, de investigações fundamentalmente quali-quantitativas.

As bases teórico-metodológicas da LC já estão substancialmente consolidadas por meio de estudos na língua inglesa. O marco inaugural desse campo remonta à publicação do *Brown Corpus*⁹, compilado em 1964, que é considerado o primeiro conjunto eletrônico de dados linguísticos textuais coletados especificamente para exploração por meio de ferramentas computacionais (BERBER SARDINHA, 2004; TOGNINI-BONELLI, 2010; TAGNIN, 2018). Linguistas de *corpus* trabalham, tipicamente, com textos autênticos em linguagem natural e, com o auxílio de ferramentas computacionais e métodos estatísticos avançados, buscam analisar a frequência de ocorrências linguísticas. Essas ocorrências fornecem evidências para a identificação e compreensão de padrões, viabilizando a descrição da língua em uso em diferentes contextos e por distintos usuários (GABLASOVA, 2020).

⁸ No original, lê-se: If the language tests we use do not provide reliable information, and if the uses we make of these test scores cannot be supported with credible evidence, then we risk making incorrect and unfair decisions that will be potentially harmful to the very individuals we hope to benefit.

⁹ *Brown University Standard Corpus of Present-Day American English*, compilado por Nelson Francis e Henry Kucera. Disponível em: <http://korpus.uib.no/icame/manuals/BROWN/INDEX.HTM>. Acesso em: 15 mar. 2023.

Os objetivos das pesquisas variam, abrangendo desde a identificação de estruturas linguísticas que discriminam gêneros discursivos para o ensino de línguas até a elaboração de dicionários. O tamanho dos *corpora*, ou seja, a quantidade de textos compilados para a análise, pode variar, bem como seu conteúdo, de acordo com os propósitos do pesquisador. Em geral, os *corpora* consistem em amostras da língua a ser investigada, e os textos são etiquetados com informações sobre a amostra (*e.g.*, local de coleta, data, nível de proficiência do usuário), em formato de *headers*, ou seja, informações descritas no cabeçalho dos arquivos, de acordo com o interesse do pesquisador. Os *corpora* podem conter amostras de língua em uso em diferentes modalidades e contextos, como escrita, oral, online, multimodal ou língua de sinais (GABLASOVA, 2020). Em Avaliação de línguas informada por corpus, o tipo de *corpus* mais utilizado é composto por respostas dos examinados a testes de larga escala, mas existem ainda *corpora* de usuários da língua como língua adicional e *corpora* compostos pelo domínio de uso da língua-alvo que o teste busca simular (EGBERT, 2017).

Considerando o exposto, para esta tese, selecionamos como *corpus* de estudo os textos de insumo das tarefas que avaliam de maneira integrada a leitura e a produção textual no Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras). Esses textos são denominados textos autênticos, sendo originalmente produzidos para circulação na sociedade brasileira e não especificamente para serem usados no teste. O Celpe-Bras tem como objetivo avaliar a proficiência em Português como Língua Adicional (PLA) e é o único exame oficialmente reconhecido pelo governo brasileiro¹⁰. Tradicionalmente, o exame é administrado duas vezes ao ano e atualmente conta com 130 postos aplicadores credenciados que estão localizados nas cinco regiões do Brasil e nos continentes americano, africano, europeu e asiático¹¹. Embora originalmente projetado, em 1993, para fins acadêmicos (SCHLATTER, 2014; SCARAMUCCI, 2016) e tendo como público-alvo estudantes estrangeiros não-lusófonos interessados em ingressar em Instituições de Ensino Superior (IES) brasileiras, o Documento base do exame Celpe-Bras (a publicação oficial mais recente que apresenta as especificações

¹⁰ Existem outros testes no Brasil para estrangeiros usados para comprovar proficiência em língua portuguesa, como, por exemplo, o Certificado Internacional da Língua Portuguesa, elaborado pela Universidade de Caxias do Sul, reconhecido pela União Latina e aplicado em postos credenciados no Brasil, Argentina, Paraguai, Portugal e Uruguai. Por outro lado, visando um público-alvo mais específico, composto por estudantes estrangeiros de programas de pós-graduação, muitas universidades elaboram e aplicam suas próprias provas de avaliação de proficiência em leitura. Além disso, conforme a Portaria nº 623, de 13 de novembro de 2020, IES credenciadas pelo MEC podem oferecer cursos de língua portuguesa na modalidade a distância com avaliação final presencial para certificação de proficiência que pode ser utilizada no processo de naturalização de estrangeiros.

¹¹ Na última edição (2023-1) antes do término desta tese, o exame foi aplicado em 48 postos no Brasil e 81 no exterior. <https://www.gov.br/inep/pt-br/assuntos/noticias/celpe-bras/encerrada-a-aplicacao-do-celpe-bras-2023>

do exame, dentre outras informações) (doravante Documento Base) destaca que o Celpe-Bras, atualmente, não é considerado um teste de língua para fins acadêmicos (BRASIL, 2020, p. 30).

Cabe destacar dois usos atuais que se afastam notavelmente da concepção original dos elaboradores em 1993: a exigência do certificado para estudantes lusófonos candidatos ao Programa de Estudantes-Convênio de Graduação (PEC-G)¹² e a utilização do certificado para o processo de naturalização de estrangeiros no Brasil¹³. Tosatti (2021) aponta que isso confirma o uso dos resultados do Celpe-Bras como um mecanismo de políticas linguísticas capaz de excluir indivíduos de determinadas esferas de atuação e de “criar espaços de poder” (TOSATTI, 2021, p. 49). Scaramucci e Diniz (2022) frisam que o uso do Celpe-Bras para o processo de naturalização não foi previsto durante a concepção inicial do exame. Consequentemente, seus instrumentos de avaliação não são adequados para esse contexto de avaliação e público-alvo. Os autores argumentam que o exame tem sido equivocadamente utilizado para esse fim e que as inferências feitas dos resultados do Celpe-Bras neste contexto se tornam não apenas inválidas, mas também prejudiciais aos migrantes, uma vez que o exame se configura “como um poderoso mecanismo de exclusão à nacionalidade brasileira” (SCARAMUCCI; DINIZ, 2022, p. 41).

O construto teórico de um exame diz respeito à conceitualização teórica que descreve a situação-alvo da língua que está sendo testada. O construto é operacionalizado por meio dos instrumentos de avaliação que viabilizam a simulação das situações de uso da língua-alvo em contexto controlado, permitindo a observação do desempenho dos candidatos. No caso do exame Celpe-Bras, o construto pode ser definido como o uso da língua portuguesa para desempenhar ações a fim de agir no mundo em práticas sócio-historicamente situadas. Ou seja,

[o] Celpe-Bras fundamenta-se em uma visão de uso da língua(gem) com propósitos sociais, construída social e localmente por seus participantes. Proficiência, nesse caso, implica ser capaz de engajar-se em diferentes situações de uso da língua portuguesa

¹² Programa do governo brasileiro instituído em 1965 que tem por objetivo oferecer a estudantes de países em desenvolvimento com os quais o Brasil mantém acordo educacional, cultural ou científico-tecnológico vagas em IES brasileiras para a realização de seus estudos de graduação gratuitamente. A partir do Decreto nº 7.948, de 12 de março de 2013, o certificado passa a ser exigido para todos os candidatos estrangeiros participantes do programa, inclusive estudantes lusófonos. A exigência do certificado do Celpe-Bras para estudantes participantes do PEC-G oriundos de Países Africanos de Língua Oficial Portuguesa (PALOP) tem sido amplamente problematizada por pesquisadores da área de PLA. Ver, por exemplo, Diniz e Bizon (2015) e Tosatti (2022).

¹³ O requerimento do certificado Celpe-Bras para o procedimento de naturalização ordinária foi estabelecido pela Portaria Interministerial nº 11, de 3 de maio de 2018, e posteriormente alterado pela Portaria Interministerial nº 16, de 3 de outubro de 2018, que estabelece o exame como uma das opções de comprovação de proficiência em língua portuguesa. Para uma discussão aprofundada desse uso do certificado Celpe-Bras, sugerimos a leitura do artigo de Scaramucci e Diniz (2022), no qual é realizada uma revisão histórica de instrumentos utilizados para a comprovação de proficiência para processos de naturalização no Brasil, bem como uma avaliação da validade do uso do Celpe-Bras nesse contexto.

no mundo, mostrando adequação às demandas dos vários contextos. Trata-se, em outras palavras, de um exame que avalia “o uso adequado da língua para desempenhar ações no mundo” (Brasil. MEC, 2006, p. 4). Ser proficiente, nessa perspectiva, significa ser capaz de construir enunciados adequados para participar de contextos de comunicação em diferentes esferas de uso da linguagem, segundo Schoffen (2009). Essa visão de proficiência leva em conta a prática de uso da língua(gem) como situada e contextualizada, e não a mera manipulação de formas gramaticais ou itens lexicais (BRASIL, 2020, p. 28).

Trata-se, portanto, de um exame de desempenho no uso da linguagem que, por meio de um único instrumento, visa avaliar distintos níveis de proficiência de forma integrada – abrangendo a compreensão e a produção textual (escrita e oral). Na Parte Oral (PO), a compreensão e a produção oral são avaliadas de modo integrado mediante uma interação face a face com o avaliador durante vinte minutos. A Parte Escrita (PE) é composta por quatro tarefas que apresentam um enunciado (comando da tarefa) e um texto de insumo. Na tarefa 1, o texto de insumo é um vídeo; na tarefa 2, um áudio; nas tarefas 3 e 4, textos escritos. Os textos de insumo utilizados no Celpe-Bras são, em geral, provenientes de emissoras de televisão, de rádio, periódicos e outras fontes midiáticas brasileiras. Por meio das quatro tarefas, realizadas em, no máximo, três horas, o examinando demonstra, de modo integrado, as habilidades de compreensão oral, leitura e produção escrita.

Testes de línguas apresentam especificações públicas que são consideradas centrais para a elaboração das tarefas. Essas especificações contêm descrições minuciosas dos objetivos de avaliação, habilidades e conteúdos avaliados, público-alvo, níveis de proficiência, apresentação dos instrumentos, critérios de avaliação e exemplos de questões (DOUGLAS, 2000; GREEN; FULCHER, 2020; JIN, 2021), bem como diretrizes que orientam a seleção de texto de insumo e os procedimentos de elaboração das tarefas (WEIGLE, 2004). No que concerne às especificações do Celpe-Bras, verifica-se a ausência de detalhamento sobre a elaboração das tarefas, constando publicamente no Documento Base (BRASIL, 2020, p. 58) apenas uma lista de gêneros discursivos e fontes que podem ser potencialmente utilizados como textos de insumo (*e.g.*, crônica, reportagem, entrevista; jornais, revistas, livros).

A trajetória desta pesquisa teve início com a compilação de um *corpus* contendo os textos de insumo de todas as tarefas da PE do Celpe-Bras já aplicadas. Esse *corpus* inclui transcrições dos vídeos, áudios e textos escritos, e *headers* com informações relativas a esses textos, tais como o gênero discursivo e a fonte do texto, além de outras características apontadas por pesquisas anteriores, tais como o gênero discursivo e o propósito comunicativo da produção textual solicitada como resposta (SCHOFFEN *et al.*, 2018). Da mesma forma, compilamos um *corpus* de referência composto pelos textos de insumo de todas as tarefas de redação do Exame

Nacional do Ensino Médio (Enem), acompanhadas por *headers* descritivos. Esse *corpus* foi utilizado em nossas análises preliminares, nas quais efetuamos comparações entre o conteúdo dos dois testes (NAGASAWA; SCHLATTER, 2021; NAGASAWA; PICORAL, 2022). Além dessas análises preliminares, nossa pesquisa incluiu a geração de palavras-chave do *corpus* de estudo em contraste a três *corpora* de referência. Os resultados dessa análise, no entanto, não revelaram informações novas que pesquisas anteriores já não tivessem identificado sobre os tópicos predominantes das tarefas do Celpe-Bras (ver, por exemplo, SCHOFFEN *et al.*, 2018; SCHOFFEN; MENDEL, 2018, para citar alguns).

No percurso da pesquisa, essas análises preliminares e o *corpus* de referência do Enem foram excluídos da versão final do estudo. Isso se deve ao fato de que, na prática, o volume de análises necessário para lidar com textos de natureza tão diversificada e o grau de profundidade das reflexões que essas análises demandariam excedem os limites de uma tese. Assim, como é típico em todo processo de pesquisa, decisões foram tomadas e exclusões foram feitas a fim de que esta tese conseguisse abranger as contribuições que julgamos mais relevantes para este momento histórico. Nesse sentido, optamos por selecionar somente os textos escritos utilizados como insumo das tarefas 3 e 4 da PE do Celpe-Bras, para analisar o conteúdo dessas tarefas e o nível de complexidade desses textos.

Por compreender que ainda não há, na literatura sobre a língua portuguesa do Brasil, noções precisas do que é a complexidade textual, neste trabalho, ela é entendida como um construto multifacetado e multidimensional (BIBER *et al.*, 2021)¹⁴. A fim de empreender as análises propostas, a complexidade textual neste estudo é operacionalizada por meio de 200 métricas relativas a fatores lexicais, sintáticos e discursivos extraídas automaticamente por meio da ferramenta NILC-Metrix¹⁵. As definições que associam essas métricas à complexidade textual foram estabelecidas pelos próprios criadores da ferramenta, e nosso propósito é interpretar os resultados de forma a permitir as inferências necessárias para atingir os objetivos desta tese¹⁶. Lançada publicamente em 2021, essa ferramenta representa o resultado de mais de uma década de trabalho realizado pelo grupo de pesquisa Núcleo Interinstitucional de

¹⁴ No campo da Linguística de *Corpus* voltado para a análise da língua inglesa, diferentes abordagens para a análise e descrição da complexidade linguística coexistem, algumas vezes com divergências entre si em relação aos traços linguísticos utilizados para defini-la. Aqui oferecemos um panorama das pesquisas no campo, sem nos afiliarmos a uma perspectiva teórica específica, o que seria prematuro no nosso caso, considerando a fase ainda inicial dos estudos sobre a língua portuguesa. Por essa mesma razão, não pretendemos aqui definir um construto de complexidade textual da língua portuguesa, mas sim discutir medidas que podem ser relevantes para essa definição.

¹⁵ <http://fw.nilc.icmc.usp.br:23380/nilcmatrix>

¹⁶ Cabe a ressalva de que algumas métricas do NILC-Metrix podem apresentar definições de complexidade textual que divergem das encontradas em pesquisas mencionadas na revisão bibliográfica.

Linguística Computacional (NILC), contando com a colaboração de pesquisadores da Universidade de São Paulo e da Universidade Federal de São Carlos¹⁷. A ferramenta retorna automaticamente um valor numérico para cada uma das 200 métricas, todas elas relativas ao texto fornecido para a análise, sendo que “o principal objetivo das métricas é calcular coesão, coerência e nível de complexidade textual” (LEAL *et al.*, 2021, p. 7)¹⁸.

Adotando uma abordagem orientada pelo uso, ou seja, partindo da observação das frequências de padrões presentes na língua natural, e não apenas de modelos teóricos, para a operacionalização do nível de complexidade textual, são levadas em consideração as medidas fornecidas pelo NILC-Metrix em relação a dois *corpora* de níveis de complexidade distintos, conforme evidenciado por pesquisas anteriores que consideraram o nível de leitura brasileiro como padrão de referência, de acordo com o Indicador de Alfabetismo Funcional. Para representar a complexidade textual alta, são utilizadas as métricas correlacionadas ao *corpus* PorSimples como *proxy* (indicador ou parâmetro substituto), que é composto por textos da esfera jornalística categorizados pelos autores como textos complexos (LEAL; DURAN; ALUÍSIO, 2018). Já como indicador de complexidade textual baixa, as métricas estão correlacionadas a um *subcorpus* do *corpus* PorPopular, composto por textos da esfera jornalística classificados por seus autores como textos que fazem uso de linguagem simples (FINATTO, 2012). Os *corpora* usados como *proxy* foram selecionados como referência para o nosso *corpus* de estudo, visto que as tarefas da PE também incorporam textos da mídia brasileira. Todas as análises foram realizadas por meio da linguagem de programação R, utilizando seu ambiente de desenvolvimento integrado (*Integrated Development Environment*) RStudio¹⁹. É importante enfatizar que as análises conduzidas neste estudo possuem um caráter unicamente descritivo e exploratório, não tendo, assim, natureza inferencial, ainda que alguns testes estatísticos tenham sido realizados com o objetivo de aprofundar a compreensão das características observadas.

Apesar de estudos da língua inglesa sinalizarem que o uso de métricas automáticas na análise de complexidade textual seja limitado (CROSSLEY *et al.*, 2019; CUNNINGHAM; MESMER, 2014), recomendando, em vez disso, o uso de ferramentas de Processamento de

¹⁷ <http://www.nilc.icmc.usp.br/nilc/index.php>

¹⁸ No original, lê-se: The main objective of these metrics is to provide proxies to assess cohesion, coherence and textual complexity.

¹⁹ Versão RStudio 2022.07.1+554 “Spotted Wakerobin” Release (7872775ebddc40635780ca1ed238934c3345c5de, 2022-07-22) for Windows Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) QtWebEngine/5.12.8 Chrome/69.0.3497.128 Safari/537.36

Linguagem Natural e técnicas de Aprendizagem de máquina para a análise de *corpora* (BALYAN *et al.*, 2017, 2018; CROSSLEY *et al.*, 2017; JI *et al.*, 2021), a situação das pesquisas em língua portuguesa se mostra distinta. Somente há dois anos é que se tornou público o acesso à ferramenta NILC-Metrix, com as 200 métricas. Entendemos que, a partir dessas métricas, muitos estudos brasileiros acerca da complexidade textual surgirão. Por isso, consideramos pertinente, neste momento, testar a validade dessa nova ferramenta através de métodos estatísticos de seleção de métricas que melhor explicam a variação de complexidade textual entre os textos jornalísticos dos nossos *corpora*. Investigações dessa natureza vêm sendo realizadas em língua inglesa desde 2004, ano em que a ferramenta Coh-Metrix 1.0 (com 200 métricas), na qual o NILC-Metrix é baseado, foi concebida (GRAESSER *et al.*, 2004). Entretanto, não temos conhecimento de estudos publicados que tenham se debruçado sobre as 200 métricas do NILC-Metrix e sua validade quanto aos resultados sobre complexidade textual, exceto pelos trabalhos desenvolvidos pelos próprios criadores, membros do grupo de pesquisa NILC.

Considerando o exposto, o principal objetivo deste estudo consiste em fornecer evidências para verificar o argumento de validade relacionado ao domínio de uso da linguagem do Celpe-Bras e a capacidade preditiva de complexidade textual do NILC-Metrix por meio da descrição do conteúdo e da complexidade dos textos de insumo das tarefas que integram leitura e produção escrita da PE do exame. Ao fazer isso, pretendemos contribuir tanto para a área de AL quanto para a LC, incentivando, por conseguinte, o avanço de estudos no campo de Avaliação de línguas informada por corpus no Brasil.

Com base no exposto, formulamos as seguintes perguntas de pesquisa:

1. Em que medida os padrões de características observados nas tarefas que integram leitura e escrita na PE do Celpe-Bras estão em consonância com as especificações publicadas no Documento Base?
2. Quais métricas do NILC-Metrix apresentam capacidade preditiva de complexidade textual em textos da esfera jornalística?
3. De acordo com as métricas do NILC-Metrix e os *corpora* de referência, qual é o nível de complexidade textual observado nos textos de insumo das tarefas que integram leitura e escrita na PE do Celpe-Bras, em relação ao nível de leitura brasileiro, de acordo com o Indicador de Alfabetismo Funcional?

Para conduzir esta pesquisa, conjugamos as abordagens metodológicas empregadas em pesquisas da língua inglesa com as contribuições de estudos prévios sobre a língua portuguesa. Isso nos permitiu desenvolver um desenho metodológico de natureza descritiva e abordagem empírica indutiva. Ou seja, ao invés de partir de variáveis aprioristicamente estabelecidas como medidas de complexidade textual com base na literatura existente, optamos por identificar medidas produtivas por meio de nossas próprias análises.

Esta tese está estruturada em sete capítulos. No capítulo 2, traçamos um panorama dos estudos localizados na esfera interdisciplinar da Avaliação de línguas informada por corpus e destacamos a carência de pesquisas dessa natureza no Brasil. O capítulo 3 aborda a área de Avaliação de línguas, focando nos conceitos de validade, leitura, letramento e complexidade textual. O capítulo 4 aprofunda o objeto de estudo, o exame Celpe-Bras, com foco nas tarefas que integram leitura e produção textual na PE, na avaliação de leitura e na descrição do nível de proficiência intermediário. O capítulo 5 explora o campo da Linguística de *Corpus*, concentrando-se na análise da complexidade textual. O capítulo 6 integra novamente as duas áreas, AL e LC, para apresentar nossa metodologia, apresentando nossa base teórico-metodológica, descrevendo os *corpora* e os índices e critérios de análise adotados nas investigações realizadas. Por fim, no capítulo 7 apresentamos os resultados obtidos como produto desse olhar interdisciplinar entre as duas subáreas da Linguística Aplicada, relacionando-os aos pressupostos teóricos discutidos nos capítulos 2, 3 e 4. Nas Considerações Finais, discorreremos sobre as questões ideológicas, sociais e éticas decorrentes das decisões sobre o desenho de testes de leitura e das informações divulgadas para os usuários, sobretudo no exame Celpe-Bras. Também destacamos possíveis contribuições e limitações desta pesquisa e perspectivas futuras.

2 AVALIAÇÃO DE LÍNGUAS INFORMADA POR CORPUS

Conforme introduzido no início deste trabalho, nosso foco se concentra na área de pesquisa interdisciplinar que abrange a Avaliação de línguas e a Linguística de *Corpus*, designada como Avaliação de línguas informada por corpus. Esse campo de pesquisa é amplamente caracterizado por estudiosos como uma zona de interseção, convergência e integração entre essas duas especialidades da Linguística Aplicada. Desde a década de 1990, têm surgido trabalhos que estabelecem conexões entre essas duas esferas, sendo publicados em livros de referência em Linguística Aplicada, como as coletâneas *The Routledge Handbooks* (O'KEEFFE; MCCARTHY, 2010; HALL, 2016; WINKE; BRUNFAUT, 2020; FULCHER; HARDING, 2021; JABLONKAI; CSOMAY, 2022), bem como em periódicos científicos proeminentes, como *Language Testing*, *Language Assessment Quarterly* e *Assessing Writing* (CUSHING, 2021).

Os estudos conduzidos nesse campo têm como objetivo fundamental aprimorar os testes de línguas, notadamente os exames de proficiência em línguas adicionais aplicados em larga escala e de alto impacto social. Para tal, eles se valem da análise de *corpora* e da utilização de ferramentas computacionais visando à investigação refinada de fenômenos linguístico-discursivos das línguas. Como explicitado previamente, essas evidências desempenham um papel crucial nas tomadas de decisões relativas ao desenvolvimento ou refinamento do construto teórico, ao desenho e elaboração de tarefas avaliativas, à criação e refinamento de rubricas de avaliação, ao balizamento e descrição de níveis de proficiência, bem como no processo de avaliação e correção de respostas. Nas palavras de Barker e colaboradores:

Uma razão convincente para usar dados de corpus para testes de línguas é que os corpora fornecem evidências empíricas que equilibram as intuições e a expertise de professores e elaboradores de testes no desenho de instrumentos de avaliação e nas correções. Além disso, os corpora fornecem uma amostra da língua em uso em uma variedade de contextos, aprimorando, desse modo, o que o teste se propõe a medir (BARKER *et al.*, 2015, p. 512)²⁰.

Cushing (2017) aponta que os estudos realizados nesse campo possuem alta relevância ao fornecerem evidências que permitem sustentar inferências acerca da descrição da língua que está sendo avaliada, contribuindo para a validação de testes. A análise comparativa entre dois

²⁰ No original, lê-se: a persuasive reason to use corpus data for language testing is that corpora provide empirical evidence that balances teachers' and test writers' intuitions and expertise in designing and rating assessments. Furthermore, corpora provide real language across a range of settings, thus positively enhancing what the test purports to measure.

ou mais *corpora*, possibilitada pelo uso de ferramentas computacionais, por exemplo, viabiliza a identificação de características linguísticas em nível lexical, de unidades de palavras, de estruturas sintáticas e discursivas, dados que informam sobre os contextos de uso da língua, bem como sobre os gêneros discursivos e os interlocutores. No que se refere ao processo de elaboração de tarefas, a análise de *corpora* contribui para que o conteúdo utilizado como insumo possa refletir, de maneira mais autêntica e consistente, a língua-alvo. No que se refere ao Celpe-Bras, tais análises são especialmente importantes, uma vez que todas as tarefas utilizam textos autênticos. Além disso, a exploração de *corpora* da língua-alvo auxilia nas tomadas de decisões relativas a características linguísticas que podem influenciar a dificuldade de leitura dos textos de insumo para o público-alvo (CUSHING, 2017). Ou seja, a abordagem permite verificar se as características das tarefas do teste são condizentes com as práticas da língua-alvo, visando assegurar que o teste esteja avaliando o que se propõe. Considerando que a língua em uso está em constante mudança, o refinamento contínuo do construto teórico e de sua operacionalização, com base em evidências empíricas e atualizadas, pode aumentar a validade do teste (CUSHING, 2017). Seguindo essa mesma linha, Egbert (2017) indica que as análises de *corpora* auxiliam especialmente nas decisões relacionadas à seleção de textos de insumo para a composição de tarefas dos testes de língua.

No entanto, para viabilizar tais análises, é fundamental realizar escolhas adequadas quanto ao tipo de *corpora* a ser analisado e ao desenho metodológico. Barker *et al.* (2015) chamam a atenção para a constituição desses *corpora*, os quais devem refletir a mesma modalidade da língua, seja ela oral ou escrita, os mesmos gêneros, formatos e até mesmo os tópicos tratados. Essas variáveis, caso não controladas, têm o potencial de impactar os resultados das análises. Egbert (2017) amplia esse raciocínio ao apontar que, mesmo dentro da mesma modalidade ou registro, podem existir distinções. Assim uma avaliação criteriosa das situações comunicativas a serem compiladas é essencial, considerando que, como afirma o autor, “um corpus inadequado é um corpus inadequado não importa quão conveniente seja seu acesso ou uso” (EGBERT, 2017, p. 559)²¹. O autor destaca que princípios sólidos e métodos de amostragem bem fundamentados na compilação dos *corpora* são cruciais, pois deles dependem as possíveis inferências e generalizações de resultados. A definição da população, o método de seleção da amostra e a avaliação da representatividade delineiam os limites do que o pesquisador poderá afirmar ou não. Assim, é imperativo que os *corpora* sejam elaborados por especialistas nos padrões comunicativos das comunidades que fazem uso da língua-alvo, com

²¹ No original, lê-se: the wrong corpus is the wrong corpus no matter how convenient it is to access and use.

o propósito de representar fielmente a língua objeto de estudo (KILGARRIFF; GREFFENSTETTE, 2003; SINCLAIR, 2004). Nessa lógica, Barker adverte que “é importante observar que cada corpus é projetado com parâmetros e propósitos diferentes que restringem o uso que pode ser feito dele” (BARKER, 2006, p. 3)²². Os teóricos enfatizam essa questão à medida que muitos *corpora* estão prontamente acessíveis online, o que pode levar a usos inadequados de *corpora* preexistentes que não atendem plenamente às exigências de outras investigações. Muitas vezes, tais usos inadequados se dão porque os *corpora* prontos proporcionam uma alternativa conveniente para pesquisadores, evitando o investimento de tempo na compilação de novos *corpora*. Assim, entende-se que os *corpora* devem ser construídos de acordo com os objetivos dos pesquisadores, que devem compilar seus próprios *corpora* de forma meticulosa ou selecionar *corpora* disponíveis, desde que esses estejam alinhados com seus objetivos de pesquisa. Afinal, como destaca Gablasova (2020), “um corpus adequado está no centro de qualquer análise significativa e elucidativa informada por corpus” (GABLASOVA, 2020, p. 47)²³.

No contexto da Avaliação de línguas informada por corpus, Egbert (2017) destaca três categorias de *corpora* frequentemente utilizados nas pesquisas. O primeiro tipo é composto por textos coletados de usuários de uma língua como língua adicional. O segundo tipo é constituído por textos produzidos por examinandos que se submeteram a testes dessa língua; e o último tipo de *corpus* é composto pelo domínio de uso da língua-alvo que o teste busca simular²⁴.

Em relação aos *corpora* formados por textos de usuários da língua como língua adicional, pode-se encontrar diferentes graus de naturalidade na produção dos textos coletados. Esses níveis se referem à medida em que os textos refletem a linguagem natural produzida espontaneamente. Citamos três exemplos com progressão no grau de naturalidade: a produção linguística em interações espontâneas, que ocorre sem intervenção de avaliadores, professores ou pesquisadores (não temos informação sobre a existência desse tipo de *corpus* de PLA); a produção linguística em contexto educacional, que é relativamente controlada e, portanto, menos espontânea, uma vez que os textos respondem a tarefas solicitadas por professores ou pesquisadores com propósitos específicos (*e.g.*, o *MACAWS* com mais de 2 mil textos escritos²⁵); e a produção linguística em situação de teste, que é altamente controlada, resultando

²² No original, lê-se: it is important to note that every corpus is designed along differing parameters and purposes which constrain what it can be used for.

²³ No original, lê-se: a suitable corpus lies at the heart of any meaningful and informative corpus-based analysis.

²⁴ No original: Learner language, test responses, and target language use domain.

²⁵ <https://macaws.corporaproject.org/>

em textos menos naturais devido à necessidade de cumprir critérios predefinidos de avaliação (e.g., o *Corpus Celpe-Bras* com mais de 70 mil textos escritos²⁶). É essencial considerar essas diferenças ao escolher os *corpora* de estudo e de referência, visto que o nível de comparabilidade dos dados pode ser impactado pelo grau de naturalidade da produção. Em outras palavras, é preciso certificar se o desenho da pesquisa inclui distintos *corpora* por erroneamente considerá-los comparáveis ou se o objetivo é precisamente observar diferenças entre eles. Nesse tipo de pesquisa, “para que os resultados sejam significativos, os *corpora* a serem comparados devem ser semelhantes em todos os aspectos principais e diferir apenas em relação à variável cujo efeito desejamos observar”²⁷ (GABLASOVA, 2020, p. 48).

Na área de Avaliação de línguas informada por corpus, o tipo de *corpus* mais utilizado, de acordo com Egbert (2017), é aquele composto por respostas de examinandos a testes de larga escala, incluindo respostas tanto orais quanto escritas. O autor cita que grandes organizações examinadoras, como a *Educational Testing Service* (ETS) e a *Cambridge Michigan Language Assessments*, têm desempenhado um papel importante ao incentivar pesquisas nesse campo, frequentemente concedendo bolsas para pesquisadores. Isso têm contribuído para aumentar o interesse em investigações sobre textos produzidos pelos examinandos em seus testes.

No caso da ETS, responsável pelo exame *Test of English as Foreign Language* (TOEFL), a organização possui centros de pesquisa com especialistas internacionais em avaliação. Nesses centros de pesquisa são conduzidos estudos constantes e rigorosos, cujos resultados são publicados em relatórios de acesso público, visando assegurar a transparência e a reputação do exame como um teste válido de proficiência de inglês em nível acadêmico. Os estudos realizados pela ETS abrangem várias áreas da avaliação, por exemplo, investigações relacionadas a evidências que contribuem para validar o argumento de interpretação e uso dos resultados do exame (ETS, 2011); características dos tipos de tarefas (BIBER; GRAY, 2013); percepção de examinandos sobre o teste (STRICKER; ATTALI, 2010); efeito retroativo no ensino da língua inglesa na Europa central e oriental (WALL.; HORÁK, 2011); impacto de conhecimento prévio na avaliação de leitura (HILL; LIU, 2012); e desenvolvimento de ferramentas de avaliação automática (XI *et al.*, 2008).

Além disso, Gablasova (2020) afirma que as comissões examinadoras *British Council*, *Cambridge Assessment English*, *Pearson* e *Trinity College London* também adotam práticas similares, encorajando pesquisas baseadas em seus *corpora*. Cabe ressaltar que, em contraste

²⁶ Para uma descrição detalhada do *corpus*, ver Schlatter, Nunes e Kunrath (2021).

²⁷ No original, lê-se: in order for the results to be meaningful, the *corpora* to be compared should be similar in all major aspects, and differ only with respect to the variable whose effect we wish to observe.

com essas práticas, o órgão responsável pela elaboração do Celpe-Bras no Brasil não incentiva investigações nesse sentido e tem restrições rígidas quanto ao acesso de pesquisadores independentes interessados em analisar as respostas dos examinandos ao exame (VICENTINI, 2022). Devido a isso, as pesquisas que utilizam *corpus* de respostas de examinandos do Celpe-Bras ainda são incipientes.

Quanto ao trabalho com *corpora* de produções de examinandos, Xi (2017) ressalta a importância de considerar cuidadosamente as análises que comparam diferentes níveis de proficiência. Nos níveis mais baixos, pode ocorrer que examinandos não compreendam o comando da tarefa, resultando em textos que falham em representar o gênero do discurso cujas características foram previstas no enunciado. Isso, por sua vez, afeta diretamente a amostragem e as interpretações possíveis de serem feitas a partir dela. Assim sendo, os textos não certificados não devem ser aprioristicamente classificados como pertencentes ao gênero discursivo solicitado na tarefa no momento da amostragem (XI, 2017).

Outro tipo de *corpus* amplamente utilizado nas pesquisas que informam testes de línguas consiste em textos compilados da língua-alvo, provenientes de contextos naturalísticos e produzidos por usuários da língua como primeira língua de socialização, representando a língua que o exame busca avaliar. Esses *corpora* costumam ser bastante extensos, pois têm o propósito de representar o uso da linguagem de modo amplo e, para tanto, incluem amostras de diferentes gêneros e esferas de comunicação (GABLASOVA, 2020). O conteúdo do *corpus* pode variar desde mais geral (e.g., *Corpus Brasileiro* com 1057.7 milhões de palavras²⁸) até o mais especializado (e.g., *Corpus Técnico-Científico de Linguística* com mais de um milhão de palavras²⁹).

Ao comparar os *corpora* da língua-alvo com as respostas dos examinandos, é crucial compreender que o teste busca simular o uso da linguagem de maneira próxima aos usos feitos nas práticas sociais da língua-alvo para que o examinando possa demonstrar sua proficiência em um ambiente controlado, enquanto seu desempenho é observado. Com isso, entende-se que não é possível realizar uma comparação sob a suposição de que os examinandos devam reproduzir fielmente as práticas da língua-alvo, utilizando exatamente os mesmos recursos linguísticos, uma vez que, como já vimos, os contextos possuem níveis de naturalidade distintos (XI, 2017). Isso posto, entendemos que as escolhas dos fenômenos linguísticos a serem considerados nas análises de produções de examinandos devem refletir não as características

²⁸ <https://www.linguateca.pt/aceso/corpus.php?corpus=CBRAS>

²⁹ <https://comet.fflch.usp.br/projeto#cortec>

gerais da língua-alvo, mas as características explicitadas como relevantes de acordo com o construto teórico do teste, o qual projeta e delimita os usos da linguagem no contexto controlado do exame; no caso do Celpe-Bras, por meio do enunciado da tarefa.

Em relação aos fenômenos linguísticos a serem observados nos *corpora*, os pesquisadores também alertam para a tentação que o acesso fácil a ferramentas computacionais e índices automáticos pode representar para análises de padrões linguísticos (CUSHING, 2017; EGBERT, 2017; XI, 2017). Diante desse cenário, a tomada de decisões embasadas em princípios éticos e integradas a julgamentos de especialistas é crucial para evitar a proliferação de pesquisas focadas em variáveis linguísticas que careçam de significado ou que sejam inúteis aos agentes envolvidos com os testes. Dentro dessa lógica, Egbert argumenta que “só porque algo pode ser medido, não significa necessariamente que deva ser medido” e que “encontrar uma correlação e encontrar uma variável significativa não são a mesma coisa” (EGBERT, 2017, p. 561). Dito de outro modo, é fundamental que o pesquisador seja capaz de fazer interpretações confiáveis dos resultados das análises de seus dados, levando em consideração seu conhecimento acerca do construto do exame em questão.

Seguindo uma linha de raciocínio semelhante, Xi (2017) reconhece que as medidas linguísticas observáveis por meio de ferramentas da LC fornecem uma perspectiva valiosa para definir as sub-habilidades avaliadas e compreender a natureza da linguagem usada em diferentes esferas de atuação, contribuindo, assim, para a validação dos testes. Entretanto, assim como Egbert (2017), Xi (2017) salienta que a análise das medidas extraídas e analisadas por programas computacionais deve estar alinhada com o construto teórico do exame. No caso de avaliações holísticas, nas quais são priorizados critérios voltados para o cumprimento dos propósitos comunicativos de maneira geral, que refletem as características discursivas do desempenho na interação, as medidas linguísticas superficiais devem ser usadas cautelosamente, levando-se em consideração em que medida suas ocorrências estão relacionadas com o cumprimento da interação comunicativa. Em suma, como o próprio nome da área sugere, o *corpus* deve servir para informar as decisões humanas, não para determiná-las. Ainda que atenta a isso, Xi (2017) indica que a análise comparativa da complexidade textual ou outras características linguísticas com *corpora* da língua-alvo “fornece uma base forte para o uso de textos de insumo para leitura e compreensão oral que estejam no nível apropriado de complexidade e tenham as características linguísticas almejadas”³⁰ (XI, 2017, p. 573).

³⁰ No original, lê-se: provides a strong basis for using reading and listening stimulus materials that are at the right complexity level and have the desired linguistic characteristics.

No que diz respeito às aplicações de *corpora* para análises com vistas ao aprimoramento de testes, exemplificamos, no quadro a seguir, algumas possibilidades que os estudos dessa área têm explorado.

Quadro 1 – Possíveis aplicações de corpora no aprimoramento de testes

Tipo de <i>corpora</i>	Possível aplicação	Possível análise
Respostas de examinandos	Informar as grades de avaliação analíticas	Análises comparativas de textos avaliados por humanos para identificação de padrões que diferem desempenhos
	Informar o corte de níveis de proficiência	Análises comparativas de textos dos diferentes níveis, avaliados por humanos, para identificação de padrões que os diferem
	Informar a descrição de níveis de proficiência	Análises de textos avaliados por humanos nos diferentes níveis para a descrição de padrões observados
	Informar os ajustes de grades holísticas singulares para cada tarefa	Análises de textos de amostragem aleatória representativa da população, avaliados por humanos, para identificação de padrões observados
	Informar algoritmos de correção automática ou semiautomática	Análises de textos de amostragem aleatória representativa da população, avaliados por humanos, para identificação de padrões observados
	Informar algoritmos de identificação de plágio/cópia do texto de insumo	Análises de textos de amostragem aleatória representativa da população, avaliados por humanos, em comparação com os textos de insumo, para identificação de padrões observados
	Informar algoritmos de geração automática de feedback	Análises de textos avaliados por humanos nos diferentes níveis para a descrição de diagnósticos individuais
	Informar argumento de validade de inferência de generalização	Análises comparativas de textos dos diferentes níveis, avaliados por humanos, com textos da língua-alvo para identificação de padrões que se aproximam ou se distanciam do desempenho-alvo previsto
Produção de usuários da língua como língua adicional	Informar a descrição de níveis de proficiência	Análises de textos produzidos em pilotagens (ou outros contextos externos ao exame), avaliados por humanos, para a descrição de padrões que diferem desempenhos
	Informar descrição de variáveis externas	Análises de textos produzidos em pilotagens (ou outros contextos externos ao exame), avaliados por humanos, com controle de diferentes variáveis não previstas, para a descrição de padrões que diferem desempenhos
	Informar a escolha de textos de insumo e a elaboração de tarefas	Análises de textos produzidos em pilotagens (ou outros contextos externos ao exame), avaliados por humanos, com controle de características das tarefas, para a descrição de padrões que diferem desempenhos
Produção de usuários da	Informar o construto teórico	Análises de textos produzidos nos contextos-alvo para a descrição de padrões observados

língua como primeira língua de socialização	Informar especificações de elaboração de tarefas integradas	Análises de textos produzidos nos contextos-alvo para a identificação de padrões relevantes para provocar a produção textual desejada
	Informar especificações de escolha de textos de insumo para tarefas integradas	Análises de textos da língua-alvo para identificação de padrões relacionados às características de textos adequados ao público-alvo

Fonte: Elaborado pela autora.

Com o intuito de investigar pesquisas que utilizam esses tipos de *corpora* para aprimorar testes de PLA no Brasil, realizamos uma busca no *Portal de Periódicos CAPES*³¹, uma biblioteca virtual que oferece serviços de busca e acesso a publicações acadêmicas. Para essa pesquisa, utilizamos os operadores booleanos e inicialmente as seguintes palavras-chave: “Avaliação de línguas informada por corpus” OU “*corpus-informed language testing*”. No entanto, não encontramos nenhum registro relevante.

Posteriormente, realizamos a seguinte busca: “linguística de *corpus*” OU “*corpus*” E “avaliação” OU “teste” OU “exame” E “língua” OU “linguagem” E “português” OU “língua portuguesa” E “validade”³². O sistema de busca nos retornou sete resultados, mas apenas um deles abordava o nosso tópico de interesse. No artigo intitulado “Um teste de verificação lexical de português brasileiro (TVLPB) como língua adicional: criação e validação”, Sá e colaboradores (2021) relatam a elaboração e validação de um teste de base psicométrica que avalia conhecimento lexical da língua portuguesa em 11 minutos.

Os autores trabalharam com três *corpora* compostos por produções de usuários da língua como primeira língua de socialização (*ptTenTen*³³, *Corpus Brasileiro*³⁴ e *C-Oral*³⁵) para identificar as palavras mais frequentes a serem incluídas em um teste de múltipla-escolha, no qual o examinando deve combinar um significado ou sinônimo ao léxico correto entre as opções fornecidas. Foram selecionados 6 adjetivos, 12 substantivos e 6 verbos para compor as respostas corretas, além de outras palavras frequentes como distratoras. Os autores utilizaram significados e sinônimos do dicionário online *Priberam*³⁶ e pilotaram o teste com 115 participantes usuários de português como primeira língua de socialização, com escolaridade variando entre ensino médio completo e pós-graduação completa, além de 23 usuários de

³¹ <https://www.periodicos.capes.gov.br>

³² Pesquisa realizada em 31 de maio de 2023.

³³ <https://www.sketchengine.eu/pttnten-portuguese-corpus/>

³⁴ <https://www.linguateca.pt/acesso/corpus.php?corpus=CBRAS>

³⁵ <https://www.c-oral-brasil.org/>

³⁶ <https://dicionario.priberam.org/>.

português como língua de herança (ou seja, usuários que têm contato com o português fora do Brasil, em contextos domiciliares) em nível de graduação. Os resultados indicaram que o tempo de realização e os itens do teste foram capazes de diferenciar os níveis de proficiência dos participantes, confirmando a validade do teste.

Além do trabalho encontrado no *Portal de Periódicos CAPES*, temos conhecimento do estudo realizado por Evers (2013), que investigou a complexidade textual dos textos produzidos por examinandos do Celpe-Bras com o objetivo de explorar a possibilidade de classificação automática em diferentes níveis de proficiência. A pesquisa utilizou 48 métricas extraídas da ferramenta Coh-Metrix-Port³⁷, as quais se relacionam com a coesão, a coerência e a complexidade textual dos textos analisados. Para isso, a autora examinou 177 textos de respostas de examinandos que foram usados para ajustar a grade de correção utilizada pela equipe de avaliadores humanos. Evers (2013) utilizou uma metodologia quantitativa, sob os preceitos da LC, realizando análises dos resultados das métricas extraídas pela ferramenta e correlacionando esses dados com as notas atribuídas pelos corretores humanos. A autora identificou índices distintivos de complexidade textual, como o número de palavras e de parágrafos, a medida de riqueza lexical, a incidência de conectivos negativos e de adjetivos e o Índice Flesch. O Índice Flesch é um algoritmo que contabiliza o número de palavras por sentenças e a quantidade de sílabas por palavra para indicar o nível de inteligibilidade do texto. Ele gera um valor numérico entre 0 e 100, sendo que valores mais altos indicam maior facilidade de leitura. O Índice Flesch é um parâmetro que considera que quanto mais curtas as palavras e as sentenças do texto, maior a facilidade da leitura e é geralmente pareado ao grau de escolaridade necessário para compreender o texto. Os resultados obtidos por Evers demonstraram que as análises quantitativas automáticas foram capazes de distinguir os textos sem certificação e com certificação, corroborando as notas holísticas atribuídas pelos corretores humanos.

Além dos estudos mencionados, o grupo de pesquisa *Avalia – Avaliação de uso da linguagem*³⁸, atuante no Instituto de Letras da Universidade Federal do Rio Grande do Sul, ao qual esta pesquisa se filia, tem se dedicado à análise de *corpora* de respostas de examinandos do Celpe-Bras por meio da abordagem teórico-metodológica da LC, visando contribuir para a validade do exame.

³⁷ <http://fw.nilc.icmc.usp.br:23380/cohmetrixport>

³⁸ <https://www.ufrgs.br/grupoavalia/>

Membra do grupo Avalia, Kunrath (2019), em sua tese de doutorado, empreendeu uma análise de 50 respostas de examinandos com o intuito de contribuir para a elaboração de especificações mais robustas de dois níveis de proficiência. Através do emprego da ferramenta Coh-Metrix-Port, a autora realizou um levantamento de métricas distintivas entre os níveis de proficiência utilizando os índices de contagens básicas (como palavras e sentenças), operadores lógicos (como “e”, “se” e “ou”), pronomes pessoais e conectivos (aditivos, causais, lógicos, etc.). Essas observações foram associadas a procedimentos de análise qualitativa dos textos, visando verificar o cumprimento da tarefa de modo holístico. Os resultados obtidos por Kunrath contribuem para uma descrição mais detalhada e precisa dos níveis de proficiência do Celpe-Bras a partir das diferenças observadas nos dados, especialmente relacionadas à recontextualização de informações dos textos de insumo e do uso de recursos linguísticos.

De modo semelhante, o trabalho de conclusão de curso de graduação de Divino (2021), também membra do grupo Avalia, igualmente trouxe contribuições relevantes a essa discussão. A autora identificou distintos padrões no uso de índices lexicais (envolvendo contagens básicas de palavras e sentenças, bem como diversidade lexical e palavras-chave) em dois níveis de proficiência do Celpe-Bras. Divino se valeu de *corpora* compreendendo um total de 865 respostas de examinandos, e fez uso de um conjunto de ferramentas de análises disponibilizadas pelo programa Sketch Engine³⁹ que permitem a observação das frequências e padrões no uso de determinados aspectos linguísticos. Os resultados da autora corroboram as conclusões de Evers (2013), destacando a distintividade do número de palavras e sentenças nos diferentes níveis de proficiência. Adicionalmente, a pesquisa apontou para diferenças no nível de diversidade lexical, ou seja, quanto mais abrangente e diversificado o uso lexical no texto, maior a nota atribuída pelos corretores humanos.

Mediante o exposto, torna-se evidente que o campo de estudos de Avaliação de línguas informada por corpus no Brasil está em estágios iniciais de desenvolvimento. Sobre a formação de pesquisadores da área de Linguística Aplicada, nas palavras de Chapelle e Voss (2017):

Mudanças mais revolucionárias provavelmente exigirão alunos de pós-graduação com formação em avaliação de línguas, além de outras áreas da linguística aplicada. Por exemplo, os alunos precisam ter formação em linguística de corpus para realizar boas análises de domínios de uso da linguagem como base para o desenvolvimento de testes (e.g., Biber 2006). A formação em aprendizagem de língua adicional também é necessária para que os alunos usem corpora de aprendizes para definir níveis de proficiência linguística (Saville e Haakey 2010). [...] Esses e outros aspectos da linguística aplicada parecem ser cruciais para ajudar a aumentar a utilidade da avaliação em todo o processo educacional, aprofundar a compreensão dos linguistas aplicados

³⁹ <https://www.sketchengine.eu/>

sobre proficiência linguística e expandir suas agendas para a validação de testes (CHAPELLE; VOSS, 2017, p. 159)⁴⁰.

Em uma direção similar, Oushiro (2022) ressalta a importância de capacitar linguistas brasileiros na análise estatística de dados empíricos, aliada a abordagens qualitativas, para uma compreensão mais sólida sobre o funcionamento das línguas naturais. A autora destaca que “até o momento, a formação em Estatística não é componente obrigatório da maior parte dos cursos de Letras e de Linguística, e tem de ser buscada por iniciativa própria de pesquisadores, professores e estudantes de pós-graduação ou graduação” (OUSHIRO, 2022, p. 9).

Conforme evidenciado pelos estudos resenhados neste capítulo, *corpora* compostos por respostas de examinandos predominam no campo da Avaliação de línguas informada por corpus (EGBERT, 2017). Por outro lado, as investigações que utilizam *corpora* constituídos por textos de insumo das tarefas, analisando seu conteúdo, investigando o processo de elaboração e avaliando o possível impacto da complexidade textual no desempenho de leitura dos candidatos aos testes, ainda são escassas (SPENCER *et al.*, 2019; SHIN, 2021; PLAKANS, 2021). No intuito de ampliar as investigações acerca do exame Celpe-Bras, este trabalho se propõe a preencher essa lacuna, analisando um *corpus* composto por textos de insumo de tarefas de leitura e escrita. Investigações dessa natureza têm o potencial de aprimorar a compreensão sobre a representatividade do conteúdo dos testes e das habilidades de leitura avaliadas, oferecendo evidências para a validade do exame (FULCHER; OWEN, 2016).

⁴⁰ No original, lê-se: More revolutionary changes will probably require graduate students educated in language testing in addition to other areas of applied linguistics. For example, students need to be educated in corpus linguistics to conduct appropriate domain analyses as a basis for test development (e.g., Biber 2006). Education in second language acquisition is needed too for students to use learner corpora for defining levels of linguistic competence (Saville and Hakey 2010). [...] These and other aspects of applied linguistics appear to be critical for helping to increase the usefulness of assessment throughout the educational process, strengthen applied linguists' understanding of language proficiency, and expand their agendas for test validation.

3 AVALIAÇÃO DE LÍNGUAS

Este capítulo inicia com uma introdução à área de Avaliação de línguas, focalizando especialmente nos testes de proficiência em línguas adicionais. A análise crítica do poder social desses testes de alto impacto é abordada, destacando a necessidade fundamental de investigar seus argumentos de validade. A seguir, exploramos a noção de leitura sob uma perspectiva de letramento sociocultural e orientada para o desempenho. A partir dessa mesma perspectiva, discutimos a noção de complexidade textual em avaliações de leitura.

3.1 AVALIAÇÃO DE LÍNGUAS E ARGUMENTOS DE VALIDADE

A área de AL é a subárea da Linguística Aplicada dedicada ao estudo de diferentes instrumentos que têm como objetivo avaliar o uso da língua em variados contextos e para variadas finalidades. Alguns visam ranquear os candidatos que competem entre si para obter uma nota mais alta, diferenciando o desempenho entre eles (*e.g.*, as provas de inglês do Concurso de Admissão à Carreira de Diplomata brasileiro); outros têm o objetivo de avaliar o desempenho dos candidatos e classificá-los em determinados níveis de proficiência sem uma comparação entre os resultados mas de acordo com descritores pré-estabelecidos (*e.g.*, o TOEFL) (DOUGLAS, 2000). As avaliações também podem se diferenciar em termos de objetivos e conteúdo da língua-alvo, podendo ser classificadas em três grandes grupos: testes de proficiência; testes de aproveitamento, em geral utilizados em cursos de línguas, para avaliar a extensão da aprendizagem em relação a determinados objetivos de ensino; e testes de nivelamento, em geral utilizados em cursos de línguas para avaliar conhecimentos prévios de estudantes para tomar decisões sobre o ingresso mais adequado em um percurso de estudo (HUGHES; HUGHES, 2020).

As avaliações de proficiência, foco desta tese, podem variar de acordo com a abordagem utilizada para medi-la, sendo que o principal fator que define o método de avaliação é a perspectiva de linguagem e de proficiência. Além disso, os testes podem se dedicar a medir habilidades e sub-habilidades de ler, escrever, ouvir e falar separadamente, geralmente por meio de questões fechadas, ou de modo integrado, através de questões abertas. Brunfaut (2021) descreve que, em questões fechadas (*selected-response items*), os examinandos precisam selecionar uma resposta dentre as opções fornecidas pelos elaboradores e, nas questões abertas

(*constructed-response items*), existe a necessidade da produção de uma resposta oral ou escrita. Por fim, para cada tipo de testagem, existem diferentes métodos de correção. A correção das questões fechadas (*e.g.*, verdadeiro ou falso, múltipla escolha) é geralmente automática, baseada nas respostas corretas definidas *a priori* pelos elaboradores. Já a correção de questões abertas (*e.g.*, produções textuais), normalmente exige a avaliação de especialistas.

Spolsky (2017) afirma que a área de AL começou a se consolidar na década de 1960, coincidindo com o surgimento da centralização e industrialização de testes de larga escala em Princeton e Cambridge. No entanto, o autor também observa que os *Exames Imperiais Chineses*, iniciados no século 7, podem ser considerados como os primeiros sistemas de testagem controlados pelo governo para selecionar indivíduos. O autor destaca que os anos 60 marcaram um momento histórico em que os testes tradicionais de línguas, que incluíam questões de produções textuais, traduções, compreensão e gramática foram substituídos por testes psicométrico-estruturalistas (*e.g.*, *Michigan Lado test of English*). Esses novos testes consistiam principalmente em itens de múltipla escolha e eram considerados confiáveis por serem replicáveis de forma objetiva em larga escala. Spolsky também observa que somente no final do século 20 a área de estudo ampliou-se para abranger discussões sociais e éticas, incorporando diferentes práticas e contextos sociais nas testagens. Essa mudança teve o objetivo de aumentar a autenticidade e a proximidade com os desempenhos em situações externas aos testes. Desde então, os teóricos têm trabalhado em colaboração para elaborar códigos de ética e prática profissional na área de avaliação, como é o caso da *International Language Testing Association*.

A preocupação com questões éticas está fortemente relacionada às consequências geradas pelos testes na sociedade. Os testes de línguas adicionais têm uma ampla aplicação na comprovação de proficiência em diversos países. Imigrantes frequentemente se submetem a esses testes não apenas para demonstrar domínio da língua oficial, mas também para evidenciar conhecimentos culturais e históricos valorizados por comunidades específicas e privilegiadas. Isso ocorre porque as classes dominantes construíram a imagem de testes de línguas como símbolos de sucesso e mobilidade, assim como critérios de mérito, qualidade e valor; conceitos que são, muitas vezes, respaldados pelas instituições educacionais que fazem uso desses testes. Some-se a isso os interesses de uma indústria multimilionária de elaboração de testes em ampla ascensão devido aos recentes movimentos migratórios⁴¹.

⁴¹ Ver, por exemplo, alguns dados sobre o mercado de exames de proficiência em: <https://www.sbs.com.au/language/punjabi/en/article/ielts-is-a-money-making-business-greens-to-push-for>

Ademais, em face de diversos eventos geopolíticos nas últimas décadas, como ataques terroristas, conflitos armados e tensões entre diferentes grupos étnicos e religiosos, os departamentos de segurança governamentais voltaram-se para a temática de imigração. Nesse contexto, os testes de línguas são percebidos como dispositivos de políticas sociais por meio dos quais certas línguas e comportamentos são impostos aos estrangeiros e, desse modo, suas vidas e direitos humanos são controlados por legisladores (MCNAMARA; SHOHAMY, 2008). Shohamy (2017) afirma que as políticas sociais que fazem uso de testes

[...] não derivam de resultados de pesquisas que demonstram que a proficiência nas línguas nacionais é relevante para uma participação social funcional; ainda assim, os testes de língua se tornam uma ferramenta de triagem, levando a decisões sobre se os imigrantes têm permissão para permanecer no país ou se serão forçados a sair. Isso também ignora situações em que os imigrantes estão em uma idade em que são incapazes de aprender um novo idioma e/ou não sabem ler ou escrever em seu próprio idioma [...] (SHOHAMY, 2017, p. 443)⁴².

Dado o poder desses instrumentos, McNamara e Shohamy (2008) fazem um apelo à comunidade envolvida com testes de línguas e aos linguistas aplicados para que estabeleçam um diálogo efetivo com os legisladores, assumindo uma consciência moral e uma responsabilidade social e política em relação a essas questões. Reforçando os apelos dos autores, Winke e Brunfaut (2020) enfatizam que os resultados dos testes não devem ser utilizados para propósitos diferentes daqueles pretendidos durante a elaboração, e que avaliadores e pesquisadores devem ser incisivos e se opor a usos indevidos que possam gerar informações imprecisas ou irrelevantes sobre os examinandos.

McNamara e Roever (2006) destacam que é responsabilidade dos elaboradores informar sobre possíveis impactos de fatores como as características das tarefas e dos corretores nos resultados observáveis dos testes, assim como sobre os esforços para controlá-los. Nas últimas décadas, diversas pesquisas têm tratado sobre a correção de tarefas em testes de línguas, e os resultados dessas investigações fornecem evidências para a validade dos argumentos dos resultados dos testes (CARRELL, 1995; CUMMING, 1990, 2001; BARKAOUI, 2010, 2011; CROSSLEY *et al.*, 2014; GEBRIL; PLAKANS, 2014; SHIN; EWERT, 2015; CHAN *et al.*, 2015; PLAKANS; GEBRIL, 2017; OHTA *et al.*, 2018; DAVIS, 2021). No entanto, como

[review-of-english-test-for-migrants/f8qtbu5i0](https://www.linkedin.com/pulse/english-proficiency-test-market-hit-big-revenues-future-pears/) e <https://www.linkedin.com/pulse/english-proficiency-test-market-hit-big-revenues-future-pears/>, acessado em 20 set. 2023.

⁴² No original, lê-se: this policy does not originate from research findings that demonstrates that proficiency in national languages is relevant for functionality; still, language tests become the tool for screening, leading to decisions as to whether immigrants are allowed to stay in the country or would be forced to leave. It also ignores situations when immigrants are at an age that they are incapable of learning the new language and/or cannot read or write in their own language.

mencionado na introdução, pesquisas sobre o processo de elaboração dessas tarefas ainda são escassas (SPENCER *et al.*, 2019; SHIN, 2021; PLAKANS, 2021).

A concepção de validade de um teste não é unânime entre teóricos da área, e, ao longo das últimas décadas, diferentes perspectivas e teorizações têm contribuído para uma compreensão mais aprofundada desse conceito, central para os testes de línguas. Xi e Sawaki (2017) apontam as contribuições de teóricos como Robert Lado e John Clark nas décadas de 1960 e 1970, que focaram em predições de desempenhos; de Samuel Messick na década de 1980, que expandiu o conceito de validade para incluir as implicações sociais dos resultados; e, a partir dos anos 2000, de Elana Shohamy, que passa a examinar questões éticas de interpretações e usos dos resultados, e de Tim McNamara, que aborda questões políticas e sociais. Atualmente, coexistem diversas concepções de validade de testes, que podem ser subdivididas em diferentes tipos de validade (validade de conteúdo, validade de face, validade de construto, validade de critério, validade de contexto, validade concorrente, validade preditiva, validade cognitiva, validade de pontuação, validade de contexto)⁴³ e tipos de evidências que contribuem para conferir validade ao teste (fundamentação e análises de especialistas sobre o conteúdo, estudos sobre o processo de construção de respostas pelos examinandos, análise estatística dos dados das respostas dos examinandos, relação do construto com outras variáveis, consequências dos testes) (CHAPELLE; LEE, 2021).

Apesar dessa diversidade, de acordo com Chapelle e Lee (2021), mais recentemente têm-se utilizado o termo “validade baseada em argumento” (*argument-based validity*), fundamentado no modelo proposto pelo teórico Michael Kane. Esse modelo é capaz de abranger as diversas concepções de validade, desde a elaboração dos testes até as consequências dos usos de seus resultados, tornando o processo de validação mais acessível para os pesquisadores. Os autores consideram que

[...] pesquisas e práticas recentes de avaliação de línguas mostram que as abordagens de argumento de validade servem como recursos eloquentes que envolvem um conjunto compartilhado de conceitos e linguagem. No campo da avaliação de línguas, em que o conceito basilar de validade tem sido difícil de ser compreendido ao longo de sua evolução, tal linguagem e lógica compartilhadas parecem ser bem-vindas [...] (CHAPELLE, LEE, 2021, p. 27)⁴⁴.

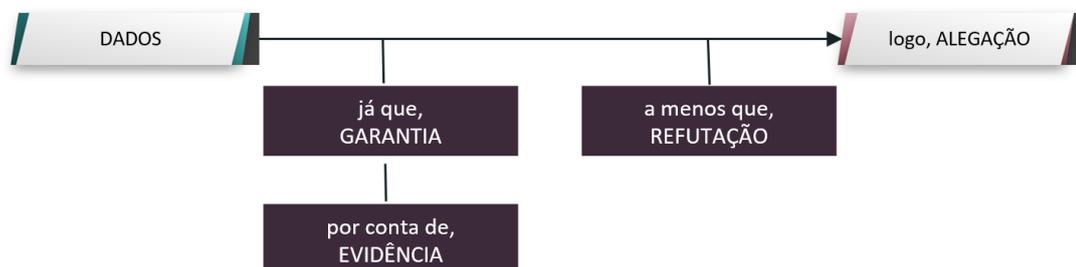
⁴³ Para uma discussão mais aprofundada sobre os tipos de validade, ver Chapelle e Lee (2021).

⁴⁴ No original, lê-se: recent language assessment research and practice demonstrate that validity argument frameworks serve as rhetorical devices with a common set of concepts and language. In the domain of language assessment, where the foundational concept of validity has been difficult to grasp throughout its evolution, such a common language and logic appear to be welcomed [...].

Kane (2021) esclarece que a validade envolve a avaliação da adequação dos usos dos resultados dos testes e da credibilidade da fundamentação de suas interpretações. Nesse sentido, o autor enfatiza que é crucial que os usos dos resultados pretendidos pelos elaboradores dos testes sejam bem definidos e claramente explicitados. Caso contrário, “qualquer validação será necessariamente incompleta e fragmentada” (KANE, 2021, p. 33)⁴⁵. Em outras palavras, “ao afirmar que o teste é adequado aos propósitos, estamos apresentando um argumento. A validação é, portanto, uma questão de apresentar esse argumento de modo suficientemente convincente para quem se interesse pelo teste” (GU, 2021, p. 5)⁴⁶.

O modelo de análise proposto por Kane (2021) é cuidadosamente estruturado e dividido em duas fases. Na primeira fase, descrevem-se as asserções relacionadas à interpretação e aos usos dos resultados dos testes (argumentos). Na segunda fase, busca-se comprovar que essas asserções são verdadeiras, baseando-se em evidências teóricas e empíricas (validação), a menos que se prove o contrário (GU, 2021). Para essa análise de argumento, é utilizado o modelo de argumentação de Toulmin (2003 *apud* GU, 2021), que é ilustrado abaixo.

Figura 1 – Modelo de argumentação de Toulmin



Fonte: Adaptado de GU (2021, p. 6).

Os tipos de argumentos podem ser construídos explicitando-se a relação entre os resultados do teste e, por exemplo, a definição do domínio de uso da linguagem focalizado, o desempenho considerado no contexto externo ao teste e na correção, a explanação do construto, o uso da linguagem no contexto externo ao teste e as consequências previstas do uso dos resultados (CHAPELLE, 2020; CHAPELLE; LEE, 2021). Altos graus de consistência nessas relações (e isso requer descrições robustas de todos esses aspectos) contribuem para conferir validade às inferências que se pode fazer a partir dos resultados e, portanto, ao teste. Dentro

⁴⁵ No original, lê-se: any validation will necessarily be incomplete and fragmentary.

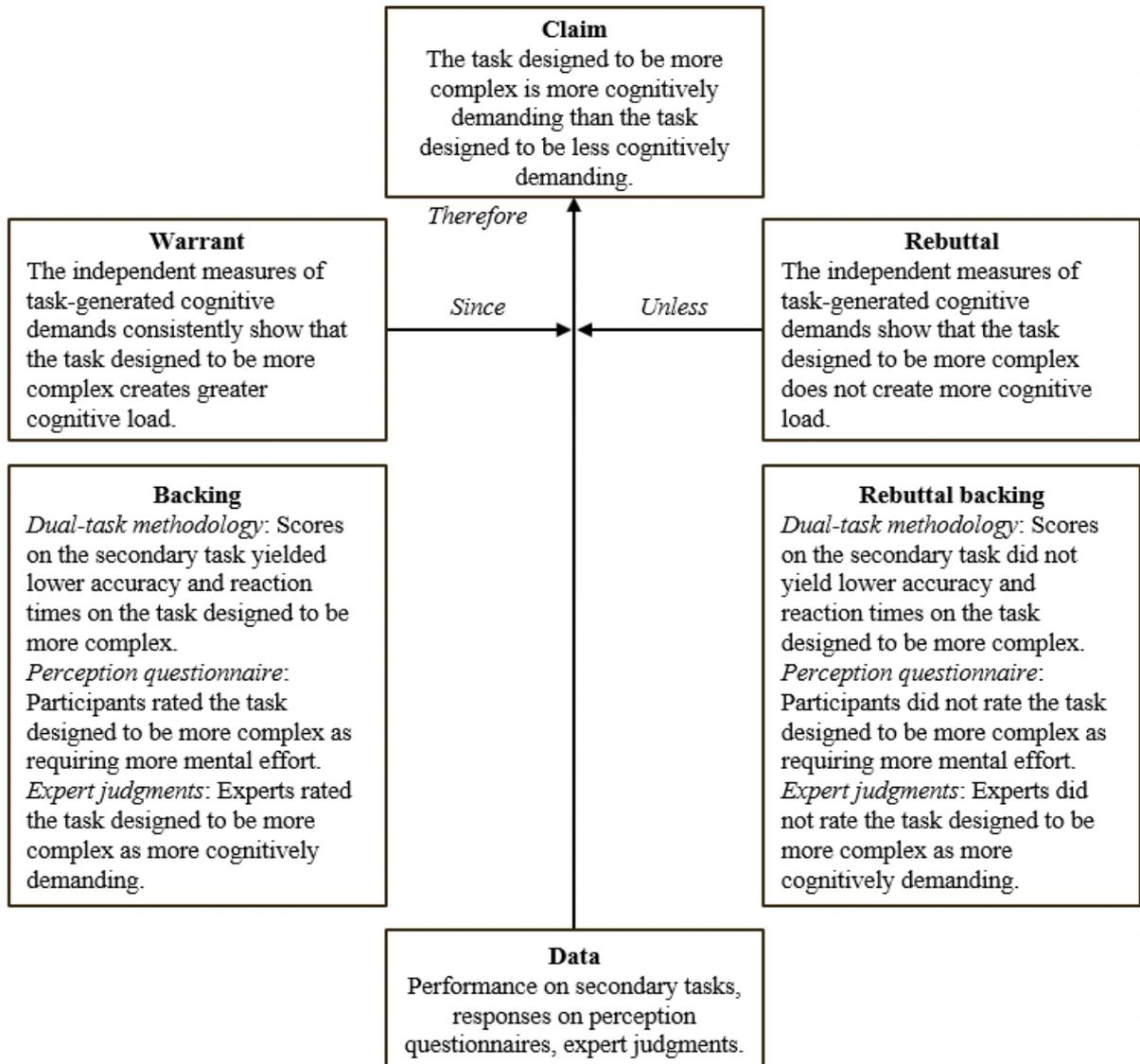
⁴⁶ No original, lê-se: claiming that our assessment is good for its purposes, we are making an argument. Validation is therefore a matter of making this argument convincing enough for people who care about our assessment.

desse leque de possibilidades, nos interessam as discussões relativas ao domínio de uso da linguagem avaliada pelos testes, que, segundo Kane (2021), deve ser definido a partir dos usos pretendidos dos resultados relativos ao desempenho avaliado.

De acordo com o autor, o domínio de uso da linguagem a ser avaliado deve ser definido da forma mais clara possível. No entanto, Kane reconhece que existem situações em que o uso da língua-alvo está relacionado a habilidades mais gerais, que envolvem diversas ações mobilizadas, por exemplo, na leitura de textos, desde leituras casuais de placas sinalizadoras até leituras cuidadosas de manuais. Nesses casos, o autor afirma que é improvável que os elaboradores possam incluir todas as diferentes atividades no teste, e o uso da língua-alvo a ser avaliado provavelmente não será definido com precisão, fragilizando, assim, a validade das inferências. Essa visão é compartilhada por Brown (2004), que ressalta que, por mais válido que um teste seja, ele apenas reflete parcelas da capacidade de comunicação e das interações sociais autênticas do usuário de uma língua, uma vez que nenhum teste é capaz de abranger todas as possibilidades de uso da linguagem em contextos externos à avaliação.

Révész e Brunfaut (2020) ilustram uma possibilidade de pesquisa de validade baseada em argumento com uma estrutura de argumento que parte da hipótese de que versões mais complexas de tarefas exigem maior carga cognitiva para serem respondidas. A figura a seguir ilustra essa proposta: analisando o gráfico de baixo para cima, os autores explicitam que os dados (*Data*), quando analisados, fornecem evidências que potencialmente podem ser usadas para sustentar a alegação de que a alta complexidade de uma tarefa demanda maior carga cognitiva (*Claim*). Com base nos resultados dos dados disponíveis, a alegação pode ser validada (*Since*) ou refutada (*Unless*).

Figura 2 – Possível estrutura de argumento sobre complexidade de tarefas



Fonte: Révész e Brunfaut (2020, p. 28)⁴⁷.

⁴⁷ Tradução nossa: Dados. Desempenho em tarefas secundárias, respostas a questionários de percepção, avaliações de especialistas. Assim Alegação. A tarefa elaborada para ser mais complexa é mais exigente do ponto de vista cognitivo do que a tarefa elaborada para ser menos exigente do ponto de vista cognitivo. Já que Garantia. As medidas independentes das demandas cognitivas geradas pela tarefa mostram consistentemente que a tarefa elaborada para ser mais complexa exige uma carga cognitiva maior. Evidência de garantia. Metodologia de dupla tarefa: As pontuações na tarefa secundária resultaram em menor precisão e tempos de reação na tarefa elaborada para ser mais complexa. Questionário de percepção: Os participantes classificaram a tarefa elaborada para ser mais complexa como exigindo maior esforço mental. Julgamentos de especialistas: Os especialistas classificaram a tarefa elaborada para ser mais complexa como mais exigente do ponto de vista cognitivo. A menos que Refutação. As medidas independentes das demandas cognitivas geradas pela tarefa mostram que a tarefa elaborada para ser mais complexa não exige uma carga cognitiva maior. Evidência de refutação. Metodologia de dupla tarefa: As pontuações na tarefa secundária não resultaram em menor precisão e tempos de reação na tarefa elaborada para ser mais complexa. Questionário de percepção: Os participantes não classificaram a tarefa elaborada para ser mais complexa como exigindo maior esforço mental. Julgamentos de especialistas: Os especialistas não classificaram a tarefa elaborada para ser mais complexa como mais exigente do ponto de vista cognitivo.

Conforme mostra a figura, a garantia e a refutação da alegação baseiam-se nas evidências resultantes das análises. No exemplo apresentado pelos autores, o procedimento metodológico incluiu análises de pontuações dos participantes em duas tarefas, um questionário de percepção dos participantes e o julgamento de especialistas sobre a complexidade das tarefas. A distinção entre validar ou refutar a alegação está baseada, essencialmente, na condição de os testes apresentarem resultados positivos ou negativos. O exemplo de Révész e Brunfaut (2020) ilustra como a validação do argumento pode ser direcionada aos objetivos de pesquisa e personalizada para o teste em foco, com base nos dados e evidências disponíveis. Isso, como afirmado por Chapelle e Lee (2021), representa um modelo de análise de validade abrangente, pois reúne argumentos de várias naturezas (*e.g.*, pontuação, conteúdo, face, etc.).

Os pesquisadores da área de avaliação concordam quanto à necessidade de evidências empíricas que sustentem a validade do teste em simular (operacionalizar) a descrição teórica (construto) da percepção do fenômeno (uso da língua-alvo) que o exame considera estar avaliando por meio de seus instrumentos. Embora seja reconhecido que a criação e a validação de questões abertas eficientes sejam processos demorados e custosos (BROWN, 2004, p. 45), que requerem especialistas em avaliação, coleta de dados e infraestrutura moderna, é responsabilidade dos elaboradores de testes garantir que esse processo seja devidamente conduzido.

Há um longo debate sobre as diferenças entre questões fechadas e questões abertas em testes de leitura, discutindo as vantagens e limitações de cada uma (SOUZA; HÜBNER, 2015). Segundo Shin (2021), embora mais facilmente validados, a perspectiva da abordagem crítica de elaboração de itens enxerga os testes com questões fechadas como concedendo excessivo poder aos elaboradores, uma vez que ao fornecerem apenas uma opção de acerto, podem determinar o que é considerado “correto” e “incorreto” em termos de determinado uso da língua-alvo. A perspectiva crítica de testes de línguas, que se preocupa com os usos indevidos de certificados de testes e seus impactos sociopolíticos, tende a se alinhar mais com questões abertas, que promovem o uso autêntico da linguagem, por entenderem que, dessa maneira, se ampliam as possibilidades de interpretações e respostas adequadas (HAMP-LYONS, 2000; SHOHAMY, 2001; MCNAMARA; SHOHAMY, 2008). Porém, justamente devido a essa ampliação de possibilidades, tais tarefas requerem mais investigação para que as asserções feitas a partir dos dados obtidos sejam consideradas válidas (SHIN, 2021). Nesse contexto, Plakans (2021) observa que uma grande motivação para o uso de questões abertas e integradas em testes de proficiência linguística é a possibilidade de uso autêntico da linguagem que é capaz

de simular de modo mais preciso o uso da língua-alvo que os testes pretendem avaliar. No entanto, a autora ratifica o fato de que a validação dessas tarefas é um processo mais complexo do que nos testes que avaliam habilidades de forma separada.

Quanto ao conceito de tarefa integrada, Plakans (2021) destaca que várias categorias de questões que integram habilidades em diferentes níveis são agrupadas dentro desse termo guarda-chuva de avaliação integrada. A autora utiliza parâmetros propostos por Knoch e Sitajalabhorn (2013) para definir tarefas integradas, caracterizando-as como tarefas que:

- 1) utilizam texto de insumo verbal, não apenas imagético;
- 2) demandam a transformação do texto de insumo em produção;
- 3) pressupõem critérios de avaliação que incorporam diferentes habilidades;
- 4) têm como base desempenhos reais de tarefas integradas.

Assim como Plakans, Mendel (2019) destaca que a literatura apresenta diversas definições para o conceito de tarefa integrada, porém, a combinação de mais de uma habilidade e o uso de texto de insumo são elementos essenciais. A autora ressalta que a leitura é uma habilidade imprescindível para a realização de tarefas integradas, e a familiaridade com esse método de avaliação pode influenciar o desempenho dos examinandos, que precisam compreender a exigência da integração das habilidades.

Kremmel e colaboradores (2021) enfatizam que a fase de pilotagem de tarefas integradas é uma etapa pré-operacional crucial para assegurar a validade das inferências feitas sobre os resultados dos testes. Além disso, eles ratificam que essa etapa requer um investimento considerável de tempo e recursos financeiros. Os autores sugerem que abordagens que combinam métodos qualitativos e quantitativos devem ser empregadas para abordar questões relacionadas ao conteúdo, método, extensão, dificuldade e percepção dos candidatos ao longo de três estágios pré-operacionais:

- 1) elaboração de um protótipo (*prototyping*) - minitriagem de propostas de modelos de tarefas;
- 2) testagem piloto (*pilot-testing*) - experimentação em maior escala do modelo de tarefa escolhido após ajustes no protótipo da minitriagem;
- 3) testagem no campo (*field-testing*) - confirmação do modelo completo do teste em um contexto que representa o público-alvo do exame. Nesse estágio, especialmente no caso de

testes de alto impacto, é essencial que as tarefas sejam testadas em campo com uma amostra significativa e representativa da população-alvo do exame.

Durante o processo de elaboração de um teste de língua, um passo crucial na validação do instrumento envolve o estudo da dimensão da dificuldade e da distinção de níveis das tarefas elaboradas. Esse aspecto é especialmente relevante quando se trata de exames padronizados de larga escala e forte impacto social, nos quais é pressuposto que o que está sendo testado é constante de uma edição para outra. Scaramucci (2016, p. 406) sublinha que, para assegurar a equivalência das edições do Celpe-Bras, cada edição deve ser uma amostra representativa dos elementos delineados nas especificações do exame, que representam o seu construto, como as sub-habilidades, os propósitos, os gêneros do discurso, os interlocutores e as temáticas abordadas. Um dos principais objetivos das especificações é possibilitar a reprodução de edições equivalentes, fornecendo detalhes suficientes para que os elaboradores possam recriar, a cada aplicação do teste, versões com conteúdo e demandas similares (GREEN; FULCHER, 2020; JIN, 2021). Por essa razão, as especificações do teste devem “conter todos os detalhes técnicos para guiar sua elaboração” (JIN, 2021, p. 281)⁴⁸.

Jin (2021) defende que as especificações devem iniciar explicitando, honesta e claramente, o objetivo e o público-alvo do teste. A autora argumenta que os usos não previstos pelos elaboradores não devem constar nas especificações, pois sua inclusão no documento sugere que o exame respalda esses usos indevidos. Como exemplo de teste que anuncia claramente o seu propósito, a autora cita o *International English Language Testing System* (IELTS), organizado pelo *British Council*. A seguir, exemplificamos, com trechos do site oficial do exame, a descrição do propósito de três tipos distintos de testes, explicitando, assim, os usos que podem ser feitos dos resultados.

IELTS Academic

O exame de inglês IELTS Academic avalia se você tem o nível de proficiência em língua inglesa exigido para ingressar em um curso acadêmico, refletindo algumas das competências da linguagem acadêmica.

IELTS General Training

A modalidade do exame de inglês IELTS General Training avalia o seu conhecimento da língua inglesa em um contexto cotidiano. As questões e etapas da prova refletem o uso do idioma tanto no trabalho como em situações sociais.

IELTS for UKVI

Existe um IELTS específico para fins de vistos e imigração no Reino Unido (UKVI). Se você deseja trabalhar, estudar ou morar no Reino Unido, pode ser necessário fazer esse exame de inglês (PROVA..., [2023?]).

⁴⁸ No original, lê-se: [...] contains all the technical details for guiding the design [...].

Ao discutir o registro de especificações relacionadas aos textos de insumo de tarefas integradas, Jin (2021) ressalta a importância do detalhamento dos textos selecionados e cita as especificações do *Scholastic Aptitude Test* (SAT), um teste padronizado utilizado nos Estados Unidos para admissão no Ensino Superior, como exemplo. A seguir, reproduzimos trechos do documento *Test Specifications for the Redesigned SAT* (THE COLLEGE BOARD, 2015), nos quais algumas características dos textos de insumo utilizados para a produção textual são delineadas, tais como a fonte, a natureza argumentativa e, especialmente, o nível de complexidade textual, que não deve representar uma barreira para a realização do teste no tempo previsto. Além disso, o documento explicita o modo como o nível de complexidade dos textos selecionados é avaliado antes da aplicação das provas.

Para garantir que o nível de complexidade dos textos do SAT seja adequado – desafiadores, mas não inacessíveis ao nível de preparo de quem está iniciando no ensino superior ou na carreira profissional – a equipe de elaboração do teste utiliza o feedback de especialistas atuantes nos níveis do ensino básico e superior, dados de desempenho dos alunos e medidas quantitativas e qualitativas de complexidade textual. [...] Todas as passagens são retiradas de fontes de alta qualidade, previamente publicadas. Embora o estilo e o conteúdo específicos dos trechos inevitavelmente variem em certa medida, dado o compromisso do College Board em utilizar textos autênticos nesta tarefa, os trechos assumem uma forma geral que pode ser considerada como textos argumentativos destinados a um público amplo. [...] A complexidade textual dos trechos é cuidadosamente monitorada para garantir que o desafio da leitura seja adequado e equivalente entre as edições, mas que não se constitua como uma barreira intransponível para que os alunos respondam ao texto de insumo dentro do tempo previsto (THE COLLEGE BOARD, 2015, p. 26)⁴⁹.

Por ser considerada uma matriz (*blueprint*) para a elaboração do teste, as especificações são elaboradas durante a fase de desenho do exame, que, segundo Brown (2004, p. 71), é o momento de lançar sólidas bases para o futuro do teste e pode demandar semanas, meses ou até anos de dedicação. Contudo, o autor ressalta que as análises estatísticas de validação do teste não devem ser encerradas após a implementação, mas sim conduzidas de forma sistemática por meio de um programa rigoroso que forneça evidências para corroborar os resultados originais da fase de desenho e promover o aperfeiçoamento constante do exame (BROWN, 2004, p. 81).

⁴⁹ No original, lê-se: To ensure that texts on the SAT are appropriately complex — challenging but not inaccessible to college- and career-ready test-takers — test development staff make use of feedback from secondary and postsecondary subject-matter experts and test data on student performance as well as quantitative and qualitative measures of text complexity. [...] All passages are taken from high-quality, previously published sources. While the specific style and content of the passages inevitably vary to some extent given the College Board’s commitment to using authentic texts with this task, the passages take the general form of what might be called arguments written for a broad audience. [...] Text complexity of the passages is carefully monitored to ensure that the reading challenge is appropriate and comparable across administrations but not an insuperable barrier to students responding to the source text under timed conditions.

Schlatter *et al.* (2021) destacam a importância do fortalecimento de pesquisas que se ocupem de análises estatísticas, visando o aperfeiçoamento do sistema de avaliação do Celpe-Bras, incluindo a pilotagem de tarefas como um dos procedimentos-chave. As autoras acrescentam que “tais pesquisas podem trazer subsídios para tomadas de decisão mais fundamentadas tanto acerca da calibragem entre construto e especificações do exame como da elaboração e do aprimoramento do próprio instrumento de avaliação” (SCHLATTER *et al.*, 2021, p. 34).

Os instrumentos elaborados pela comissão do Celpe-Bras foram testados antes da primeira aplicação oficial da prova, ocorrida em 1998. Os testes-piloto das tarefas elaboradas foram conduzidos em 1994, contendo a amostragem de duas populações distintas: estudantes brasileiros do último ano do Ensino Médio de escolas públicas e estudantes estrangeiros de universidades públicas brasileiras. Essa etapa de testagem constituiu apenas um dos procedimentos pré-operacionais que a comissão empreendeu ao longo de cinco anos, com o intuito de coletar evidências para validar as inferências feitas a partir do desempenho observável dos examinandos por meio dos instrumentos elaborados. Outros procedimentos incluíram reuniões sistemáticas para deliberar sobre questões relacionadas à natureza do teste, à concepção de avaliação de língua em uso, à estrutura em duas partes (Escrita e Oral) e à integração de habilidades. Após a fase inicial de deliberações, três versões completas do exame foram elaboradas, e validações internas dessas versões foram realizadas mediante análise estatística e revisão crítica por três especialistas da área antes da aplicação dos dois testes-piloto⁵⁰. Desde então, não temos conhecimento de outra publicação ou documento oficial que faça referência a testagens sistemáticas de tarefas do Celpe-Bras. Conhecemos apenas a descrição das testagens realizadas na fase de concepção do exame em 1994, que tinham claramente estabelecidos o público-alvo, estudantes estrangeiros não-lusófonos interessados em ingressar em IES brasileiras, e o objetivo do exame, que era aferir o potencial para participar de interações da vida cotidiana e estudantil (SCHLATTER, 2014).

Reconhecemos que a experiência e a reflexão contínua da Comissão Técnica do Celpe-Bras, que mantém sua composição relativamente estável há mais de 20 anos, assim como as pesquisas desenvolvidas durante esse período⁵¹, contribuem para a validade do exame. No entanto, após três décadas sem a condução sistemática de procedimentos de validação dos

⁵⁰ Para uma descrição mais completa do histórico de elaboração do Celpe-Bras e a listagem de chamadas públicas para a composição da Comissão Técnica, ver Nagasawa (2018).

⁵¹ <https://www.ufrgs.br/acervocelpebras/pesquisas/>

instrumentos, consideramos relevante que análises do conteúdo das tarefas aplicadas ao longo desse tempo fossem realizadas.

Por um lado, reconhecemos que a pilotagem de todas as tarefas do Celpe-Bras é uma medida pouco viável, principalmente devido ao histórico de investimento limitado nesse sistema de avaliação ao longo dos anos por parte do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), órgão vinculado ao Ministério da Educação (MEC), que é responsável pelo exame desde 2009. Por outro lado, aprofundar a análise dos textos selecionados para compor as tarefas pode ser realizada de forma robusta, por meio de especificações detalhadas e claras sobre esse material, suas características linguísticas e seu nível de complexidade (CUSHING, 2017; GREEN; FULCHER, 2020). Portanto, acreditamos que uma maneira de contribuir para a validade do teste é investigar as tarefas integradas de leitura e escrita, a fim de verificar se o conteúdo utilizado como insumo ao longo dos anos é representativo da língua-alvo que o teste busca simular, segundo suas especificações, e se o nível de complexidade textual é compatível com o nível de leitura de seu público-alvo.

3.2 LEITURA E LETRAMENTO

Alexander e Fox (2018) traçam um panorama histórico do desenvolvimento da pesquisa sobre leitura ao longo das últimas décadas, começando em 1956, ano de fundação da *International Reading Association*. Ainda que os autores reconheçam que delimitaram as fronteiras entre os períodos de forma aproximada, uma vez que elas se entrelaçam e se sobrepõem, a narrativa histórica é estruturada em sete períodos, nos quais eles identificam princípios norteadores distintos característicos de cada um. Os períodos descritos são: (1) aprendizagem condicionada (1950–1965); (2) aprendizagem natural (1966–1975); (3) processamento de informações (1976–1985); (4) aprendizagem sociocultural (1986–1995); (5) aprendizagem participativa (1996–2005); (6) aprendizagem orientada ao desempenho (2006–2015); e (7) aprendizagem personalizada (2016–presente).

De acordo com Alexander e Fox (2018), a primeira década se inicia após a Segunda Guerra Mundial, impulsionada pela entrada de um grande número de crianças no sistema público de educação dos Estados Unidos (*baby boom*). A transformação no perfil da aquisição de leitura nas escolas chama a atenção dos pesquisadores, que buscam sanar as dificuldades enfrentadas pelas crianças. Esses estudiosos adotam uma perspectiva behaviorista que possibilita a definição e separação dos processos e sub-habilidades envolvidas na leitura,

permitindo um ensino sistemático e a prática isolada dessas etapas. O segundo período começa em meados da década de 1960, quando os avanços na neurologia e na inteligência artificial redirecionam o foco dos pesquisadores para uma abordagem na qual a leitura é entendida como um processo mental natural e inato do ser humano. Já o terceiro período traz uma mudança teórica significativa para a área, estimulada por investimentos governamentais que permitiram a criação de centros interdisciplinares de pesquisa em leitura. Os estudos se concentram no processamento de informações durante a leitura, considerando o papel dos conhecimentos prévios e das estratégias utilizadas pelos leitores no desempenho da leitura.

Os pressupostos teóricos que orientam a elaboração do exame Celpe-Bras parecem refletir discussões teóricas ocorridas entre o quarto e o sexto período, de 1986 até 2015, conforme a estrutura histórica proposta por Alexander e Fox (2018). Ao que parece, a primeira comissão responsável pelo desenho do Celpe-Bras não apenas estava atualizada sobre as discussões internacionais contemporâneas sobre letramento na década de 1990, mas estava na vanguarda dessas reflexões, incorporando noções ao construto teórico do exame que ainda estariam em pauta nos dois períodos subsequentes.

O período da aprendizagem sociocultural foi impulsionado por pesquisas do campo da antropologia sociocultural (HEATH, 1983), quando novas perspectivas de leitura começaram a integrar a agenda de discussões. Essa abordagem considera textos autênticos e seus contextos de produção, bem como a relação entre as comunidades e suas práticas sociais, em oposição ao foco no letramento individual de cada sujeito. A aprendizagem da leitura passou a ser vista como uma experiência sociocultural que acontece em colaboração entre indivíduos que formam uma comunidade de aprendizagem.

No período seguinte, da aprendizagem participativa, a noção de texto foi ampliada devido à rápida expansão do uso de computadores e, conseqüentemente, de textos multimodais distintos do formato linear tradicional da leitura de textos escritos em papel. Assim, as pesquisas sobre leitura se estenderam além da aquisição na primeira infância, reconhecendo o ato de ler como uma habilidade de natureza progressiva, que se amplia de acordo com o desenvolvimento e as necessidades do leitor. Desse modo, a participação ativa do indivíduo no processo de leitura, reagindo ao texto com base nas demandas das práticas sociais, pré-conhecimentos, estratégias, motivações e reflexões, tornou-se parte do processo de aprendizagem (ALEXANDER; FOX, 2018).

De acordo com os autores, no período da aprendizagem orientada ao desempenho, sexto período, a visão de proficiência em leitura passou a incluir o pensamento crítico e analítico dos

leitores. A noção de proficiência em leitura ampliou-se para ser capaz de demonstrar o que compreendeu por meio de uma (re)ação ao texto, envolvendo a expressão oral ou escrita. A competência leitora passa a incluir conhecimentos sobre gêneros discursivos, estrutura e composição, tópicos e argumentos presentes no texto. Avançando para o sétimo e último período descrito por Alexander e Fox (2018), percebeu-se uma fase em que a saturação de informações, a inevitabilidade da tecnologia e a individualização dos processos e produtos da aprendizagem humana moldam a experiência de leitura, resultando em características e experiências únicas para cada leitor diante das diversas formas de acesso e comunicação de informações.

Essas concepções de leitura permeiam o construto teórico do Celpe-Bras e sua operacionalização, que, como vimos, “fundamenta-se em uma visão de uso da língua(gem) com propósitos sociais, construída social e localmente por seus participantes” (BRASIL, 2020, p. 28). Portanto, ser proficiente “implica ser capaz de engajar-se em diferentes situações de uso da língua portuguesa no mundo, mostrando adequação às demandas dos vários contextos” (BRASIL, 2020, p. 28). No que diz respeito à leitura, essa perspectiva se concretiza no exame por meio de textos de insumo apresentados, em sua grande maioria, na íntegra. A compreensão desses textos é definida pelos objetivos de escrita que demandam uma compreensão específica do material lido (BRASIL, 2020, p. 35), conforme detalhado mais adiante.

Em consonância com esse construto teórico, neste estudo, entendemos o texto como o próprio lugar da interação entre sujeitos ativos que se constroem e são construídos dialogicamente nessa interação, ou seja, o texto não é concebido apenas como uma representação mental ou um artefato linguístico, mas como um construto histórico-social, complexo e multifacetado (KOCH, 2018). A partir dessa ótica, o texto é a manifestação de uma atividade de linguagem que não se limita à modalidade escrita, mas se refere também “à fala e a objetos de linguagens não-verbais ou híbridas” (RIO GRANDE DO SUL, 2009, p. 44). No caso do texto escrito, não basta ser um conjunto de palavras, ele precisa ser compreendido como uma unidade significativa global e, para tanto, deve apresentar características reconhecíveis pelos interlocutores envolvidos no contexto de comunicação, como coesão, coerência, informatividade e intertextualidade (ANTUNES, 2005), combinadas de modo a (re)construir situações comunicativas específicas.

Martins (1982) reconhece que ao ler, ocorre tanto o processo de decodificação mecânica de signos linguísticos (perspectiva behaviorista-skinneriana) quanto o processo de compreensão abrangente (perspectiva cognitivo-sociológica). A autora inclui no ato de ler, o indivíduo e suas

capacidades de interação social, política, econômica e cultural. Para Martins (1982), o processo de leitura é dialético, pois se, por um lado, decodificar sem compreender é inútil, por outro lado, não decodificar restringe a apreensão de sentidos do texto. A autora introduz a expressão “o pulo do gato” para ilustrar o momento em que o sujeito se torna um leitor autônomo por meio da fruição na decodificação e da apreensão dos sentidos do texto. Nessa etapa, o leitor é capaz de perceber a não neutralidade do que o autor expressa e a possibilidade de relacionar o que está sendo lido com suas próprias vivências de mundo, refletindo, questionando e reagindo ao processo de leitura ao responder às práticas sociais das quais participa.

Corroborando as afirmações de Martins sobre o aspecto social da leitura, Kleiman (1995) utiliza o conceito de letramento social, relacionado às práticas socioculturais de diversos usos e usuários da língua escrita, que se distancia da dicotomia de sujeitos alfabetizados e não-alfabetizados. A autora assinala que “alfabetização” se refere à capacidade individual de aquisição de códigos, não se estendendo ao conceito ampliado de Paulo Freire de “levar o analfabeto a organizar reflexivamente seu pensamento, desenvolver a consciência crítica, introduzi-lo num processo real de democratização da cultura e de libertação” (FREIRE, 1980 *apud* KLEIMAN, 1995, p. 16). Isso porque a proposta de alfabetização Freireana se aproxima do modelo ideológico de letramento de Street (1984) devido à determinação de Paulo Freire “em tornar explícita a natureza política e ideológica da prática da alfabetização” (STREET, 1984, p. 202)⁵².

Street (1984) propõe o letramento ideológico, entendendo que o uso escrito da língua é condicionado pelas estruturas socioculturais de poder e contextos sociodiscursivos específicos em que ocorrem. Ele contrasta esse letramento sociohistórico com o modelo autônomo de letramento, que, grosso modo, está circunscrito apenas aos aspectos lógicos da língua, descolados da natureza interacional da linguagem humana. Antes de Street, Heath (1983) definiu “eventos de letramento” como “eventos em que a linguagem escrita é parte essencial da natureza das interações entre os participantes e de seus processos e estratégias interpretativas” (1983, p. 319)⁵³, por meio da observação de que as práticas letradas domiciliares orientam o modo como as práticas letradas escolares são elaboradas por diferentes comunidades. Essa noção de práticas sócio-histórico-culturais, à qual nos afiliamos, é corroborada por Gee (2019), que as nomeia como “Discursos”, diversos papéis sociais que assumimos nos diferentes contextos dos quais participamos. Nesse sentido, os “Discursos”, além da linguagem, envolvem

⁵² No original, lê-se: [his willingness] to make explicit the political and ideological nature of literacy practice.

⁵³ No original, lê-se: occasions in which written language is integral to the nature of participants’ interactions and their interpretive processes and strategies.

modos de usar a língua, agir, interagir, crer, valorizar e sentir, a fim de produzir identidades e ações significativas e socialmente situadas. Gee (2019) afirma que

[...] se as discussões sobre a leitura não são sobre línguas sociais (e, portanto, também sobre ação e interação concretas no mundo, perspectivas imbuídas de valoração e identidades socialmente situadas), então não são, na realidade, sobre a leitura como um processo semiótico de produção de sentido (e é difícil saber o que é leitura se não for isso) (GEE, 2019, p. 110)⁵⁴.

Assim, entendemos que são múltiplas as leituras possíveis de um texto, considerando-se quem lê, a partir de qual posicionalidade, com que propósito e em qual prática social. É crucial reconhecer, porém, que um determinado texto pode permitir várias leituras, mas não qualquer leitura (JOUVE, 2002). Segundo Jouve (2002), há um espaço de tensão entre a liberdade do leitor e a delimitação do texto escrito. Dado que autor e leitor estão separados por espaço e tempo, a cooperação dialógica para a construção de sentidos, presente nas interações síncronas, não ocorre. Desse modo, ao ler, se espera que o leitor não transgrida as regras do “jogo” da leitura, mas sim construa a recepção do texto em seu imaginário, realizando interpretações plausíveis e legítimas com base na coerência interna e externa do texto.

Entendemos, portanto, que o leitor proficiente é o que sabe reagir ao texto conforme a expectativa criada pelo gênero discursivo ao qual o texto se circunscreve (SCHLATTER, 2009), realizando as interpretações preferíveis tendo em vista os propósitos de leitura compartilhados pela comunidade de leitores que o está mobilizando em práticas sociais específicas e situadas (SCHLATTER, 2018). Seguindo essa perspectiva, o sujeito letrado não é apenas alguém alfabetizado, capaz de usar um determinado sistema simbólico para ler e escrever textos, mas também um participante ativo socialmente por meio de textos, que “responde adequadamente às demandas sociais de leitura e de escrita” (SOARES, 2009, p. 40).

Sousa e Hübner (2015) ainda citam a motivação como um fator que pode influenciar o desempenho na avaliação de leitura. Apesar de reconhecerem que a motivação é um fator subjetivo e desafiador de se controlar, as autoras sugerem que

[...] modificações na tarefa como extensão do texto, formas de instruções, apresentação (papel x computador) e escolha dos temas, ou mesmo na situação do teste como duração, horário e local, podem contribuir para diminuir a tensão e motivar os participantes para seu engajamento na realização da tarefa (SOUZA; HÜBNER, 2015, p. 42).

⁵⁴ No original, lê-se: If discussions about reading are not about social languages (and thus, too, about embodied action and interaction in the world, value-laden perspectives, and socially situated identities), then they are not, in reality, about reading as a semiotic meaning-making process (and it is hard to know what reading is if it is not this).

À luz do exposto, entendemos que durante a leitura, o leitor desempenha de maneira integrada e não sequencial diversos papéis, aplicando habilidades distintas. Luke e Freebody (1999) propõem a expressão *família de práticas* para descrever quatro habilidades consideradas fundamentais para a leitura. Os autores argumentam que essas famílias de práticas não são autossuficientes para assegurar o acesso às outras famílias ou ao letramento do sujeito, mas são práticas dinâmicas que se reelaboram, recombina e articulam mutuamente, embora sejam independentes entre si. À medida que o leitor se envolve no ato de ler, ele coloca em prática as habilidades de **decodificar o código**, acionando conhecimentos sobre o funcionamento da língua escrita; de **participar da compreensão**, relacionando o conteúdo do texto com experiências prévias; de **usar funcionalmente**, reagindo ao texto de acordo com a expectativa daquela prática social; e de **analisar criticamente**, posicionando-se criticamente, reconhecendo que cada texto é um ponto de vista e que não é ideologicamente neutro (LUKE; FREEBODY, 1990; SCHLATTER, 2009).

No Celpe-Bras, as habilidades de leitura são avaliadas por meio da produção escrita, na qual o candidato deve demonstrar a compreensão do texto de insumo. Ou seja, nas tarefas que integram leitura e produção textual da PE do exame, o candidato necessita, além de mobilizar as quatro habilidades de leitura citadas acima, ter competências para desenvolver um texto escrito que demonstre a leitura e atenda aos propósitos da escrita delineados no enunciado das tarefas. A seguir, apresentamos um exemplo da tarefa 3 da PE.

Figura 3 – Tarefa 3 da Parte Escrita de 2016-2

2016/2 **Celpe Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 3 | Cadastro de hortas

Página 6

Você tem um *blog* sobre alimentação saudável e, depois de ler a notícia sobre o cadastro de hortas em São Paulo, decidiu implementar essa iniciativa em sua cidade. Escreva um texto em seu *blog* para pedir a participação dos leitores no cadastramento, explicando o funcionamento e a utilidade das informações coletadas.

À procura de hortas em plena São Paulo



Até esta manhã, eram 25 hortas cadastradas, sendo 20 de caráter “educativo” – ou seja, comunitárias e escolares – e 5 “comerciais” – quer dizer, voltadas à produção e venda de alimentos. O que elas têm em comum? Estão todas dentro do município de São Paulo. Várias – olhem só! – em densas áreas urbanas, como nos bairros de Vila Madalena, Vila Mariana, Pompeia, Mooca, Cambuci, Casa Verde, Sapopemba, Vila Nova Esperança, Jardim Ângela...

Já as “comerciais” inserem-se sobretudo na zona rural paulistana. Desde 2014 a cidade de São Paulo “ganhou” uma área rural, ou melhor, porções do território paulistano situadas principalmente nos extremos sul e leste da capital passaram a ser reconhecidas como “rurais”.

Esses dados vêm de uma pesquisa colaborativa em pleno andamento. Quem souber da existência de alguma horta urbana e orgânica sendo cultivada em algum “rincão”

da metrópole pode entrar na plataforma Wikiversidade e inserir os dados ali. Tem de preencher o nome da horta; se ela é educativa ou comercial; a região em que está; o bairro; o endereço e, por fim, o *link* para contato.

O cadastramento dessas hortas urbanas foi iniciado pela dupla Regiane Nigro, que trabalha no Instituto Kairós, em São Paulo, e a jornalista e uma das fundadoras do grupo Hortelões Urbanos, Cláudia Visoni. Ambas têm forte atuação em duas dessas hortas comunitárias, a da Praça das Corujas, na Vila Madalena, e também a Horta do Ciclista, na Praça do Ciclista, região da Avenida Paulista. Regiane atua também, pelo Kairós, com os produtores da Associação de Agricultores da Zona Leste, que contempla bairros do extremo leste paulistano, como São Mateus, Guaianases, Cidade Tiradentes e São Miguel Paulista.

A ideia de cadastrar todas as hortas paulistanas – embora a plataforma também

permita a inclusão de cultivos urbanos em todo o país – já era antiga, conta o blog Regiane Nigro. E foi finalmente posta em prática há cerca de 15 dias, para servir como base de uma reunião que Cláudia Visoni teria com pessoas de outros países interessadas no tema hortas urbanas. “Foi este o motivador para o projeto começar a andar”, conta Regiane.

Se inicialmente, há cerca de cinco anos, quando a “onda” do resgate urbano da natureza começou aqui em São Paulo, as hortas cultivadas em plena metrópole eram raras, agora pode-se dizer que elas já têm animado bastante gente disposta a inserir mais verde na selva de pedra. Na maioria dessas hortas, dá-se preferência ao cultivo orgânico.

Regiane conta que a plataforma é importante também para estimular os produtores comerciais orgânicos a incluírem suas hortas ali e se tornarem mais conhecidos – afinal, muitos deles fornecem, em sistema de compra coletiva, alimentos para a metrópole, entre outras formas de comercialização.

Atualmente, nos “rincões” paulistanos, um dos grupos mais organizados de agricultores – e que já estão lá no cadastro – é o da Cooperapas, no extremo sul da capital, região de Parelheiros. Outra grande utilidade deste cadastro é atrair mais pessoas interessadas em se tornarem voluntárias no cultivo orgânico em plena cidade. Cláudia Visoni, ao chamar, num grupo do Facebook, as pessoas a colaborarem com o cadastramento das hortas, comenta que o registro é “importante para estudos acadêmicos, para os voluntários acharem uma horta perto de casa, para possíveis apoiadores (dos agricultores profissionais) e também para quem não está no Facebook”. E continua: “Quem puder, ajude a incluir as hortas que ainda não estão na tabela. Quem quiser criar tabelas para outras cidades, é só fazer. A plataforma Wiki é totalmente colaborativa e fácil de editar”.

Disponível em: <http://vida-estilo.estadao.com.br/blogs/alimentos-organicos/a-procura-de-hortas-em-plena-sao-paulo/>. Acesso em: 5 ago. 2016 (adaptado).

Fonte: Acervo Celpe-Bras.

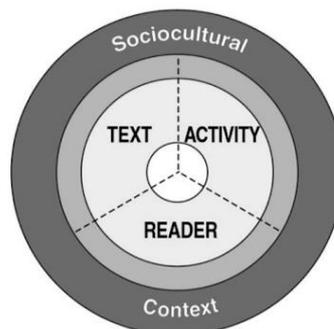
Essa tarefa foi aplicada em 2016 e, como é possível observar, é composta por um título e um enunciado que dá o comando de escrita com base na leitura do texto de insumo fornecido junto à tarefa. Para cumprir adequadamente o que é solicitado no enunciado, o examinando necessita fazer uso de suas habilidades de leitura e de produção textual de modo integrado.

Partindo dessa perspectiva de leitura como uma interação entre texto, tarefa e leitor, que possibilita a participação de sujeitos em práticas sociais coconstruídas social e situadamente com os demais membros de suas comunidades, passamos a discutir o processo de compreensão leitora e o impacto da complexidade textual nesse processo, um dos objetivos desta pesquisa.

3.3 LEITURA E COMPLEXIDADE TEXTUAL

Na esteira do crescente interesse no tema da leitura como prática social desde a virada do século, em 2002, nos Estados Unidos, foi publicado o relatório do *RAND Corporation* (RAND, 2002). Nesse relatório, foi introduzida a ilustração abaixo, que representa o processo de compreensão leitora, envolvendo quatro elementos: leitor, texto, tarefa e contexto sociocultural. A figura mostra que a interação texto-tarefa-leitor é central para a noção de compreensão leitora, na qual esses três componentes simultaneamente influenciam e são influenciados entre si, no âmbito de um contexto sociocultural específico (VALENCIA *et al.*, 2014).

Figura 4 – Modelo de Compreensão Leitora RAND



Fonte: RAND Reading Study Group, 2002.

Seguindo do modelo RAND, mas com interesse em refletir sobre modos de avaliar o produto da interação entre esses componentes, ou seja, o desempenho na leitura, o Modelo de Medida de Dificuldade Textual em Três Partes (NATIONAL GOVERNORS ASSOCIATION

CENTER FOR BEST PRACTICES, COUNCIL OF CHIEF STATE SCHOOL OFFICERS, 2010), desenvolvido para o Common Core State Standards for English Language Arts (CCSS-ELA)⁵⁵, une tarefa-leitor em um mesmo eixo, enquanto decompõe o texto em duas partes distintas, resultando no triângulo ilustrado abaixo.

Figura 5 – Modelo de Medida de Dificuldade Textual em Três Partes



Fonte: Appendix A Common Core State Standards. NATIONAL GOVERNORS ASSOCIATION CENTER FOR BEST PRACTICES, COUNCIL OF CHIEF STATE SCHOOL OFFICERS, 2010.

Mesmo que o contexto sociocultural não esteja representado na figura proposta para o CCSS-ELA, os três elementos essenciais permanecem: texto-tarefa-leitor. No entanto, o enfoque no texto é privilegiado nessa perspectiva, ocupando dois terços da área total da figura e combinando abordagens qualitativas e quantitativas de análise. Esse modelo de três partes chama a atenção para a importância de analisar a complexidade dos textos e sua adequação aos diferentes graus de proficiência leitora nos níveis escolares dos Estados Unidos da América. Para tal, parte de um entendimento de pesquisa sobre a dificuldade textual dos textos, unindo uma análise qualitativa dos textos conduzida por leitores humanos especialistas da área, uma análise quantitativa tipicamente realizada por meio de programas de computador, além da dimensão associada ao leitor e às particularidades da tarefa, tratada por professores especializados, levando em conta as variáveis relacionadas ao leitor e às especificidades da tarefa.

Nesta tese, focalizamos principalmente a esfera quantitativa, pois lidamos com dados numéricos e análises tipicamente realizadas por meio de programas computacionais. Contudo, é fundamental destacar que esse tipo de pesquisa também envolve análises qualitativas e é

⁵⁵ O título original utiliza o termo “complexidade textual”, mas para fins de equivalência com a terminologia adotada neste trabalho, traduzimos para “dificuldade textual”. No original, lê-se: A Three-Part Model for Measuring Text Complexity of Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects.

conduzido por especialistas da área. Os programas são desenvolvidos por profissionais com experiência em Ciência da Computação e Linguística, sendo necessário o conhecimento da língua, como coesão, coerência, léxico, sintaxe e semântica. Além disso, uma vez criados, os programas são utilizados por pesquisadores que devem ser capazes de compreender o funcionamento dos algoritmos que geram os dados numéricos, fazer inferências qualitativas a partir dos resultados obtidos e relacioná-los aos textos analisados. As três partes são descritas na figura abaixo, conforme consta no documento do CCSS-ELA.

Figura 6 – Descrição do modelo de Medida de Dificuldade Textual em Três Partes



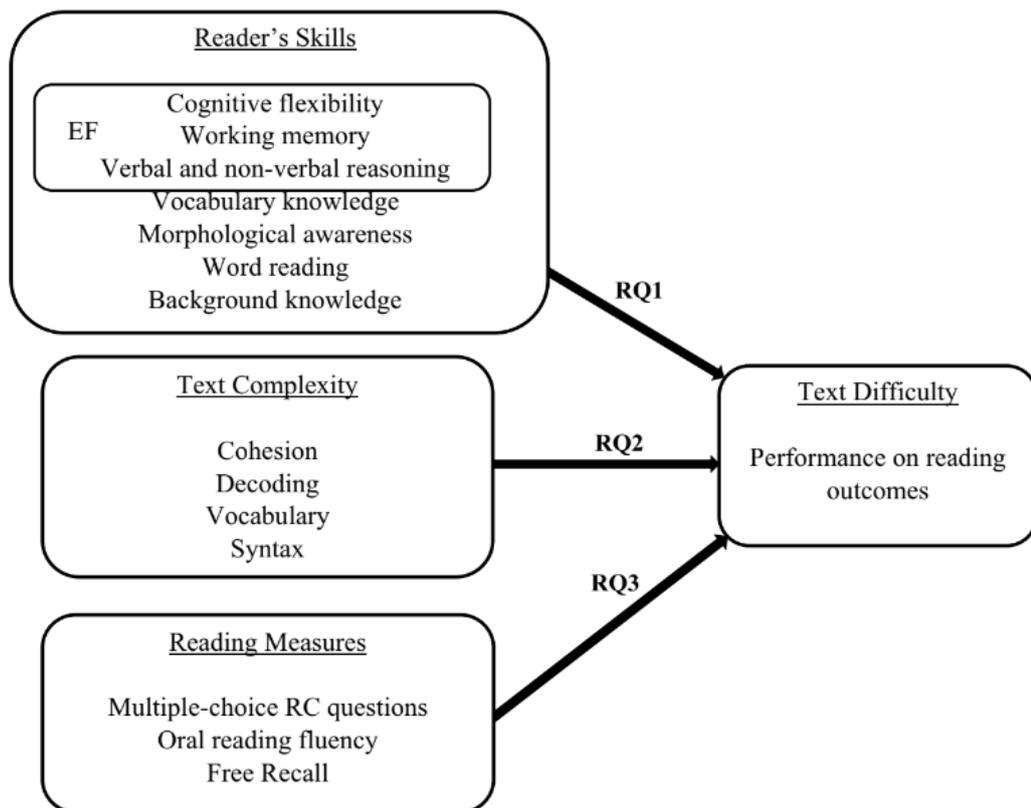
Fonte: Appendix A Common Core State Standards (2010).

A figura ilustra o modo como cada parte da análise é igualmente importante e complementa o todo da investigação da dificuldade textual. Cada uma das três partes desempenha um papel essencial e, juntas, compõem uma abordagem abrangente da análise. Sem considerar todas essas partes, uma metodologia de análise não conseguiria abranger todos os elementos que se interligam para determinar o nível de desafio de compreensão de um texto por um determinado público-alvo, com propósitos de leitura específicos. Entretanto, a figura

do CCSS-ELA fornece apenas diretrizes vagas e genéricas sobre como proceder para investigar esses três fatores e avaliar a dificuldade textual (PEARSON; HIEBERT, 2014; VALENCIA *et al.*, 2014). É importante notar que grande parte das pesquisas se concentram apenas em uma das três partes e desenvolvem métodos próprios que contribuem de forma geral para a compreensão da dificuldade textual.

Um exemplo de pesquisa focada na interação entre as variáveis específicas do leitor e da tarefa, além do impacto da complexidade textual no desempenho na leitura, é a pesquisa conduzida por Spencer *et al.* (2019). A figura a seguir ilustra as variáveis que os autores examinaram em cada um desses níveis e que incluem diferentes aspectos relacionados a habilidades do leitor, à complexidade textual e ao instrumento de avaliação utilizado, todos eles podendo causar variação na dificuldade textual ou no desempenho do leitor.

Figura 7 – Relações entre leitor, texto e tarefa



Fonte: Spencer *et al.*, 2019, p. 3⁵⁶.

⁵⁶ Tradução nossa: Habilidades do leitor: funções executivas, flexibilidade cognitiva, memória de trabalho, raciocínio verbal e não verbal; conhecimento de vocabulário; consciência morfológica; decodificação de palavras; conhecimentos prévios. Complexidade textual: coesão, decodificação, vocabulário, sintaxe. Instrumentos de avaliação de leitura: itens de múltipla escolha; fluência de leitura em voz alta; recordação livre. RQ: perguntas de pesquisa. Dificuldade textual: resultado do desempenho em leitura.

Spencer *et al.* (2019) coletaram dados com a participação de 181 estudantes adolescentes falantes de inglês e conduziram análises estatísticas visando prever o desempenho na leitura por meio da manipulação de 11 variáveis associadas aos estudantes e suas funções cognitivas. Eles também consideraram 4 características de complexidade textual presentes nos textos de insumo, bem como 3 tipos diferentes de tarefas de leitura. Além de examinar conhecimento prévio, lexical, morfológico e capacidade de decodificação, os autores analisaram funções executivas, avaliando flexibilidade cognitiva, memória de trabalho e raciocínio verbal e não verbal dos participantes. Entre outros resultados, os pesquisadores corroboram estudos anteriores que indicam que a relação entre medidas de complexidade textual, como coesão e estrutura do texto, e características particulares do leitor, como funções cognitivas e repertório lexical, são preditivas do desempenho na compreensão leitora. Os achados da pesquisa também sugerem que o desempenho na leitura pode ser influenciado tanto pela complexidade textual do texto quanto pelo tipo de questão proposta. Assim, os autores defendem que os estudantes devem ser avaliados por meio de diferentes instrumentos, incluindo diferentes textos e tipos de tarefas. Corroborando os autores, Winke e Brunfaut (2020), ao refletirem sobre o que os testes de língua pretendem avaliar, declaram que

[u]m ponto importante a ser lembrado, conforme demonstrado pela ciência, é que nenhuma medida única de um traço latente é uma medida perfeitamente precisa e confiável desse traço. A recomendação geral é testar mais de uma vez (já que mais estimativas representarão melhor o construto subjacente) e quantas vezes forem viáveis e benéficas. No entanto, o número de vezes a ser testado deve ser determinado por informações que sejam baseadas em questões educacionais, cognitivas, sociais e éticas (WINKE; BRUNFAUT, 2020, p. 3)⁵⁷.

Ainda segundo Spencer *et al.* (2019), para que a compreensão leitora possa ser avaliada, a decodificação do texto é fundamental. Contudo, se a decodificação não for o foco principal da avaliação, essa deve ocorrer de maneira eficiente e sem demandar esforço excessivo, pois um esforço intenso demais pode interromper o processo mental do leitor, prejudicando seu desempenho nas outras habilidades envolvidas no processo da leitura. Sendo assim, os autores reiteram a importância de que a interpretação dos resultados aferidos leve em consideração as habilidades leitoras que estão sendo testadas, o formato do teste e o nível de complexidade textual do texto de insumo.

⁵⁷ No original, lê-se: An important point to keep in mind, as science has shown, is that no single measure of a latent trait is ever a perfectly accurate and reliable measure of that trait. The general call is to test more than once (as more estimates will better represent the underlying construct), and as many times as would be practical and beneficial. However, the number of times to test should be driven by information that is educationally, cognitively, socially, and ethically construed.

Como já explanado no capítulo introdutório, a complexidade textual é entendida neste estudo como um construto multifacetado e multidimensional. A dificuldade textual emerge da interação dinâmica entre diversos fenômenos linguísticos que ocorrem em níveis textuais e de interação com o leitor e os propósitos de leitura. Essa perspectiva está alinhada com as discussões apresentadas por Spencer e colaboradores (2019), que estabelecem uma distinção entre complexidade textual do texto de insumo e a dificuldade textual, e afirmam que o desempenho na leitura é o resultado da interação entre o leitor, o texto e a tarefa.

Tendo como base a distinção proposta por Spencer *et al.* (2019), nesta pesquisa, fazemos menção a dois níveis de análise dos textos. O primeiro nível é a *complexidade textual*, que diz respeito estritamente às características internas do texto relacionadas aos aspectos linguísticos a nível textual (no caso deste estudo, representados pelas métricas do NILC-Matrix, apresentadas no capítulo 6). O segundo nível trata da *dificuldade textual*, que se refere à interação entre o leitor, o texto e a tarefa, indicando o quão desafiador o texto pode ser em contextos específicos de leitura.

Não é uma tarefa simples diferenciar com precisão todos os termos empregados por pesquisadores que estudam a complexidade textual, especialmente considerando as diversas áreas do conhecimento que investigam a temática da leitura (*e.g.*, Linguística, Psicolinguística, Psicologia, Ciência da Computação, Educação e Design). No entanto, neste estudo, enfatizamos a distinção entre *complexidade* e *dificuldade*, uma vez que isso é importante para que o leitor possa acompanhar a construção dos argumentos apresentados. Cabe ressaltar que termos como “leitabilidade”, “inteligibilidade” e “compreensibilidade” também têm sido usados em pesquisas da língua portuguesa que examinam o grau de desafio que um texto representa para a compreensão do leitor (SOUZA; HÜBNER, 2015). Ademais, é relevante destacar que a “legibilidade”, diferentemente dos outros termos mencionados, refere-se aos aspectos tipográficos dos textos, como tamanho e tipo de fonte, espaçamentos, layout, ilustrações, cor, entre outros, que tornam o texto mais ou menos confortável para ser lido (LOURENÇO, 2016)⁵⁸.

Neste estudo, também abordaremos um terceiro nível de complexidade, ao qual denominamos *complexidade de tarefa*. Esse nível se refere às tarefas que compõem exames e aborda o quão desafiadoras essas tarefas são, independentemente de seu tipo e composição. A análise da complexidade da tarefa leva em consideração os critérios de avaliação, os propósitos

⁵⁸ Para uma leitura mais aprofundada sobre o intercâmbio entre os termos legibilidade e inteligibilidade nos estudos brasileiros, sugerimos a leitura de Santos (2013).

do teste e o público-alvo projetados pelos elaboradores. Ou seja, essa dimensão de complexidade tem potencial efeito no grau de dificuldade envolvido em responder corretamente a uma tarefa e, conseqüentemente, na possibilidade de distinguir diferentes níveis de proficiência dos examinandos, de acordo com o construto teórico do exame.

A esse respeito, cabe observar que, entre as justificativas para a opção de um único exame para testar distintos níveis de proficiência, o Documento Base considera que “as situações de interação por meio da linguagem não são, na sociedade, classificadas por nível” (p. 33). Para o Celpe-Bras, “a diferença entre os níveis certificados espelha o desempenho do participante na realização das tarefas”, isto é, “o que distingue os níveis de proficiência são os recursos mobilizados pelo participante nas situações de interação propostas” (BRASIL, 2020, p. 33). Tendo em vista que as tarefas 3 e 4 do exame propõem uma produção escrita pressupondo a leitura de um texto autêntico com um propósito específico proposto, pode-se afirmar que a complexidade dessas tarefas resulta da relação entre os traços linguísticos do texto de insumo (complexidade textual), a capacidade leitora demandada a partir da ação proposta ao participante no enunciado (dificuldade textual) e a produção do texto solicitado. Pode-se supor, portanto, que a complexidade da tarefa poderia ser calibrada de diferentes modos, seja pela gradação da complexidade do texto de insumo, seja pela gradação da capacidade leitora projetada (tendo-se em vista a leitura exigida pela ação proposta no enunciado), seja pelo texto a ser produzido, em diversas combinações desses elementos. O que não fica claro, no entanto, é de que modo esses elementos são combinados para estabelecer gradações de complexidade da tarefa que sejam preditivas para distinguir níveis de proficiência. Do mesmo modo, embora esteja explicitado que “o objetivo da produção escrita pressupõe a compreensão do texto de [insumo]” (BRASIL, 2020, p. 30), não temos conhecimento de estudos de tarefas do Celpe-Bras que tenham analisado de que modo diferentes propósitos de leitura aumentam ou diminuem a dificuldade de leitura de um mesmo texto ou de que modo diferentes complexidades textuais poderiam interferir nas respostas considerando um mesmo propósito de leitura, a ponto inclusive de, caso o texto de insumo seja altamente complexo, tornar-se potencialmente inacessível para, por exemplo, um usuário com nível de proficiência intermediário, público-alvo deste exame. É sobre esse último ponto que esta tese se debruça, entendendo que cabe analisar cada um dos aspectos que se inter-relacionam no exame, começando por compreender de que modo se pode medir a complexidade textual e como isso pode se relacionar com outros elementos para estabelecer gradações preditivas de dificuldade textual.

Na figura abaixo ilustramos as três definições, entendendo que neste estudo:

- complexidade textual é a inter-relação entre traços linguísticos em nível textual;
- complexidade da tarefa é o potencial da tarefa em distinguir níveis de proficiência, com base em um planejamento informado de inter-relações entre os aspectos que a compõem (texto de insumo, propósito de leitura, propósito de escrita);
- dificuldade textual é a medida que resulta da inter-relação entre complexidade da tarefa (texto, tarefa) e leitor.

Figura 8 – Três níveis de análise de complexidade



Fonte: Elaborado pela autora.

No decorrer deste texto, os três termos são utilizados. Entretanto, como ilustra a figura, é imperativo salientar que não são intercambiáveis; são empregados de maneira consciente, correspondendo às especificidades do objeto em discussão.

No Brasil, a Guten Educação⁵⁹ coordenou um estudo intitulado *Complexidade textual: contribuição do tema para a progressão dos estudantes brasileiros em leitura* (GUTEN, 2018), que apontou que a temática da complexidade textual tem sido negligenciada nas discussões sobre leitura no contexto educacional brasileiro. O relatório do estudo ressalta que os resultados de testes de leitura, como o PISA e a Prova Brasil, revelam que 70% dos estudantes concluem o Ensino Fundamental sem possuírem habilidades básicas de compreensão leitora em língua portuguesa. Diante desse panorama, o documento sugere que o sistema educacional brasileiro, ao assumir a perspectiva discursiva no ensino de língua e ter o texto como objeto central no ensino de leitura, deve realizar análises da complexidade textual como forma de selecionar textos que garantam uma progressão leitora dos estudantes. Essa progressão deve iniciar com textos em um nível simples e ir incorporando elementos textuais complexificadores para desafiar o estudante sem desencorajá-lo do processo de aprendizagem da leitura.

A Guten Educação contou com a colaboração de especialistas das áreas de Psicopedagogia, Psicologia, Pedagogia e Linguística para conduzir o estudo. No entanto, o relatório final concentra-se principalmente nos resultados de pesquisas que investigam o processo de leitura por meio de análises de ondas cerebrais e movimentos oculares, obtidos por meio de aparelhos que geram neuroimagens, como o eletroencefalograma, e *eyetrackers*, que capturam os movimentos dos olhos durante a leitura.

Para discutir as variáveis inerentes ao texto, o relatório cita estudos que analisam, em nível frasal, a complexidade textual a partir de elementos que provocam estranhamento ao leitor e demandam maior ou menor gasto de processamento cognitivo e, para tanto, analisam fenômenos como combinação de palavras e estruturas sintáticas (FRANÇA, 2018; MOTA, 2018; MAIA, 2018; RODRIGUES, 2018). Tais estudos apontam níveis de complexidade textual baseados, portanto, na percepção de estranhamento e no gasto cognitivo, entendendo a leitura como circunscrita na materialidade de pequenos trechos de texto, em geral frases, sem levar em conta que, por exemplo, essas frases podem estar combinadas de modo a construir uma coerência textual com base em um propósito de comunicação. Isso pode resultar na alteração da aparente dificuldade de leitura quando uma frase é lida isoladamente ou quando é inserida em uma situação em que tanto o propósito do texto quanto o propósito da leitura sejam

⁵⁹ <https://gutennews.com.br/>

explicitados e levados em consideração. Sob essa ótica, o relatório Guten afirma que “uma tarefa mais simples torna o nível de compreensão leitora exigido mais simples, porém não altera a complexidade do texto, que é inerente a ele” (GUTEN, 2018, p. 21).

Souza e Hübner (2015) diferenciam as abordagens que investigam o nível de dificuldade da leitura a partir de processos internos que ocorrem na mente do leitor no momento em que a leitura está sendo processada e as teorias que se concentram no produto da compreensão leitora. A primeira categoria, na qual o relatório Guten se insere, se interessa por fatores como funções executivas, atenção, reconhecimento de palavras e estruturas sintáticas, memória de trabalho e monitoramento⁶⁰. Seu foco é descrever as operações cerebrais e os processos cognitivos que ocorrem no exato momento em que o processo de leitura está ocorrendo, independente do contexto. Nessa perspectiva, também se encontram os estudos da área da Psicolinguística, que, entre outras variáveis, operacionaliza a habilidade de leitura a partir da medida de tempo de reação ao texto, ou seja, medindo quanto tempo o leitor leva para demonstrar compreensão do significado de palavras ou frases (RIGATTI; FONTES, 2019). Já a segunda categoria busca interpretar o produto da leitura de textos integrais, avaliando assim a compreensão leitora enquanto resposta do sujeito leitor ao texto lido. Isso envolve analisar o desempenho de indivíduos específicos frente a determinados propósitos de leitura de um texto autêntico e completo, abordagem de estudo à qual esta pesquisa se filia.

Contudo, vale a importante ressalva de que não estamos com isso desconsiderando os estudos que adotam outras concepções teóricas de leitura e dificuldade textual, tampouco pretendemos argumentar que o viés adotado neste trabalho seja preferível. Independentemente da abordagem adotada, reconhecemos que as diferentes pesquisas fornecem subsídios complementares para o avanço da ciência. Apesar de ainda estarmos longe de desvendar inteiramente os múltiplos processos mentais que ocorrem simultaneamente durante o processo de leitura, as pesquisas que utilizam neuroimagem estão revelando processos fisiológicos que ocorrem enquanto lemos. Essas pesquisas têm sido especialmente valiosas no contexto clínico, permitindo o diagnóstico de dificuldades específicas e facilitando intervenções educacionais personalizadas para sujeitos que enfrentam desafios na compreensão leitora, como afásicos, disléxicos ou pessoas com doenças neurodegenerativas. Contudo, é importante lembrar que essas pesquisas não abordam a leitura sob o construto de interação social, portanto seus

⁶⁰ As pesquisas indicadas no relatório Guten que realizaram experimentos de monitoramento de leitura por meio de *eyetracking* observam os locais das orações onde os leitores focam o olhar, tais pesquisas buscam analisar a diferença entre o processamento de períodos longos com orações subordinadas e períodos curtos com orações coordenadas ou analisar o processamento de leitura de estruturas sintáticas que utilizam voz passiva.

resultados não são suficientes para se traçarem inferências que levam a tomadas de decisões relacionadas ao ensino e à avaliação da leitura em contextos escolares ou acadêmicos. Nessas esferas, estamos tratando de leitura a partir do construto sócio-histórico de interação entre indivíduos por meio da língua escrita. Sob essa ótica, a verificação da compreensão leitora está intrinsecamente integrada com uma resposta ao texto por meio de uma produção textual (oral, escrita ou multimodal), e inúmeros fatores extramentais estão imbricados durante o processo da leitura.

De maneira geral, as concepções de texto e leitura coadunam-se com a forma como o pesquisador compreende o processo de investigação da dificuldade textual. Nesse sentido, para levar a cabo a tarefa que nos propomos nesta tese, adotamos uma abordagem teórico-metodológica que não considera a observação de processamentos cerebrais e cognitivos, uma vez que tais análises não nos forneceriam os dados necessários para alcançar nossos objetivos de pesquisa. Por esse motivo, adotamos a LC, uma vez que entendemos que essa base teórica nos permite observar e interpretar padrões de fenômenos linguísticos através da análise empírica de textos autênticos utilizados para a leitura em contextos específicos de interações sociais, permitindo que os resultados estejam em consonância com o construto teórico do exame Celpe-Bras, que abordamos no capítulo a seguir.

4 O CELPE-BRAS

Este capítulo é dedicado a contextualizar o leitor no escopo do objeto de investigação desta tese: os textos de insumo que compõem as tarefas que buscam avaliar leitura e produção textual de modo integrado no exame Celpe-Bras. Nesse contexto, apresentaremos o referencial teórico que norteia a elaboração do exame, descrevendo as concepções de avaliação de leitura e de proficiência em nível intermediário.

No Brasil, em 1993, um grupo de especialistas em linguagem recebeu a demanda de criação de um exame nacional padronizado para selecionar estudantes estrangeiros candidatos ao Programa de Estudantes-Convênio de Graduação. Naquele mesmo ano, foi nomeada a primeira comissão técnica para dar início às discussões teóricas, elaboração e testagem piloto dos instrumentos, além da execução de tratamentos estatísticos que informaram sobre a validade e a confiabilidade do teste (SCHLATTER, 1999). Desde então, o Celpe-Bras é o único teste que avalia proficiência em Português como Língua Adicional oficialmente reconhecido pelo governo brasileiro. O Documento Base indica que os elaboradores tinham a motivação de desenvolver um instrumento de avaliação com potencial para redirecionar o ensino de PLA; e

essa motivação foi, juntamente com o atendimento das necessidades do público-alvo inicial – estudantes estrangeiros de graduação e de pós-graduação nas instituições de ensino superior (IES) brasileiras –, responsável pela definição da natureza do Exame, de seus conteúdos, das habilidades avaliadas e dos pesos dados a cada uma delas (BRASIL, 2020, p. 17).

Schlatter *et al.* (2009) explicitam que os materiais didáticos para o ensino de PLA, antes da instituição do Celpe-Bras, priorizavam o ensino de conhecimentos referentes a gramática e vocabulário. O uso de gêneros discursivos e a integração de habilidades no exame teriam o potencial de promover uma mudança para uma perspectiva de ensino de uso da linguagem para realizar ações no mundo e participar de interações sociais mais variadas (SCHLATTER *et al.*, 2009).

A fase de desenho do Celpe-Bras abrangeu um período de cinco anos, desde o estabelecimento do objetivo do exame até o início de sua aplicação oficial. Como mencionado anteriormente, apesar de ter sido desenvolvido para fins acadêmicos, o Documento Base menciona que atualmente o público-alvo vai além dos estudantes interessados em participar da vida acadêmica brasileira. Também abrange estrangeiros que desejam validar diplomas profissionais obtidos no exterior para poder exercer suas profissões no Brasil, indivíduos interessados no registro em entidades de classe profissionais brasileiras (*e.g.*, Conselho Federal

de Medicina) e estudantes lusófonos de países africanos participantes do PEC-G. O documento afirma também que o Celpe-Bras é aceito internacionalmente por empresas e instituições de ensino como comprovação de proficiência em PLA. Ademais, o Documento Base adverte que o exame não é apropriado para pessoas com menos de 16 anos de idade ou com escolaridade inferior ao ensino fundamental brasileiro completo. Pressupõe-se que os candidatos estejam familiarizados com as práticas de letramento de um cidadão escolarizado, ou seja, capazes de participar de práticas sociais que envolvem leitura, escrita e interações orais (BRASIL, 2020, p. 30).

Embora o Documento Base não estabeleça explicitamente o objetivo do teste, a partir da delimitação do público-alvo e dos usos que esse público faz do certificado, é possível inferir o contexto da língua-alvo que o exame tem como objetivo avaliar. De acordo com o Documento Base, “o Celpe-Bras avalia o desempenho em relação ao uso da língua portuguesa em diferentes práticas de linguagem de que seu público-alvo necessita participar, no Brasil ou no exterior” (BRASIL, 2020, p. 26). Além disso, o documento também esclarece que “o Celpe-Bras avalia vários níveis de proficiência, que levam em conta as especificidades e necessidades da situação de uso da língua em contextos diversos” (BRASIL, 2020, p. 27). Nessa visão de uso da linguagem, também se circunscreve a noção de língua e cultura como indissociáveis, pois “ao aprender uma língua, aprende-se também um modo de ser e de viver, que é marcado culturalmente, uma vez que ela passa a ser um vetor de interpretação do mundo para o sujeito que a aprende” (BRASIL, 2020, p. 29). Portanto, é possível inferir um contexto abrangente no qual o objetivo é avaliar o desempenho em distintas práticas de uso da língua portuguesa não específicas, em contextos variados, que ocorrem tanto no Brasil quanto no exterior.

Para operacionalizar o construto teórico do exame e avaliar o desempenho dos participantes, conforme mencionado anteriormente, o Celpe-Bras é composto por uma parte escrita e outra oral, ambas contendo instrumentos específicos que visam a avaliação integrada de compreensão e produção. Na Parte Oral, o construto é operacionalizado por meio de uma interação presencial entre um avaliador e o candidato. De acordo com o Documento Base, a interação tem como objetivo simular uma situação de uso da língua falada, similar àquelas que os participantes podem encontrar fora da situação do teste: a interação “parte de situações de comunicação próximas às reais, procurando simular, na medida do possível, conversas do dia a dia” (BRASIL, 2020, p.41). A Parte Oral não é foco desta pesquisa, mas cabe observar que, embora haja um esforço para que o avaliador interaja com o candidato com base em suas respostas, buscando simular características de uma conversa, a avaliação envolve posições

sociais de maior e menor poder – uma pessoa avalia e a outra é avaliada –, e a interação é, de fato, realizada no formato de uma entrevista guiada por um roteiro de perguntas, resultando em uma interação inflexível e assimétrica⁶¹.

Na Parte Escrita do Celpe-Bras, foco deste estudo, o construto é operacionalizado através de tarefas que solicitam a produção de um texto escrito como resposta a um enunciado e a um texto de insumo. As tarefas são concebidas como um “convite” para que o examinando utilize a língua portuguesa no contexto social indicado pelo enunciado da tarefa, cumprindo os propósitos comunicativos solicitados, ao assumir um determinado papel social e se dirigir aos interlocutores projetados. Com a finalidade de integrar compreensão e produção, esse “convite” circunscreve-se a um evento comunicativo no qual o candidato deve expressar-se por escrito para alcançar os propósitos da tarefa, e “uma das implicações desse procedimento é que os objetivos de compreensão levam em conta os textos usados como insumo e são determinados pelos propósitos de escrita” (BRASIL, 2020, p. 31). Em outras palavras, o comando da tarefa delimita as leituras possíveis do texto de insumo, indicando a leitura preferida e a expectativa de reação ao texto prevista pelos elaboradores. Portanto, não é desejável que o examinando transgrida as regras do “jogo”, que não assuma o papel social designado ou não mobilize os recursos linguísticos adequados de acordo com a demanda social específica descrita no enunciado. Uma das etapas do processo de correção da PE envolve a análise de uma amostra aleatória de respostas de examinandos, com o intuito de verificar se os participantes não fizeram, de um modo geral, interpretações das tarefas que também sejam possíveis, mas que não tenham sido previstas, o que implicará ajustes e refinamentos na grade de avaliação por parte da equipe de correção.

Cada edição do exame é composta por quatro tarefas, e cada uma delas é constituída por um enunciado e um texto de insumo diferentes. A tarefa 4 da segunda edição de 2016, a seguir, ilustra como as situações comunicativas são delineadas no exame.

⁶¹ Para detalhes adicionais sobre a Parte Oral do exame, sugerimos consultar o Documento Base do exame Celpe-Bras (BRASIL, 2020).

Figura 9 – Tarefa 4 da Parte Escrita de 2016-2

2016/2 **Celpe Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 4 | Shopping em casa Página 8

Depois de ler o texto “Shopping em casa”, você decidiu escrever para a seção Cartas do Leitor do jornal O Estado de São Paulo, defendendo seu ponto de vista em relação às questões levantadas por Denise Fraga.

Shopping em casa

— Que bonita esta saia!
— Pois é, estava perdida — ela me respondeu.

— Você perdeu uma saia?

— Sim, dentro do meu armário.

Arrumando as coisas pra mudança, eu achei.

Esta conversinha banal com minha amiga aumentou a minha angústia. Será que preciso de uma mudança pra conseguir achar o que anda perdido diante dos meus olhos? Muitas vezes, tenho a sensação de não ter o que vestir diante do meu guarda-roupa lotado e fico doida cobiçando um novo vestido na vitrine do shopping. No outro dia, comprei uma blusa que tinha amado e, quando fui pendurá-la no armário, vi que sua vizinha de cabide era sua prima. O curioso é que, se fosse a blusa antiga que estivesse na loja, eu a compraria de novo. Mais curioso: ela estava lá, eu sabia que tinha e pouco a usava. A blusa nova me devolvera a velha, que passei a usar com prazer. Mas por que comprei uma nova se havia uma quase igual pendurada? Que fenômeno é esse de precisar do que já se tem? Que cegueira é esta para o que está ao seu lado? Será a galinha do vizinho sempre mais gorda? Adquirir é melhor do que ter?

Não sou exatamente consumista, mas olho a minha casa e, apesar de morrer de amores por meus objetos, sinto que há uma infinidade de coisas que eu realmente poderia viver sem. O curioso é que, se descuidar, se me forem oferecidas, estou arriscada a comprá-las de novo.

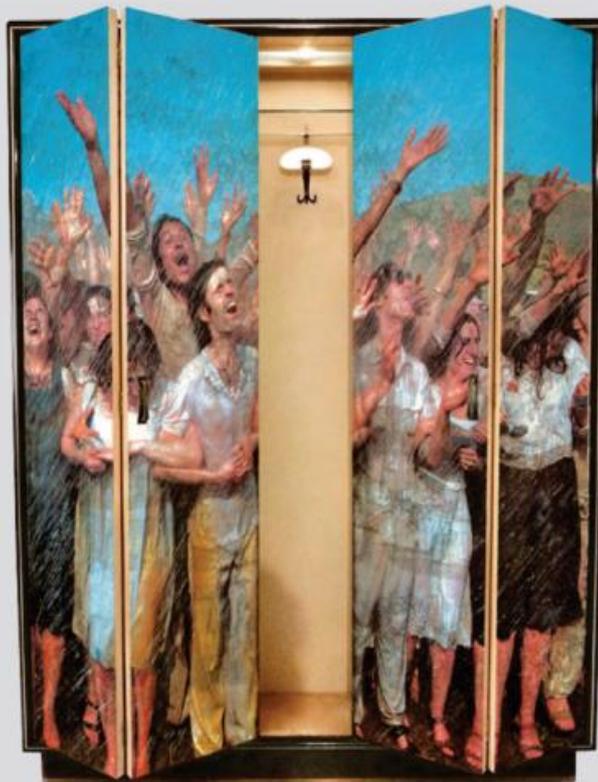


Ilustração: Zé Vitorino

É realmente muito violento o impulso consumista a que fomos doutrinados. Nem precisamos ir ao shopping. Vivemos em uma feira constante. Meu e-mail está lotado de propostas de vendas que nem desconfio como chegaram ali. A internet virou, como previsto, um mar de publicidade que, a um clique, desfalca nossa conta bancária. Tudo pela estrela maior: o dinheiro, o circular da moeda. Cada vez mais produtos, cada vez mais necessidades desnecessárias. Não entendo nada de economia, mas gostaria de viver num mundo com menos coisas.

Gostaria de morar numa casa com menos coisas, ter um armário com menos roupas.

Resolvi fazer um exercício: promover as vitrines a uma exposição de arte. A arte de um artista poderoso e midiático que se chama Mercado e de quem estou longe de poder adquirir uma obra. Mesmo sem ter que mudar de apartamento, vou tirar as coisas de meu armário, colocar sobre a cama para comprar com meu novo olhar os meus velhos vestidos. Liberdade é viver com pouco.

Disponível em: www1.folha.uol.com.br/solunas/denisefraga/2016/02/1739053-shopping-em-casa.shtml. Acesso em: 5 ago. 2016 (adaptado).

Conforme é possível observar, a tarefa é composta por um título, “Shopping em casa”, seguido pelo enunciado, uma crônica com título e imagem, e, adicionalmente, é indicada a fonte do texto de insumo, que informa ser uma adaptação do texto original. O enunciado dessa tarefa propõe a escrita de um texto destinado à seção “Cartas do Leitor” do jornal *O Estado de São Paulo*. A carta do leitor deve ser elaborada com o propósito de se posicionar a respeito das questões levantadas por Denise Fraga, e para isso, o candidato deve ser capaz de identificar, a partir da leitura da crônica, quais são as questões abordadas por ela.

Para responder adequadamente a essa tarefa, o candidato precisa dominar um determinado repertório linguístico em português e ter alguma experiência com a leitura de jornais ou revistas. Essa familiaridade com a esfera jornalística é necessária para que o candidato seja capaz de compreender o significado da expressão “seção Cartas do Leitor” e conhecer marcas linguístico-discursivas típicas de textos publicados nesse espaço jornalístico. Além disso, é fundamental que o candidato compreenda que o texto a ser produzido poderá ter como possíveis interlocutores os outros leitores do jornal, bem como a própria autora do texto e demais colaboradores do veículo de comunicação. Ao projetar essa interlocução pública, o candidato precisa lançar mão de um determinado repertório linguístico específico, sem perder de vista que o texto produzido deve ser apropriado ao contexto de avaliação. Schoffen (2009) esclarece que o Celpe-Bras transpõe gêneros primários das práticas sociais brasileiras para as tarefas, o que também demanda do candidato uma familiaridade com a esfera de avaliação. Isso se dá ao fato que “os textos produzidos pelo candidato fazem parte de um gênero secundário existente e operante no mundo real, que é o gênero prova de proficiência em língua estrangeira” (SCHOFFEN, 2009, p. 135).

A leitura proficiente da crônica também requer que o candidato seja capaz de identificar Denise Fraga como a autora e compreender sua narrativa e opinião por meio da leitura do texto, o qual utiliza recursos linguísticos e estilísticos específicos para construir um discurso reflexivo e crítico sobre o tema do consumismo. Todos esses elementos são levados em consideração durante a elaboração da tarefa e a avaliação dos textos produzidos pelos candidatos. Logo, não basta apenas ler o texto de insumo; é preciso ler de acordo com a expectativa criada pela prática social compartilhada por uma determinada comunidade que existe e é atuante no mundo real, mas que está sendo transposta ao gênero prova.

Na correção da PE, ao invés de medir o conhecimento e a acurácia de uso de recursos linguísticos descontextualizados, o exame utiliza parâmetros de avaliação holística para operacionalizar o construto teórico de uso da língua portuguesa para cumprir os propósitos

sociais solicitados nas tarefas. Essa concepção de avaliação alterou, ao longo dos anos, a atuação e a formação docente em português como língua adicional, especialmente dos profissionais envolvidos diretamente com o exame, visto que se afasta da perspectiva estruturalista (SCARAMUCCI; SCHLATTER, no prelo). Entretanto, algumas pesquisas mostram que, apesar de mais de duas décadas de aplicação do Celpe-Bras, em determinados contextos, professores ainda não mostram uma compreensão sólida do construto de uso da linguagem para desempenhar ações no mundo e, em muitos casos, o ensino descontextualizado de estruturas gramaticais do português foi substituído pelo ensino descontextualizado de estruturas de gêneros do discurso (DIONÍSIO, 2022).

O processo de capacitação dos avaliadores da PE é detalhadamente descrito no Documento Base⁶², no entanto, a capacitação de elaboradores de tarefas da PE é apenas citada como exemplo de efeito retroativo positivo na formação dos profissionais que atuam na área de Português como Língua Estrangeira (PLE)⁶³. Ou seja, é considerada como uma possibilidade de que o processo de elaboração possa influenciar de forma positiva a comunidade envolvida, moldando suas atitudes e práticas no contexto de ensino e avaliação (BRASIL, 2020, p. 20)⁶⁴.

Considerando a ausência de informações sobre o processo de elaboração das tarefas nas especificações públicas do Celpe-Bras, para esta pesquisa, buscamos obter acesso ao documento oficial que orienta o procedimento de elaboração das tarefas. Para isso, foi encaminhada uma solicitação de acesso através da Plataforma Integrada de Ouvidoria e Acesso à Informação da Controladoria-Geral da União⁶⁵. Junto à solicitação foram enviados o resumo do nosso projeto de pesquisa, o comprovante de matrícula da doutoranda no Programa de Pós-Graduação em Letras e um artigo já publicado em revista acadêmica, como desdobramento do processo de doutoramento. No entanto, nossa solicitação de acesso foi recusada pela equipe técnica do Inep após uma subsequente solicitação de recurso. A justificativa apresentada foi a de que “como explicado anteriormente, apenas os documentos publicados são de acesso livre. As discussões internas são confidenciais, realizadas em ambiente sigiloso e restrito, com ônus para o Inep e, ademais, segundo a equipe técnica, não há registro documental dessas

⁶² Os cursos de capacitação são conduzidos para os avaliadores a cada edição do Celpe-Bras, sob a coordenação de profissionais experientes e a supervisão da equipe técnica do Inep. A correção dos textos é realizada por profissionais da área de PLA com formação em Letras, e o desempenho dos avaliadores é monitorado por relatórios diários durante o processo de avaliação.

⁶³ Embora ciente de que os termos revelam diferentes concepções teóricas e políticas, neste trabalho, não diferenciamos os termos PLE e PLA para nos referirmos à área de ensino, avaliação e pesquisa da língua portuguesa para falantes inicialmente socializados em outras línguas. Para uma reflexão terminológica mais aprofundada sobre esses dois termos e outros cinco comumente utilizados no Brasil, ver Bulla e Kuhn (2020).

⁶⁴ Para uma reflexão mais aprofundada sobre o fenômeno de efeito retroativo, ver Scaramucci (2011).

⁶⁵ <https://falabr.cgu.gov.br/>

discussões”⁶⁶. Por essa razão, recorreremos ao documento disponibilizado em Mendel (2019), que obteve acesso a esse material para a realização de sua pesquisa de dissertação, como mostraremos a seguir. Cabe ressaltar que Vicentini (2022) aponta que o órgão responsável pela elaboração do Celpe-Bras tem restrições rígidas aos materiais relacionados ao exame, em particular aos textos produzidos pelos examinandos. Em outras palavras, apesar de a literatura da área de Avaliação de Línguas apontar que as orientações para a elaboração das tarefas devam constar de modo transparente nas especificações do teste, o órgão responsável não possibilitou o acesso a essa área restrita, encerrando o nosso processo de solicitação alegando a inexistência do documento⁶⁷.

Em relação à inexistência desse registro, é válido salientar que Weigle (2004) já advertia que especificações cuidadosas sobre a escolha dos textos de insumo e do processo de elaboração de tarefas precisam ser registradas por escrito a fim de minimizar as diferenças entre as edições. Green e Fulcher (2020, p. 75) contribuem ao afirmar que “é importante não só manter documentos detalhados e claramente escritos que definam os procedimentos a serem seguidos, mas também manter registros sobre até que ponto cada etapa do ciclo funcionou em cada edição”⁶⁸. Mais recentemente, Shin (2021) enfatiza que respostas às indagações “quem são os elaboradores das tarefas?”, “quem treina esses indivíduos?”, “quais imagens vêm à nossa mente quando pensamos em elaboradores de tarefas?”, “o que fazem os elaboradores?”, “análises estatísticas estão sendo feitas de modo a oferecer *feedback* aos elaboradores para o aprimoramento do processo?” fornecem subsídios para fortalecer os argumentos de validade dos testes.

Dentre os documentos públicos do Celpe-Bras acessíveis a nós, encontramos o edital mais recente de chamada pública para a seleção e cadastramento de colaboradores para compor a Comissão Técnico-Científica do Celpe-Bras (CTC), publicado em 25 de julho de 2022⁶⁹. No artigo 10 do referido edital, “Das obrigações e responsabilidades”, inclui o encargo da CTC em prestar apoio à elaboração e validação dos instrumentos de avaliação, bem como apoiar e subsidiar as distintas fases do processo de elaboração de tarefas. Dessa maneira, nossa

⁶⁶ Protocolo número 23546.027198/2022-21, processado no dia 19 de abril de 2022.

⁶⁷ A autora ressalta que tais restrições acarretaram prejuízos irreparáveis ao desenvolvimento de sua pesquisa, mesmo tendo como objetivo analisar evidências para fortalecer o argumento de validade do exame. Contudo, Vicentini recebeu autorização para acessar as respostas dos examinandos durante um período de uma semana, diretamente no ambiente sigiloso e restrito no qual os documentos relativos ao exame são processados.

⁶⁸ No original, lê-se: it is important not only to have detailed and very clearly written documents setting out the procedures to be followed, but also to keep records of how well each step in the cycle has functioned on each occasion.

⁶⁹ Edital nº 65, de 25 de julho de 2022. <https://in.gov.br/web/dou/-/edital-n-65-de-25-de-julho-de-2022-417804839>

interpretação é de que a responsabilidade da CTC reside em oferecer suporte ao processo de elaboração, já que o documento não afirma que a equipe seja encarregada pela elaboração das tarefas. Esse entendimento é corroborado pelo edital de chamada pública para o cadastramento de docentes e pesquisadores visando à composição do Banco de Colaboradores do Inep, publicado em 28 de junho de 2012, o qual estabelece como responsabilidade da equipe elaborar, revisar e corrigir itens para o Celpe-Bras⁷⁰. Isso posto, causa estranhamento que a comunidade não tenha informação pública e transparente sobre o perfil dos membros do Banco de Colaboradores encarregados pela elaboração das tarefas. Tampouco há informações sobre o processo de elaboração das tarefas ou a existência de monitoramento do desempenho desses colaboradores.

Shin (2021), citando Kim *et al.* (2010), diz que tipos de personalidade e o *status* de uso da língua-alvo não representam qualificações soberanas dos elaboradores de questões dos testes, mas que definir certas características é fundamental para compor e capacitar uma equipe, além de facilitar discussões construtivas entre os responsáveis pela elaboração. Por outro lado, o Documento Base (BRASIL, 2020), além de minuciosamente elucidar ao leitor somente o processo de avaliação do exame, apresenta apenas o perfil dos avaliadores, que são

[...] professores e pesquisadores da área de PLE, selecionados pela empresa contratada para operacionalizar o Exame. Esses professores devem ter formação em Letras, em nível de graduação e de pós graduação, experiência no ensino de português para estrangeiros e experiência em pesquisa na área de PLE e/ou de avaliação (BRASIL, 2020, p. 71).

Shin (2021) ressalta que comissões de avaliação frisam a necessidade de “procedimentos claros de elaboração de tarefas, guias ou especificações que orientem a elaboração de tarefas de modo que o processo linear (ou iterativo) de seleção de textos de insumo, edição, construção da tarefa e revisão seja cumprido com êxito” (2021, p. 347)⁷¹. A autora vai além ao afirmar que “elaboradores de tarefas são simultaneamente oprimidos (mal remunerados, contratos informais, pouco reconhecimento de suas habilidades) e opressores (selecionando e decidindo o que e como os examinandos serão avaliados)” (2021, p. 349)⁷². Por essa razão, Shin (2021) destaca a importância de as pesquisas abordarem esse tópico,

⁷⁰ Edital nº 4, de 28 de junho de 2012. <https://in.gov.br/web/dou/-/edital-n-65-de-25-de-julho-de-2022-417804839>

⁷¹ No original, lê-se: explicit item development procedures, item construction guidelines, or test specifications to guide item writers to successfully complete the linear (or iterative) processes of source text selection, editing, item writing, and revising.

⁷² No original, lê-se: Item writers are both oppressed (underpaid, casual contracts, little recognition of their skills) and oppressors (selecting and deciding what test takers will be tested on and how).

considerando que a descrição das atribuições dos elaboradores de tarefas é uma das áreas menos articuladas no campo da Avaliação de línguas.

Conforme Mendel (2019), as especificações que orientam a elaboração das tarefas da PE do Celpe-Bras existem, mas não estão disponíveis nos documentos oficiais do exame. A versão do documento “Especificações para a elaboração de tarefas do exame Celpe-Bras - Parte Escrita” (anexo A), consultada por Mendel (2019), foi redigida pela CTC em 2012, com revisões e atualizações realizadas nos anos 2013 e 2016. A existência do documento e as revisões documentadas evidenciam que a CTC está atenta à importância do processo de elaboração das tarefas e à responsabilidade de assegurar que as versões das provas sejam similares a cada edição, como preconizado pela literatura para salvaguardar a validade do teste (WEIGLE, 2004; SCARAMUCCI, 2016; GREEN; FULCHER, 2020; SHIN, 2021). No entanto, apesar de o documento conter uma seção específica sobre a seleção dos textos de insumo e outra seção sobre as características desses textos, as orientações são limitadas e não incluem um detalhamento dos elementos linguísticos relacionados à complexidade textual dos textos selecionados. A única diretriz que aborda de maneira tangencial a complexidade textual está relacionada à extensão dos textos de leitura, recomendando que seja “desejável que o texto utilizado como insumo para cada tarefa não exceda uma página” (MENDEL, 2019, p. 181), conforme pode ser constatado abaixo.

Quadro 2 – Especificações para a elaboração de tarefas do Celpe-Bras - Parte Escrita

II. Seleção dos textos-base

- Evitar textos que exijam uma grande carga de conhecimento cultural prévio.
- Evitar temas que possam ser tabus para algumas culturas ou sujeitos (exemplo: morte, tragédias, guerras, pena de morte);
 - Não utilizar textos que reforcem estereótipos. Por outro lado, textos que os questionem podem ser utilizados de maneira produtiva para a elaboração de tarefas.
- Evitar textos que possam privilegiar certos grupos, de forma que a leitura possa ser dispensada e que a tarefa possa ser realizada com base em conhecimentos prévios (textos sobre tratamentos de doenças ou problemas de saúde, em que os examinandos médicos nem precisem ler o texto para realizar a tarefa, por exemplo). Não se trata apenas do tema, mas de como o tema é tratado. Buscar textos em que os assuntos sejam tratados de maneira original, não previsível.
- Evitar textos que tratem de episódios datados (circunscritos em um momento específico), que ficam desatualizados facilmente, perdendo seu sentido após poucos meses.
- Buscar textos que possibilitem a criação de uma situação-problema para ser resolvida, criticada ou discutida pelo examinando.

III. Características dos textos-base

A Tarefa I deve ter como insumo um segmento de vídeo, a Tarefa II um segmento de áudio e as Tarefas III e IV devem ter como insumo texto(s) escrito(s).

- Extensão do vídeo/áudio: 2' a 2'30" (se o vídeo/áudio original for maior, os cortes devem ser indicados e não devem prejudicar a qualidade e a unidade da imagem e do texto).

- Qualidade do vídeo/áudio: a dicção deve ser clara e o volume das vozes adequado. Os ruídos não devem prejudicar a compreensão/audição do texto (se houver ruídos, o material deve passar por filtro ANTES de se fazer a tarefa).

- Se houver mais de uma pessoa falando, como no caso de entrevistas, deve-se cuidar para que não haja muita oscilação de volume entre os falantes ou interposição de vozes.

- Extensão do(s) texto(s) escrito(s): é desejável que o texto utilizado como insumo para cada tarefa não exceda uma página. No caso de dois textos, cuidar para que os dois juntos não excedam uma página.

- É preferível que os textos escritos utilizados como insumo tenham título interessante, fotos ilustrativas, gráficos, mapas ou outros elementos que possam ser explorados na tarefa. É preciso cuidar, porém, para que as imagens e fotos ou gráficos não apresentem toda a informação do texto, para não haver risco de o examinando ler apenas as imagens.

[...]

Gêneros que podem servir de insumo para as tarefas III e IV:

- Textos de periódicos (jornais e revistas) – editorial, notícia, entrevista, reportagem, anúncio classificado, publicidade, cartas de leitores, horóscopo, cartuns, quadrinhos, etc.;

- Textos de livros – crônica, conto, poema, texto didático, receita etc.;

- Textos de panfletos, cartazes, avisos, placas de trânsito;

- Textos de telegramas, cartas, bilhetes, e-mails, cartões-postais;

- Textos de diários, agendas, notas, listagens, resenhas, relatórios, currículos, biografias;

- Textos de documentos, formulários, questionários, instruções; de mapas, roteiros, quadro de horários, calendários, programas, cardápios, recibos;

- Textos de dicionários, catálogos, listas telefônicas, letras de música, legendas de filme etc.

Fonte: Schoffen *et al.*, 2012, *apud* Mendel, 2019, p. 181.

É evidente a atenção ética direcionada à escolha de tópicos e conteúdo não tendenciosos, evitando favorecer grupos específicos e prevenindo possíveis constrangimentos culturais. Essa precaução deriva do uso do mesmo texto de insumo para todos os examinandos, objetivando fornecer uma fonte de informação comum que garanta um ponto de conhecimento base de modo homogêneo (WEIGLE, 2004; SCARAMUCCI, 2016). Ademais, encontram-se orientações relacionadas à qualidade desses materiais, com destaque para a seleção de vídeos e áudios. No tocante aos textos escritos, percebemos a sugestão de busca por textos multimodais, ainda que

o documento não expanda em maiores detalhes quanto à qualidade e ao nível de complexidade dos textos não verbais, bem como seus possíveis impactos no desempenho da leitura e na posterior avaliação. Os gêneros que podem ser utilizados como texto de insumo listados no documento são análogos aos gêneros que contam nas especificações do Manual do Examinando (BRASIL, 2011), mas reúnem uma lista mais ampla do que os gêneros mencionados nas especificações do Documento Base (BRASIL, 2020), documento orientador mais recente sobre o exame.

No que tange à correção das respostas dos examinandos, uma resposta esperada é elaborada pela equipe de correção de modo singular para cada tarefa. A resposta esperada concentra-se no desempenho geral do examinando em relação ao cumprimento da tarefa, considerando a adequação contextual e discursiva do texto produzido. Esses elementos são avaliados considerando a composição do gênero discursivo apropriado à esfera comunicativa, bem como o uso de informações relevantes do texto de insumo e de recursos linguísticos apropriados para a organização de um texto claro e coeso. A nota final de cada tarefa é determinada pela média aritmética das notas independentes de dois avaliadores⁷³ e a nota final da PE é obtida por meio da média aritmética das notas finais das quatro tarefas. O resultado final do examinando é baseado na nota mais baixa entre as duas partes, PO e PE, sendo que o Celpe-Bras certifica a partir do nível intermediário, que corresponde a notas entre 2,00 e 2,75 pontos. Além desse, são conferidos ainda os certificados Intermediário Superior (2,76 a 3,50), Avançado (3,51 a 4,25) e Avançado Superior (4,26 a 5,00).

Durante o processo de avaliação da PE, a resposta esperada de cada tarefa é formulada com base em uma resposta prevista pelos elaboradores, como no exemplo da tarefa “Shopping em casa”, em que a resposta esperada foi a seguinte:

⁷³ Para mais informações, ver o Documento Base do exame Celpe-Bras (BRASIL, 2020).

Quadro 3 – Resposta esperada para a tarefa “Shopping em casa”

Resposta Esperada
Gênero discursivo: carta do leitor
Enunciador: leitor da coluna
Interlocutor(es): leitores do jornal
Propósito(s): opinar sobre o tema consumismo, considerando as questões levantadas no texto.
Conteúdo informacional
Opiniões do texto: <ul style="list-style-type: none"> • Não sou exatamente consumista, mas olho a minha casa e, apesar de morrer de amores por meus objetos, sinto que há uma infinidade de coisas que eu realmente poderia viver sem. O curioso é que, se descuidar, se me forem oferecidas, estou arriscada a comprá-las de novo. • É realmente muito violento o impulso consumista a que fomos doutrinados. Nem precisamos ir ao shopping. Vivemos em uma feira constante. Meu e-mail está lotado de propostas de vendas que nem desconfio como chegaram ali. A internet virou, como previsto, um mar de publicidade que, a um clique, desfalca nossa conta bancária. Tudo pela estrela maior: o dinheiro, o circular da moeda. Cada vez mais produtos, cada vez mais necessidades desnecessárias. • Não entendo nada de economia, mas gostaria de viver num mundo com menos coisas. Gostaria de morar numa casa com menos coisas, ter um armário com menos roupas. • Mesmo sem ter que mudar de apartamento, vou tirar as coisas de meu armário, colocar sobre a cama para comprar com meu novo olhar os meus velhos vestidos. • Liberdade é viver com pouco.
Observar: tese defendida + consistência da argumentação Clareza dos argumentos (mais de um viés) Exploração do tema de maneira genérica Presença de “Lugar-comum”/ Circularidade/ ideias esparsas, ‘espalhadas’ no texto Exemplos apresentados devem contribuir para o desenvolvimento do texto Narrativas a serviço da argumentação

Fonte: Sirianni, 2020, p. 112.

É possível notar que há a expectativa de composição de um texto que demonstre uma leitura específica do texto de insumo, reformulando as informações e construindo um novo texto que, neste caso, é uma carta do leitor que visa opinar sobre o tema em pauta. Conforme Schoffen *et al.* (2018) destacam, desde 2002, as tarefas da PE têm consistentemente apresentado características que permitem uma categorização das tarefas integradas. A tarefa “Shopping em casa”, por exemplo, é categorizada pelas autoras do seguinte modo: Esfera de atuação - jornalística; Temática - consumo; Propósito - posicionar-se; Relação de interlocução - de leitor/ouvinte para veículo de comunicação; Natureza de interlocução - pública, simétrica (quando o enunciador escreve para interlocutor que está em posição de hierarquia social similar à sua); Gênero do discurso - carta do leitor; e Suporte - jornal. A partir do trabalho de categorização realizado pelas autoras é possível dizer que as tarefas integradas já aplicadas desde 2002 solicitam que:

O texto a ser escrito esteja inserido em determinada esfera da comunicação humana, verse sobre determinado tema, cumpra determinado propósito comunicativo, configure uma determinada relação de interlocução, apresente, portanto, determinadas características informacionais e composicionais, constituindo-se assim dentro de determinado gênero do discurso, e seja passível de ser publicado em determinado suporte (SCHOFFEN *et al.*, 2018, p. 12).

Com exceção de três edições aplicadas entre 1998 e 2001, que utilizaram tarefas constituídas apenas por perguntas de compreensão ou tarefas híbridas, combinando perguntas de compreensão com comandos de produção textual, e, portanto, não se enquadram na atual definição de tarefa integrada, a operacionalização do construto do exame em tarefas que integram compreensão e produção é estável desde 2002 (SCHOFFEN *et al.*, 2018).

Na primeira edição do manual do candidato, em 2002, foram delineados os conceitos e as características do exame, incluindo a integração de habilidades, que embasam a elaboração dos instrumentos de avaliação e definem o seu construto teórico, os quais, nas décadas subsequentes, foram revisados e aprimorados. A ampliação do Manual do Exame, de 26 páginas na publicação de 2002 (BRASIL, 2002), para 131 páginas do Documento Base publicado em 2020 (BRASIL, 2020), evidencia o comprometimento da CTC em detalhar o construto teórico e em tornar o processo de avaliação mais transparente. Ainda que os parâmetros de avaliação e alguns conceitos teóricos, como a noção de natureza comunicativa⁷⁴ do exame, tenham sofrido modificações, as concepções teóricas que fundamentam as visões de proficiência, uso da linguagem, cultura e tarefas integradas permanecem consistentes em todas as publicações desde 2002.

A noção de proficiência e a integração de habilidades respaldam os conceitos de letramento e de uso da linguagem que orientam o construto do exame. O Documento Base afirma que “o objetivo da produção escrita pressupõe a compreensão do texto de entrada e é condição fundamental para a realização da tarefa proposta” (BRASIL, 2020, p. 30). Logo, se a leitura do texto de insumo é fundamental para o sucesso da tarefa, entende-se que as tarefas da PE são elaboradas considerando um nível de complexidade textual compatível com o público-alvo previamente definido, a fim de evitar a criação de barreiras intransponíveis para os examinandos e garantir a realização da tarefa, conforme orientação de THE COLLEGE BOARD (2015), mencionada anteriormente. Dado que o Celpe-Bras abrange atualmente um público-alvo amplo e diversificado, é desafiador definir precisamente as necessidades de uso

⁷⁴ O Manual do Exame apresenta que “o exame é de natureza comunicativa” (BRASIL, 2002, p. 4) e o Documento Base distancia-se dessa terminologia que é associada ao conceito de competência, esclarecendo que o Celpe-Bras não busca avaliar a competência, mas sim “o desempenho comunicativo” (BRASIL, 2020, p. 28).

da língua desse público. No entanto, podemos inferir que esse público é constituído por indivíduos com proficiência intermediária ou superior em língua portuguesa, de acordo com os descritores de proficiência do exame, uma vez que ele certifica a partir do nível intermediário. Portanto, o nível de complexidade textual dos textos de insumo das tarefas deve ser compatível com o nível de leitura desse público. Dessa forma, nas subseções seguintes, exploraremos com maior detalhe as tarefas que integram compreensão e produção escrita, a avaliação de leitura e o descritor do nível intermediário do Celpe-Bras.

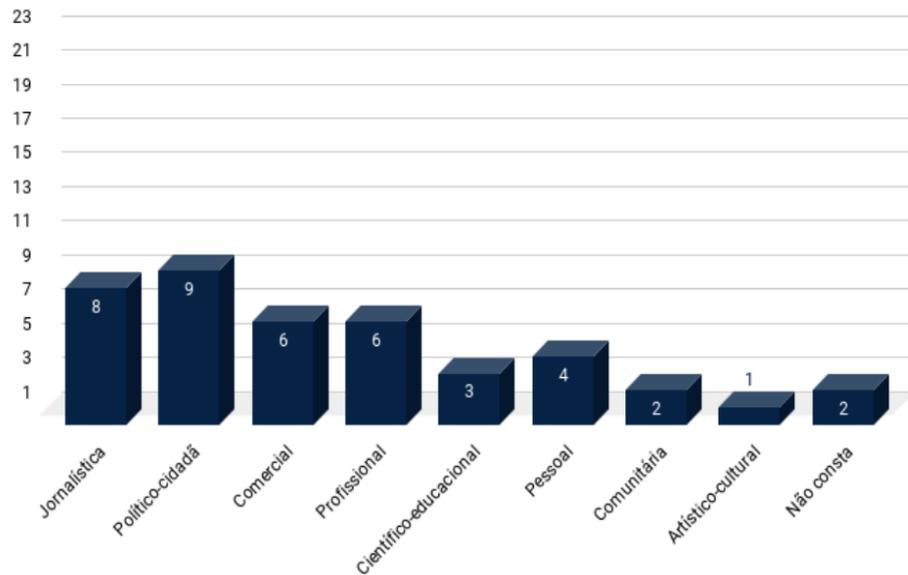
4.1 TAREFAS 3 E 4

Para a condução desta pesquisa, as tarefas 3 e 4 foram selecionadas para compor o *corpus* de estudo, visto que são as tarefas da Parte Escrita que têm como propósito avaliar de modo integrado a leitura e a produção textual e, para tanto, utilizam textos na modalidade escrita como texto de insumo. A decisão de focar exclusivamente na integração entre leitura e produção escrita, excluindo as tarefas 1 e 2, que utilizam textos de insumo orais, foi tomada devido à necessidade operacional de definir um conjunto homogêneo de textos como *corpus* de estudo. Essa delimitação foi essencial para permitir escolhas metodológicas bem fundamentadas e proporcionar maior objetividade e rigor às análises e argumentos que serão apresentados no capítulo 6.

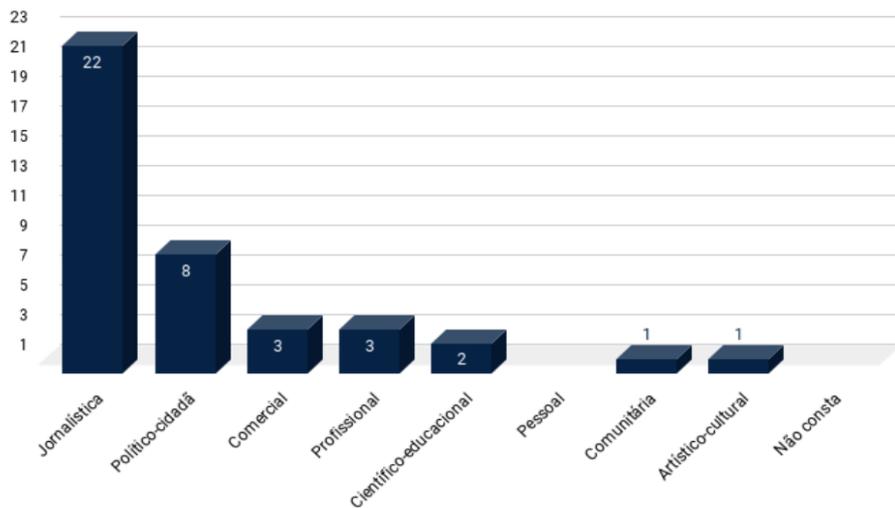
De acordo com os resultados apresentados por Schoffen *et al.* (2018), em um estudo que analisou as características de todas as edições da PE aplicadas entre 1998 e 2017, é notável que a tarefa 3 apresenta uma ampla variedade de *Esferas de atuação*, definidas como “espaços sócio-discursivos nos quais as práticas de uso da linguagem propostas pelos enunciados das tarefas acontecem” (SCHOFFEN *et al.*, 2018, p. 15), em contraste com a tarefa 4. As figuras a seguir revelam que a esfera de atuação *Jornalística*, englobando interações sociais que acontecem nas mídias de informação como jornais e revistas, compõem mais da metade das 40 tarefas 4 analisadas pelas autoras. A segunda esfera de atuação mais recorrente na tarefa 4 é a *Político-cidadã*, que é composta “por interações relativas à coletividade, direitos, deveres, infraestrutura e iniciativas em prol do bem comum” (SCHOFFEN *et al.*, 2018, p. 16). As esferas menos recorrentes nas tarefas 3 e 4 incluem as denominadas *Artístico-cultural*, *Científico-educacional*, *Comunitária* e *Pessoal*.

Figura 10 – Ocorrências de esferas de atuação nas tarefas 3 e 4 de 1998 a 2017

Esfera de atuação - III



Esfera de atuação - IV



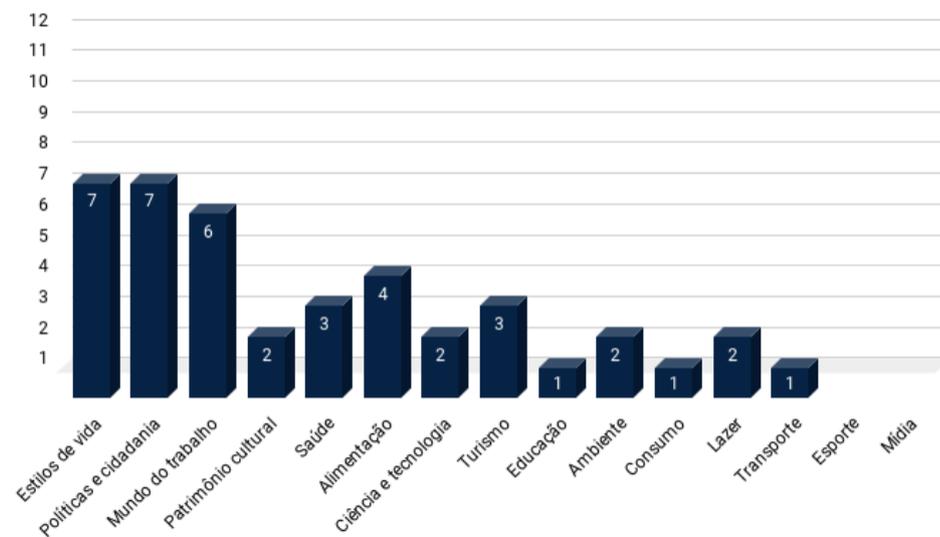
Fonte: Schoffen *et al.*, 2018, p. 44.

A temática, de acordo com as autoras, está relacionada ao tema central em torno do qual a tarefa foi elaborada e desempenha um papel importante, uma vez que gera impacto no repertório linguístico necessário para a realização da tarefa, sobretudo em relação ao repertório lexical do candidato. Schoffen *et al.* (2018) concluem que a tarefa 4 possui uma temática que se destaca em relação às demais, enquanto a tarefa 3 tende a propor três principais temáticas: *Estilos de vida*, *Políticas e cidadania* e *Mundo do trabalho*. No caso da tarefa 4, a temática particularmente proeminente é *Estilos de vida*, que abrange “hábitos e rotina, perfil de segmentos da sociedade, etapas e mudanças de vida, tendências de comportamento em relação

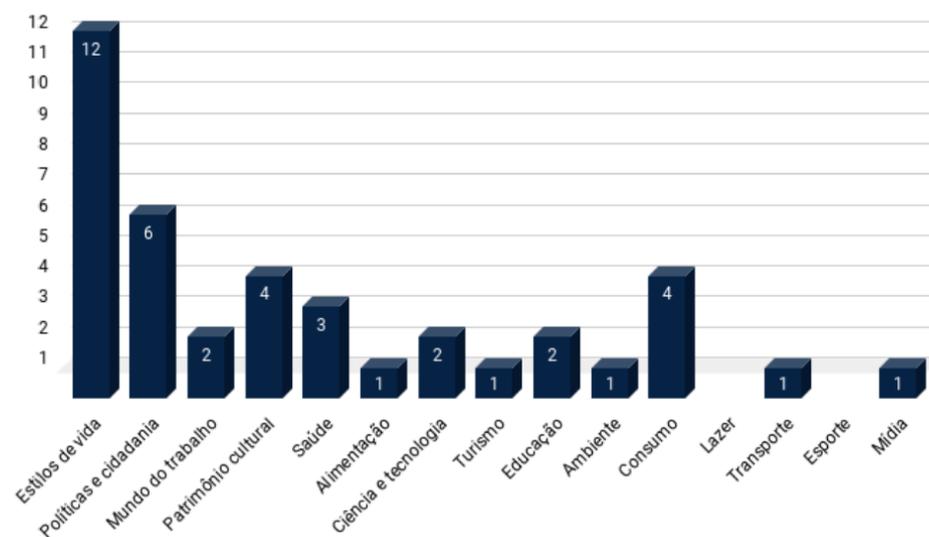
ao uso de novas tecnologias, padrões estéticos, moda, moradia, relacionamentos, convivência” (SCHOFFEN *et al.*, 2018, p. 20). Ainda assim, os resultados, como mostra a figura a seguir, ilustram uma variedade ampla de outras temáticas presentes em ambas as tarefas.

Figura 11 – Ocorrências de temáticas nas tarefas 3 e 4 de 1998 a 2017

Temática - III



Temática - IV



Fonte: Schoffen *et al.*, 2018, p. 46.

As demais categorias descritivas propostas por Schoffen *et al.* (2018) estão vinculadas à produção textual projetada em resposta ao enunciado da tarefa. As autoras propõem outras cinco categorias que descrevem a produção textual: *Propósito*, *Relação de interlocução*,

*Natureza da interlocução, Gênero do discurso e Suporte*⁷⁵. As autoras observam que a tarefa 3 mostrou uma tendência maior em solicitar interlocuções nas quais a posição social do examinando é inferior na hierarquia social em relação aos seus interlocutores.

Quanto às características da tarefa 4, as autoras identificaram um perfil de tarefa que predominantemente solicita produções textuais argumentativas, nas quais os examinandos são convidados a expressar opiniões e argumentar acerca do tópico abordado no texto de leitura. Schoffen *et al.* (2018) reconhecem que o perfil da tarefa 4 se difere das outras três tarefas quanto à produção textual. As autoras apontam que, nas tarefas 1, 2 e 3, em geral, os examinandos necessitam compreender o texto de insumo e reorganizar as informações em um formato textual diferente. No entanto, a tendência da tarefa 4 é solicitar que os examinandos compreendam o texto de maneira a usar o conteúdo do texto de insumo como base para construir argumentos que sustentem um posicionamento em relação ao tema abordado, indo além da simples reorganização das informações em outro formato textual.

Schlatter *et al.* (2021) realizaram uma análise da distribuição das notas finais atribuídas aos examinandos que participaram das edições 2015-2, 2016-1, 2016-2 e 2017-1 do Celpe-Bras, visando ao levantamento de indícios sobre o nível de complexidade das tarefas. As autoras chegaram à conclusão de que, ao comparar o comportamento do desempenho dos examinandos nessas quatro edições, a tarefa 4 parece ter sido a mais difícil de ser respondida. Apenas na edição 2015-2, a tarefa 3 apresentou uma média de desempenho inferior à média da tarefa 4. As autoras ressaltam que a análise das médias e da distribuição das notas dos examinandos serve apenas como um indicador dos níveis de complexidade das tarefas, sendo necessário estudos qualitativos adicionais para que asserções mais conclusivas sejam feitas. Alguns desses dados serão retomados mais adiante na discussão dos resultados desta tese.

Nessa mesma perspectiva, para investigar a complexidade de tarefas, além da necessidade de conduzir investigações de natureza quantitativa e qualitativa sobre o texto de insumo, é fundamental adotar uma abordagem que englobe também o enunciado da tarefa. Isso porque é no enunciado que a situação comunicativa proposta direciona o modo como determinadas informações do texto de insumo devem ser mobilizadas, a fim de cumprir o propósito comunicativo através da produção textual. Nas palavras de Gomes (2009, p. 98), “não é possível separar complexidade das condições de recepção e de produção propostas na tarefa”.

⁷⁵ Apesar de esta pesquisa estar centrada na análise da leitura, uma vez que o objeto de estudo são tarefas que integram as habilidades de leitura e escrita, todas as categorias propostas por Schoffen *et al.* (2018) foram incluídas nas análises, como será detalhado posteriormente.

Determinar o grau de complexidade da tarefa, como aponta Brown (2004), é uma chave para a validade e o sucesso da tarefa, sendo que a dificuldade em respondê-la pode estar intrinsecamente relacionada ao enunciado: enunciado confuso, tópico obscuro, dados imprecisos e informações culturais tendenciosas podem levar a uma dificuldade de nível superior ao desejado. No contexto do Celpe-Bras, um enunciado considerado claro “é aquele que explicita ao examinando quem é o enunciador, o interlocutor, o propósito comunicativo a ser cumprido, o gênero do discurso e o suporte e que não apresenta incongruências entre esses elementos” (NAGASAWA, 2018, p. 83). Além disso, o enunciado deve apresentar comandos objetivos e explícitos, utilizando léxico e estruturas linguísticas acessíveis ao nível mínimo de compreensão leitora do público-alvo do exame.

Lobo (2021), por meio de abordagem qualitativa de análise documental, investigou elementos semânticos, léxico-sintáticos e paralinguísticos de 47 enunciados de tarefas 3 já aplicadas no Celpe-Bras e concluiu que esses enunciados carecem de uniformidade ou consistência no que diz respeito ao formato, extensão e nível de dificuldade do léxico utilizado. O estudo apontou possíveis impactos negativos dessa falta de padronização, especialmente no que se refere à preparação dos examinandos. De modo semelhante, Nagasawa e Schlatter (2021) analisaram os enunciados das tarefas 4 utilizando recursos metodológicos da LC, partindo do pressuposto de que um dos fatores preditivos do sucesso em responder a tarefa é a compreensão adequada de seu comando. Com foco na investigação da complexidade textual dos enunciados e observando padrões lexicais e sintáticos, os resultados indicaram que os enunciados apresentam uma média de complexidade textual adequada, explicitando claramente as características da produção textual solicitada. Apesar disso, as autoras ressaltam que os resultados indicaram uma variação ampla de complexidade textual que deve ser mais bem controlada para assegurar a equidade entre uma edição e outra do Celpe-Bras.

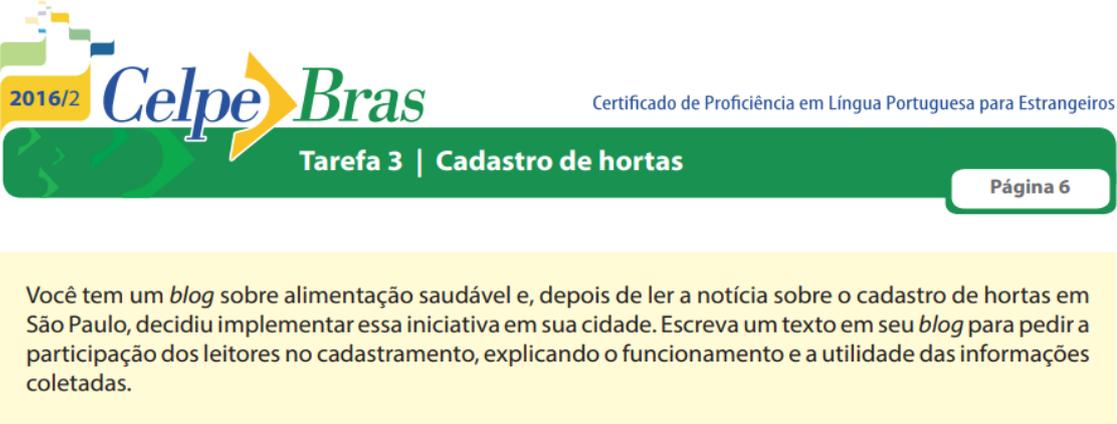
Por se tratar de tarefas que integram as habilidades de compreensão e produção, os enunciados também precisam explicitar a importância da leitura do texto de insumo, a fim de deixar claro para o examinando a necessidade de utilizar as informações contidas nesse material. Vicentini (2019) adverte que a ausência da menção do texto de insumo no enunciado de uma tarefa traz o risco de comprometer a validade do construto do exame, visto que os candidatos podem interpretar o enunciado apenas como comando de escrita e não como uma instrução de integralização das habilidades de compreensão e produção.

Sobre essa questão, por meio de uma abordagem qualitativa de análise documental, a pesquisa de Dell’Isola e Pordeus (2021) examinou 34 tarefas que integram leitura e escrita no

Celpe-Bras, identificando que 30% dos enunciados dessas tarefas, aplicadas nas edições de 2009 a 2017, não mencionam explicitamente a necessidade de leitura do texto de insumo para a produção do texto de resposta. De acordo com as autoras, os enunciados que falham em apresentar instruções diretivas de uso do texto de insumo permitem que a tarefa seja cumprida sem o resgate e a recontextualização do conteúdo informacional presentes no texto e, conseqüentemente, a avaliação integrada de leitura e escrita pode se tornar inviável. Corroborando as autoras, Lobo (2021) sustenta que as tarefas integradas devem apresentar enunciados que também esclareçam os usos esperados do texto de insumo, a fim de que o seu uso não se configure como opcional.

O enunciado da tarefa 3 da segunda edição de 2016, apresentado a seguir, conforme analisado por Lobo (2021), solicita explicitamente a recontextualização de certas informações do texto de insumo para o adequado cumprimento da tarefa. A autora observa que a instrução de leitura do texto de insumo é explicitada na contextualização da situação comunicativa proposta na tarefa, na qual o examinando assume o papel social de um blogueiro que leu o material e se interessou pela iniciativa de cadastramento de hortas urbanas. Lobo destaca que o propósito comunicativo da tarefa é incentivar o cadastramento de hortas urbanas e, para isso, o examinando precisa escrever um texto que explique o funcionamento e a utilidade das informações coletadas pelo cadastro, o que implica a leitura do texto de insumo.

Figura 12 – Enunciado da tarefa 3 da Parte Escrita de 2016-2



2016/2 **Celpe** **Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 3 | Cadastro de hortas **Página 6**

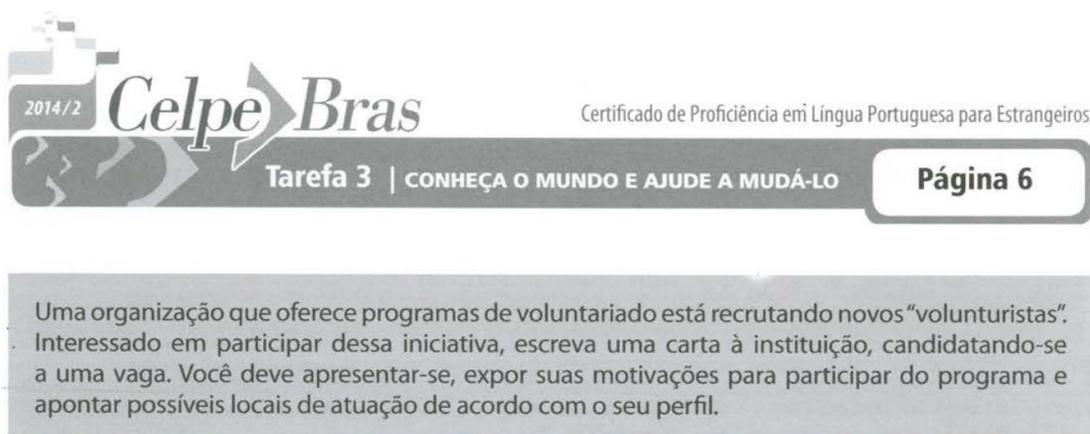
Você tem um *blog* sobre alimentação saudável e, depois de ler a notícia sobre o cadastramento de hortas em São Paulo, decidiu implementar essa iniciativa em sua cidade. Escreva um texto em seu *blog* para pedir a participação dos leitores no cadastramento, explicando o funcionamento e a utilidade das informações coletadas.

Fonte: Caderno de Questões de 2016-2. Disponível em Acervo Celpe-Bras.

Por outro lado, Lobo (2021) analisa o enunciado da tarefa 3 da primeira edição de 2014, ilustrada a seguir, e chega à conclusão de que se trata de um enunciado problemático, uma vez que é possível produzir o texto solicitado sem a leitura integral do texto de insumo. Segundo a autora, o examinando é convidado a escrever uma carta para cumprir o propósito comunicativo

de candidatar-se a uma vaga de “volunturista”, sendo exigido que o texto contenha uma apresentação pessoal, motivações para participar do programa e possíveis locais de atuação. Lobo argumenta que o candidato pode se apresentar, indicar motivações e possíveis locais de atuação livremente, independente do texto de insumo. Além disso, o título da tarefa pode auxiliar no cumprimento do propósito de expor as motivações, como exemplificado por ela: “sempre quis participar do programa de voluntariado, porque quero mudar o mundo” (LOBO, 2021, p. 116).

Figura 13 – Enunciado da tarefa 3 da Parte Escrita de 2014-2



2014/2 **Celpe Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 3 | CONHEÇA O MUNDO E AJUDE A MUDÁ-LO **Página 6**

Uma organização que oferece programas de voluntariado está recrutando novos “volunturistas”. Interessado em participar dessa iniciativa, escreva uma carta à instituição, candidatando-se a uma vaga. Você deve apresentar-se, expor suas motivações para participar do programa e apontar possíveis locais de atuação de acordo com o seu perfil.

Fonte: Caderno de Questões de 2014-2. Disponível em Acervo Celpe-Bras.

As pesquisas mencionadas acima destacam, portanto, a indispensabilidade da clareza do enunciado na avaliação da leitura do texto de insumo nas tarefas da PE. Quanto ao material utilizado, conforme já mencionado, o Celpe-Bras faz uso de textos autênticos, majoritariamente, provenientes das mídias de informação tradicionais brasileiras em todas as tarefas da PE (SCHOFFEN; MENDEL, 2018). Essa abordagem pode ser interpretada como uma sub-representação do construto, já que os materiais pertencem a uma esfera de atuação específica que não é representativa das amplas práticas de escrita escolarizadas (VICENTINI, 2019). Uma vez que o Documento Base explicita que o Celpe-Bras avalia, por meio tarefas integradas, o uso da língua portuguesa em contextos escolarizados de nível superior (BRASIL, 2020, p. 30), a predominância de texto de insumo oriundo das mídias de informação pode não ser abrangente o suficiente em relação à representação das práticas sociais previstas. Killner (2021) e Schoffen e Mendel (2018) afirmam que até mesmo os textos literários utilizados como texto de insumo na PE pertencem à esfera jornalística. Conforme Killner descreve, a crônica jornalística “tende a ser um texto que não tem grande extensão, apesar de haver exemplares

mais longos. [...] É um texto literário produzido para figurar na imprensa, em jornais ou revistas” (KILLNER, 2021, p. 197).

As especificações atuais do Documento Base detalham as seguintes características dos textos de insumo selecionados para a PE do exame:

Quadro 4 – Especificações dos textos de insumo da Parte Escrita do Celpe-Bras

	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4
Habilidades avaliadas	Compreensão oral e imagética e produção escrita	Compreensão oral e produção escrita	Leitura multimodal e produção escrita	
Características dos textos de insumo				
Modalidade do texto de insumo	Vídeo	Áudio	Texto escrito	
Gêneros dos textos de insumo	Reportagens, entrevistas, documentários, esquetes cênicas, vídeos publicitários, entre outros.	Programas de rádio, reportagens, entrevistas, peças radiofônicas, áudios publicitários, entre outros.	Reportagens, entrevistas, artigos de opinião, editoriais, crônicas, colunas assinadas, notícias, informativos, panfletos, propagandas, cartuns, quadrinhos, anúncios, poesias, resenhas, relatórios, currículos, biografias, editais, formulários, questionários, instruções, mapas, roteiros, cardápios, letras de canções, manifestos, cartas do leitor, entre outros.	
Fontes dos textos de insumo	Emissoras de televisão ou produtores de vídeo e outras fontes.	Emissoras de rádio e outras fontes.	Jornais, revistas, panfletos, informativos, livros, redes sociais, mídia eletrônica em geral, entre outras.	

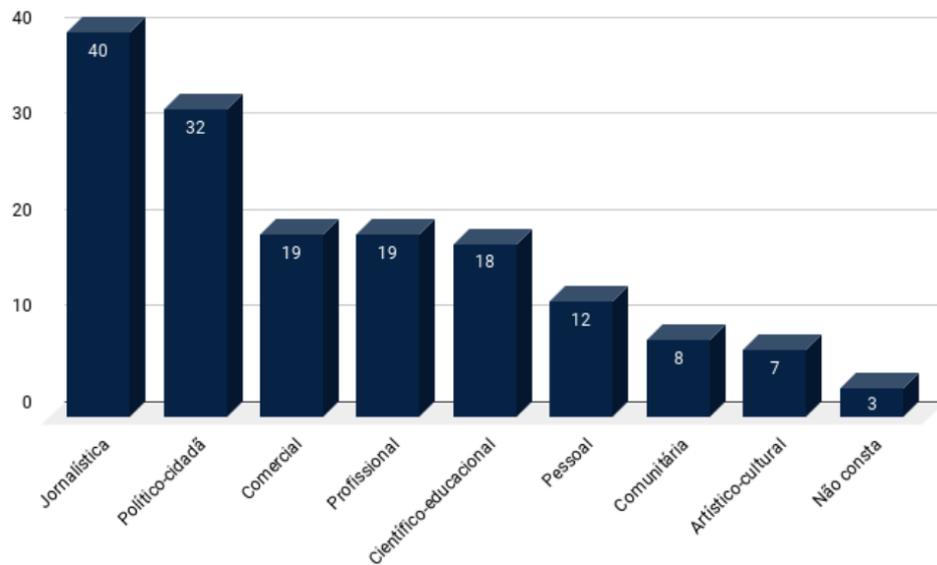
Fonte: Brasil, 2020, p. 58.

Como pode-se observar no quadro, diferente dos resultados das pesquisas, a lista dos gêneros discursivos dos textos de insumo previstos nas especificações não se restringe à esfera jornalística, o que está em consonância com a proposta do exame de incorporar “em sua estrutura, gêneros discursivos, verbais e multissemióticos, que circulam em diferentes domínios da vida social brasileira” (BRASIL, 2020, p. 29). Embora o exame tenha sido originalmente concebido para suprir uma demanda de certificação de proficiência de estudantes estrangeiros ingressantes no Ensino Superior, o Documento Base esclarece que:

[...] não se está afirmando, contudo, que o Celpe-Bras seja um exame com fins acadêmicos, mesmo porque os gêneros e os conteúdos – tanto dos textos de insumo quanto dos textos a serem produzidos pelos participantes – não circulam prioritariamente em contextos acadêmicos (BRASIL, 2020, p. 30).

Ainda que as pesquisas anteriores já tenham constatado que os materiais da PE são predominantemente do domínio jornalístico (SCHOFFEN; MENDEL, 2018; VICENTINI, 2019; KILLNER, 2021), as esferas de atuação das práticas de uso da linguagem propostas nas tarefas são abrangentes e diversificadas. Os dados de Schoffen *et al.* (2018) ilustram essa diversidade de esferas sociais, conforme mostra a figura a seguir:

Figura 14 – Ocorrências de esferas de atuação na PE de 1998 a 2017 em todas as tarefas



Fonte: Schoffen *et al.*, 2018, p. 16.

Podemos observar que, embora a esfera de atuação jornalística continue predominante, a PE abrange também outras esferas, como político-cidadã, comercial, profissional, científico-educacional, entre outras, que parecem abranger um domínio mais amplo de práticas letradas⁷⁶. Em relação a essa questão, o Documento Base enfatiza que são priorizadas as relações de interlocução que acontecem em esferas públicas, uma vez que estão relacionadas às necessidades do público-alvo e possibilitam a avaliação de habilidades representativas desses usos públicos (BRASIL, 2020).

As principais fontes de textos de insumo da PE, identificadas por Schoffen e Mendel (2018), como a *Folha de São Paulo*, *Zero Hora*, *O Globo* e *Jornal do Brasil*, são consideradas por Amaral (2006) como “jornais de referência” no Brasil. Esses jornais são associados aos principais veículos jornalísticos e abrangem “os grandes jornais consagrados econômica e politicamente ao longo da história, que dispõem de prestígio no país e são dirigidos às classes A e B”, além de serem “considerados veículos de credibilidade entre os formadores de opinião” (AMARAL, 2006, p. 3). Portanto, embora as interlocuções projetadas nas tarefas da PE não representem exclusivamente “gêneros que circulam no meio acadêmico e profissional de possível atuação dos participantes” (BRASIL, 2020, p. 30), os leitores projetados para os textos publicados nos jornais de referência da mídia tradicional brasileira podem ser considerados

⁷⁶ Pesquisas que investigaram a PO do Celpe-Bras, do mesmo modo, indicam a abrangência das temáticas abordadas nos textos de insumo utilizados para a interação face a face que se relacionam a diferentes domínios da vida social brasileira (FARIA, 2017; NEGRI, 2019; NATER; MARQUES, 2020).

como pessoas familiarizadas com as práticas de letramento de um cidadão escolarizado que faz uso da norma culta brasileira.

Carvalho e Schlatter (2022) se referem a essa linguagem de prestígio, como a norma privilegiada, uma forma linguística específica legitimada pela classe historicamente dominante, utilizada nos contextos educacionais e acadêmicos. As autoras apontam que outros repertórios comunicativos são hierarquicamente considerados inferiores em relação à língua hegemônica do europeu colonizador que hoje constitui a elite branca brasileira.

A partir da análise detalhada das características das tarefas 3 e 4 da PE, a próxima seção buscará apresentar a noção de leitura subjacente às tarefas integradas com base em resultados de estudos recentes e, sobretudo, nas especificações publicadas no Documento Base do exame.

4.2 AVALIAÇÃO DE LEITURA NA PARTE ESCRITA

Conforme delineado no Documento Base, as habilidades de leitura avaliadas não se restringem meramente a “localizar informações explícitas no texto, mas também ler nas entrelinhas, fazer inferências de implícitos, sintetizar e relacionar informações, parafrasear, sumarizar, entre outras” (BRASIL, 2020, p. 30). Conforme discutido na seção anterior, todas essas habilidades de leitura são avaliadas por meio da produção escrita que responde ao enunciado da tarefa, no qual a perspectiva de reação ao texto de insumo é estabelecida. Portanto, a capacidade de sintetizar, parafrasear e sumarizar também são habilidades que o exame destaca como intrínsecas à habilidade de leitura. Nas especificações do exame, o Documento Base lista as sub-habilidades de compreensão avaliadas na PE, conforme reproduzido a seguir (quadro completo no anexo B).

Quadro 5 – Sub-habilidades de compreensão avaliadas na Parte Escrita do Celpe-Bras

Sub-habilidades de compreensão avaliadas	<ul style="list-style-type: none"> ● Reconhecer a situação de comunicação (quem fala ou escreve, para quem, em que contexto, em que veículo, com que objetivo, em que registro, em que gênero, entre outros) ● Reconhecer marcas linguísticas características de diferentes gêneros do discurso ● Identificar a ideia principal do texto ● Identificar detalhes e ideias secundárias ● Localizar informações explícitas no texto ● Inferir palavras desconhecidas e implícitos no texto ● Selecionar informações do texto necessárias para o cumprimento do propósito da tarefa ● Reconhecer atitudes, emoções e pontos de vista do autor quando expressos explícita ou implicitamente no texto ● Interpretar gráficos e outros materiais apresentados visualmente
--	---

Fonte: BRASIL, 2020, p. 61.

Embora não esteja explicitado no Documento Base, pode-se depreender que as nove sub-habilidades de compreensão não necessitam ser avaliadas integralmente em cada uma das quatro tarefas da PE, mas sim como um todo em cada edição. É possível presumir, no entanto, que algumas das sub-habilidades sejam inevitavelmente mobilizadas, em diferentes medidas, em todas as tarefas, como reconhecer a situação de comunicação, reconhecer marcas linguísticas características de diferentes gêneros do discurso, inferir palavras desconhecidas e implícitos no texto e selecionar informações do texto necessárias para o cumprimento do propósito da tarefa.

Em contrapartida, outras sub-habilidades provavelmente sejam mais relevantes em tarefas específicas, dependendo do texto de insumo e da natureza da leitura solicitada na tarefa. Por exemplo, na tarefa 1, cujo texto de insumo é em formato de vídeo, é necessário interpretar os materiais apresentados visualmente; a tarefa 2, que utiliza material de áudio como insumo, pode demandar o reconhecimento de atitudes, emoções e pontos de vista a partir dos usos de entonação, ritmo da fala e outras características da linguagem oral; na tarefa 3, que frequentemente requer a identificação e reorganização de informações específicas do texto de insumo, a localização de informações explícitas pode ser mais recorrente; e na tarefa 4, que solicita majoritariamente a produção de um texto de opinião, a identificação de ideias principais e secundárias no texto de insumo pode ser mais recorrente para a construção de uma argumentação que considere os pontos de vista apresentados.

Uma vez que compreendemos que o certificado Intermediário requer uma média aritmética das quatro tarefas da PE que atinja no mínimo 2,0 pontos (nota mínima para certificação), podemos pressupor também que, nesse nível de proficiência, o examinando é

suficientemente capaz de mobilizar as nove sub-habilidades listadas nas especificações. Entretanto, é pertinente frisar, que pesquisas, como a apresentada nesta tese, que investigam tarefas de modo individualizado, não pretendem, nem fornecem resultados generalizáveis para a totalidade das sub-habilidades avaliadas na PE.

Como já mencionado, a avaliação das sub-habilidades de compreensão requer que o candidato demonstre seu entendimento do texto por meio da produção, conforme delineado no enunciado da tarefa. Isso reforça a importância da elaboração de enunciados claros, visto que isso contribui também para uma avaliação mais precisa da leitura, minimizando ambiguidades e interpretações não previstas. Scaramucci (2016) argumenta que a explicitação dos propósitos nos enunciados das tarefas é fundamental para a avaliação da leitura, uma vez que “leituras distintas são produzidas quando propósitos de leitura distintos são especificados” (SCARAMUCCI, 2016, p. 409).

Pesquisas que analisaram as respostas dos candidatos evidenciaram que tanto a quantidade quanto a qualidade das informações selecionadas, assim como a consistência e o modo como ocorre a recontextualização dessas informações, são determinantes para distinguir os níveis de proficiência (MENDEL, 2019; KUNRATH, 2019; SIRIANNI, 2020). A recontextualização das informações depende, em primeiro lugar, da decodificação, interpretação e atribuição de significados, e é durante esse processo que as nove sub-habilidades de compreensão são empregadas: o texto a ser produzido “demanda uma compreensão específica dos textos de insumo. Para cumprir adequadamente a tarefa proposta, não basta que o examinando compreenda o texto de insumo; é preciso saber selecionar as informações necessárias” (BRASIL, 2020, p.36). Nas tarefas 1 e 2, a avaliação ainda é influenciada por variáveis como o uso da memória para recuperar sentidos veiculados pelo texto oral e, possivelmente, por recursos multimodais e pela geração de anotações produtivas.

Em conformidade com a perspectiva de proficiência, que envolve a capacidade de “realizar ações em língua portuguesa, configurando adequadamente a interlocução solicitada na tarefa e cumprindo o propósito solicitado, dentro de um gênero e de um contexto de produção específicos, determinados pela tarefa” (BRASIL, 2020, p. 38) e, considerando a avaliação integrada das habilidades, cada nível de proficiência é descrito incluindo as habilidades de compreensão e de produção. Isso é evidenciado no descritor da nota 2 (nota mínima para o nível intermediário) da grade de parâmetros de avaliação holística da PE (a grade completa está no anexo C). No que diz respeito à compreensão do texto de insumo, o examinando pode apresentar equívocos de compreensão e recontextualizar informações insuficientes para

configurar adequadamente a interlocução solicitada na tarefa, ou ainda não articular as informações de forma clara em sua produção, conforme pode-se verificar nos trechos marcados na figura a seguir.

Figura 15 – Descritor da nota 2 do quadro de parâmetros de avaliação da Parte Escrita do Celpe-Bras

2 – Configura a relação de interlocução de forma pouco consistente, realizando superficialmente a ação solicitada. Pode estabelecer uma relação de interlocução próxima à solicitada, não cumprir propósito(s) menor(es) e/ou apresentar problemas na construção do gênero. Pode apresentar trechos do texto que remetem a um gênero diferente, comprometendo a relação de interlocução. A relação entre o propósito do texto e a interlocução configurada não é clara ou não é totalmente adequada. **Pode não recontextualizar informações que seriam necessárias** para a configuração adequada da interlocução ou **não articular claramente essas informações**. **Equívocos de compreensão** podem comprometer parcialmente o cumprimento do propósito. Os recursos linguísticos acionados são limitados e/ou inadequados, podendo prejudicar parcialmente a configuração da relação de interlocução no gênero solicitado. Problemas de clareza e coesão podem ocasionar, em diferentes momentos, dificuldades na leitura.

Fonte: BRASIL, 2020, p. 39.

Pode-se observar que não se espera que as respostas que recebem a nota 2 “cumpram perfeitamente o que foi solicitado pelo enunciado da tarefa, permitindo equívocos tanto na compreensão do texto de insumo quanto na produção do texto em resposta à tarefa” (SIRIANNI, 2020, p. 67). É importante observar que a figura 15 ilustra o descritor da nota 2 do parâmetro de avaliação, que não se relaciona a uma tarefa em particular, mas serve como ponto de partida para a elaboração das respostas esperadas (vide quadro 3), documentos que são efetivamente utilizados para a correção de cada tarefa em cada edição, e que não são de acesso público.

Mendel (2019) investigou a recontextualização de informações em 236 produções escritas em resposta a duas tarefas que integram leitura e produção textual no Celpe-Bras. Mediante a análise das respostas, a autora mostra que a avaliação da leitura transcende a simples recuperação de informações dos textos de insumo. Ela se estende ao uso de outras informações de repertório pessoal associado aos recursos linguísticos mobilizados para a recontextualização. Os resultados indicaram que a cópia direta de trechos do texto de insumo, que a autora relaciona com a habilidade de decodificação do texto, é mais prevalente nas respostas de candidatos não certificados. Mendel conclui que, no Celpe-Bras, a recontextualização não se limita apenas à habilidade de compreensão, mas envolve, em grande medida, a habilidade de escrita, não sendo possível dissociá-las para avaliar somente leitura nas tarefas integradas. Ou seja, para demonstrar proficiência em leitura, o candidato deve não somente extrair informações do texto de insumo, mas também incorporar seu repertório pessoal para estruturar essas informações de

maneira singular, de acordo com o contexto social projetado pelo enunciado da tarefa. Nesse sentido, recontextualizar “diz respeito não apenas à habilidade de leitura como decodificação, que seria suficiente para selecionar informações do texto de insumo, mas também à sua integração com a habilidade de escrita” (MENDEL, 2019, p. 152). Portanto, não basta realizar a leitura preferível na situação de comunicação proposta e selecionar informações relevantes, é preciso ser capaz de escrever sobre o que foi lido de acordo com as expectativas delineadas no enunciado.

A maioria das tarefas 3 e 4 inclui textos não verbais que podem ser considerados como suporte pré-leitura. Serafini (2012) aborda a leitura de textos multimodais e destaca que elementos visuais não são neutros, ou seja, eles requerem conhecimento de mundo para que a construção de sentidos ocorra, demandando assim repertório pessoal do leitor. A autora argumenta que, ao ler textos não verbais, o leitor se baseia “em uma ampla gama de experiências com outras imagens e textos durante seu ato de interpretação” (SERAFINI, 2012, p. 156). Essas experiências pessoais singulares de cada leitor ampliam as possibilidades de variação das interpretações.

Nessa mesma linha de pensamento, a relação entre texto escrito e imagem na construção de uma narrativa verbo-visual é categorizada por Linde (2011 *apud* LACERDA; FARBIARZ, 2018) como redundante, colaborativa ou disjuntiva. A relação é considerada redundante quando a linguagem verbal e visual remetem ao mesmo sentido, enquanto a relação é disjuntiva quando ambos os tipos de linguagem apontam para sentidos divergentes. Por outro lado, na relação colaborativa, o texto verbal e o não verbal trabalham juntos de maneira articulada e cooperativa para construir um sentido em comum.

Tendo dito isso, podemos supor que os textos não verbais são utilizados nas tarefas tanto como recursos de suporte pré-leitura, quanto como textos que agregam sentidos e estabelecem uma relação cooperativa com o texto verbal. Sirianni (2020), em sua análise da edição de 2016-2 do Celpe-Bras, aponta que as imagens podem desempenhar um papel importante na facilitação da compreensão tanto do texto de insumo quanto da tarefa como um todo. No entanto, também ocorre o oposto, casos em que a imagem pode se colocar como obstáculo para a compreensão, como ocorreu na tarefa 4 (ver figura 9 apresentada na página 74), na qual a “imagem pouco ajuda para a compreensão do texto ou da temática; pelo contrário, parece que para entendê-la é necessário primeiro compreender o texto e do que ele está tratando” (SIRIANNI, 2020, p. 112). Dessa forma, podemos concluir que, devido à natureza multimodal

dos textos, a leitura que o enunciado projeta é influenciada por todos os elementos presentes no texto de insumo, não se limitando apenas ao aspecto linguístico.

A próxima seção busca descrever o desempenho de examinandos em nível de proficiência intermediário, de acordo com os descritores do Documento Base, bem como as contribuições de pesquisas que investigaram os níveis de proficiência do Celpe-Bras. Nossa consideração em relação a esse nível de proficiência é de que ele representa o limiar mínimo necessário para que o público-alvo do exame seja capaz de compreender tanto o enunciado quanto o texto de insumo das tarefas.

4.3 NÍVEL DE PROFICIÊNCIA INTERMEDIÁRIO

Conforme exposto previamente, o primeiro nível de proficiência certificado pelo Celpe-Bras é o intermediário (com nota final entre 2,00 e 2,75 pontos). Esse é o nível exigido para aprovação no teste e, portanto, está sendo considerado como o nível mínimo do público-alvo projetado pelo exame. Dito de outro modo, entendemos que o nível intermediário constitui o ponto de corte que define o candidato que apresenta o potencial de acessar o conteúdo das tarefas, diferenciando-o do candidato que não alcança a certificação, possivelmente, por não ser capaz de acessá-lo. Com isso, inferimos que, para um candidato com proficiência abaixo do nível intermediário, a complexidade do conteúdo das tarefas se coloque como uma barreira intransponível, obstaculizando fortemente ou mesmo impedindo a realização da prova.

Sendo a primeira certificação, entende-se que o examinando em nível intermediário cumpre minimamente as ações solicitadas que, como previamente abordado, podem englobar as seguintes sub-habilidades de leitura: reconhecer o gênero do discurso, identificar a ideia principal e secundárias do texto, localizar informações explícitas e implícitas, selecionar informações, reconhecer atitudes, emoções e pontos de vista e interpretar textos não verbais. O nível intermediário é utilizado pelos estudantes do PEC-G, os quais, por meio desse certificado, demonstram que possuem o potencial para ingressar nas IES brasileiras. Isso está em consonância com o objetivo original do exame: “uma certificação de uso da língua portuguesa para participar da vida na universidade, [...] um instrumento de avaliação que aferisse o potencial dos candidatos para ler, escrever, ouvir e falar em interações da vida cotidiana e estudantil” (SCHLATTER, 2014, sem paginação).

Ao analisar o número de certificações já emitidas, torna-se evidente que o nível intermediário é efetivamente o mais utilizado por seus usuários e o mais exigido atualmente

(SIRIANNI, 2020). Entre as edições de 1999-1 e 2017-1, esse foi o nível de proficiência que registrou o maior número de certificados emitidos, de acordo com os dados disponíveis no Acervo Celpe-Bras⁷⁷. Das 36 edições aplicadas durante esse período, o intermediário foi o nível com maior número de certificados em 20 delas⁷⁸, totalizando 29.457 certificações, que representa 38% do total de candidatos homologados (77.305 examinandos).

Os resultados do estudo realizado por Mendel (2019) indicam que o examinando com proficiência intermediária é capaz de ler os textos de insumo, porém a recontextualização das informações na produção escrita não é articulada com recursos informacionais e linguísticos de repertório próprio e apropriados para o cumprimento das tarefas. Nesse nível, as produções textuais revelam equívocos de compreensão e incluem um mínimo de recontextualização de informações, suficiente para atender parcialmente aos propósitos comunicativos e à configuração da interlocução. Esses resultados são compatíveis com o descritor de proficiência intermediário do Celpe-Bras, conforme pode-se observar nos trechos marcados da figura a seguir. De acordo com o descritor (quadro completo no anexo D), o examinando em nível intermediário é capaz de selecionar algumas informações a partir da interpretação do texto de insumo, porém apresenta problemas de compreensão e dificuldades para a recontextualização das informações em sua produção textual.

Figura 16 – Descritor do nível de proficiência intermediário do Celpe-Bras

Intermediário

O examinando que atinge o nível Intermediário é capaz de produzir textos escritos sobre assuntos variados que, com dificuldade, podem ser reconhecidos como pertencentes a determinados gêneros discursivos, podendo não configurar adequadamente a interlocução. Os recursos lexicais e gramaticais mobilizados são limitados, podendo apresentar problemas de clareza e coesão e/ou inadequações frequentes que comprometem mais frequentemente a fluidez da leitura. **É capaz de selecionar algumas informações a partir da interpretação de textos de diferentes gêneros orais e escritos, evidenciando problemas de compreensão e dificuldades no trabalho de recontextualização** que podem levar ao cumprimento parcial dos propósitos dos textos produzidos. É capaz de interagir oralmente para a expressão de ideias e opiniões sobre assuntos variados. Apresenta poucas hesitações, com algumas interrupções no fluxo da conversa. Seu vocabulário pode apresentar limitações que podem comprometer o desenvolvimento da interação. Utiliza variedade limitada de estruturas, com algumas inadequações em estruturas complexas e poucas inadequações em estruturas básicas. Sua pronúncia contém inadequações e/ou interferências frequentes de outras línguas. Demonstra alguns problemas de compreensão do fluxo da fala, com necessidade frequente de repetição e/ou reestruturação ocasionada por palavras de uso frequente em nível normal de fala.

Fonte: BRASIL, 2020, p. 68.

⁷⁷ Gráficos elaborados pela membra do grupo Avalia Luiza Sarmento Divino. Disponível em: <https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2021/12/Numero-de-Examinandos-por-Nivel-de-Certificacao-nas-Edicoes-do-Celpe-Bras.pdf> Acesso em: 15 jul. 2022.

⁷⁸ Nas sete primeiras edições contabilizadas, havia apenas dois níveis de certificação: Intermediário e Avançado.

Visto que os textos de insumo são iguais para todos os candidatos, entendemos que o examinando em nível intermediário é capaz de abordar assuntos variados e pertencentes a diversos domínios da vida social brasileira. Nesse contexto de esferas sociais letradas, o examinando em nível intermediário produz com dificuldades, mobilizando recursos lexicais e gramaticais limitados. Como pode-se ver na figura, tanto na oralidade quanto na escrita, o descritor do nível intermediário prevê inadequações linguísticas e interferências de outras línguas, que podem comprometer a interlocução, embora não impeçam que os propósitos comunicativos sejam parcialmente cumpridos. Nas palavras de Sirianni “aqueles que recebem certificado de nível Intermediário demonstram domínio parcial da língua portuguesa, no entanto são capazes de compreender e produzir em português, possibilitando a comunicação, mesmo que limitada, na língua” (SIRIANNI, 2020, p. 66).

Como argumenta Kleiman (2003, p. 12), “o foco de uma prova de leitura deveria ser a capacidade de compreender, não um texto em particular, escolhido para a prova, mas qualquer texto enquanto representativo de outros do gênero e do nível de dificuldade”. Assim, com o objetivo de investigar o conteúdo e a complexidade textual dos textos de insumo das tarefas da PE do Celpe-Bras, é necessário considerar parâmetros de textos similares, ou seja, é essencial compreender os níveis de complexidade de textos da esfera jornalística já estabelecidos por pesquisas anteriores. Nesse contexto, operacionalizamos a complexidade textual por meio de 200 métricas da ferramenta NILC-Metrix e em relação aos níveis de complexidade dos *corpora* PorSimples e PorPopular. Ambos os *corpora* tomam como referência o nível de leitura brasileiro e o público-alvo dos jornais que os constituem para estabelecer o grau de complexidade textual que representam. O *corpus* PorSimples é considerado de complexidade textual alta, enquanto o *corpus* PorPopular é classificado como de baixa complexidade. Isso justifica a escolha desses indicadores como parâmetro de complexidade para a nossa pesquisa. As métricas e os *corpora* serão detalhadamente apresentados a seguir, contudo, abordaremos, primeiramente, a abordagem teórico-metodológica adotada nesta pesquisa, revisando pressupostos teóricos da Linguística de *Corpus*.

5 LINGUÍSTICA DE CORPUS

O objetivo deste capítulo é introduzir a área da Linguística de *Corpus* com um foco específico na complexidade textual. Iniciamos, portanto, estabelecendo um panorama da disciplina, para em seguida concentrar nossas discussões na temática da complexidade textual. Para tanto, organizamos as discussões sobre a complexidade textual em duas subseções. Na primeira seção, abordamos estudos conduzidos na língua inglesa que informaram nossa pesquisa, enquanto, na segunda, oferecemos uma síntese de trabalhos prévios sobre a língua portuguesa, que dialogam com nosso trabalho.

Conforme já mencionado, a LC é uma abordagem teórico-metodológica que se fundamenta na manipulação de extensas coleções de textos de linguagem autêntica processáveis por programas computacionais, visando descrever a língua em uso. Apesar de o nome da área focar em *corpus*, os próprios *corpora* não evidenciam características das línguas, *per se*, é por meio das análises computacionais que se torna possível obter perspectivas que não seriam facilmente observáveis pelo olhar humano.

McCarthy e O’Keeffe (2010) mencionam que a necessidade de analisar um grande volume de textos para identificar padrões da linguagem escrita já existia muito antes da era tecnológica, remontando ao século 13, quando estudiosos da Bíblia precisavam indexar de forma alfabética as ocorrências das palavras e suas respectivas passagens no livro. Foi somente no século 20 que o interesse pela análise da linguagem transcrita em formato legível por computadores cresceu, envolvendo linguistas e estudiosos da literatura. Os autores observam que, a partir de 1980, com a revolução tecnológica, o uso de *hardware*, *software* e internet permitiu o surgimento das análises de *corpora* que edificaram os pressupostos teóricos e metodológicos da LC como a conhecemos hoje. À medida que a área evoluiu, as aplicações da análise empírica da linguagem natural por meios computacionais se expandiram, levando os autores a afirmar que a LC “é um meio para um fim e não um fim em si mesmo” (MCCARTHY; O’KEEFFE, 2010, p. 7)⁷⁹, destacando que as abordagens e técnicas podem ser aplicadas para abordar diversas questões de pesquisa. A seguir, apresentamos um resumo não exaustivo de aplicações da LC em diferentes áreas de estudo.

⁷⁹ No original, lê-se: CL is a means to an end rather than an end in itself.

Quadro 6 – Aplicações da Linguística de Corpus por área de estudos

Área de estudo	Aplicações
Ensino e aprendizagem de línguas	Descrever níveis de proficiência, informar currículo e avaliação de línguas, elaborar materiais didáticos, capacitar professores, elaborar gramáticas.
Análise do discurso	Identificar padrões de uso da linguagem que apontem a constituição de sentidos e discursos ideológicos.
Estudos literários e de tradução	Comparar diferentes obras literárias, comparar traduções de obras literárias em diferentes línguas, identificar estratégias de tradução.
Linguística forense	Identificar padrões para autenticação de autoria de textos.
Pragmática	Descrever o uso de marcadores pragmáticos na língua oral e escrita, comparar o uso de marcadores pragmáticos em diferentes línguas.
Sociolinguística	Identificar ideologias, identificar palavras-chave culturalmente significativas e variação entre registros.
Lexicografia	Elaborar dicionários.
Terminologia fraseológica	Elaborar glossários para especialistas.
Ciência da informação e biblioteconomia	Informar sistemas de recomendação, de extração de conhecimento e mineração de textos.

Fonte: Adaptado de McCarthy e O’Keeffe (2010) e Tagnin (2018).

Apesar das áreas e aplicações distintas, as técnicas empregadas são bastante semelhantes; fundamentalmente, trabalha-se com a frequência de fenômenos linguísticos, seja em nível de morfemas, palavras ou padrões gramaticais, observados nos *corpora* (CUSHING, 2021). O uso dos *corpora* tem diferentes propósitos nos distintos estudos, podendo ser classificado como um *corpus* de estudo, composto por textos da língua-alvo que se pretende descrever, ou como um *corpus* de referência, usado para comparação e contraste com o *corpus* de estudo. Além disso, na criação de programas computacionais, os *corpora* podem ser utilizados para treinar e avaliar ferramentas automáticas (BERBER SARDINHA, 2004). Para uma breve introdução à tipologia de *corpora*, baseada em Tognini-Bonelli (2010), apresentamos no quadro a seguir uma síntese dos tipos de *corpora* comumente utilizados para análises de LC.

Quadro 7 – Tipologia de corpora

Corpora de amostragem (corpora geral)	São amostras de recortes de usos das línguas em determinados momentos históricos. Para ser considerada confiável para a representatividade dessa língua, a amostragem precisa apresentar características linguísticas de todos os principais usos da língua em diversos contextos e em quantidade substancial (100 milhões de palavras é o tamanho típico).
--	---

Corpora de comparação	São dois ou mais <i>corpora</i> compilados com base nos mesmos critérios e que apresentam tamanhos semelhantes.	
	Geográfico	São dois ou mais <i>corpora</i> da mesma língua, mas de variedades diferentes de acordo com critérios geográficos.
	Histórico	São dois ou mais <i>corpora</i> da mesma língua compilados em diferentes tempos históricos.
	Tópico	São dois ou mais <i>corpora</i> compilados em diferentes contextos, mas que tratam da mesma temática.
	Contrastivo	É um <i>corpus</i> compilado com o objetivo específico de possibilitar a observação da variação interna da língua a partir de traços distintivos.
Corpora especializados	São compilados sem o propósito de serem representativos de uma língua geral ou variedade de língua; ao contrário, são compilados para que as singularidades de determinados usos da língua possam ser observadas.	
Corpora de dimensão temporal (corpora de monitoramento)	Podem ser diacrônicos, contendo amostras de diferentes recortes de tempo, ou de monitoramento, um tipo de <i>corpus</i> temporal que é atualizado continuamente com marcações de tempo para cada amostra compilada.	
Corpora bilíngues e multilíngues (corpora paralelos)	Podem ser compostos por traduções literais ou por traduções mais livres, que buscam ser mais fiéis às convenções fraseológicas da língua-alvo.	
	Multilíngue comparável	São <i>corpora</i> com amostras de mais de uma língua natural, não traduzidas, e compilados de acordo com os mesmos critérios relacionados à quantidade e ao tipo de textos coletados.
	Contrastivo	São dois <i>subcorpora</i> da mesma língua, um contendo textos traduzidos e o outro contendo textos não traduzidos.
Corpora normativos	São compilados especificamente com o propósito de representar a norma padrão da língua e, para tanto, são compostos por textos escritos por usuários da língua como primeira língua de socialização e considerados proficientes.	
Corpora de falantes da língua como língua adicional	São compilados para representar a língua usada por usuários da língua como língua adicional.	
Corpora oral	Quando não fazem parte de <i>corpora</i> de amostragem ou normativos, geralmente são <i>corpora</i> especializados, uma vez que representam contextos específicos e são compilados com base nos objetivos dos pesquisadores e podem, inclusive, ser coletados em contexto controlado de laboratório.	

Fonte: Adaptado de Tognini-Bonelli (2010)⁸⁰.

Mesmo que haja uma ampla variedade e que algumas definições possam se sobrepor quando se trata de caracterizar os tipos de *corpora*, cada área de estudo e suas aplicações

⁸⁰ No original, a tipologia de *corpora* é: sample corpora; corpora for comparison, geographical, historical, topic, contrastive; special corpora; corpora along the time dimension; bilingual and multilingual corpora, the multilingual comparable corpus, the contrastive corpus; normativeness; non-native speaker corpora; spoken corpora.

demandam o uso de diferentes coleções de textos. Como já discutido, esses textos devem ser criteriosamente compilados ou selecionados para representar a língua-alvo do estudo e possibilitar investigações que forneçam respostas às perguntas específicas de cada pesquisa. Além da importância da representatividade, que se refere à “relação entre o corpus e a população (de textos) de onde são coletados” (HUNSTON, 2022, p. 33)⁸¹, os linguistas de *corpus* também se atentam ao equilíbrio dos *corpora*. Isso significa que os pesquisadores buscam trabalhar com *corpora* que apresentem quantidades próximas de textos ou de palavras.

Após a compilação dos *corpora*, o linguista de *corpus* pode iniciar sua investigação com o objetivo de identificar fenômenos linguísticos desde uma perspectiva indutiva (*corpus-driven*) ou com o intuito de observar os padrões linguísticos de variações e uso dos fenômenos da língua previamente apontados por outras pesquisas (*corpus-based*) (BIBER, 2012). A partir das informações de frequência das palavras dos *corpora*, é possível realizar análises de palavras-chaves, que podem indicar temas dominantes relacionados ao conteúdo, estrutura e estilo dos textos; linhas de concordância, que possibilita analisar a palavra em seus contextos de uso; e colocações, instâncias em que é possível analisar relações de recorrência entre palavras, dentre outras abordagens analíticas (EVISON, 2010). Ferramentas como Sketch Engine e WordSmith⁸², para citar algumas, podem ser empregadas para a condução desses tipos de análises.

Para investigações lexicais, alguns pesquisadores utilizam a técnica de lematização (*lemmatisation*), que envolve “combinar qualquer forma da palavra, com todas as suas flexões, a uma forma base, geralmente, a forma dicionarizada da palavra” (WALTER, 2010, p. 432)⁸³. Por exemplo, no português, palavras como “meu”, “meus”, “minha” e “minhas” seriam lematizadas para a forma base “meu”, enquanto verbos como “saiu” e “sairemos” seriam lematizados como “sair”, para serem contabilizadas sob essas formas base. Esse procedimento é indicado especialmente para a análise de línguas com sistemas ricos em flexões, como o português, pois permite a inclusão de todas as formas de uma palavra durante a análise de distribuição de frequência com maior desempenho. Isso permite assegurar um nível alto de precisão na escolha das palavras investigadas e na identificação de todas as instâncias dessa palavra (*precision e recall*), ou seja, certificando que o processo incluía as palavras relevantes sem excluir nenhuma ocorrência (ANDERSEN, 2010).

⁸¹ No original, lê-se: the relationship between the corpus and the population (of texts) it is taken from.

⁸² <https://www.lexically.net/wordsmith/>

⁸³ No original, lê-se: matching any form of a word, with all its inflections, to a base form, usually the form that would be found as a dictionary headword.

No que diz respeito à geração de listas de palavras-chave, entre outros métodos existentes, Gabrielatos e Marchi (2012) sugerem a análise de tamanho do efeito da chavicidade (*keyness effect size*) das palavras-chave. Esse cálculo envolve a diferença percentual entre a frequência de uma palavra no *corpus* de estudo com sua frequência em um *corpus* de referência, normalizando as frequências em ambos os *corpora* antes de calcular o tamanho do efeito (*ranking by frequency difference – %DIFF*). Os autores argumentam que a observação isolada da diferença estatisticamente significativa das palavras pode levar a interpretações equivocadas, já que o resultado pode ser influenciado por tamanhos distintos de *corpora*. Por outro lado, o %DIFF da chavicidade oferece uma medida mais abrangente da magnitude da diferença ou da similaridade observada, além da significância estatística e da dispersão das ocorrências nos textos do *corpus* de estudo.

Por meio da contagem das ocorrências de palavras, também é possível calcular a medida de *type-token ratio* (TTP), que indica a diversidade de uso de palavras no *corpus*, ou seja, a densidade lexical. Essa medida resulta da divisão do número de tipos de palavras diferentes (*types*), independente da frequência de ocorrência, pelo número total de palavras (*tokens*). Essas técnicas, quando combinadas com análises qualitativas dos textos, são utilizadas para análises preliminares dos *corpora* de estudo e, em outros casos, podem responder perguntas de pesquisa (ver, por exemplo, REBECHI, 2018; NAGASAWA; SCHLATTER, 2021).

Outras abordagens de análises de *corpora* podem incluir a inclusão de informações diretamente nos textos, não nos *headers*, como a anotação de formas inadequadas no uso da língua (*error tagging*). Os pesquisadores podem ter interesse em observar as diferenças na ocorrência desses fenômenos em diferentes *corpora*, como em *corpora* contendo respostas de examinandos de diferentes níveis de proficiência em testes, com o intuito de identificar correlações entre a quantidade de inadequações e o nível de proficiência (CUSHING, 2021). No entanto, esse tipo de estudo apresenta desafios significativos para os pesquisadores, já que determinar o que é considerado inadequado ou adequado na língua-alvo sem comprometer a autenticidade e a naturalidade do uso da língua é uma tarefa complexa. Além disso, automatizar essa marcação por meio de programas computacionais é um desafio adicional, e a análise de recursos linguísticos minuciosos, como desvios lexicais, pode entrar em conflito com avaliações holísticas baseadas no desempenho da interação comunicativa (XI, 2017; CUSHING, 2021).

Por outro lado, programas desenvolvidos por especialistas nas áreas de Ciência da Computação e Linguística permitem a automatização de diversas etapas de processamentos

linguísticos. Esses programas realizam a divisão do texto em palavras individuais (tokenização), a etiquetagem morfossintática (*Part-of-speech*) e a análise sintática (*parsing*). Existem ainda etiquetadores semânticos, que atribuem definições de sentido ou categorias semânticas, bem como etiquetadores discursivos, que identificam referentes anafóricos, tópicos ou marcadores discursivos (BERBER SARDINHA, 2004). Os resultados desses processamentos podem ser utilizados em conjunto com outros programas para análises mais aprofundadas, que fazem uso da exploração estatística dos dados.

Em suma, a LC pode ser considerada uma abordagem metodológica empírica que observa padrões reais da linguagem natural em grandes coleções de textos, fazendo uso extensivo de tecnologias computacionais e técnicas automáticas e interativas, que variam desde as mais simples até as mais complexas. Isso permite que os pesquisadores possam conduzir análises quantitativas e qualitativas para compreender diversos aspectos da linguagem (BIBER, 2012).

5.1 COMPLEXIDADE TEXTUAL EM LÍNGUA INGLESA

Dado que os objetivos das investigações da LC são amplos, não caberia aqui esgotar todas as abordagens já desenvolvidas para o processamento da língua inglesa, portanto, nos voltamos às discussões relacionadas à complexidade textual. Nosso propósito é esclarecer ao leitor o escopo de estudos relevante para esta pesquisa, dado que são referenciais teórico-metodológicos que informaram e inspiraram as decisões adotadas aqui.

Hiebert e Pearson (2014) relatam que investigações sobre a complexidade textual começaram a ser realizadas sistematicamente por linguistas no final do século XIX, com foco exclusivo em aspectos qualitativos. Essas investigações envolviam extensas descrições de características textuais que poderiam impactar a compreensão dos textos⁸⁴. Somente no início do século XX é que métodos quantitativos começaram a aparecer no campo educacional e rapidamente dominaram os esforços de pesquisadores em descrever os aspectos de complexidade textual, especialmente pensando em progressão curricular de leitura de textos escolares⁸⁵. Os autores apontam a publicação do modelo RAND e, posteriormente, a tríade do CCSS-ELA, apresentados na seção 3.3, como momentos importantes a partir dos quais a

⁸⁴ Os autores citam como exemplo: Nikolai Aleksandrovic Rubakin (1889).

⁸⁵ Os autores citam como exemplo: Lively e Pressey (1923) e Gray e Leary (1935).

atenção dos pesquisadores voltou-se para o papel central das características específicas do leitor e seus propósitos de leitura na avaliação da compreensão leitora.

Na literatura que trata da língua inglesa, é possível encontrar um grande volume de estudos que se valem de metodologias quantitativas para investigar aspectos textuais relacionados a níveis de leiturabilidade, extensão de parágrafos, orações e palavras, tipos sintaxe e léxico e medidas de coesão e coerência (CUNNINGHAM; MESMER, 2014; GRAESSER *et al.*, 2014; SHEEHAN *et al.*, 2014; MESMER; HIEBERT, 2015; DASCALU *et al.*, 2018). Sobre esses estudos, Hiebert e Pearson reforçam que as pesquisas quantitativas, embora lidem com variáveis numéricas, exigem que o pesquisador interprete os resultados obtidos por meio das ferramentas analíticas, visto que as “características quantitativas não têm validade até e a menos que sejam compreendidas qualitativamente” (HIEBERT; PEARSON, 2014, p. 157)⁸⁶. Apesar de estarem fora do escopo desta tese, vale ressaltar que também há um grande número de pesquisas sobre dificuldade textual em língua inglesa que trabalham especificamente com metodologias qualitativas, buscando compreender como elementos como níveis de sentido, uso de textos não verbais e demanda de conhecimentos prévios do leitor influenciam no desempenho em leitura (PEARSON; HIEBERT, 2014; SIERSCHYNSKI *et al.*, 2015; OHTA *et al.*, 2018).

Em virtude de esta pesquisa se constituir de análises quantitativas, direcionaremos nossa atenção para os estudos dessa área. Em relação ao tópico da complexidade textual, Biber *et al.* (2021) destacam que “a ‘complexidade’ não é um construto singular e uniforme e, portanto, não é prudente supor que uma única métrica representará adequadamente esse construto” (BIBER *et al.*, 2021, p. 308)⁸⁷. Biber *et al.* (2021c) afirmam que “o construto de complexidade linguística pode ser concebido como um continuum em termos de número de elementos linguísticos e das inter-relações entre esses elementos” (BIBER *et al.*, 2021b, p. 469)⁸⁸.

Com interesse voltado para a complexidade gramatical, Biber *et al.* (2021) conduziram uma análise de períodos compostos e orações subordinadas em artigos acadêmicos e interações orais. Os resultados revelaram padrões de complexidade sintática específicos da modalidade escrita e da modalidade falada e, em vista disso, os autores defendem uma abordagem de pesquisa *bottom-up* que, ao invés de assumir aprioristicamente que certas estruturas gramaticais

⁸⁶ No original, lê-se: In fact, quantitative features have no validity until and unless they are ‘realized’ qualitatively.

⁸⁷ No original, lê-se: “complexity” is not a single unified construct, and it is therefore not reasonable to suppose that any single measure will adequately represent this construct.

⁸⁸ No original, lê-se: the construct of linguistic complexity can be conceptualized as a cline in terms of the number of linguistic elements and the interrelationships among elements.

são mais complexas do que outras, se baseia na análise empírica do *corpus* em questão, levando em consideração a modalidade e os gêneros que o compõem.

Quanto às análises realizadas por meio de fórmulas propostas para designar automaticamente índices de complexidade textual, as primeiras fórmulas de leitura superficial da língua inglesa, que contabilizam número de sílabas, palavras e sentenças para indicar nível de complexidade do texto, datam de meados do século XX (HIEBERT; PEARSON, 2014). Os índices mais utilizados até hoje, como o Índice Flesch⁸⁹ e o Lexile Framework⁹⁰, são adotados por instituições como escolas e universidades para a análise de leitura e a escolha de textos (CUNNINGHAM *et al.*, 2018). No entanto, inúmeras limitações dessas fórmulas já foram evidenciadas por pesquisadores da língua inglesa. Entre essas limitações, Crossley *et al.* (2019) mencionam a ausência de análise da coesão textual, o desempenho inferior desse modelo em comparação a outros que consideram o contexto de leitura, a tarefa e o leitor, e a fragilidade na validade do construto. Em outras palavras, essas fórmulas não são elaboradas com base em conceitos teóricos sobre leitura, levando em conta apenas as correlações estatísticas a nível lexical e sintático para fazer previsões sobre a complexidade textual.

Nesse contexto, a fim de possibilitar uma análise quantitativa mais profunda e confiável dos fenômenos linguísticos que influenciam a complexidade textual, é recomendada a utilização de ferramentas de Processamento de Linguagem Natural (PLN) e métodos de Aprendizagem de máquina (AM) (BALYAN *et al.*, 2017, 2018; CROSSLEY *et al.*, 2017; CROSSLEY *et al.*, 2019; JI *et al.*, 2021). De maneira geral, as análises estatísticas mais robustas, que trabalham com traços linguísticos mais amplos, têm se mostrado mais precisas do que as fórmulas superficiais de leitura, reportando maior acurácia em análises preditivas de complexidade textual em comparação aos modelos clássicos de leitura (BALYAN *et al.*, 2020).

Diante disso, algumas fórmulas foram ampliadas para incorporar fatores relacionados à coesão e coerência, considerando fatores como léxico, sintaxe e aspectos discursivos. Um exemplo é o Coh-Matrix, uma ferramenta desenvolvida por pesquisadores da Universidade de Memphis, que emprega análises geradas por meio de técnicas provenientes das áreas de PLN e AM, como etiquetadores morfossintáticos (*Part-of-speech taggers*), analisadores sintáticos (*parsers*), algoritmos de seleção de atributos e modelos de classificação, com base em uma

⁸⁹ Métrica original do inglês (FLESCH, 1948) adaptada ao português brasileiro em 1996 (MARTINS *et al.* 1996).

⁹⁰ Métrica original do inglês sem adaptação ao português (STENNER *et al.*, 2006).

ampla base de dados (GRAESSER *et al.*, 2003). O Coh-Metrix analisa automaticamente aspectos linguístico-discursivos de textos em língua inglesa, fornecendo 106 métricas⁹¹ em sua versão gratuita e, de acordo com Barker (2010), mais de 600 métricas na versão completa.

É relevante destacar que algumas dessas métricas, embora classificadas como índices quantitativos devido a sua natureza numérica, são resultados de pesquisas experimentais que envolveram testagens com seres humanos. Por exemplo, a medida de Idade de aquisição (*Age of acquisition - AoA*) estima a idade média em que determinada palavra é aprendida por um usuário do idioma e é comumente utilizada para as análises de complexidade lexical, uma vez que as pesquisas já mostram correlação com baixa margem de erro na predição de dificuldade textual (BOTARLEANU *et al.*, 2022). Outra abordagem de investigação lexical decorre de observações sobre a interação do leitor com o texto, como a análise de sofisticação lexical, que leva em consideração fatores como a frequência de uso e a concretude das palavras. Esses traços linguísticos são importantes para o construto teórico de leitura, que engloba tanto a decodificação quanto a construção de sentidos como parte essencial da compreensão leitora (CROSSLEY *et al.*, 2017).

Transcendendo as fórmulas superficiais e os extratores automáticos de métricas, em 1988, Douglas Biber desenvolveu a teoria de Análise Multidimensional (AMD), uma técnica que utiliza o procedimento estatístico de análise fatorial para descrever variações linguísticas em textos orais e escritos em língua inglesa. Seu trabalho, publicado no livro “*Variation across Speech and Writing*”, marcou o início de investigações altamente computacionais e estatísticas aplicadas a grandes *corpora*, possibilitando a análise de múltiplas variáveis para compreender traços linguísticos comuns que podem se relacionar com características situacionais das produções textuais investigadas. Com ênfase na investigação de padrões linguísticos que coocorrem com alta frequência em diferentes registros, a análise começa pela observação da frequência de um grande número de traços linguísticos que, por meio de uma análise fatorial, são reduzidos a fatores que podem ser interpretados para entender a relação funcional entre a variação observada e o contexto situacional. Embora a complexidade textual não seja o foco dessa abordagem metodológica, a AMD também tem sido aplicada em análises de complexidade textual (STAPLES *et al.*, 2016; ANSARIFAR *et al.*, 2018; NESI; GARDNER, 2019) e forneceu a base para uma nova abordagem teórico-metodológica de análise de complexidade gramatical conhecida como abordagem Registro-Funcional (*Register-Functional*

⁹¹ Na versão Coh-metrix 3.0 online gratuita da ferramenta. Disponível em: <http://cohmetrix.com/>. Acesso em: 7 jan. 2023.

– *RF*). A *RF* se interessa em analisar características de complexidade relacionadas ao uso de orações e períodos na construção de discursos, sem partir de conceitos estáticos preconcebidos, como a noção de que unidades longas seriam mais complexas ou que os gêneros acadêmicos escritos teriam mais cláusulas encaixadas do que a língua falada (BIBER *et al.*, 2021c). Os autores argumentam que essas noções subjacentes aos estudos de complexidade textual nem sempre são evidentes e fundamentadas, mas são aceitas como fato, uma vez que parece óbvio que quanto maior a estrutura do período, maior será a complexidade. Contudo, eles mostram que pesquisas recentes relatam resultados contrários a essa operacionalização de complexidade gramatical simplista e destacam que diferentes modalidades manifestam complexidade de maneiras distintas. As variações de complexidade se estendem até escalas muito específicas, como entre diferentes áreas acadêmicas, artigos mais ou menos especializados e tipos de produções de estudantes, entre outros (BIBER *et al.*, 2021c). A modalidade oral exhibe traços de complexidade mais relacionados à natureza da produção em tempo real e espontânea, própria da fala, como as orações dependentes. Por outro lado, a escrita também apresenta suas complexidades linguísticas, visto que as “possibilidades de produção cuidadosa, revisão e edição associadas à modalidade escrita permitem o uso de recursos frasais complexos” (BIBER *et al.*, 2021b, p. 464)⁹². Em vista disso, os autores consideram que a preconização de uma descrição empírica e indutiva da complexidade, baseada no uso e sem restringir o escopo da análise a uma definição *a priori*, tem sido uma grande contribuição da *RF* para os estudos de complexidade, notadamente, no âmbito da complexidade gramatical (BIBER *et al.*, 2021b).

Na mesma linha, Egbert afirma que “a operacionalização da complexidade não é uma tarefa fácil, em parte porque a língua em diferentes registros é complexa de maneiras distintas” (EGBERT, 2017, p. 561)⁹³. O autor critica o uso de medidas de complexidade que carecem de construtos bem definidos, não são motivadas linguisticamente e não são claramente operacionalizadas. Em tais estudos, os pesquisadores partem da hipótese de que, em textos mais complexos, um determinado recurso linguístico é mais frequente. Eles usam as ferramentas automáticas para explorar *corpora* considerados complexos e, ao encontrar o padrão investigado, concluem que existe correlação entre a variável e a complexidade. Isso resulta em uma definição circular de complexidade textual, na qual uma métrica é considerada como uma variável significativa para a complexidade porque sua frequência é recorrente em textos

⁹² No original, lê-se: possibilities for careful production, revision, and editing associated with the written mode, which enable the use of phrasal complexity devices.

⁹³ No original, lê-se: operationalizing complexity is non-trivial in part because the language in different registers is complex in different ways.

complexos, e os textos são considerados complexos porque apresentam alta frequência daquela variável (EGBERT, 2017).

Em relação às motivações linguísticas, Biber *et al.* (2021a) enfatizam a importância de reconhecer as características funcionais dos traços linguísticos que se inter-relacionam de forma holística, em vez de analisá-los sob uma mesma perspectiva sem distinguir seus usos nas diferentes modalidades e gêneros, ou interpretá-los de maneira individual. Para tanto, a análise de trechos dos textos se torna crucial para ilustrar e fundamentar as análises e asserções sobre a complexidade. Os autores afirmam que a prática de análise de exemplos subjaz à convicção de que

[...] os dados quantitativos por si só não capturam toda a extensão da variação e padrões de uso; que características funcionais só podem ser observadas ao examinar a complexidade do uso dentro de seus contextos discursivos; e que qualquer variável linguística quantitativa deve ser facilmente observável e interpretável ao examinar a língua em uso. Surpreendentemente, muitos estudos de complexidade em pesquisas de Linguística Aplicada incluem poucas ou nenhuma análise de trechos de texto que identificam pontualmente as características de complexidade que estão sendo medidas. Tais análises textuais são cruciais para demonstrar o que as medidas específicas de complexidade refletem em termos do uso efetivo da língua (BIBER *et al.*, 2021b, p. 45)⁹⁴.

Em geral, em pesquisas que trabalham com um número grande de traços linguísticos, é necessário reduzir o número elevado de variáveis observadas a um grupo menor de fatores, que são variáveis altamente relacionadas, de modo a possibilitar uma interpretação adequada das dimensões dos dados. Essa característica dos dados, representada coletivamente pelos fatores, é o cerne de abordagens que fazem uso do método estatístico de análise fatorial (HAIR, *et al.*, 2009). Diferentemente dessa abordagem, a nossa pesquisa não analisa diretamente os fenômenos linguísticos de etiquetagem morfosintática dos *corpora*, mas trabalha com as multivariáveis extraídas da ferramenta NILC-Metrix.

Apesar da existência de trabalhos de AMD da língua portuguesa (KAUFFMANN, 2005; BERBER SARDINHA *et al.*, 2014) e da possibilidade de realização de uma AMD aditiva com o nosso *corpus* (KAUFFMANN, 2022), o nosso objetivo não é observar a variação entre os

⁹⁴ No original, lê-se: quantitative data alone fails to capture the full extent of variation and patterns of use; that discourse function can only be observed by examining the use of complexity features within their discourse contexts; and that any quantitative linguistic variable should be readily observable and interpretable by looking at actual language in use. Surprisingly, many studies of complexity in Applied Linguistics research include few if any analyses of text excerpts that directly identify the complexity features being measured. Such textual analyses are crucial for demonstrating what particular complexity measures reflect in terms of actual language use.

registros jornalísticos que compõem nossos *corpora*. Sendo assim, a escolha dessa metodologia de análise não seria produtiva para o nosso propósito.

A AMD aditiva é empregada quando o *corpus* de estudo é de tamanho pequeno. Nesse método, baseando-se em resultados de pesquisas de AMD anteriores que foram realizadas com *corpora* mais extensos, os novos textos a serem adicionados são etiquetados e as frequências das variáveis são normalizadas para se adequar aos escores médios da pesquisa original. A seguir, as dimensões do *corpus* original são calculadas para esses novos textos, possibilitando a interpretação da variação dos novos dados (BERBER SARDINHA *et al.*, 2019). Essa abordagem se justifica pelo fato de que o método estatístico de análise fatorial é sensível ao tamanho da amostra.

Conforme Hair e colaboradores (2009) apontam, para o desenho de uma análise fatorial, o tamanho da amostra, idealmente, deve conter pelo menos 100 observações (textos) e uma proporção de 10 observações para cada variável (métrica linguística) (*10:1 ratio*). Caso contrário, há o risco de que exista uma alta probabilidade de as correlações ocorrerem devido ao acaso, comprometendo a confiabilidade da análise. Isso ocorre porque, em situações com um alto número de variáveis, torna-se desafiador distinguir, por meio de métodos estatísticos tradicionais, as variáveis genuinamente relevantes das que apresentam correlações aleatórias (KURSA *et al.*, 2010). Tal fato constitui um obstáculo no estudo dos dados desta tese por meio do método de análise fatorial, já que nosso *corpus* de estudo é composto por um número baixo de textos e, ao mesmo tempo, um número alto de variáveis a serem observadas.

Como resultado da busca por maior confiabilidade no desenho de análises com amostras pequenas e um grande número de variáveis, Kursá e Rudnicki (2010) desenvolveram o algoritmo Boruta. Esse algoritmo representa “um novo método de seleção de variáveis baseado em random forest, que possibilita uma seleção imparcial e estável de características relevantes e irrelevantes em um sistema de informação” (KURSA; RUDNICKI, 2010, p. 11)⁹⁵.

O funcionamento desse algoritmo envolve a geração de amostras aleatórias das variáveis originais, cada uma com valores distintos. Essas amostras, conhecidas como “variáveis sombra”, são incorporadas ao conjunto de dados, ampliando a quantidade de dados. O classificador compara repetidamente a relevância das variáveis originais com as variáveis sombra, e somente as variáveis originais com uma importância maior do que a melhor variável sombra são aceitas como relevantes. As variáveis originais são classificadas em três níveis:

⁹⁵ No original, lê-se: a novel random forest based feature selection method, which provides unbiased and stable selection of important and non-important attributes from an information system.

“recusa”, que compreende as variáveis consideradas como ruído; “aceite”, que engloba as variáveis preditivas; e “indecisão”, sendo que nessa categoria, o pesquisador confronta variáveis que requerem uma investigação qualitativa para serem classificadas como relevantes ou não (KURSA; RUDNICKI, 2010).

Portanto, o algoritmo Boruta não apenas resolve a limitação relacionada à sensibilidade do método estatístico de análise fatorial ao tamanho da amostra, mas também nos oferece a possibilidade de analisar variáveis duvidosas de forma qualitativa. Isso nos permite incorporar nosso julgamento como especialistas e considerar os resultados de estudos prévios. Além disso, o algoritmo, como veremos mais adiante, já foi utilizado com métricas do NILC-Matrix para a seleção de variáveis que melhor explicam a complexidade textual (LEAL *et al.*, 2021). O funcionamento desse método de redução de multivariáveis será explorado em maior detalhe no capítulo dedicado à metodologia.

Compreende-se, portanto, que a abordagem de trabalhar com múltiplas variáveis, empregar procedimentos estatísticos robustos e examinar exemplares textuais é aconselhada para formular asserções válidas e confiáveis acerca da complexidade textual. Além disso, é necessário considerar as potenciais ameaças às análises, como a variedade dos gêneros dos textos analisados ou um conjunto de dados muito pequeno (SHEEHAN *et al.*, 2014). Tais considerações orientaram as decisões na formulação do desenho metodológico empregado na condução desta pesquisa.

5.2 COMPLEXIDADE TEXTUAL EM LÍNGUA PORTUGUESA

Apesar de os estudos na língua inglesa possuírem uma história consolidada no campo da investigação da complexidade textual ancorada nas bases metodológicas da LC, os índices de leiturabilidade e as ferramentas computacionais para análise de *corpora* do português brasileiro ainda podem ser considerados restritos, como será exposto nesta seção.

Podemos considerar que a LC brasileira foi sedimentada com o auxílio de dois precursores que começaram a trabalhar com *corpora* no final dos anos 1990: o Prof. Dr. Tony Berber Sardinha, da Pontifícia Universidade Católica de São Paulo (PUCSP), e a Profa. Dra. Stella Tagnin, da Universidade de São Paulo (USP). Esses dois linguistas desempenharam um papel fundamental na coordenação de projetos de pesquisa que capacitaram linguistas de *corpus* brasileiros. Além disso, eles realizaram várias iniciativas que contribuíram para a introdução e

o estabelecimento da LC no país, e continuam a contribuir para o fortalecimento dessa área (FONSECA *et al.*, 2021; CAMARGO, 2008; TEIXEIRA, 2023).

Após duas décadas, a LC no Brasil já possui uma rede de pesquisadores e professores universitários especializados na área, presentes em universidades e instituições de ensino superior em todo o país (FONSECA *et al.*, 2021). Nas palavras de Finatto e colaboradoras (2018), no Brasil,

[...] a Linguística de Corpus (LC) tem se associado a diferentes perspectivas de investigação, seja nos Estudos da Linguagem, com destaque para Linguística Aplicada e Estudos do Texto, seja na área de Ciência da Computação, pela via do Processamento da Linguagem Natural (PLN), sem praticamente nada rejeitar em termos de parcerias de trabalho e de trocas de conhecimentos. A disposição ao diálogo disciplinar e interdisciplinar e o princípio de compartilhamento de materiais de pesquisa têm sido marcas constantes da comunidade brasileira de LC (FINATTO *et al.*, 2018, p. 6).

As pesquisas brasileiras empregam as mesmas técnicas já mencionadas para a compilação, tratamento e análise de *corpora*, abordando principalmente investigações relacionadas a estudos de gêneros discursivos, tradução, lexicografia, terminologia fraseológica, variação linguística e ensino de línguas. Conforme destacado por Tagnin (2018), a LC contribui significativamente no avanço dessas áreas de estudo, permitindo a obtenção de resultados objetivos e confiáveis, tendo em vista o contínuo desenvolvimento e especialização das tecnologias das ferramentas utilizadas. Entretanto, é válido mencionar que os estudos recentes no campo da Avaliação de línguas informada por corpus, que abordam testes de língua portuguesa, ainda podem ser considerados pioneiros no país.

A área da LC no Brasil dedicada ao estudo da complexidade textual em português concentra-se, em grande parte, na simplificação textual e na testagem de modelos automáticos de classificação de textos em diferentes níveis de complexidade. Essas investigações buscam compreender quais características textuais simplificam um texto, ajustando-o ao nível de leitura de diversos públicos-alvo, considerando seu nível de letramento, escolaridade e perfil sociodemográfico (PASQUALINI, 2018), bem como possíveis deficiências cognitivas (ALUÍSIO *et al.*, 2016). A simplificação textual resulta na ampliação da acessibilidade desses textos. Os estudos realizados sob essa perspectiva tendem a entender a complexidade textual como uma “propriedade ou condição do texto e, como tal, pode ser mensurada mediante a análise de seus atributos de construção verbal - pelo vocabulário (termos e léxico geral) presente no texto e também por sua configuração gramatical” (PARAGUASSU, 2018, p. 127). Desde essa ótica,

[...] sentenças longas, coordenadas e subordinadas, ordenações sintáticas incomuns, palavras difíceis e de pouco uso, informações irrelevantes, entre outros aspectos, contribuem para formar o conceito de “complexidade da língua”. Assim, “simplicidade”, neste contexto, diz respeito à sintaxe das sentenças, dando preferência à ordem “sujeito-verbo-objeto”, com cada sentença sendo formada por apenas uma oração, na voz ativa (FINATTO *et al.*, 2015, p. 24).

Um dos centros de pesquisa que tem desempenhado um papel crucial no avanço dessa área no Brasil se encontra no Instituto de Letras da Universidade Federal do Rio Grande do Sul, sob coordenação da Profa. Dra. Maria José Bocorny Finatto⁹⁶. A professora tem conduzido projetos e orientado pesquisas relacionadas à simplificação textual e à acessibilidade textual e terminológica em diferentes domínios, como jornalismo, saúde, agronomia, direito, entre outros.

Alguns dos estudos brasileiros relacionados à simplificação textual partem de pesquisas bibliográficas sobre a *Plain Language*⁹⁷, uma abordagem de simplificação da língua inglesa. Esses estudos brasileiros adaptam e testam os preceitos da *Plain Language* para a simplificação do português brasileiro (FINATTO; MOTTA, 2017). Ou seja, os resultados obtidos em pesquisas sobre a língua inglesa são aplicados à língua portuguesa para simplificá-la. A título de exemplo, Paraguassu (2018), em seu trabalho sobre a tradução especializada acessível da língua portuguesa, destaca a relevância das métricas de complexidade descritas pela *Federal Plain Language Guidelines* do Governo Federal dos Estados Unidos, que se aplicam ao inglês americano, como exemplo para o contexto do português brasileiro.

Em relação ao Celpe-Bras, destacam-se dois estudos que abordam a complexidade textual. Primeiramente, mencionamos o trabalho de Evers (2013), detalhado no capítulo 2, no qual as respostas de candidatos do Celpe-Bras foram analisadas usando a ferramenta Coh-Matrix-Port. O objetivo era investigar a viabilidade de correções automáticas baseadas nas métricas de complexidade textual. Sob outra perspectiva, focando na relação texto-tarefa-leitor, ou seja, na noção de dificuldade textual, Gomes (2009) examinou a complexidade das tarefas da PE do Celpe-Bras. A autora utilizou um *corpus* de respostas de estudantes a quatro tarefas que integravam leitura e escrita, produzidas em um contexto de simulação do exame em sala de aula. O procedimento metodológico adotado foi de natureza qualitativa, na qual Gomes triangulou a descrição dos enunciados e dos textos de insumo das tarefas, juntamente com as grades de correção, com o resultado do desempenho observado em 144 produções textuais dos

⁹⁶ <http://lattes.cnpq.br/0180606646463169>

⁹⁷ Estudos brasileiros utilizam termos como Linguagem Simples, Linguagem Simplificada, Linguagem Objetiva e Linguagem Acessível (PARAGUASSU, 2018).

estudantes e com os dados coletados por meio de entrevistas com os autores dessas produções. As respostas dos estudantes foram avaliadas de modo holístico, de acordo com o construto do exame e os critérios de avaliação estabelecidos nas grades, visando verificar o cumprimento da tarefa. Com base nessa análise, Gomes concluiu que a complexidade das tarefas de leitura e escrita não pode ser prevista de antemão, uma vez que ela resulta da interação entre a tarefa, o participante e a avaliação, sendo determinada após a execução da tarefa, o que aponta para a necessidade de testagem das tarefas antes da aplicação, conforme discutido anteriormente.

Além dos dois estudos mencionados sobre a complexidade na PE, também há duas pesquisas que investigaram o conteúdo dos textos de insumo da PO do Celpe-Bras e que, de certa forma, dialogam com o escopo da nossa pesquisa. Candido (2015) conduziu uma investigação qualitativa que associou uma análise documental dos textos de insumo da interação face a face com uma abordagem interpretativista das transcrições das interações orais ocorridas durante a avaliação da PO, com o objetivo de examinar o uso desses materiais pelos avaliadores. A autora analisou as interações de cinco avaliadores na primeira edição de 2014 do Celpe-Bras, considerando tanto os textos de insumo quanto o modo como foram utilizados durante essas interações. A pesquisa apontou que materiais contendo textos verbais longos podem prejudicar o desempenho de candidatos com níveis de proficiência mais baixos, uma vez que demandam um tempo significativo para leitura e interpretação antes e durante a interação com o avaliador. Entre outras considerações, Candido destaca que esses materiais deveriam apresentar um padrão de composição mais controlado para que o construto do exame não seja ameaçado.

Seguindo uma linha semelhante, Faria (2017) também adotou uma abordagem qualitativa e interpretativa para investigar o papel dos textos de insumo na interação oral do Celpe-Bras. O estudo foi conduzido durante a segunda edição de 2014 do exame e envolveu a participação de 19 candidatos e 5 avaliadores. A autora triangulou resultados da coleta de questionários preenchidos por candidatos e avaliadores sobre suas percepções em relação aos textos de insumo da PO, descrições documentais desses materiais e transcrições das interações face a face originadas a partir desses materiais. Dentre os resultados, Faria apontou que o desempenho dos candidatos na PO pode ser comprometido quando a linguagem utilizada nos textos de insumo permite múltiplas interpretações, o que requer explicações por parte do avaliador para que a interação seja estabelecida. Além disso, a autora ressaltou que incongruências entre o tópico e o conteúdo não verbal do texto de insumo podem se colocar como obstáculos para a integração de compreensão e produção do examinando, potencialmente ameaçando a validade do exame.

Em contexto de estudos que abordam a complexidade textual a partir de métricas automáticas extraídas das ferramentas Coh-Metrix-Port e NILC-Metrix, diversos trabalhos têm se dedicado ao processo de criação, testagem e aprimoramento das ferramentas (SCARTON; ALUÍSIO, 2010; LEAL *et al.*, 2021, para citar alguns), e outros trabalhos utilizam essas ferramentas para a análise de complexidade textual (FINATTO *et al.*, 2011; GAZZOLA *et al.*, 2019, para citar alguns). Considerando a importante relação desses trabalhos com esta tese, antes de proceder à apresentação dos procedimentos metodológicos desta pesquisa, trazemos aqui alguns trabalhos prévios sobre a língua portuguesa relacionados ao nosso estudo.

Iniciamos destacando que os estudos de Scarton e Aluísio (2010) e Leal *et al.* (2021) não apenas se relacionam, mas viabilizaram a realização desta investigação, pois são os desenvolvedores da ferramenta NILC-Metrix. O quadro a seguir reúne os traços linguísticos de complexidade textual identificados por diferentes autores, por meio de diferentes metodologias. Uma vez que o NILC-Metrix é uma ferramenta recente e, excetuando Leal *et al.* (2021), não foi empregado nos estudos anteriores, propomos uma correspondência entre cada métrica de complexidade textual indicada nos trabalhos e as métricas disponíveis no NILC-Metrix. No quadro apresentamos o ano da publicação do trabalho, o título e a autoria, os objetivos dos autores, os índices de complexidade textual apontados como resultado de suas pesquisas e as nossas correspondências para o número da métrica e o grupo do NILC-Metrix a qual o índice pertence.

Quadro 8 – Métricas de complexidade textual de trabalhos relacionados

ANO	TÍTULO E AUTORIA	OBJETIVOS	ÍNDICES DE COMPLEXIDADE TEXTUAL INDICADOS	ID NILC (correspondência)	GRUPO NILC (correspondência)
2010	Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. SCARTON, C.; ALUÍSIO, S. M.	Relatar o processo de adaptação de 41 métricas automáticas da ferramenta em língua inglesa, Coh-Metrix, para a língua portuguesa, Coh-Metrix-Port.	Incidência de substantivos Incidência de pronomes por sintagmas Incidência de verbos Número de parágrafos Número de sentenças Sentenças por parágrafos Sílabas por palavras Número de palavras Índice Flesch	106 117 125 152 154 155 157 158 198	1. Medidas Descritivas 11. Informações Morfossintáticas de Palavras 14. Índices de Leiturabilidade
2010	Readability Assessment for Text Simplification. ALUÍSIO, S. M.; SPECIA, L.; GASPERIN, C.; SCARTON, C.	Treinar e testar um classificador automático binário de complexidade textual utilizando métricas automáticas, considerando os níveis de alfabetismo do Inaf como parâmetro de complexidade textual.	Incidence of apposition Incidence of clauses Incidence of relative clauses Words before main verb Number of positive additive connectives Number of negative causal connectives Sentences per paragraph Syllables per word Words per sentence Flesch index	21 22 32 44 47 49 155 157 159 198	1. Medidas Descritivas 7. Conectivos 9. Complexidade Sintática 14. Índices de Leiturabilidade
2011	Características do jornalismo popular: avaliação da inteligibilidade e auxílio à descrição do gênero. FINATTO, M. J.	Identificar índices preditivos de inteligibilidade capazes de distinguir dois <i>corpora</i> , um constituído de textos jornalísticos de veículos tradicionais e	Sobreposição de palavras de conteúdo em sentenças adjacentes Sobreposição do radical de palavras Palavras antes de verbos principais Incidência de adjetivos Frequência de palavras de conteúdo Pronomes por sintagmas	2 5 44 87 96 117	1. Medidas Descritivas 3. Coesão Referencial 9. Complexidade Sintática 11. Informações Morfossintáticas de Palavras

	B.; SCARTON, C. E.; ROCHA, A.; ALUÍSIO, S. M.	outro de jornalismo popular, utilizando métricas automáticas.	Incidência de verbos Número de sentenças Número de parágrafos Sílabas por palavras Número de Palavras Palavras por sentenças Sobreposição do argumento Modificadores por sintagmas	125 154 155 157 158 159 N/A N/A	
2012	Leitura para o neoleitor. FISCHER, L. A.; GARCEZ, P. M.	Indicar índices de simplificação textual de natureza textual e extratextual, ancorados em estudos sociolinguísticos e psicolinguísticos, para tornar a literatura brasileira clássica acessível a neoleitores.	Reiteração ou sintagma anafórico Orações relativas Vocabulário complexo Futuro do pretérito Futuro do presente Pronomes pessoais oblíquos “-lo, -lhe” Pronomes pessoais possessivos “seu, sua” Pretérito mais que perfeito Extensão	18 32 82 101 102 110 121 143 158	1. Medidas Descritivas 3. Coesão Referencial 8. Léxico Temporal 9. Complexidade Sintática 11. Informações Morfossintáticas de Palavras 13. Frequência de Palavras
2013	Processamento de língua natural e níveis de proficiência do português: estudo de produções textuais do exame Celpe-Bras. EVERS, A.	Identificar índices preditivos de complexidade capazes de distinguir diferentes níveis de proficiência utilizando métricas automáticas.	Incidência de conectivos aditivos negativos Incidência de adjetivos Incidência de verbos Número de parágrafos Número de palavras Índice Flesch	46 87 125 152 158 198	1. Medidas Descritivas 7. Conectivos 11. Informações Morfossintáticas de Palavras 14. Índice de Leiturabilidade
2016	Automatic classification of the complexity of nonfiction texts in Portuguese for early	Testar um algoritmo classificador de complexidade textual de textos escolares utilizando métricas	Incidence of additive negative connectives Diversity of punctuation Incidence of Indicative mood (Preterit perfect tense)	46 73 145	1. Medidas Descritivas 6. Diversidade Lexical 7. Conectivos 8. Léxico Temporal 14. Índices de Leiturabilidade

	school years. HARTMANN, N.; CUCATTO, L.; BRANTS, D.; ALUÍSIO, S. M.	automáticas.	Number of paragraphs Flesch	152 198	
2019	Predição da complexidade textual de recursos educacionais abertos em Português. GAZZOLA, M.; LEAL, S. E.; ALUÍSIO, S. M.	Testar algoritmos de seleção de variáveis para verificar o efeito do gênero textual na predição de complexidade textual para diferentes etapas escolares utilizando métricas automáticas.	Densidade de conteúdo Pronomes de primeira pessoa Proporção de pronomes Proporção de palavras de conteúdo do texto com familiaridade entre 4 e 5.5 Proporção de palavras de conteúdo do texto com imageabilidade entre 4 e 5.5 Proporção de palavras de conteúdo do texto com imageabilidade entre 5.5 e 7 Desvio padrão da imageabilidade Desvio padrão do comprimento da sentença Gênero textual	63 98 114 168 180 181 183 185 N/A	1. Medidas Descritivas 5. Medidas Psicolinguísticas 6. Diversidade Lexical 11. Informações Morfosintáticas de Palavras
2021	NILC-Matrix: assessing the complexity of written and spoken language in Brazilian Portuguese. LEAL, S. E.; DURAN, M. S.; SCARTON, C. E.; HARTMANN, N. S.; ALUÍSIO, S. M.	Relatar o processo de desenvolvimento da ferramenta NILC-Matrix e apresentar as 200 métricas de complexidade textual. Além disso, os autores apresentam exemplos de aplicações das métricas.	adj_arg_ovl arg_ovl cross_entropy sentences_with_one_clause content_density content_word_max adverbs_standard_deviation content_words first_person_pronouns function_words non-inflected_verbs prepositions_per_clause prepositions_per_sentence pronouns_min ratio_function_to_content_words verbs_max	1 4 6 35 63 65 95 96 98 99 105 112 113 116 119 126	2. Simplicidade Textual 3. Coesão Referencial 4. Coesão Semântica 5. Medidas Psicolinguísticas 6. Diversidade Lexical 9. Complexidade Sintática 11. Informações Morfosintáticas de Palavras 14. Índices de Leiturabilidade

			idade_aquisicao_mean	176	
			long_sentence_ratio	189	
			dalechall_adapted	197	

Fonte: Elaborado pela autora.

Conforme observado no quadro, os estudos em questão são relativamente recentes, com publicações realizadas na última década. Dois desses trabalhos abordam a criação das ferramentas destinadas à extração automática de métricas de complexidade textual, Coh-Metrix-Port e NILC-Metrix. A maioria dos demais estudos tem o objetivo de testar algoritmos para a classificação automática de textos com base na complexidade textual, utilizando tais métricas. Um aspecto positivo a ser destacado é a diversidade de *corpora* utilizados nas análises, englobando textos jornalísticos, escolares e respostas de candidatos ao exame Celpe-Bras, embora os autores não realizem análises qualitativas de excertos desses textos. Diferenciando-se desses trabalhos, o estudo de Fischer e Garcez (2012) adota uma metodologia qualitativa para indicar traços linguístico-discursivos de complexidade textual em textos literários.

Ao analisar a última coluna do quadro, embora os resultados variem, é possível observar uma indicação frequente dos grupos “Medidas Descritivas”, “Informações Morfosintáticas de Palavras” e “Índices de Leitabilidade” nos diferentes estudos. Isso parece sugerir que as métricas mais preditivas estão relacionadas a cálculos estatísticos básicos de frequência de características textuais em nível lexical e frasal. Embora os estudos citados não correlacionem os resultados das métricas a análises qualitativas, tal como em nossa abordagem metodológica da Avaliação de línguas informada por corpus, os índices indicados pelos autores serão considerados na análise de nossos dados e, por essa razão, consideramos essencial delineá-los aqui.

Isso posto, um primeiro aspecto que se evidenciou para nós, ao buscar um modelo metodológico que sustentasse este estudo, foi a distinção entre as concepções de leitura e de complexidade textual presentes nos estudos de LC no Brasil e as nossas próprias. Desde o início, percebemos que encontrar uma única abordagem que fosse capaz de abranger a especificidade de nossos dados e objetivos de pesquisa, em consonância às nossas concepções teóricas, sem limitar ou enviesar as análises e resultados, seria um desafio. Por outro lado, ao aprofundar as leituras no campo da LC, tornou-se evidente que replicar procedimentos analíticos *ipsis litteris* em *corpora* distintos não é uma abordagem indicada, uma vez que a variedade e a modalidade da linguagem, os propósitos de investigação, o tamanho da amostra e o número de variáveis impactam diretamente os resultados obtidos.

Reiteramos que, nesta tese, partiremos de observações empíricas do *corpus* de estudo, evitando apoiar-nos exclusivamente em pressupostos teóricos, mas valendo-nos da literatura da área como referência no momento de interpretar os resultados de nossas análises. Assim, para

este estudo, implementamos um desenho metodológico amalgamado que combina técnicas propostas por pesquisadores da língua inglesa com a adaptação e a expansão de modelos já existentes, que foram desenvolvidos a partir de pesquisas sobre a língua portuguesa. Esse enfoque permite uma abordagem mais flexível e adaptada às nuances de nossa análise, considerando tanto os aspectos teóricos subjacentes quanto às particularidades dos nossos dados e tópicos de interesse.

Por fim, com base nas possibilidades de análises e contribuições produtivas indicadas pelos teóricos do campo da Avaliação de línguas informada por corpus, bem como em suas advertências, no próximo capítulo apresentaremos o desenho da pesquisa empreendida nesta tese. Todas as considerações tecidas até aqui acerca do poder social de testes de proficiência, da responsabilidade dos elaboradores pela validação dos instrumentos, e das concepções de texto, leitura e complexidade textual proporcionam reflexões teóricas valiosas que orientaram as decisões metodológicas a serem detalhadas no próximo capítulo. Contudo, na prática, ao refletirem uma concepção ampla de uso da língua portuguesa como prática sócio-historicamente situada, reconhecemos que o número de variáveis que podem influenciar no desempenho da leitura dos textos examinados é incomensurável, se considerarmos os inúmeros contextos de leitura e de leitores possíveis desses textos, ainda que sob a perspectiva empírica e probabilística da LC. Sem embargo, ao ponderar sobre a importância de ampliar as investigações nessa área e os desdobramentos desta tese em pesquisas futuras, reforça-se a relevância do enfrentamento do desafio de implementar um desenho de pesquisa que, desde seu início, sabia-se que abordaria apenas a primeira camada das questões elucidadas.

6 METODOLOGIA

O objetivo deste capítulo é apresentar os passos metodológicos adotados para a implementação desta pesquisa. Caracterizada pela análise de dados linguísticos por meio de programas computacionais e tratamentos estatísticos, a LC viabiliza a interpretação de traços distintivos da língua que compõem um dado conjunto de textos em comparação com outro, exigindo a manipulação de múltiplos conjuntos de dados. Isso não é diferente neste estudo, que busca descrever o conteúdo das tarefas integradas de leitura e produção escrita da PE do Celpe-Bras e a complexidade textual de seus textos de insumo. Para conduzir nossas análises, foram necessários mais dois conjuntos de dados que foram empregados como *corpora* de referência, considerados representativos de textos da esfera jornalística e indicadores de complexidade textual. Esses *corpora* foram selecionados e processados de acordo com os preceitos da LC, que preconiza sólidos princípios na compilação e utilização de *corpora* que atendam aos objetivos de pesquisa de maneira ética e válida (EGBERT, 2017).

Na primeira seção deste capítulo, discorreremos sobre os *corpora* utilizados neste estudo. Nas seções seguintes detalhamos as faixas de alfabetismo do Indicador de Alfabetismo Funcional (Inaf), usadas como referência de complexidade textual para diferentes públicos-alvo e as 200 métricas da ferramenta NILC-Matrix, utilizadas para operacionalizar a complexidade textual. Por fim, apresentamos os procedimentos metodológicos adotados. Vale ressaltar novamente que todas as análises realizadas neste estudo são de caráter descritivo e exploratório, não sendo, portanto, de natureza inferencial. Todas as análises foram executadas por meio da linguagem de programação R, utilizando seu ambiente de desenvolvimento integrado RStudio. Esse ambiente único permitiu a manipulação e observação dos textos, a geração de gráficos e

tabelas e a realização de testes estatísticos. Para a condução das análises, utilizamos os pacotes `knitr`⁹⁸, `tidyverse`⁹⁹, `ggrepel`¹⁰⁰, `ggthemes`¹⁰¹, `boot`¹⁰² e `boruta`¹⁰³.

6.1 OS CORPORA

O *corpus* de estudo utilizado neste trabalho é composto por todos os textos de insumo das tarefas 3 e 4 aplicadas na história do exame Celpe-Bras até 2020, totalizando 45 edições. As análises finais contaram com 79 tarefas, visto que 11 tarefas foram excluídas porque não apresentavam texto de insumo ou somente apresentavam perguntas de leitura (sem integrá-las à escrita) ou ainda porque a quantidade de palavras do texto desviou da média normal de quantidade de palavras do texto de insumo, como será detalhado na seção de resultados. Nosso *corpus* de estudo é considerado especializado, isto é, conforme definido por Sinclair (2004), é um *corpus* de menor tamanho em comparação com os tipicamente utilizados em estudos que buscam um olhar mais amplo de uso da linguagem. Nesse sentido, portanto, está fora de seu escopo fornecer evidências sobre a natureza mais geral da linguagem (PHILIP, 2008). Um *corpus* especializado “contém textos de um determinado tipo para que se possa estudar a linguagem relativa a um determinado tema, registro ou situação” (HUNSTON, 2022, p. 21)¹⁰⁴. Com relação aos nossos objetivos de pesquisa, compilamos todos os textos disponíveis no

⁹⁸ XIE, Yihui. `knitr`: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.42., 2023. XIE, Yihui. *Dynamic Documents with R and knitr*. 2nd edition. Chapman and Hall/CRC, 2015. ISBN 978-1498716963 XIE, Yihui. `knitr`: A Comprehensive Tool for Reproducible Research in R. In: STODEN, Victoria; LEISCH, Friedrich; PENG, Roger D. (editors). *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2014. ISBN 978-1466561595

⁹⁹ WICKHAM, H.; AVERICK, M.; BRYAN, J.; CHANG, W.; MCGOWAN, L. D.; FRANÇOIS, R.; GROLEMUND, G.; HAYES, A.; HENRY, L.; HESTER, J.; KUHN, M.; PEDERSEN, T. L.; MILLER, E.; BACHE, S. M.; MÜLLER, K.; OOMS, J.; ROBINSON, D.; SEIDEL, D. P.; SPINU, V.; TAKAHASHI, K.; VAUGHAN, D.; WILKE, C.; WOO, K.; YUTAI, H. Welcome to the tidyverse. *Journal of Open Source Software*, 4, 43, 1686, 2019. doi:10.21105/joss.01686 <<https://doi.org/10.21105/joss.01686>>.

¹⁰⁰ SLOWIKOWSKI, K. `ggrepel`: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.3, 2023. <<https://CRAN.R-project.org/package=ggrepel>>.

¹⁰¹ ARNOLD, J. `ggthemes`: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 4.2.4, 2021. <<https://CRAN.R-project.org/package=ggthemes>>.

¹⁰² CANTY, Angelo; RILEY, Brian. `boot`: Bootstrap R (S-Plus) Functions. R package version 1.3-28.1, 2022. DAVISON, A. C.; HINKLEY, D. V. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge, 1997. ISBN 0-521-57391-2

¹⁰³ KURSA, M. B.; RUDNICKI, W. R. Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36, 11, 1–13, 2010. <https://doi.org/10.18637/jss.v036.i11>

¹⁰⁴ No original, lê-se: [A specialised corpus] contains texts of one particular type so that the language relating to a particular topic, register or situation can be studied.

contexto de uso focalizado, até o momento em que os dados foram coletados; sob esse viés, estamos trabalhando com a totalidade da linguagem a ser investigada, e não uma amostra.

Para a compilação do *corpus* de estudo, denominado neste trabalho “*corpus* Celpe”, foram baixadas todas as provas aplicadas no período de 1998 a 2020 a partir da página do Acervo Celpe-Bras (www.ufrgs.br/acervocelpebras). Essas provas foram obtidas em formato .pdf e seus textos foram digitados manualmente em formato .txt, utilizando a codificação de caracteres em UTF-8¹⁰⁵, que é compatível com os programas computacionais destinados à análise linguística. Foram transcritos não apenas os textos de insumo, mas também os títulos e os enunciados de todas as quatro tarefas que compõem as edições das provas. Apesar de nosso foco de análise estar nos textos de insumo das tarefas 3 e 4, é importante ressaltar que o material compilado pode oferecer contribuições para futuras investigações do grupo de pesquisa Avalia, visando avançar nas investigações relacionadas à composição das provas e à complexidade das tarefas. As transcrições foram realizadas por mim e revisadas por dois membros do grupo de pesquisa Avalia, assegurando assim a precisão do trabalho e a padronização do processo. Cabe mencionar que, nessa etapa da pesquisa, todos os membros envolvidos participaram de reuniões, durante as quais foram alinhadas as concepções teóricas que embasaram as decisões metodológicas tomadas acerca das transcrições.

O conteúdo foi transcrito em um único documento, portanto, tornou-se necessário separar os textos de forma que cada arquivo representasse apenas uma tarefa. A separação dos textos foi executada pela coorientadora usando código Python, nomeando os arquivos com o nome do exame, ano, edição, tarefa e quantidade de textos usados (*e.g.*, CELPE_2010_2_3_1). Subsequentemente, em cada arquivo do *corpus* de estudo foram incluídos os *headers* contendo dados específicos de cada texto, os quais foram inseridos entre colchetes angulares. Esses cabeçalhos incluem as variáveis relacionadas às categorias descritivas das tarefas propostas pelo grupo Avalia e apresentadas em Schoffen *et al.* (2018), conforme apresentado na seção 4.1. Importante destacar que nenhuma forma de anotação foi realizada diretamente no corpo dos textos.

Além das tradicionais edições 1 e 2 que ocorrem anualmente, é importante destacar que em 2008 houve uma reaplicação da segunda edição¹⁰⁶ do ano, e em 2016, ocorreu uma

¹⁰⁵ Para a codificação UTF-8, utilizamos a ferramenta *Corpus Text Processor* disponibilizada pelo grupo de pesquisa Corpus & Repository of Writing (Crow). Disponível em: <https://github.com/writecrow/ciabatta/wiki/6a.-Automatic-processing-with-our-Corpus-Text-Processor> Acesso em: 7 fev. 2023.

¹⁰⁶ Edição aplicada na Universidade Estadual de Campinas.

reaplicação da primeira edição¹⁰⁷ do ano. Nos nossos estudos, ambas edições foram categorizadas como edição 3. Para ilustrar, apresentamos a seguir um exemplo de transcrição com *headers* das tarefas pertencentes à segunda edição de 2016.

Quadro 9 – Exemplo de transcrição do corpus Celpe

<p><ano:2016> <edição:2> <arefa:3> <revisores:Mariana Alves,Fernanda Xavier> <número de textos:1> <texto não verbal:sim> <cor:sim> <avalia temática:políticas e cidadania> <avalia esfera de atuação:comunitária> <avalia propósito:incentivar> <avalia relação de interlocução:integrante de grupo social para seus pares> <avalia natureza da interlocução:pública, assimétrica descendente> <avalia gênero do discurso:artigo> <avalia suporte:blog> <título> Cadastro de hortas </título> <enunciado> Você tem um blog sobre alimentação saudável e, depois de ler a notícia sobre o cadastro de hortas em São Paulo, decidi implementar essa iniciativa em sua cidade. Escreva um texto em seu blog para pedir a participação dos leitores no cadastramento, explicando o funcionamento e a utilidade das informações coletadas. </enunciado> <texto l:reportagem,jornal,completo,Estadão> À procura de horas em plena São Paulo Até esta manhã, eram vinte e cinco hortas cadastradas, sendo vinte de caráter “educativo” – ou seja, comunitárias e escolares – e cinco “comerciais” – quer dizer, voltadas à produção e venda de alimentos. O que elas têm em comum? Estão todas dentro do município de São Paulo. Várias – olhem só! – em densas áreas urbanas, como nos bairros de Vila Madalena, Vila Mariana, Pompéia, Mooca, Cambuci, Casa Verde, Sapopemba, Vila Nova Esperança, Jardim Ângela... Já as “comerciais” inserem-se sobretudo na zona rural paulistana. Desde 2014 a cidade de São Paulo “ganhou” uma área rural, ou melhor, porções do território paulistano situadas principalmente nos extremos sul e leste da capital passaram a ser reconhecidas como “rurais”. Esses dados vêm de uma pesquisa colaborativa em pleno andamento. Quem souber da existência de alguma horta urbana e orgânica sendo cultivada em algum “rincão” da metrópole pode entrar na plataforma Wikiversidade e inserir os dados ali. Tem de preencher o nome da horta; se ela é educativa ou comercial; a região em que está; o bairro; o endereço e, por fim, o link para contato. O cadastramento dessas hortas urbanas foi iniciado pela dupla Regiane Nigro, que trabalha no Instituto Kairós, em São Paulo, e a jornalista e uma das fundadoras do grupo Hortelões Urbanos, Claudia Visoni. Ambas têm forte atuação em duas dessas hortas comunitárias, a da Praça das Corujas, na Vila Madalena, e também a Horta do Ciclista, na Praça do Ciclista, região da Avenida Paulista. Regiane atua também, pelo Kairós, com os produtores da Associação de Agricultores da Zona Leste, que contempla bairros do extremo leste paulistano, como São Mateus, Guaianases, Cidade Tiradentes e São Miguel Paulista. A ideia de cadastrar todas as hortas paulistanas – embora a plataforma também permita a inclusão de cultivos urbanos em todo o país – já era antiga, conta ao blog Regiane Nigro. E foi finalmente posta em prática há cerca de 15 dias, para servir como base de uma reunião que Claudia Visoni teria com pessoas de outros países interessadas no tema hortas urbanas. “Foi este o motivador para o projeto começar a andar”, conta Regiane.</p>
--

¹⁰⁷ Edição aplicada na Universidade Federal do Pará.

Se inicialmente, há cerca de cinco anos, quando a “onda” do resgate urbano da natureza começou aqui em São Paulo, as hortas cultivadas em plena metrópole eram raras, agora pode-se dizer que elas já têm animado bastante gente disposta a inserir mais verde na selva de pedra. Na maioria dessas hortas, dá-se preferência ao cultivo orgânico.

Regiane conta que a plataforma é importante também para estimular os produtores comerciais orgânicos a incluírem suas hortas ali e se tornarem mais conhecidos – afinal, muitos deles fornecem, em sistema de compra coletiva, alimentos para a metrópole, entre outras formas de comercialização.

Atualmente, nos “rincões” paulistanos, um dos grupos mais organizados de agricultores – e que já estão lá no cadastro – é o da Cooperapas, no extremo sul da capital, região de Parelheiros. Outra grande utilidade deste cadastro é atrair mais pessoas interessadas em se tornarem voluntárias no cultivo orgânico em plena cidade.

Claudia Visoni, ao chamar, num grupo do Facebook, as pessoas a colaborarem com o cadastramento das hortas, comenta que o registro é “importante para estudos acadêmicos, para os voluntários acharem uma horta perto de casa, para possíveis apoiadores (dos agricultores profissionais) e também para quem não está no Facebook”. E continua: “Quem puder, ajude a incluir as hortas que ainda não estão na tabela. Quem quiser criar tabelas para outras cidades, é só fazer. A plataforma Wiki é totalmente colaborativa e fácil de editar”.

Disponível em: <http://vida-estilo.estadao.com.br/blogs/alimentos-organicos/a-procura-de-hortas-em-plena-sao-paulo/>. Acesso em: 5 ago. 2016 (adaptado)

</texto1>

<ano:2016>

<edição:2>

<tarefa:4>

<revisores:Mariana Alves,Fernanda Xavier>

<número de textos:1>

<texto não verbal:sim>

<cor:sim>

<avalia temática:consumo>

<avalia esfera de atuação:jornalística>

<avalia propósito:posicionar-se>

<avalia relação de interlocução:leitor/ouvinte para veículo de comunicação>

<avalia natureza da interlocução:pública, simétrica>

<avalia gênero do discurso:carta do leitor>

<avalia suporte:jornal>

<título>

Shopping em casa

</título>

<enunciado>

Depois de ler o texto “Shopping em casa”, você decidiu escrever para a seção Cartas do Leitor do jornal O Estado de São Paulo, defendendo seu ponto de vista em relação às questões levantadas por Denise Fraga.

</enunciado>

<texto1:crônica,jornal,completo,Folha de São Paulo>

— Que bonita esta saia!

— Pois é, estava perdida — ela me respondeu.

— Você perdeu uma saia?

— Sim, dentro do meu armário. Arrumando as coisas pra mudança, eu achei.

Esta conversinha banal com minha amiga aumentou a minha angústia. Será que preciso de uma mudança pra conseguir achar o que anda perdido diante dos meus olhos? Muitas vezes, tenho a sensação de não ter o que vestir diante do meu guarda-roupa lotado e fico doida cobiçando um novo vestido na vitrine do shopping. No outro dia, comprei uma blusa que tinha amado e, quando fui pendurá-la no armário, vi que sua vizinha de cabide era sua prima. O curioso é que, se fosse a blusa antiga que estivesse na loja, eu a compraria de novo.

Mais curioso: ela estava lá, eu sabia que tinha e pouco a usava. A blusa nova me devolvera a velha, que passei a usar com prazer. Mas por que comprei uma nova se havia uma quase igual pendurada? Que fenômeno é esse de precisar do que já se tem? Que cegueira é esta para o que está ao seu lado? Será a galinha do vizinho sempre mais gorda? Adquirir é melhor do que ter?

Não sou exatamente consumista, mas olho a minha casa e, apesar de morrer de amores por meus objetos, sinto que há uma infinidade de coisas que eu realmente poderia viver sem. O curioso é que, se descuidar, se me forem oferecidas, estou arriscada a comprá-las de novo.

É realmente muito violento o impulso consumista a que fomos doutrinados. Nem precisamos ir ao shopping. Vivemos em uma feira constante. Meu e-mail está lotado de propostas de vendas que nem desconfio como

chegaram ali. A internet virou, como previsto, um mar de publicidade que, a um clique, desfalca nossa conta bancária. Tudo pela estrela maior: o dinheiro, o circular da moeda. Cada vez mais produtos, cada vez mais necessidades desnecessárias. Não entendo nada de economia, mas gostaria de viver num mundo com menos coisas. Gostaria de morar numa casa com menos coisas, ter um armário com menos roupas.

Resolvi fazer um exercício: promover as vitrines a uma exposição de arte. A arte de um artista poderoso e midiático que se chama Mercado e de quem estou longe de poder adquirir uma obra. Mesmo sem ter que mudar de apartamento, vou tirar as coisas de meu armário, colocar sobre a cama para comprar com meu novo olhar os meus velhos vestidos. Liberdade é viver com pouco.

Disponível em: www1.folha.uol.com.br/colunas/denisefraga/2016/02/1739053-shopping-em-casa.shtml.
 Acesso em: 5 ago. 2016 (adaptado)

</texto1>

Fonte: Elaborado pela autora.

Neste exemplo, podemos identificar informações relacionadas ao número de textos destinados à leitura, a presença de elementos não verbais nos textos e se a edição é colorida. Também estão presentes as sete categorias de descrição das tarefas elaboradas pelo grupo Avalia: temática, esfera de atuação, propósito, relação de interlocução, natureza da interlocução, gênero do discurso e suporte. Além disso, são fornecidos o título, enunciado e texto de insumo, bem como detalhes acerca desse material, incluindo a indicação do gênero e suporte do texto, se ele foi reproduzido integralmente ou se representa apenas um fragmento, e a fonte de publicação do texto original. As informações contidas nesses *headers* serão explanadas na próxima seção.

As transcrições foram realizadas seguindo as normas do Acordo Ortográfico da Língua Portuguesa de 1990. Foram particularmente observados aspectos como o uso de trema, acento agudo, crase e hífen. Como resultado, alguns textos de insumo sofreram adaptações dos originais publicados nas provas, uma vez que as novas regras ortográficas entraram em vigor somente em 2009, dez anos após a instituição do exame. A escolha de padronizar as transcrições de acordo com o novo acordo ortográfico visou aprimorar as análises relacionadas ao léxico.

Para a análise do *corpus* de estudo, utilizamos dois *corpora* de referência compostos por textos provenientes da esfera de comunicação jornalística. Portanto, esses *corpora* compartilham características semelhantes com os textos de insumo presentes no *corpus* de estudo. Os dois *corpora* são similares entre si no que diz respeito aos propósitos comunicativos, à modalidade da língua e ao modo de circulação. Contudo, conforme mencionamos anteriormente, eles divergem quanto ao nível de complexidade da linguagem utilizada, de acordo com o público-alvo a que se destinam. Tal decisão metodológica está em consonância, entre outras, com as orientações apresentadas por Kübler e Aston (2010), que recomendam que a composição do *corpus* de referência leve em consideração a esfera e o gênero dos textos, a fim de estabelecer correspondência adequada. Como já mencionado anteriormente, neste

estudo, os dois *corpora* de referência desempenham a função de parâmetros substitutos de linguagem jornalística complexa e simples: o *corpus* PorSimples é utilizado como substituto da linguagem complexa (ALUÍSIO *et al.*, 2010), enquanto um *subcorpus* do *corpus* PorPopular serve como substituto da linguagem simples (FINATTO *et al.*, 2011).

O *corpus* utilizado como *proxy* da língua jornalística complexa, denominado em nosso estudo como “*corpus* PorSimples”, foi compilado pelo grupo de pesquisa NILC em 2009, com a finalidade de servir como base para o treinamento do algoritmo de classificação de textos complexos e simples, como explicado anteriormente. Tivemos acesso a 148 textos do *corpus* PorSimples, os quais foram disponibilizados por Leal *et al.* (2018) na página <https://github.com/sidleal/porsimplessent>. Esses pesquisadores explicam que o *corpus* é constituído por textos extraídos dos jornais *Zero Hora*¹⁰⁸ e *Folha de São Paulo*¹⁰⁹. Embora não tenhamos conseguido obter uma descrição detalhada da composição do *corpus* PorSimples, Leal *et al.* (2021) indicam que o *corpus* original era composto por 154 textos. Destes, 104 textos eram provenientes do jornal *Zero Hora* e 50 textos do *Caderno de Ciência* do jornal *Folha de São Paulo*. Esses textos foram considerados complexos de acordo com o público-alvo dos veículos originais, os quais, como discutido previamente, pertencem a classes socioeconômicas de maior poder aquisitivo e maior grau de escolaridade.

O jornal *Zero Hora* foi fundado em 1964 e é gerido pelo *Grupo RBS*. Desde sua criação, o jornal tem tido publicações impressas diariamente na cidade de Porto Alegre, e, desde 2007, oferece uma versão digital que disponibiliza parte de seu conteúdo de forma gratuita. Fundado em 1921, o jornal *Folha de São Paulo*, atualmente organizado pelo *Grupo Folha*, é uma publicação diária da cidade de São Paulo. A partir de 1995, passou a disponibilizar acesso ao seu conteúdo pela internet, sendo que uma parcela é gratuita e outra é exclusiva para assinantes. Como destacado anteriormente, ambos os jornais são considerados referências da mídia brasileira devido à sua sólida reputação e credibilidade entre os leitores e formadores de opinião do país.

Scarton e Aluísio (2010) caracterizam o *corpus* PorSimples como constituído por textos do “gênero jornalístico”. As autoras indicam que os textos provenientes da *Zero Hora* foram coletados entre os anos de 2006 e 2007. Por sua vez, Aluísio *et al.* (2010) nos informam que os textos compilados da *Folha de São Paulo* foram coletados randomicamente no período entre 1994 e 2005. Devido à ausência de descrições mais detalhadas sobre os gêneros que compõem

¹⁰⁸ <https://gauchazh.clicrbs.com.br>

¹⁰⁹ <https://www.folha.uol.com.br>

o *corpus*, optamos por considerá-los e nomeá-los aqui como gêneros da esfera jornalística. Os autores esclarecem que o nível de complexidade textual dos textos é direcionado ao público com o mais alto nível de funcionalmente alfabetizados do Inaf, ou seja, o nível Proficiente (*advanced*). Conforme apresentaremos mais detalhadamente na seção 6.2, a seguir, esse nível abrange somente 12% dos indivíduos entrevistados na amostragem representativa da população brasileira conduzida em 2018 pelo Inaf. Eis um exemplo de um texto do *corpus* PorSimples categorizado como sendo do nível Proficiente do Inaf:

Quadro 10 – Exemplo de transcrição do corpus PorSimples

O trabalho de pesquisadores da USP está revelando uma série de novas espécies de um tipo todo especial de fungo: Pequenos cogumelos que emitem uma misteriosa luminosidade verde no escuro. As criaturas, antes desconhecidas no Brasil, podem ajudar a elucidar o mecanismo bioquímico que leva à produção de luz em fungos. Além disso, com um pouco mais de estudo, poderiam servir como sensores vivos de poluição ou mesmo fontes de moléculas úteis para a biotecnologia.

Segundo Cassius Vinicius Stevani, químico da USP que coordena os estudos, é possível que o material recolhido abranja pelo menos dez espécies novas. Não é pouca coisa, já que no mundo todo se conhecem só 42 espécies do fungo, quase todas restritas ao Sudeste Asiático. Já temos uma descrita, a *Gerronema viridilucens*, e seis já estão em fase de descrição, contou Stevani à Folha.

Mistério antigo.

Em fungos, a bioluminescência, trocando em miúdos, a luminosidade produzida por seres vivos, é um fenômeno conhecido há milênios, tendo sido descrito pelo filósofo grego Aristóteles (século 4º a.C.). Os cientistas, no entanto, continuam sem saber como ela funciona. Foi na esperança de entender o fenômeno que Stevani e seus colaboradores se puseram a procurar cogumelos do tipo no Brasil. É provável que nenhum pesquisador tivesse identificado espécies no Brasil porque é preciso procurá-los à noite. Além do mais, a luz é muito fraca, até sob o luar, explica o químico.

Com base em relatos de moradores e muita paciência na coleta, a equipe chegou aos cogumelos em dois locais: O Parque Estadual Turístico do Alto Ribeira, em Iporanga (SP) e uma área de mata ciliar no município de Costa Rica, Mato Grosso do Sul. Outros relatos sugerem que fungos do mesmo tipo podem ser encontrados no Tocantins e no Amazonas. A identificação e descrição está sendo feita com a ajuda de Marina Capelari, do Instituto de Botânica de São Paulo, e Dennis Desjardin, da Universidade Estadual de San Francisco (Unidos). A emissão de luz é efêmera. Os cogumelos, que na verdade são apenas uma espécie de órgão sexual do fungo, duram entre 24 e 48 horas. Eles é que emitem luz, que aparentemente pode vir tanto da parte inferior do chapéu quanto do “cabo” ou do cogumelo inteiro. Algumas espécies surgem da casca de árvores vivas, enquanto outras se alimentam de matéria vegetal em decomposição.

Apesar da rapidez do processo, os pesquisadores descobriram como cultivar o *G. Viridilucens* em laboratório e verificaram que ele pode funcionar como uma espécie de sensor de poluição. É que a luz parece variar de acordo com a concentração de poluentes aplicados sobre o fungo.

Dessa forma, você teria um parâmetro rápido de medir e fácil de avaliar, ao contrário de outros biossensores com fungos, que exigem tempo maior de aplicação e tratamento estatístico mais complicado, explica Stevani. Como os fungos ajudam a reciclar nutrientes, têm um importante papel na manutenção da saúde dos ecossistemas, outra razão para usá-los como indicadores.

Finalmente, com informações mais detalhadas sobre o mecanismo da luminescência, uma série de aplicações úteis pode surgir. É que os estudos de biologia molecular dependem muito de reações que geram luz e de proteínas fluorescentes de diversas cores. Quando um pesquisador quer saber se determinado gene está ativo ou não, por exemplo, gruda nele o trecho de DNA correspondente à GFP (proteína fluorescente verde). Se o gene for ativado, a proteína “acende”, dedurando-o.

Como as técnicas desse tipo têm limitações, é sempre bom contar com novas opções, avalia Stevani, com a ressalva de que os estudos ainda estão no começo.

Fonte: Elaborado pela autora.

O exemplo apresentado acima ilustra exatamente como os textos são disponibilizados publicamente pelo grupo de pesquisa NILC, estando isento de qualquer cabeçalho e organizados em arquivos individuais numerados de 1 a 164, embora o número total de arquivos seja 148. Para utilizar o *corpus* PorSimples em nossas análises, também nesses textos realizamos a revisão para o novo acordo ortográfico da língua portuguesa. Ademais, realizamos correções relacionadas à digitação, tais como: espaço em branco extra, falta de espaço em branco, letras trocadas e falta de capitalização em nome próprio.

A escolha de incluir a totalidade dos textos disponíveis no *corpus* PorSimples respondeu ao nosso anseio de trabalhar com o maior número de textos possíveis de modo a assegurar a validade dos testes estatísticos realizados. Paralelamente, preocupamo-nos em estabelecer equilíbrio entre os *corpora* utilizados como parâmetros substitutos para a classificação de nível de complexidade. Nesse sentido, ainda que o *corpus* PorPopular contemple um número maior de textos, deliberadamente optamos por selecionar a mesma quantidade de arquivos para assegurar a simetria com o *corpus* PorSimples.

Nosso segundo *proxy* é referente à linguagem jornalística simples e é composto por partes de *subcorpora* do *corpus* PorPopular. O *corpus* PorPopular é resultado de um projeto de pesquisa coordenado pela Professora Doutora Maria José Bocorny Finatto, do Instituto de Letras da Universidade Federal do Rio Grande do Sul, e foi organizado em duas fases. A primeira fase, iniciada em 2008, inclui textos compilados do jornal popular *Diário Gaúcho*, do Rio Grande do Sul, abrangendo 7 meses do ano de 2008 e 1 mês de 2010, todos organizados em um arquivo anual. A segunda fase do projeto, que teve início no final de 2012, visou expandir o *corpus*. Para isso, textos do jornal popular *Massa!*, do estado da Bahia, dos anos 2011, 2012, 2014 e 2015 foram compilados, também organizados em um arquivo por ano.

Ambos os jornais são classificados como “jornais populares brasileiros” e são definidos por Finatto e colaboradoras (2011) como pertencentes a um segmento de jornais comerciais “que visa uma aproximação com as amplas camadas da população urbana de menor poder econômico, consumidores com nível de escolaridade relativamente baixo e pouco hábito de leitura” (FINATTO *et al.*, 2011, p. 49). O *Diário Gaúcho*, tal como a *Zero Hora*, também é organizado pela empresa jornalística *Grupo RBS* e publicado na cidade de Porto Alegre, porém, é direcionado a leitores das classes sociais C, D e E (FINATTO *et al.*, 2011). Fundado em 2000, o *Diário Gaúcho* se apresenta como um jornal de aquisição avulsa, disponível em bancas, mercados e outros pontos, além de oferecer conteúdo gratuito na internet. Por sua vez, o jornal *Massa!* é publicado em Salvador pelo conglomerado de mídia *Grupo A Tarde* desde 2010,

focalizando as classes C e D da cidade. Assim como o *Diário Gaúcho*, o *Massa!* está disponível em formato impresso e digital, oferecendo conteúdo gratuito na versão digital (RIBEIRO, 2013).

Finatto *et al.* (2011) e Finatto (2012) reportam resultados de análises contrastivas de inteligibilidade entre textos do *corpus* PorPopular e o *Banco do Português*, assim como textos do jornal *Zero Hora*. Por meio de procedimentos da LC, os autores evidenciam diferenças lexicais e de índices do Coh-Metrix-Port relacionados à coesão e coerência, que permitem a distinção automática entre esses textos. Essas descobertas fundamentam o reconhecimento de um uso particular da língua portuguesa denominado por Finatto (2012) de “Português Popular Escrito”. O estudo de Zilio (2015) acrescenta que os textos do *corpus* PorPopular tendem a empregar sentenças simples, evitando extensões preposicionais entre as estruturas sintáticas, bem como fazem uso limitado de voz passiva e adjuntos, quando comparados a um *corpus* de artigos científicos da área da saúde. Segundo Pasqualini (2018, p. 67), “o contingente da maioria dos leitores brasileiros é composto por indivíduos entre 15 e 64 anos, das classes B, C e D/E, em níveis elementar e intermediário de letramento”; dentro desse contexto, o PorPopular faz parte de um conjunto de textos em linguagem simples que são “potencialmente adequados para esse perfil de leitores” (PASQUALINI, 2018, p. 67). Como veremos mais adiante, a faixa Elementar do Inaf compreende 34% e a Intermediário abrange 25% da população brasileira, conforme o relatório mais recente de 2018. Pasqualini conduziu análises lexicais ancoradas na LC e pesquisas de percepção de complexidade textual com estudantes de um curso popular, confirmando sua hipótese de que os textos considerados como simples apresentam características linguísticas distintivas e adequadas a leitores com letramento na faixa média de letramento no Brasil.

O número total de textos do PorPopular é significativamente maior em comparação aos 148 textos obtidos do *corpus* PorSimples. Portanto, optamos por utilizar os mesmos 80 textos previamente empregados na pesquisa de Finatto *et al.* (2011), sendo 40 textos do ano de 2008 e outros 40 textos do ano de 2010, todos referentes a notícias com *lead* do *Diário Gaúcho*. A fim de alcançar o total de 148 textos, selecionamos aleatoriamente mais 68 textos de dois meses do ano de 2014 do jornal *Massa!*. A escolha do ano de 2014 foi feita de maneira a obter textos mais atuais. Similarmente ao *corpus* PorSimples, não dispomos de descrições detalhadas dos gêneros presentes no *corpus* PorPopular. Consideramos, portanto, que o *corpus* é formado por gêneros que circulam na esfera jornalística. A seguir, um exemplo do conjunto de textos do que

neste trabalho denominamos como “*corpus* PorPopular”, embora composto por apenas uma parte do *corpus* original do Projeto PorPopular¹¹⁰.

Quadro 11 – Exemplo de transcrição do corpus PorPopular

NOVIDADE NO RS

Justiça inaugura audiências por vídeo

Hoje, em uma sala do Presídio Central, dois presos serão ouvidos pela juíza, que estará na Vara de Execuções Criminais, no Bairro Praia de Belas.

CARLOS WAGNER

carlos.wagner@zerohora.com.br

O dia de hoje deverá entrar para história da Justiça do Rio Grande do Sul. Às 14h, estão previstas para ocorrer as primeiras audiências à distância. Dois presos vão entrar em uma sala no Presídio Central e prestar depoimento para a juíza Adriana da Silva Ribeiro, da Vara de Execuções Criminais de Porto Alegre, que estará a várias quadras dali, em uma sala no Foro Central.

A magistrada e os réus se comunicarão, em tempo real, por meio de câmeras de vídeo acopladas a um computador. No início da semana, Adriana esteve no Presídio Central e conversou com os presos a respeito da participação deles na videoconferência.

Eles são dois jovens que cumpriam pena no regime semiaberto e fugiram. Foram presos em flagrante praticando um crime. Na audiência, a juíza irá ouvi-los para decidir se retornam ao semiaberto ou ao regime fechado.

Sistema para presos perigosos

A audiência por vídeo foi regulamentada por lei em janeiro de 2009. Será usada para ouvir presos perigosos, cujo deslocamento até o tribunal representa perigo. Ou no depoimento de testemunhas, que estejam presas, em casos considerados de alto risco. Na audiência de hoje, os dois presos não representam alto risco, mas a Justiça decidiu fazer um teste, com valor jurídico.

O projeto gaúcho custou R\$ 500 mil e é inspirado em experiências bem-sucedidas em São Paulo e no Distrito Federal, onde a videoconferência existe desde 2003.

Uma economia de até R\$ 16 milhões

No Estado, ocorrem 300 audiências de presos diariamente. Em um ano, os veículos da Superintendência dos Serviços Penitenciários (Susepe) percorrem 5 milhões de quilômetros transportando detentos.

A mudança significará uma redução de R\$ 16 milhões nos gastos anuais de transporte de presos (combustível, diárias dos agentes e desgaste dos veículos).

Fonte: Elaborado pela autora.

Os textos selecionados do Projeto PorPopular foram tratados para que cada arquivo contivesse somente um texto; a separação foi novamente conduzida pela coorientadora por meio do código Python. Decidimos remover os cabeçalhos que continham informações sobre a edição do jornal e a data por compreendermos não fazerem parte do texto de leitura. Entretanto, mantivemos o título, o lead, o nome e o endereço de e-mail do autor, além do próprio texto. Assim como nos outros *corpora*, aplicamos a padronização para o novo acordo ortográfico da língua portuguesa e corrigimos erros de digitação, provavelmente, causados por limitações de reconhecimento automático de palavras, como espaços em branco extras, falta de espaços em branco, substituição de letras, quebras de linha desnecessárias, e a omissão das letras “fi” e “t” no meio das palavras.

¹¹⁰ <https://www.ufrgs.br/textecc/porlexbras/porpopular/index.php>

Dessa forma, os dados analisados neste estudo estão organizados em três *corpora*, e a composição inicial de cada *corpus* é apresentada na tabela abaixo. O *corpus* Celpe está constituído de 45 tarefas 3 e 45 tarefas 4, sendo que cada arquivo é considerado um “texto” (n), mesmo nas situações em que as tarefas fazem uso de mais de um texto de insumo.

Tabela 1 – Composição inicial dos corpora

	<i>Corpus</i>	n	Total de <i>tokens</i>	Média de <i>tokens</i> (sd)	Total de <i>types</i>	Média de <i>types</i> (sd)
<i>Corpus</i> de estudo	Celpe	90	52.668	585 (268)	27.220	302 (85)
<i>Corpus</i> de referência (<i>proxy</i> linguagem jornalística complexa)	PorSimples	148	61.525	415 (144)	33.183	224 (65)
<i>Corpus</i> de referência (<i>proxy</i> linguagem jornalística simples)	PorPopular	148	68.334	461 (137)	35.644	240 (59)

Fonte: Elaborado pela autora.

6.2 INDICADOR DE ALFABETISMO FUNCIONAL

Nesta seção apresentamos uma descrição do Indicador de Alfabetismo Funcional (Inaf)¹¹¹, um teste de aferição de letramento da sociedade brasileira que utiliza textos escritos da esfera jornalística em suas testagens de leitura. Esse indicador é uma referência para a classificação do nível de complexidade dos nossos *corpora* de referência, PorSimples (*proxy* de complexidade alta) e PorPopular (*proxy* de complexidade baixa). Isso implica que os criadores desses *corpora* partiram de hipóteses teóricas sobre a complexidade dos textos jornalísticos selecionados, com base no nível de letramento do Inaf e do público-alvo ao qual os jornais são direcionados. Com base nessa hipótese, os criadores conduziram testes que validaram suas hipóteses (SCARTON; ALUÍSIO, 2010; ALUÍSIO *et al.*, 2010; FINATTO *et al.*, 2011; PASQUALINI, 2018; LEAL *et al.*, 2021). Para a nossa pesquisa, adotamos a prática de considerar as faixas de letramento do Inaf como uma referência, uma vez que os nossos *corpora proxies* são baseados nelas e porque, pelo que foi possível verificar, não há atualmente outro parâmetro de complexidade textual da língua portuguesa que seja adequado aos nossos objetivos de pesquisa.

¹¹¹ <https://alfabetismofuncional.org.br/>

O Inaf é um parâmetro de aferição de letramento da sociedade brasileira, relacionado a nível de escolaridade, que utiliza textos escritos provenientes da esfera jornalística em suas testagens de leitura. Os resultados do Inaf têm sido utilizados por pesquisas brasileiras que versam sobre a complexidade textual (ALUÍSIO *et al.*, 2010; PASQUALINI, 2018; LEAL, 2021, para citar alguns), dado que a complexidade está diretamente relacionada ao leitor, obrigando os pesquisadores a questionar: “complexo para quem?” (LEAL, 2021, p. 3). Os níveis do Inaf, portanto, são aplicados a fim de que essa resposta possa ser respondida e o nível de complexidade textual possa levar em consideração públicos-alvo específicos com base em seu letramento. Nessa perspectiva, consideramos oportuno fornecer uma breve descrição sobre o Indicador e suas categorias de letramento, considerando que serão essenciais para as interpretações de nossos resultados. É imprescindível compreender os diferentes níveis de letramento da sociedade brasileira, bem como suas relações com as classes sociais e o nível de escolaridade para os quais os textos jornalísticos analisados são direcionados.

O relatório mais recente do Indicador, publicado em 2018, mostra que, quanto mais alta a escolaridade, maior a capacidade de compreender, utilizar e refletir sobre informações escritas. O relatório do Inaf apresenta dados de uma amostra estratificada, composta por 2.002 adultos com idades entre 15 e 64 anos, representando a população adulta brasileira¹¹². Esses dados foram gerados por meio de entrevistas domiciliares realizadas entre fevereiro e abril de 2018, envolvendo instrumentos de coleta como questionários e testes cognitivos. A proficiência em leitura foi avaliada por meio de perguntas orais feitas pelo avaliador após a leitura de textos escritos, sendo que as respostas eram fornecidas oralmente pelos participantes. As informações coletadas foram base para fazer inferências sobre as competências de letramento e numeramento da população-alvo, que, conjuntamente, compreendem o conceito de Alfabetismo. Ressaltamos, portanto, a diferença entre a metodologia de avaliação de leitura do Inaf (responder oralmente perguntas orais sobre o texto de insumo) e do exame Celpe-Bras (usar e recontextualizar informações do texto de insumo relevantes para a produção de um determinado texto escrito). Embora cientes de que essa diferença traga limitações à comparação entre os níveis de proficiência em leitura de ambos os exames, é importante frisar que nosso foco neste estudo é refletir sobre a complexidade textual como um dos fatores preditivos para planejar, na inter-relação com outros componentes da tarefa (propósitos de leitura e de escrita),

¹¹² A amostra da edição de 2018 foi coletada de modo proporcional conforme a população brasileira de interesse: “43% vivem na região Sudeste, 27% no Nordeste, 15% no Sul, 8% no Norte e 8% no Centro-Oeste. As mulheres representam 52% e os homens, 48%. Quanto à faixa etária, 24% têm entre 15 e 24 anos, 23% de 25 a 34, 31% entre 35 e 49 anos e 23% entre 50 até 64 anos” (AÇÃO EDUCATIVA; INSTITUTO PAULO MONTENEGRO, 2018, p. 9).

uma gradação da complexidade das tarefas, com implicações importantes para a projeção da dificuldade textual, levando-se em conta o público-alvo do exame, de nível intermediário. Nesse sentido, cabe utilizar a escala de proficiência do Inaf (única existente até o momento para a leitura em língua portuguesa) para que, a partir dos resultados aqui obtidos, se possa fazer novas hipóteses e testagens com leitores em diferentes níveis de proficiência e com diferentes instrumentos de avaliação.

Abaixo, apresentamos a escala de proficiência do Inaf, com suas cinco faixas de alfabetismo.

Quadro 12 – Escala de proficiência do Inaf

Grupos	Escala especial para estudo Alfabetismo e mundo do trabalho
Analfabeto ($0 < x \leq 50$)	<ul style="list-style-type: none"> Corresponde à condição dos que não conseguem realizar tarefas simples que envolvem a leitura de palavras e frases ainda que uma parcela consiga ler números familiares (de telefone, preços etc.).
Rudimentar ($50 < x \leq 95$)	<ul style="list-style-type: none"> Localiza uma ou mais informações explícitas, expressas de forma literal, em textos muito simples (calendários, tabelas simples, cartazes informativos) compostos de sentenças ou palavras que exploram situações familiares do cotidiano doméstico. Compara, lê e escreve números familiares (horários, preços, cédulas/moedas, telefone) identificando o maior/menor valor. Resolve problemas simples do cotidiano envolvendo operações matemáticas elementares (com ou sem uso da calculadora) ou estabelecendo relações entre grandezas e unidades de medida. Reconhece sinais de pontuação (vírgula, exclamação, interrogação etc.) pelo nome ou função.
Elementar ($95 < x \leq 119$)	<ul style="list-style-type: none"> Seleciona uma ou mais unidades de informação, observando certas condições, em textos diversos de extensão média realizando pequenas inferências. Resolve problemas envolvendo operações básicas com números da ordem do milhar, que exigem certo grau de planejamento e controle (total de uma compra, troco, valor de prestações sem juros). Compara ou relaciona informações numéricas ou textuais expressas em gráficos ou tabelas simples, envolvendo situações de contexto cotidiano doméstico ou social. Reconhece significado de representação gráfica de direção e/ou sentido de uma grandeza (valores negativos, valores anteriores ou abaixo daquele tomado como referência).
Intermediário ($119 < x \leq 137$)	<ul style="list-style-type: none"> Localiza informação expressa de forma literal em textos diversos (jornalístico e/ou científico) realizando pequenas inferências. Resolve problemas envolvendo operações matemáticas mais complexas (cálculo de porcentagens e proporções) da ordem dos milhões, que exigem critérios de seleção de informações, elaboração e controle em situações diversas (valor total de compras, cálculos de juros simples, medidas de área e escalas); Interpreta e elabora síntese de textos diversos (narrativos, jornalísticos, científicos), relacionando regras com casos particulares com o reconhecimento de evidências e argumentos e confrontando a moral da história com sua própria opinião ou senso comum. Reconhece o efeito de sentido ou estético de escolhas lexicais ou sintáticas, de figuras de linguagem ou sinais de pontuação.
Proficiente (>137)	<ul style="list-style-type: none"> Elabora textos de maior complexidade (mensagem, descrição, exposição ou argumentação) com base em elementos de um contexto dado e opina sobre o posicionamento ou estilo do autor do texto. Interpreta tabelas e gráficos envolvendo mais de duas variáveis, compreendendo elementos que caracterizam certos modos de representação de informação quantitativa (escolha do intervalo, escala, sistema de medidas ou padrões de comparação) reconhecendo efeitos de sentido (ênfases, distorções, tendências, projeções). Resolve situações-problema relativos a tarefas de contextos diversos, que envolvem diversas etapas de planejamento, controle e elaboração, que exigem retomada de resultados parciais e o uso de inferências.

Fonte: Ação Educativa; Instituto Paulo Montenegro, 2018, p. 21.

O descritor da primeira faixa, cuja pontuação se estende até 50, corresponde ao nível de proficiência relacionado ao sujeito considerado analfabeto. Isso significa que, de maneira geral, essa pessoa não é capaz de compreender textos escritos simples para a realização de tarefas. O

grupo Rudimentar corresponde à faixa entre 51 e 95 pontos na avaliação e inclui os sujeitos capazes de ler gêneros pertencentes ao contexto doméstico cotidiano, os quais fazem uso de linguagem explícita e literal, como por exemplo, calendários e informativos. Adicionalmente, esses indivíduos demonstram compreender a função da pontuação na língua portuguesa escrita. Tanto os sujeitos da faixa de pontuação até 50 quanto os da faixa de 51 a 95 são categorizados como parte do grupo de Analfabetos Funcionais. Esse grupo é composto por pessoas que enfrentam considerável dificuldade em “fazer uso da leitura e da escrita e das operações matemáticas em situações da vida cotidiana, como reconhecer informações em um cartaz ou folheto ou ainda fazer operações aritméticas simples com valores de grandeza superior às centenas” (AÇÃO EDUCATIVA; INSTITUTO PAULO MONTENEGRO, 2018, p. 8).

As faixas seguintes são consideradas como parte do grupo Funcionalmente Alfabetizado e englobam a população que demonstra maior domínio das habilidades de letramento, relacionadas ao uso da linguagem verbal, e de numeramento, que envolvem noções e operações matemáticas. A primeira faixa nesse grupo, denominada Elementar compreendida entre a pontuação 96 e 119, abrange a população que demonstra capacidade de ler textos de extensão média, além dos gêneros do contexto doméstico cotidiano. Esses indivíduos são capazes de realizar pequenas inferências e selecionar uma ou mais unidades de informação dos textos escritos, e de ler gráficos e tabelas simples.

Na faixa Intermediário do Inaf, situada entre 120 e 137 pontos, indica-se que, nesse nível de alfabetismo, a pessoa é capaz de localizar informações explícitas, realizar algumas inferências e interpretar textos da esfera jornalística. Além disso, em nível intermediário, o sujeito também é capaz de expressar sua própria opinião sobre os assuntos, elaborando sínteses de textos narrativos, jornalísticos e científicos. Também consegue reconhecer efeitos de sentido de escolhas lexicais e sintáticas e de uso de figuras de linguagem e uso de pontuação. Vale observar que o intermediário é o segundo nível dentro do grupo de Funcionalmente Alfabetizados, e apenas 25% dos entrevistados adultos representantes da população brasileira foram classificados nesse nível em 2018.

A descrição do nível Proficiente começa por enfatizar a capacidade de produzir textos de maior complexidade, inclusive textos argumentativos, de acordo com um contexto específico, levando em consideração o posicionamento e o estilo do autor do texto lido. Isso implica que um indivíduo que se enquadra na faixa Proficiente do Inaf é capaz de ler um texto de insumo e responder a ele por meio da argumentação. Dentro da faixa Proficiente do Inaf, o

sujeito é capaz de resolver situações-problema em diversos contextos, demonstrando habilidades para planejar, controlar e elaborar.

Vale destacar novamente que o Inaf não avalia as respostas aos textos na modalidade escrita da língua portuguesa, como no Celpe-Bras; os participantes da pesquisa fornecem suas respostas oralmente. Portanto, o objetivo do indicador não abrange a avaliação da integração das habilidades de compreensão e produção escrita, como acontece no caso do Celpe-Bras. Contudo, reiteramos que o uso dos níveis do Inaf se justifica por ser o parâmetro utilizado pelas pesquisas anteriores que versam sobre complexidade textual da língua portuguesa e que balizaram os níveis dos *corpora proxies* utilizados nesta tese.

O relatório do Inaf não disponibiliza exemplos dos textos utilizados para a avaliação de leitura, tampouco detalha o tipo de perguntas conduzidas pelos avaliadores. No entanto, Kleiman (2003) apresenta o seguinte texto como exemplo de insumo da avaliação de leitura do indicador nacional sobre o letramento do povo brasileiro realizado pela Ação Educativa.

Quadro 13 – Texto de insumo da avaliação de leitura do Inaf

<p>O CABOCLO, O PADRE E O ESTUDANTE</p> <p>Um estudante e um padre viajavam pelo sertão, tendo como bagageiro um caboclo. Deram-lhes numa casa um pequeno queijo de cabra. Não sabendo como dividi-lo, mesmo porque chegaria um pequenino pedaço para cada um, o padre resolveu que todos dormissem e o queijo seria daquele que tivesse, durante a noite, o sonho mais bonito, pensando engabelar todos com seus recursos oratórios. Todos aceitaram e foram dormir. À noite, o caboclo acordou, foi ao queijo e comeu-o.</p> <p>Pela manhã, os três sentaram à mesa para tomar café e cada qual teve de contar o seu sonho. O frade disse ter sonhado com a escada de Jacob e descreveu-a brilhantemente. Por ela, subia triunfalmente para o céu. O estudante, então, narrou que sonhara já dentro do céu à espera do padre que subia. O caboclo sorriu e falou:</p> <p>— Eu sonhei que via seu padre subindo a escada e seu doutor lá dentro do céu, rodeado de amigos. Eu ficava na terra e gritava:</p> <p>— Seu doutor, seu padre, o queijo! Vosmincês esqueceram, o queijo.</p> <p>Então, vosmincês respondiam de longe, do céu:</p> <p>— Come o queijo, caboclo! Nós estamos no céu, não queremos queijo.</p> <p>O sonho foi tão forte que eu pensei que era verdade, levantei-me, enquanto vosmincês dormiam, e comi o queijo...</p> <p style="text-align: right;">Colhida no Ceará. GUSTAVO BARROSO. AO SOM DA VIOLA, RIO DE JANEIRO, 1921.</p>

Fonte: Kleiman, 2003, p. 19.

A autora explica que se trata de um conto sobre um padre, um estudante e um caboclo, o qual os participantes da pesquisa precisam ler para responder a duas perguntas: “(a) ‘Que acordo o padre, o estudante e o caboclo fizeram’ e (b) ‘Quem tem mais estudo sempre se dá melhor na vida.’ A história confirma essa frase? Por quê?” (KLEIMAN, 2003, p. 8). Vemos que se trata de um conto do início do século XX, o que se enquadra na descrição do texto de

insumo da escala de proficiência do Inaf, a partir do nível de letramento intermediário, que indica o uso de textos narrativos, jornalísticos e científicos. Ao passo que a primeira pergunta solicita a localização de uma informação expressa no texto de forma mais explícita, a segunda demanda compreensão de sentidos implícitos.

Ao confrontar o exemplo fornecido com os do Celpe-Bras apresentados anteriormente, evidencia-se a diferença entre as metodologias utilizadas entre os dois testes para avaliar a proficiência em leitura dos participantes. Como já vimos, ao passo que, no Inaf, os participantes são convidados a responder perguntas de leitura sobre os textos oralmente, no Celpe-Bras sabemos que as sub-habilidades de leituras indicadas nas especificações são avaliadas mediante a demonstração da compreensão no texto escrito produzido em resposta ao enunciado da tarefa.

A distribuição percentual dos 2.002 participantes do Inaf 2018 nas cinco faixas de alfabetismo é apresentada a seguir.

Tabela 2 – Distribuição da população por níveis de alfabetismo do Inaf 2018

Níveis	Porcentagem da população
Analfabeto	155 (8%)
Rudimentar	433 (22%)
Elementar	688 (34%)
Intermediário	495 (25%)
Proficiente	231 (12%)
Base	2.002

Fonte: Adaptado de Ação Educativa; Instituto Paulo Montenegro, 2018.

A distribuição revela um comportamento em que há maior concentração de participantes classificados no nível Elementar. À medida que nos afastamos dessa faixa, observa-se uma diminuição nos números, abarcando as faixas Analfabeto e Proficiente como as menos populosas. Observando a tabela, torna-se evidente que o público-alvo de nossos *corpora* de referência distancia-se significativamente em proporção: o PorSimples, composto por textos voltados às classes A e B, direcionados ao nível Proficiente, representa apenas 12% da população (ALUÍSIO *et al.*, 2010; VIGEL, 2013); o PorPopular, constituído por textos direcionados às classes sociais C, D e E, acessíveis às faixas Elementar e Intermediário, em conjunto, representam 59% da população (FINATTO *et al.*, 2011; PASQUALINI, 2018).

Como podemos observar, os criadores dos *corpora* de referência descrevem a complexidade textual do público-alvo pela classe social e nível de alfabetização do Inaf. Contudo, o Inaf não indica as classes sociais dos leitores. Com o objetivo de aprofundar nossa compreensão sobre os indivíduos classificados nos níveis de letramento relacionando-os com as classes sociais, conduzimos uma pesquisa sobre a estratificação socioeconômica brasileira, a fim de obter uma compreensão mais clara do público para o qual os textos são direcionados.

Embora o conceito de estratificação socioeconômica de A a E não esteja presente nos dados do Inaf, a coleta de dados demográficos inclui a renda como uma das variáveis e classifica os participantes em cinco níveis: Não remunerado; Até 1 salário-mínimo; Mais de 1 e até 2 salários-mínimos; Mais de 2 e até 5 salários-mínimos; Mais de 5 salários-mínimos. Para estabelecer uma correspondência entre esses níveis de renda e as classes sociais A, B, C, D e E, usadas na descrição dos *corpora*, foi necessário realizar esse pareamento. Para isso, recorremos à estratificação por renda média domiciliar mensal estabelecida pela Associação Brasileira de Empresas de Pesquisa (ABEP, 2022), embora os valores de renda utilizados pelo Inaf não coincidam perfeitamente. A ABEP é um classificador socioeconômico que, de acordo com Kamakura e Mazzon (2016), possui forte sustentação teórica e metodológica que valida as análises e categorizações com base no poder aquisitivo, levando em consideração diferenças regionais e composição familiar. Com o intuito de permitir uma exploração dos níveis de Alfabetismo nas classes sociais apontadas como público-alvo dos *corpora* de referência, sugerimos a seguinte correspondência:

Tabela 3 – Proposta de equiparação de estratos socioeconômicos entre ABEP e Inaf

ABEP (renda média)	Estrato Socioeconômico ABEP	Renda Inaf	Correspondência proposta
\$900,60	D e E	Não remunerado	D e E
\$1.965,87	C2	Até 1 salário-mínimo (até \$1.320)	D e E
\$3.276,76	C1	Mais de 1 e até 2 salários-mínimos (até \$2.640)	D e E
\$5.755,23	B2	Mais de 2 e até 5 salários-mínimos (até \$6.600)	C
\$10.361,48	B1	Mais de 5 salários-mínimos (mais de \$6.600)	A e B
\$21.826,74	A	-	-

Fonte: Adaptado de Inaf, 2018? e ABEP, 2022.

Para a equivalência proposta, levamos em consideração que o valor da tabela ABEP refere-se à renda média. Portanto, até a terceira faixa do Inaf, com salários entre R\$1.320,00 e R\$2.640,00, consideramos improvável que se enquadrem nas faixas C do ABEP como renda média. A quarta faixa do Inaf abrange salários entre R\$2.640,00 e R\$6.600,00, dessa forma, aproximamos essa faixa aos valores médios de C1 e C2 da ABEP. Por fim, a última faixa do Inaf só pode ser aproximada aos dois últimos estratos da ABEP.

Ao considerar a estratificação socioeconômica proposta, podemos cotejar os níveis das classes com público-alvo dos nossos *corpora* de referência. No caso do jornal *Zero Hora*, que integra o *corpus* PorSimples juntamente com os textos da *Folha de São Paulo*, conforme mencionado anteriormente, seus leitores são identificados como pertencentes às classes A e B, portanto, mais de 5 salários-mínimos pelo Inaf (FINATTO *et al.*, 2011; VIGEL, 2013). Por outro lado, o PorPopular é composto por textos do jornalismo popular, representado pelos jornais *Diário Gaúcho* e *Massa!*, direcionados aos leitores das classes sociais C, D e E (FINATTO *et al.*, 2011; RIBEIRO, 2013), logo, estão entre o nível não remunerado até 5 salários-mínimos pelo Inaf. Vejamos como se correlacionam essas classes socioeconômicas e os níveis de alfabetismo do Inaf.

Tabela 4 – Níveis de Alfabetismo Inaf por estrato socioeconômico

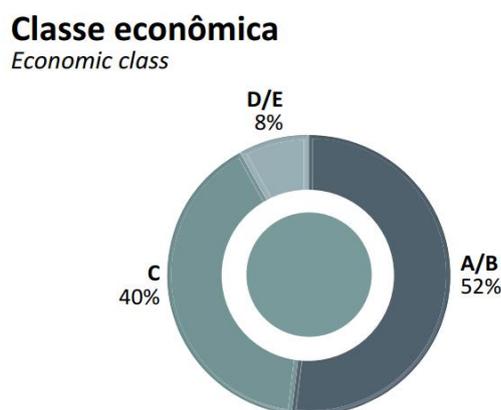
Grupo Inaf	Estrato socioeconômico	Base	Analfabeto	Rudimentar	Elementar	Intermediário	Proficiente
NR	D e E	196 (10%)	24 (12%)	49 (25%)	57 (29%)	45 (23%)	21 (11%)
Até 1 SM	D e E	501 (25%)	75 (15%)	167 (15%)	165 (33%)	71 (14%)	23 (5%)
Mais de 1 a 2 SM	D e E	589 (29%)	42 (7%)	132 (22%)	222 (38%)	140 (24%)	53 (9%)
Mais de 2 a 5 SM	C	540 (27%)	12 (2%)	74 (14%)	184 (34%)	189 (35%)	81 (15%)
Mais de 5 SM	A e B	176 (9%)	2 (1%)	11 (6%)	60 (28%)	50 (28%)	53 (30%)
		2.002	155 (8%)	433 (22%)	688 (34%)	495 (25%)	231 (12%)

Fonte: Elaborado pela autora.

Ao analisarmos os dados, percebe-se uma relação direta e proporcional: quanto mais alto o estrato socioeconômico, maior o nível de alfabetismo. Isso corrobora as afirmações de que os jornais de referência são direcionados a leitores com maior escolaridade e poder aquisitivo, enquanto os jornais populares são direcionados a leitores com baixa escolaridade e menor capacidade de consumo. Além disso, os dados revelam que apenas 9% da população do Inaf se enquadra nas classes A e B, enquanto o restante 91% nas classes C, D e E.

Levando em consideração a base de 2.002 participantes da pesquisa, os indivíduos enquadrados nas classes A e B, com nível de alfabetismo Proficiente, público-alvo do PorSimples, representam apenas 3% do total da população. Em contrapartida, as classes C, D e E, associadas aos níveis Elementar e Intermediário, correspondem a 54% da população, teoricamente, o público-alvo do PorPopular. Sem embargo, isso não se traduz no fato de que os jornais populares sejam mais consumidos do que os jornais de referência no Brasil, e sim o oposto. O relatório que apresenta informações demográficas, mercadológicas e comerciais do mercado de mídia brasileiro e sul-americano, publicado pelo Grupo de Mídia São Paulo, indica que os jornais com maior circulação nacional, tanto na versão impressa quanto na digital, são a *Folha de S. Paulo*, *O Globo*, *O Estado de São Paulo* e *Zero Hora* (GRUPO DE MÍDIA SÃO PAULO, 2021). Em relação à classe econômica dos consumidores de jornais no Brasil, o Grupo Mídia reporta os seguintes dados:

Figura 17 – Perfil dos consumidores de jornais por classe econômica



Fonte: GRUPO DE MÍDIA SÃO PAULO, 2021, p. 208.

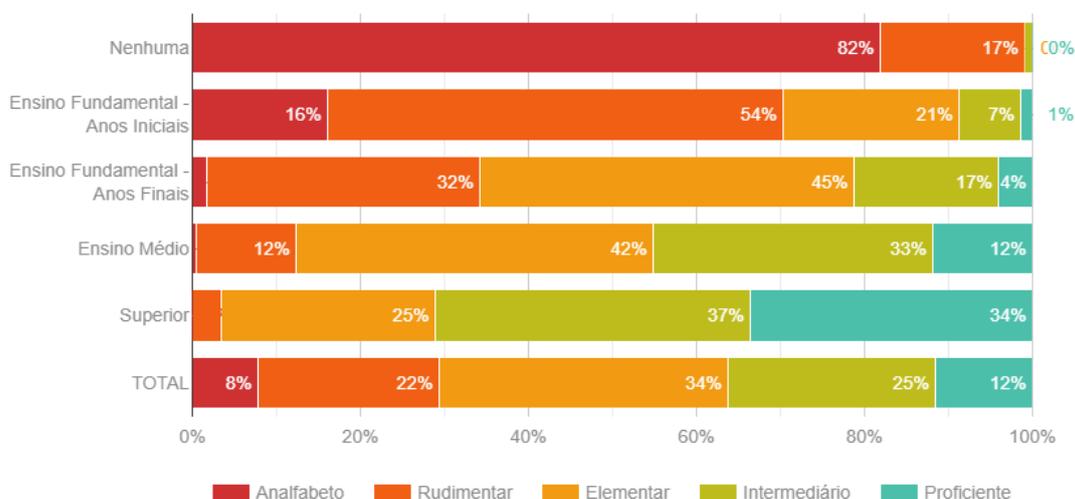
O dado revela que as classes sociais mais altas compreendem o perfil preponderante dos consumidores de jornais no país. Conforme o relatório do Grupo de Mídia, a *Folha de São Paulo* registrou uma circulação de quase 346 mil exemplares aos sábados durante o ano de 2020, enquanto a *Zero Hora* alcançou 146 mil. Por outro lado, o *Diário Gaúcho* obteve 74,8 mil e o jornal *Massa!* apenas 5,7 mil exemplares (GRUPO DE MÍDIA SÃO PAULO, 2021). De modo similar, o relatório indica que entre as revistas com maior circulação, tanto impressa quanto digital, estão a *Veja*, com uma média de 432,7 mil exemplares por edição em 2020, a *Época*, com 122,6 mil exemplares e a *Veja Saúde*, com 112,4 exemplares. Embora a revista *IstoÉ* também tenha ampla circulação nacional, sua distribuição não é auditada e, por isso, o

relatório não apresenta dados de sua tiragem. Segundo o Grupo de Mídia, o perfil do consumidor de revistas brasileiras, de maneira semelhante ao perfil do consumidor de jornais, é constituído por 58% de leitores das classes A e B, 36% da classe C e 6% das classes D e E.

Com base nos dados apresentados, podemos inferir que, embora os jornais populares tenham como público-alvo a grande massa da população do país, essa parcela da sociedade representa a minoria dos consumidores da imprensa brasileira. Adicionalmente, informações atuais acerca do consumo de mídia de informação revelam que os leitores brasileiros têm gradualmente substituído a leitura de mídias impressas pelo acesso a conteúdo digital. Conforme o relatório da *Reuters Institute Digital News*, apenas 12% da população brasileira ainda lê jornais e revistas em formato impresso, enquanto 64% buscam informações digitalmente, por meio de redes sociais como YouTube, WhatsApp, Telegram, Facebook, Instagram, Twitter e TikTok (CARRO, 2022). As informações do relatório corroboram o atual período de aprendizagem personalizada, descrito por Alexander e Fox (2018), no capítulo 3, no qual a experiência de leitura apresenta características e experiências mais individualizadas para cada leitor, tendo em vista as diversas formas de acesso e comunicação de informações proporcionadas pela tecnologia.

Além disso, é relevante destacar outra relação presente nos dados do Inaf, que diz respeito à correlação entre os níveis de alfabetismo e o grau de escolaridade dos participantes, conforme ilustrado na figura a seguir. Como pode ser observado, o nível de escolaridade e o nível de alfabetismo também são diretamente proporcionais, pois quanto mais elevado o nível de escolaridade, mais elevado o nível de alfabetismo.

Figura 18 – Alfabetismo por níveis de escolaridade do Inaf



Fonte: Inaf, [2018?].

Com base no exposto, pode-se depreender, portanto, que o grau de letramento em nível Proficiente, referente ao PorSimples, apresenta 34% no nível Superior e 12% no Ensino Médio de escolaridade. Por outro lado, referentes ao PorPopular, os níveis de letramento Elementar e Intermediário se destacam desde os Anos Finais do Ensino Fundamental até o nível Superior de escolaridade. Os dados são indicativos de que o PorSimples é composto por textos que fazem uso de linguagem de complexidade alta e direcionados para brasileiros com nível Superior. Por outro lado, o PorPopular é constituído por textos que fazem uso de uma linguagem de complexidade mais acessível aos brasileiros com escolaridade entre Fundamental completo até Superior completo.

6.3 NILC-METRIX

Após discorrer sobre as perspectivas teórico-metodológicas, este capítulo apresentará as métricas de complexidade textual do NILC-Metrix. Por ser a publicação inaugural da ferramenta NILC-Metrix, Leal *et al.* (2021) descrevem as 14 categorias que agrupam as 200 métricas de complexidade textual. Baseamos nossa apresentação dos índices nas descrições feitas pelos autores, nos trabalhos citados por eles e na documentação disponível no website da ferramenta. A documentação detalha as 200 métricas, oferecendo uma tabela com o número de identificação da métrica, o grupo ao qual ela pertence, o nome e uma breve descrição. Após essa tabela, a documentação também fornece a interpretação da métrica em relação à complexidade textual, definições de termos utilizados na descrição e a exposição das limitações da métrica e da testagem realizada pelos autores. Cada métrica é acompanhada por um exemplo de sentença utilizado nos testes, contagem, resultado esperado e resultado obtido.

Cabe esclarecer que a versão da documentação que analisamos foi capturada e armazenada, em 27 de julho de 2023, na página <https://web.archive.org/web/20230726121840/http://fw.nilc.icmc.usp.br:23380/metrixdoc>, caso outros pesquisadores necessitem ou queiram verificar as citações e referências apresentadas nesta tese. Vale ressaltar que, de acordo com as descrições da documentação, alguns índices não apresentam uma relação clara ou direta com a complexidade textual. Ao longo desta seção, faremos referência aos membros do grupo de pesquisa NILC, responsáveis pela criação das métricas, bem como aos autores da publicação inaugural (LEAL *et al.*, 2021), como autores do NILC-Metrix e, portanto, de sua documentação. Salientamos, mais uma vez,

que as definições de complexidade textual associadas às métricas foram estabelecidas pelos seus criadores e não são de nossa autoria.

O NILC-Metrix é uma ferramenta de extração de métricas de complexidade textual da língua portuguesa que realiza uma análise automatizada de textos e tem por objetivo principal “fornecer indicadores para avaliar a coesão, a coerência e a complexidade textual”¹¹³ (LEAL *et al.*, 2021, p. 5). Os dados extraídos são quantitativos e objetivos, embora algumas medidas sejam provenientes de pesquisas experimentais e possuam propriedades subjetivas, como as métricas psicolinguísticas. Os índices resultam de uma análise multinível do texto, abrangendo aspectos léxicos, sintáticos e discursivos. Essa ferramenta é o resultado de pesquisas conduzidas entre os anos de 2008 e 2021 por membros do grupo de estudos NILC. A partir de 2021, o grupo disponibiliza publicamente o acesso ao extrator, contendo 200 métricas linguísticas e psicolinguísticas, por meio de uma interface web acompanhada de documentação¹¹⁴. A partir do final de 2022, o grupo de pesquisa NILC passou a conceder acesso ao código fonte aberto da ferramenta, sob a licença AGPLv3¹¹⁵.

Conforme indicado por Leal *et al.* (2021) e pela documentação, a implementação das métricas do NILC-Metrix contou com a utilização de recursos linguísticos provenientes de banco de dados e repositórios tais como o *Banco do Português*¹¹⁶, o *Corpus Brasileiro*¹¹⁷, o *Brazilian Portuguese Web as Corpus (BrWac)*¹¹⁸, o *Dicionário Brasileiro LIWC*¹¹⁹, o *Thesaurus Eletrônico do Português*¹²⁰, o *Repositório de propriedades psicolinguísticas do Português Brasileiro*¹²¹ e o *Wordnet.Br*¹²². Para as análises, o NILC-Metrix faz uso da ferramenta de PLN *nlpnet*¹²³, a qual executa tarefas como segmentação do texto em palavras individuais (*tokenizer*), a etiquetagem morfossintática (*Part-of-Speech tagger*), a etiquetagem de papéis semânticos (*semantic role labeling*) e a análise de dependência sintática (*dependency parsing*). Para a análise sintática, a ferramenta também utiliza internamente os analisadores *LX-*

¹¹³ No original, lê-se: the main objective is to provide proxies to assess cohesion, coherence and textual complexity.

¹¹⁴ <http://fw.nilc.icmc.usp.br:23380/nilcmetrix>

¹¹⁵ <https://github.com/nilc-nlp/nilcmetrix>

¹¹⁶ <http://www2.lael.pucsp.br/corpora/bp/index.htm> [link corrompido, acesso em 31 jul. 2023]

¹¹⁷ <http://corpusbrasileiro.pucsp.br/> [link corrompido, acesso em 31 jul. 2023]

<https://www.linguateca.pt/acesso/corpus.php?corpus=CBRAS>

¹¹⁸ <https://www.inf.ufrgs.br/pln/wiki/index.php?title=BrWaC>

¹¹⁹ <http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>

¹²⁰ <http://www.nilc.icmc.usp.br/tep2/>

¹²¹ http://143.107.183.175:21380/portlex/index.php/en/?option=com_content&view=article&layout=edit&id=23

¹²² [http://143.107.183.175:21380/wordnetbr/#:~:text=A%20WordNet.Br%20%C3%A9%20a,Princeton%20\(WordNet.Pr\).](http://143.107.183.175:21380/wordnetbr/#:~:text=A%20WordNet.Br%20%C3%A9%20a,Princeton%20(WordNet.Pr).)

¹²³ <https://pypi.org/project/nlpnet/>

*Parser*¹²⁴, *Malt*¹²⁵ e *Palavras*¹²⁶ (*parsers*). As métricas resultantes são categorizadas nos 14 grupos listados abaixo e que serão explicados nas subseções a seguir.

1. Medidas Descritivas
2. Simplicidade Textual
3. Coesão Referencial
4. Coesão Semântica
5. Medidas Psicolinguísticas
6. Diversidade Lexical
7. Conectivos
8. Léxico Temporal
9. Complexidade Sintática
10. Densidade de Padrões Sintáticos
11. Informações Morfossintáticas de Palavras
12. Informações Semânticas de Palavras
13. Frequência de Palavras
14. Índices de Leiturabilidade

As explicações que sintetizamos a seguir tiveram como fonte a publicação de Leal *et al.* (2021) e a documentação da ferramenta, como mencionado anteriormente. Apresentamos uma definição das categorias explicitando as métricas que as compõem e trazendo suas limitações de acordo com os autores. Cabe dizer que, além das limitações, a documentação descreve cada índice, interpretando-o em relação à complexidade textual e fornecendo resultados de todos os testes que evidenciam a acurácia de cada um deles. Conforme apontado pela revisão bibliográfica, a atenção dada ao detalhamento de cada métrica mostra que os autores estão cientes de que a precisão das medidas depende de múltiplos fatores, como o desempenho das ferramentas de PLN utilizadas pelo algoritmo e a representatividade dos recursos linguísticos presentes nas bases de dados consultadas.

O resultado da extração automática de medidas que analisam os traços linguísticos disponibilizado pela ferramenta deve ser interpretado como um indicador, e não como uma evidência de complexidade textual, uma vez que, como amplamente discutido, os dados

¹²⁴ <http://lxcenter.di.fc.ul.pt/tools/pt/conteudo/LXParser.html>

¹²⁵ <https://www.maltparser.org/download.html>

¹²⁶ <https://visl.sdu.dk/visl/pt/credits.html>

numéricos não têm validade até que sejam interpretados qualitativamente pelo pesquisador. Por ser, até o momento, o instrumento de análise automática de complexidade textual da língua portuguesa mais acessível¹²⁷, o NILC-Metrix vem sendo utilizado por pesquisadores brasileiros e foi selecionado para operacionalizar a complexidade textual nesta pesquisa. Precisamente por se tratar de uma ferramenta relativamente nova, identificamos a ausência de investigações que se debrucem sobre as 200 métricas com o propósito de avaliar a validade de seus resultados em contextos específicos. Nesse sentido, nossos resultados poderão contribuir com dados referentes às métricas utilizadas neste estudo.

Nosso trabalho se configura como apenas uma das possibilidades de testagem de validade da ferramenta, visto que o mesmo rigor de análise deve ser aplicado sempre que a ferramenta for utilizada. Cabe ao usuário estar atento às informações fornecidas pelos autores para a tomada de decisão consciente de incluir ou excluir determinadas métricas conforme os gêneros discursivos dos textos e os objetivos da análise.

6.3.1 Medidas Descritivas

Na categoria “Medidas Descritivas”, estão reunidas 10 métricas com características estatísticas básicas que descrevem o texto analisado. As métricas descritivas fazem uso de ferramentas computacionais, como o *tokenizer* e o segmentador sintático, para delimitar as palavras, as sentenças e os parágrafos, e retornam as frequências, médias, valores mínimos, máximos e desvio padrão. Segundo os autores, a interpretação das medidas está relacionada com a complexidade textual, conforme o esforço requerido durante o processamento da leitura do texto, ou seja, quanto maior a extensão das palavras, sentenças e textos, maior a complexidade textual. O desvio padrão das medidas, em conjunto com a quantidade máxima e mínima de palavras por sentença, indica o nível de uniformidade do texto; quanto maior a uniformidade, menor a complexidade, de acordo com os autores. Dentro dessa categoria, a ferramenta também calcula a proporção de subtítulos em relação ao total de sentenças no texto, assumindo que o uso de subtítulos contribui para a clareza do texto; entretanto, é necessário que o *header* `<subtittle> ... </subtittle>` esteja presente para que a ferramenta contabilize essa frequência.

¹²⁷ A ferramenta Coh-Metrix-Port 3.0 ainda está disponível, apesar de seus índices serem contemplados pelo NILC-Metrix.

6.3.2 Simplicidade Textual

As 8 métricas de “Simplicidade Textual” indicam o quão simples o texto é para a leitura, com base em fatores como o uso de pronomes pessoais em primeira pessoa, que denota a proximidade do autor com o leitor; o uso de conjunções de fácil ou difícil entendimento; o uso de palavras de conteúdo simples; e a extensão das sentenças (até 11 palavras indicam sentenças curtas; entre 12 e 13, médias; entre 14 e 15, longas; e acima de 15, muito longas). É válido lembrar que, nas “Medidas Descritivas”, os autores partem do pressuposto de que quanto mais extenso o texto, mais complexo ele é. A extensão das sentenças é interpretada de maneira similar, mesmo que a diferença entre as categorias seja de apenas 1 a 2 palavras.

Os textos que estabelecem um diálogo com o leitor são considerados menos complexos. Para essa métrica, são considerados os pronomes pessoais “eu”, “tu”, “você”, “vocês” e “a gente”. No entanto, a métrica não contabiliza as instâncias em que os pronomes são suprimidos, mesmo quando a flexão verbal em português denota a pessoa, como por exemplo em “virei porta-voz da causa em que acredito”, na qual o pronome “eu” é omitido duas vezes.

Em relação às conjunções, os itens incluídos e a categorização em fácil ou difícil foram definidos por um linguista da Guten Educação, cujas informações adicionais não estão disponíveis na documentação. A medida informa a proporção de conjunções em relação ao total de palavras do texto, em vez do total de conjunções. Dessa forma, entendemos que essa métrica é sensível ao tamanho do texto, o que pode gerar variações que interferem na interpretação dos resultados. Adicionalmente, devido à ausência de informações sobre a autoria e o processo de pesquisa que resultaram na categorização, não foi possível averiguar a confiabilidade dos critérios empregados para a extração dessa métrica.

Para a métrica de palavras de conteúdo (substantivos, verbos, adjetivos e advérbios), o código se baseia em dois bancos de dados para determinar a simplicidade das palavras. Os autores mencionam a utilização de um conjunto de 909 palavras concretas, resultantes de uma pesquisa conduzida por Janczura *et al.* (2007). Conferimos a referência desse estudo para compreender a natureza da métrica. O estudo de Janczura *et al.* (2007) visou avaliar o nível de abstração e concretude das palavras conforme a percepção dos participantes da pesquisa. O estudo incluiu 909 palavras da língua portuguesa selecionadas aleatoriamente para a amostra e envolveu 719 estudantes universitários de Brasília, escolhidos por conveniência. Foi utilizado um instrumento que apresentava as palavras e uma Escala Likert de 7 pontos, com opções entre “altamente abstrata” e “altamente concreta”. O estudo identificou 479 palavras abstratas e 430

palavras concretas, que não são disponibilizadas na publicação, embora conste uma tabela de concretude média e desvio padrão das 909 palavras em apêndice. Contudo, mesmo considerando as 430 palavras concretas resultantes do estudo, do ponto de vista da LC, o *corpus* não é tido como representativo da língua portuguesa devido à metodologia utilizada. Além disso, a ferramenta não deveria considerar para o cálculo da métrica o total de 909 palavras, pois o resultado da pesquisa indicou apenas 430 palavras concretas.

Quanto ao segundo banco de dados utilizado para medir a simplicidade de palavras de conteúdo, Leal *et al.* (2021) citam o uso de um dicionário de autoria da Lexicógrafa Maria Tereza Biderman, referindo-se a ele como *Dicionário de Palavras Simples* ou dicionário infantil (*children's dictionary*), e o listam na referência bibliográfica como *Dicionário Didático de Português* (BIDERMAN, 1998). O *Dicionário Didático de Português*, publicado em 1998 pela editora Ática, compreende cerca de 25.000 verbetes e tem como público-alvo estudantes (MURAKAWA, 2009). Dada essa discrepância nas informações, procuramos pelo *Dicionário de Palavras Simples* no currículo lattes de Biderman¹²⁸ e em redes de livrarias online, mas sem êxito. Portanto, devido à impossibilidade de identificar o dicionário usado e seu conteúdo, concluímos que não foi viável averiguar a confiabilidade dos critérios aplicados na extração da métrica de uso de palavras de conteúdo simples.

6.3.3 Coesão Referencial

Na categoria “Coesão Referencial”, a ferramenta apresenta 9 métricas que analisam elementos que contribuem para a formação de cadeias de referência e suas posições no interior do parágrafo e em relação ao texto como um todo. As cadeias referenciais estão associadas à complexidade textual, pois seu uso auxilia o leitor a realizar conexões entre partes do texto durante a leitura, resultando em uma leitura menos complexa. Os índices de proporção de candidatos a referentes nas sentenças são derivados da identificação de pronomes pessoais do caso reto, pronomes anafóricos e pronomes demonstrativos anafóricos. A partir dessa identificação, busca-se os possíveis referentes nas sentenças precedentes. Quanto maior o número de possíveis referentes, maior a complexidade do texto. Para a extração dessa métrica, é utilizado o banco de dados *Dicionário de Palavras Simples Flexionadas para o Português Brasileiro*¹²⁹, resultado de projeto do NILC (MUNIZ, 2004), que abrange cerca de 880 mil

¹²⁸ <http://lattes.cnpq.br/8915508885677915>

¹²⁹ <http://www.nilc.icmc.usp.br/nilc/projects/unitex-pb/web/dicionarios.html>

palavras. A ferramenta só considera como possíveis referentes as entradas contidas nesse banco de dados. As métricas associadas à frequência de referentes que se repetem nas sentenças adjacentes dependem do desempenho do sentenciador, que identifica os pares de sentenças; do etiquetador, que identifica os referentes ou as palavras de conteúdo; e, para detectar as repetições, do processo de stemização (*stemming*), que reduz a palavra à sua raiz, como em “ferro” e “ferradura”, que seriam reduzidos a “ferr”. Outra limitação indicada pelos autores é que a métrica somente reconhece os candidatos a referentes que estejam no mesmo gênero, número e grau do pronome demonstrativo. A documentação da ferramenta ilustra essa questão com a sentença “Comprei a obra ‘Sapiens’, que é um best-seller desde que foi lançado. Gosto desse livro.”, na qual o pronome “desse” não seria associado à “obra” devido à divergência de gênero.

6.3.4 Coesão Semântica

A categoria “Coesão Semântica” compreende 11 métricas calculadas com base na Análise Semântica Latente¹³⁰ (ASL), que considera a ocorrência de palavras semanticamente relacionadas nas sentenças, parágrafos e no texto integral. A similaridade entre as partes do texto está associada com a manutenção do tópico, o que contribui para a compreensão do texto global, de acordo com os autores. Neste grupo de métricas, são avaliadas as similaridades da sentença em relação ao parágrafo, do parágrafo em relação aos demais parágrafos e da sentença em relação ao texto global. A ASL considera palavras como similares se compartilharem o mesmo contexto de uso. Como exemplificado na documentação, “carteira” é altamente associada a “lousa”, “giz”, “professor”, “aluno” e “escola”. A ASL usa um método estatístico complexo e o modelo foi treinado no *corpus BrWac*, composto por 3,53 milhões de textos em português coletados da internet. Os autores apontam como limitação da métrica a análise de gêneros de outros domínios distintos da internet. Entretanto, vale ressaltar que, de acordo com Wagner Filho *et al.* (2018), o *BrWac* é um *corpus* constituído por páginas de domínio .br, abrangendo conteúdo de 121.075 websites. Esse *corpus* é diversificado e equilibrado, além do seu tamanho substancial permitir considerá-lo imparcial quanto a vieses de conteúdo e gênero.

Nesta mesma categoria, encontra-se a análise de entropia cruzada, que avalia em que medida a estrutura de uma sentença causa surpresa por diferir do modelo linguístico estatístico treinado pela ferramenta. Isso indica combinações pouco convencionais de palavras, o que

¹³⁰ <http://wordvec.colorado.edu/>

implica uma maior complexidade textual, segundo os autores. O resultado dessa métrica é obtido através de um cálculo estatístico, cujo modelo foi treinado em quatro *corpora*, totalizando 120.813.620 *tokens*. Esses *corpora* abrangem diferentes gêneros, incluindo enciclopédico (*Corpus Wikipedia* - 96.183.280 *tokens*), jornalístico (*Corpus PLN-BR* - 18.003.126 *tokens*), artigo, contrato, crônica, lei, notícia, reportagem, livro didático, acordo, manual, relatório, verbete de dicionário, receita, carta, editorial, memorando e ensaio (ALUISIO *et al.*, 2004) (*Corpus Lácio-Web* - 3.723.109 *tokens*) e de divulgação científica (*Corpus Revista Pesquisa FAPESP* - 2.904.105 *tokens*). Os autores também apontam que a métrica de entropia cruzada é sensível ao gênero do texto analisado, com a ressalva de que o modelo só é capaz de identificar com confiabilidade as características linguísticas presentes nos gêneros que foram utilizados para o seu treinamento.

6.3.5 Medidas Psicolinguísticas

As 24 métricas de “Medidas Psicolinguísticas” analisam as palavras considerando sua idade de aquisição (a estimativa da idade em que a palavra é aprendida), concretude (um valor alto da métrica indica palavras concretas, enquanto um valor baixo indica palavras abstratas), familiaridade (o nível de conhecimento e uso da palavra no cotidiano) e imageabilidade (a facilidade e rapidez com que a palavra evoca uma imagem mental). De acordo com os autores, a relação entre essas métricas e a complexidade textual é a seguinte: quanto mais elevada a idade de aquisição e quanto mais baixos os níveis de concretude, familiaridade e imageabilidade, mais complexo o texto é considerado. Os autores indicam que os índices derivam de outro projeto do grupo de pesquisa NILC, especificamente o trabalho de Santos *et al.* (2017), cujo repositório abriga 26.874 palavras anotadas com essas quatro propriedades psicolinguísticas.

Ao examinar a pesquisa de Santos *et al.* (2017), percebe-se que os resultados foram obtidos por meio de um modelo computacional de previsão automática das propriedades psicolinguísticas, usando regressão estatística. Para tanto, os autores utilizaram valores previamente indicados pela literatura para treinar o modelo. Recursos como o *corpus* de 909 palavras de Janczura *et al.* (2007), que trata da concretude e já foi discutido anteriormente, e outro *corpus* com 3.800 palavras do português europeu contendo informações sobre imageabilidade, concretude e familiaridade (SOARES *et al.*, 2017) foram adaptados por Santos *et al.* (2017) para o português brasileiro (*e.g.*, ação/ação). O trabalho de Soares *et al.* (2017)

contou com a participação de 2,357 estudantes universitários de todos os distritos de Portugal, que responderam a um questionário online utilizando a Escala Likert de 7 pontos. Esse questionário visava coletar respostas relacionadas à percepção dos participantes acerca das três propriedades das palavras.

Para as previsões da idade de aquisição, Santos *et al.* (2017) trabalharam com dois *corpora* do português europeu oriundos de pesquisas experimentais realizadas com estudantes universitários. Um dos *corpora* é da Universidade de Porto (CAMEIRÃO; VICENTE, 2010) e o outro da Universidade de Lisboa (MARQUES *et al.*, 2007). Cameirão e Vicente (2010) tiveram a participação de 685 estudantes usuários de português como primeira língua de socialização para a análise de 1,749 palavras, indicando em uma escala Likert de 9 pontos a idade em que acreditavam ter aprendido cada palavra (entre 2 e mais de 13 anos de idade). De modo semelhante, Marques *et al.* (2007) investigaram 834 substantivos com a participação de 110 estudantes de graduação em psicologia, usuários de português europeu como primeira língua de socialização. Os sujeitos da pesquisa indicaram, em uma Escala Likert de 7 pontos, a idade que acreditavam ter aprendido a palavra (entre 0 e mais de 13 anos de idade).

Isso posto, reconhecemos que ferramentas automáticas de predição de complexidade têm sido utilizadas na predição de propriedades psicométricas de complexidade textual, e o desenvolvimento desses modelos se baseia em critérios e variáveis de referência resultantes de estudos qualitativos. Entretanto, sabe-se também que as medidas utilizadas no desenvolvimento do modelo de predição devem ser relevantes, consistentes e calibradas para os usos pretendidos, a fim de garantir que seus resultados sejam válidos (CUNNINGHAM, MESMER, 2014). Os autores do NILC-Matrix indicam como limitação das métricas psicolinguísticas a possibilidade de as palavras que constituem o texto a ser analisado não estarem presentes no repositório de 26.874 itens e, desse modo, a análise fica sujeita a vieses.

6.3.6 Diversidade Lexical

Na categoria “Diversidade Lexical”, estão agrupadas as métricas que calculam a proporção entre o total de palavras, desconsiderando as repetições (*types*), e o total de palavras, considerando as repetições (*tokens*), resultando no, já mencionado, *type-token ratio* (TTR). São 15 métricas de TTR que analisam diferentes categorias, incluindo todas as palavras, palavras de conteúdo, palavras funcionais (artigos, conjunções, interjeições, numerais, pronomes,

preposições e advérbios com função coordenativa e subordinativa), substantivos, verbos, adjetivos, pronomes, pronomes indefinidos, preposições e pontuação.

As medidas de TTR são utilizadas para avaliar a densidade lexical de textos, considerando-se que quanto maior a riqueza lexical e a diversidade de pontuação, maior será a complexidade textual, segundo os autores da ferramenta. A diversidade lexical está inversamente relacionada à coesão textual e à extensão do texto, ou seja, quanto menor a diversidade lexical, maior tende a ser a coesão do texto, e quanto mais extenso o texto, menor tende a ser a diversidade lexical (LEAL *et al.*, 2021). A precisão das métricas dessa categoria depende do desempenho do *parser*, do tokenizador, do sentenciador e do *tagger*. Além disso, os autores indicam que não está clara a relação de 5 dessas métricas com a complexidade textual.

6.3.7 Conectivos

O grupo de “Conectivos” engloba três classes de conectivos: aditivos, causais e lógicos. Apenas a métrica que considera a proporção total de conectivos em relação à quantidade de palavras do texto inclui na análise os conectivos temporais positivos e negativos. As métricas que analisam conectivos temporais estão categorizadas separadamente na seção “Léxico Temporal”, descrita a seguir.

De acordo com os autores, o uso de conectivos auxilia a estabelecer relações entre as partes do texto, o que tende a reduzir a complexidade textual. São 12 métricas que incluem análises de conectivos positivos (*e.g.*, bem como, assim, similarmente) e negativos (*e.g.*, mas, embora, ainda). Esta categoria também inclui análises específicas para os conectivos “e”, “ou” e “se”, bem como palavras que denotam negação (*e.g.*, não, nem, nunca). Em relação a essas categorias, os autores argumentam que a negação é mais complexa que a afirmação, e que o maior uso dos operadores lógicos “e”, “ou” e “se” resulta em maior complexidade textual. Os autores também destacam que uma limitação das métricas dessa categoria é a possibilidade de imprecisão na identificação das palavras que podem funcionar como conectivos ou não, como no caso da palavra “como”, o que depende do desempenho do *tagger*.

6.3.8 Léxico Temporal

Na categoria “Léxico Temporal”, existem 11 métricas relativas à análise de ocorrência de conectivos temporais e da diversidade de tempos e modos verbais que se relacionam com a

coesão temporal do texto. De maneira semelhante à categoria anterior, os autores indicam que o uso de conectivos contribui para a interpretação do texto e, portanto, tende a diminuir a complexidade textual. No que diz respeito aos tempos e modos verbais, é observado que quanto maior a variedade e a quantidade de tempos e modos verbais utilizados no texto, maior será a complexidade textual.

O uso do presente e do pretérito imperfeito do Indicativo é considerado marca de menor complexidade textual devido à sua alta frequência de uso. Verbos no particípio são considerados menos complexos do que as formas flexionadas, porém, a acurácia dessa métrica depende do desempenho do *parser* em distinguir a função do particípio (verbo, substantivo ou adjetivo), que é limitada à identificação do uso de artigo precedente (*e.g.*, a aposentada, o vestido). Além disso, o pretérito mais-que-perfeito do Indicativo e os tempos do modo Subjuntivo são de uso mais raro e, portanto, elevam a complexidade textual, segundo os autores. É importante mencionar que a métrica de proporção de verbos no presente do Subjuntivo apresenta a limitação de que o *parser* não diferencia quando a palavra está sendo usada no modo imperativo (*e.g.*, cante, cantem, cantemos). Da mesma forma, o *parser* não distingue a terceira pessoa do plural nos tempos pretérito perfeito e pretérito mais-que-perfeito, o que interfere no resultado da métrica (*e.g.*, cantaram). Em geral, os autores indicam que a precisão das métricas dessa categoria depende do desempenho do *parser*.

6.3.9 Complexidade Sintática

A “Complexidade Sintática” conta com 27 métricas que analisam a estrutura e a composição sintática das orações por meio de analisadores por dependência e por constituição (*dependency and constituency parsers*). Essas métricas envolvem a análise dos nódulos das árvores sintáticas e das relações entre as palavras e seus dependentes, e estão ligadas à complexidade textual, partindo do pressuposto de que há uma relação entre complexidade sintática e a demanda de memória de trabalho, segundo os autores. Os dados de árvores sintáticas são utilizados para calcular várias medidas de proporção das orações, além de dois índices de complexidade sintática derivados de métodos matemáticos resultantes de pesquisas da língua inglesa. Um desses índices é o Yngve, que é uma fórmula que parte do pressuposto de que o padrão de uma estrutura sintática é se ramificar para a direita, e desvios desse padrão tornam as sentenças mais complexas. A fórmula Frazier fundamenta-se na noção de que quanto maior a profundidade da árvore sintática, mais complexa será a sentença.

Nesta categoria também se encontram índices de contagens de palavras antes dos verbos principais, frequência de adjuntos adverbiais de tempo, a ocorrência de advérbios antes do verbo principal, apostos, orações, conjunções coordenativas, orações subordinadas reduzidas, estruturas sintáticas não canônicas como sujeito-verbo-objeto, voz passiva analítica, sujeitos pospostos, conjunções subordinativas, orações relativas e a quantidade de orações na mesma sentença. Nesses casos, os autores partem do princípio de que quanto maior a ocorrência, mais demanda de memória é necessária e, conseqüentemente, maior é a complexidade textual. Adicionalmente, há a métrica que avalia o desvio padrão do tamanho dos sintagmas nominais, indicando a uniformidade do nível de complexidade textual. Os autores correlacionam a maior extensão do sintagma à maior complexidade. Por outro lado, o índice que contabiliza sentenças sem verbo considera tais sentenças como menos complexas. Os autores observam que a precisão das métricas desta categoria está sujeita ao desempenho do tokenizador e do *parser*, os quais apresentam fragilidades na distinção do uso do participio em sentenças na voz passiva, quebras inadequadas de sentenças muito longas quando há uso de dois pontos e ponto e vírgula, e identificação apenas de apostos etiquetados.

6.3.10 Densidade de Padrões Sintáticos

A categoria “Densidade de Padrões Sintáticos” abarca 4 métricas que analisam os sintagmas nominais e são relacionadas à complexidade textual pelos autores, uma vez que consideram que quanto maior o sintagma, maior a dificuldade no processamento do texto. Essas métricas englobam a média, o valor máximo e o valor mínimo dos tamanhos dos sintagmas nominais e a proporção de verbos no gerúndio em relação a todos os verbos do texto, sendo que a forma nominal do verbo é considerada menos complexa. De acordo com os autores, a precisão dessas métricas depende do desempenho do *parser* na identificação dos verbos no gerúndio e na delimitação dos sintagmas nominais.

6.3.11 Informações Morfossintáticas de Palavras

Na categoria “Informações Morfossintáticas de Palavras”, encontram-se 42 métricas relacionadas à densidade de palavras funcionais e de conteúdo. Os autores enfatizam que, embora as informações individuais não sejam suficientes para determinar a complexidade textual, elas podem contribuir para identificar fatores que complexificam o texto. As métricas proporcionam dados sobre a proporção, a proporção máxima e mínima, e o desvio padrão de

ocorrência de adjetivos, advérbios, verbos, substantivos, preposições e pronomes em relação à quantidade total de palavras do texto, bem como em relação às palavras por sentenças.

Segundo os autores, a precisão dessas métricas depende do desempenho do *tagger*. É importante notar que, para os índices de proporção de verbos, o *parser* não distingue a função de verbos no particípio. No caso do índice de pronomes oblíquos, as funções do pronome “se” não são discriminadas pelo *parser*. A proporção de palavras funcionais em relação às palavras de conteúdo é afetada pelo método de tokenização adotado, considerando se há ou não contração, além da junção de partes de nomes próprios e multipalavras. Além disso, o cálculo da proporção de pronomes nas segundas e terceiras pessoas é influenciado pela dificuldade do *parser* em distinguir pronomes e flexões de verbos como “você”, “vocês” e “a gente” dos pronomes que se referem à terceira pessoa. Também é importante mencionar que o *parser* não identifica a primeira pessoa em casos de supressão de pronome.

6.3.12 Informações Semânticas de Palavras

Na categoria “Informações Semânticas de Palavras”, encontram-se agrupadas 11 métricas que abordam aspectos semânticos, tais como análise de sentimentos, ambiguidade, concretude, hiperônimos e entidades nomeadas. Diferentemente das “Medidas Psicolinguísticas”, essas métricas fazem uso de informações derivadas de repositórios linguísticos, dicionários e tesouros. A exceção fica por conta das entidades nomeadas, que são identificadas pelo *parser* através do uso de letras maiúsculas no início das palavras. Pressupõe-se que quanto maior a ambiguidade, abstração, e a quantidade de hiperônimos e nomes próprios, maior a complexidade textual, segundo os autores. No entanto, a relação entre a medida de polaridade e a complexidade textual não é claramente estabelecida, de acordo com a documentação das métricas de polaridade.

Um dos bancos de dados utilizados é resultado de projeto desenvolvido pelo grupo de pesquisa NILC, o Thesaurus Eletrônico para o Português do Brasil (TeP), que possui 19.888 conjuntos de unidades lexicais consideradas sinônimas em um determinado contexto e pertencentes à mesma categoria sintática (*synset*). Além disso, o TeP contém relações de antonímia e um total de 44.678 itens lexicais (MAZIERO *et al.*, 2008). Esse recurso é utilizado na extração de índices relacionados à ambiguidade, os autores consideram que quanto mais sinônimos uma palavra possui no tesouro, maior será a ambiguidade e, por consequência, a complexidade textual.

Outro repositório usado é a WordNet.Br, uma *wordnet* elaborada pelo grupo NILC, que consiste em um conjunto de *synsets* com relações léxico-semânticas de sinonímia e hiperonímia traduzidas e alinhadas com a base de dados da língua inglesa da WordNet.Pr¹³¹ (SCARTON; ALUÍSIO, 2009). A WordNet.Br conta com 5.860 verbos em 3.713 *synsets* que fornecem informações sobre os hiperônimos para a métrica “quantidade média de hiperônimos por verbo nas sentenças”. De acordo com os autores, quanto menos hiperônimos, menos complexa é a palavra e o desempenho dessa métrica depende da identificação dos hiperônimos na *wordnet*. Considerando o número de unidades lexicais, esses dois bancos de dados podem ser considerados pequenos, o que pode prejudicar o desempenho da métrica, visto que só serão considerados os léxicos existentes nos repositórios.

Outro índice desta categoria refere-se aos substantivos abstratos, considerados mais complexos do que os substantivos concretos. A métrica mede a concretude utilizando a identificação de sufixos (*e.g.*, -mento, -ção) para calcular a proporção de substantivos abstratos (*e.g.*, sentimento, aprovação) em relação ao total de palavras. Entretanto, os autores apontam para a fragilidade conceitual dessa métrica, uma vez que parte do pressuposto equivocado de que há sufixos exclusivos de palavras abstratas na língua portuguesa (*e.g.*, documento, razão).

Duas métricas desta categoria operam com o reconhecimento de letras maiúsculas pelo *parser* para contabilizar nomes próprios, pois, de acordo com a documentação, quanto maior a quantidade de entidades nomeadas, maior é a carga de memória exigida durante a leitura do texto.

Por fim, há duas métricas de características psicolinguísticas que indicam polaridade negativa (*e.g.*, lento) e polaridade positiva (*e.g.*, boa) com base em informações extraídas do Dicionário Brasileiro LIWC (*Linguistic Inquiry and Word Count*) para análise de sentimentos. Segundo Balage Filho e colaboradores (2013), o LIWC contém 127,149 verbetes, dos quais 12,878 são categorizados com sentimento positivo e 15,115 com sentimento negativo. A documentação do NILC-Matrix ressalta que não há uma relação clara entre a análise de sentimentos e a complexidade textual.

6.3.13 Frequência de Palavras

O conjunto de 10 métricas denominado “Frequência de Palavras” extrai medidas que comparam o uso de palavras no texto analisado com aquelas presentes nos *corpora Banco do*

¹³¹ <https://wordnet.princeton.edu/>

Português, Corpus Brasileiro e BrWaC. Os autores indicam que, teoricamente, quanto maior a frequência da palavra nos *corpora* de referência, menor é a sua complexidade.

Via *Banco do Português*, uma métrica calcula a média da frequência de palavras de conteúdo presentes no texto, enquanto a outra estima a média da frequência das palavras de conteúdo mais raras no nível de sentenças. Para a primeira métrica, o algoritmo identifica as palavras de conteúdo no texto em análise e contabiliza a frequência delas no *corpus* de referência; essa soma é então dividida pelo total de palavras no texto analisado. A frequência da segunda métrica é calculada a partir da menor frequência entre as palavras de conteúdo em cada sentença.

O *Banco do Português*, de acordo com a documentação, possuía 700 milhões de *tokens* na versão de 2010, sendo considerado atualmente um *corpus* pequeno. Apesar disso, os autores optaram por manter essas métricas na ferramenta por motivos históricos que não são detalhados. Em Leal *et al.* (2021), essas duas métricas são apresentadas como sendo as mais antigas e o *Banco do Português* é mencionado como *Corpus do Português*.

Segundo a documentação, através do *Corpus Brasileiro*, que abrange cerca de um bilhão de *tokens*, e do *BrWaC*, com aproximadamente 2.68 bilhões de *tokens*, são extraídas oito métricas, sendo quatro de cada *corpora*. Essas métricas reportam a média da frequência das palavras no texto, das palavras de conteúdo e das palavras de conteúdo mais raras das sentenças e do texto. As médias das palavras de conteúdo mais raras das sentenças estabelecem a relação entre a frequência de palavras mais raras com a complexidade textual, segundo os autores. A precisão das métricas depende do desempenho do *tagger* e da coletânea de textos nos *corpora*. No caso do *BrWaC*, com base em Wagner Filho *et al.* (2018), o *corpus* é composto por textos de websites brasileiros, e o *Corpus Brasileiro*, de acordo com Berber Sardinha (2010?), contém textos escritos e orais distribuídos em treze *subcorpora*: *Academia, Cinema e TV, Educação, Enciclopédia, Esporte, Informática, Jornalismo, Legislação, Literatura, Medicina, Política, Religião e Técnico*.

6.3.14 Índices de Leiturabilidade

Finalmente, a última categoria agrupa 5 índices de fórmulas de leiturabilidade superficiais. De acordo com Thomas *et al.* (2005), o Índice Brunet ($N^{V-0.165}$) e a Estatística de Honoré ($\frac{100 \log N}{1 - \frac{1}{V}}$) são medidas de densidade lexical que fazem uso de fórmulas mais sofisticadas

do que simplesmente dividir *tokens* por *types* ($\frac{V}{N}$), e essas medidas não são sensíveis ao tamanho do texto. As fórmulas Dale Chall e Gunning Fog reportam medidas sobre palavras de difícil compreensão e não familiares. A Fórmula Dale Chall busca estabelecer uma correspondência entre níveis escolares e a complexidade textual, utilizando a combinação da quantidade de palavras não familiares com a média de palavras por sentença. No contexto dessa fórmula, “palavras não familiares” são aquelas que não estão presentes no *Dicionário de Palavras Simples* que, como discutido antes, é uma fonte de dados que buscamos localizar, mas não encontramos.

Por outro lado, a fórmula Gunning Fog classifica como “palavras difíceis” aquelas que possuem mais de duas sílabas, o que não é diretamente aplicável ao português, como indicado pelos próprios autores como uma limitação da métrica. Por fim, o Índice Flesch busca estimar a quantidade de anos de escolarização necessários para compreender um nível específico de complexidade textual. Para essa estimativa, são analisadas a média de sílabas por palavra e a média de palavras por sentença. De acordo com Martins e colaboradores (1996), em inglês a média de sílabas por palavra é 1.8, enquanto em português essa média sobe para 2.2. Assim, a fórmula original do inglês foi adaptada ao português levando em consideração a extensão das palavras e os níveis de ensino brasileiro. Tanto o Índice Flesch quanto a Fórmula Dale Chall foram adaptados para o português como resultado de projetos de pesquisa conduzidos por membros do grupo NILC.

Dado o exposto, ainda que as 200 métricas como um todo busquem reportar níveis de complexidade textual, reiteramos que o pesquisador que utiliza essa ferramenta deve estar atento à singularidade de cada métrica como indicativa, e não determinante, de complexidade ao interpretar os resultados e formular asserções. No âmbito da nossa pesquisa, realizamos uma investigação da complexidade textual em textos da esfera jornalística, relacionado as variáveis numéricas a uma abordagem de análise qualitativa. Assim sendo, nossos resultados não podem ser generalizados para outros gêneros discursivos. Embora cada pesquisador necessite estar atento à adequação dos índices como indicativos de complexidade textual em seu próprio contexto de pesquisa, consideramos que o delineamento dos 14 grupos de métricas apresentado neste capítulo pode auxiliar em investigações futuras que desejem utilizar os índices de complexidade textual do NILC-Matrix.

6.4 DESCRIÇÃO DAS TAREFAS

Após a compilação do *corpus* de estudo e o processamento dos *corpora* PorSimples e PorPopular, iniciamos as descrições das tarefas por meio da exploração minuciosa dos textos e dos *headers* do *corpus* de estudo. A primeira etapa dessa exploração teve como objetivo a descrição das medidas de tendência central e a análise da distribuição de palavras por texto, bem como a detecção e exclusão de observações atípicas, a fim de evitar vieses que pudessem influenciar as análises.

Após a análise de distribuição de palavras nos três *corpora*, voltamos nossa atenção para os *headers* anotados no *corpus* Celpe, examinando a frequência de ocorrências das variantes e suas distribuições em relação aos anos de aplicação do exame, além da distribuição de palavras por variante. Adicionalmente, visando conferir se as diferenças observadas nas distribuições eram estatisticamente significativas, recorreremos a um teste não paramétrico de reamostragem com reposição para obter estimativas mais precisas. Devido ao tamanho reduzido de nossos dados, optamos pelo modelo de simulação *Bootstrapping*. Esse método utiliza uma força bruta da computação, repetindo mil vezes o mesmo teste com a reamostragem randomizada dos dados originais com repetição. Isso significa que o modelo considera o nosso conjunto original de dados como a população total, gerando aleatoriamente 1.000 amostras diferentes para cada testagem. Os valores da média e dos intervalos de confiança são calculados a partir desse conjunto de 1.000 resultados (PICORAL *et. al*, 2021).

O primeiro *header*, <número de textos: >, consiste em uma variável com apenas dois níveis: 1 e 2. Esse *header* indica o número de textos utilizados como insumo em cada tarefa. No contexto do exame Celpe-Bras, não se verificou nenhuma tarefa que fizesse uso de mais de dois textos. Abaixo, apresentamos exemplos de tarefas de ambos os tipos.

Figura 19 – Tarefas 3 e 4 da edição 2013-2 com 1 e 2 textos de insumo

2013/2 **Celpe Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 3 | TURISMO DE INCLUSÃO **Página 6**

Você é o dono de uma agência de viagens e, após ler o texto *Turismo de Inclusão*, decidiu oferecer programas de turismo de aventura destinados a pessoas com deficiência física. A fim de divulgar esses programas, elabore o texto que fará parte da propaganda de divulgação da agência a ser publicada em uma revista de turismo, descrevendo os serviços de apoio oferecidos pela sua agência aos deficientes físicos e estimulando esse público a viajar mais.

TURISMO DE INCLUSÃO

Agências investem em viagens para deficientes físicos que incluem rafting e rapel e até guias preparados para carregá-los no colo

Claudia Jordão

Ricardo Shimosakai sempre foi do tipo que adorava fazer as malas e se aventurar pelo mundo. Mas, aos 24 anos, ficou paraplégico. Preso numa cadeira de rodas, encontrou forças para superar as suas limitações físicas exatamente no prazer de viajar. Hoje, quase 20 anos depois do acidente, é o nome brasileiro à frente de um movimento internacional para inclusão de deficientes físicos no turismo. Cada vez mais, governos, organizações não governamentais e órgãos do setor se preocupam em tornar hotéis, restaurantes, meios de transportes, guias locais e cidades inteiras mais bem preparados para receber esse público. Shimosakai é o diretor da Turismo Adaptado, uma empresa que monta e vende pacotes de viagem para deficientes, que vão desde passeios convencionais até experiências de ecoturismo e esportes de aventura, além de oferecer consultoria e treinamento para profissionais da área. "Ao montar as minhas viagens, depois de ser recusado numa agência convencional por causa da minha deficiência, notei um interesse de pessoas nessa situação, e percebi que isso poderia ser uma oportunidade de trabalho", diz o empresário, que desde o acidente já visitou dez países e se aventurou em atividades como salto de paraquedas, rafting e rapel.

DESAVIAÇÕES Shimosakai em cima em Itaipava (BA) e Adriana, acompanhada de Inira, no Chapadão Diamantina (BA)

Quando trabalhava numa agência de ecoturismo, Adriana tentou guias para lidar com os cadeirantes. "Muitas vezes o deficiente precisa ser carregado por que não há como a cadeira de rodas passar", diz ela, que já foi levada no colo, de cavalinho e de liteira. "É importante que guias locais estejam treinados para transferir-lo de sua cadeira para barcos ou canoas, ou para transportá-lo num terreno cheio de obstáculos". Shimosakai e Adriana chegaram a locais como Machu Picchu e Chapadão Diamantina (BA), respectivamente, com ajuda humana, por exemplo. "Pode-se dizer que, no ecoturismo, o homem tem de se adaptar ao meio, inclusive a pessoa com deficiência", diz o empresário. A adaptação também é necessária na prática de esportes radicais. Ao saltar de paraquedas, por exemplo, Shimosakai precisou que suas pernas fossem amarradas para evitar distensões musculares. São pequenos ajustes que tornam sonhos realidade.

Viagens Possíveis

Quando a geografia e a topografia local facilitam a viagem de deficientes físicos, como no caso de suas trilhas, rapel e canoa do forte (BA), onde a vida é mais fácil.

São as grandes cidades, como Nova York e Paris, que têm arquitetura, hospedagem e transporte planejados para esse público. É destino, por isso, que não podem ser alcançados por deficientes físicos, como Machu Picchu, no Peru.

Destinos acessíveis Destinos adaptados

Revista IstoÉ, 15 jun. 2011 (adaptado).

2013/2 **Celpe Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 4 | MORAR SOZINHO **Página 8**

Um jornal de grande circulação está preparando uma edição especial sobre tendências contemporâneas de comportamento. Como colunista desse jornal, você ficou responsável por escrever um artigo de opinião sobre o fenômeno *morar sozinho*. Em seu artigo, discuta as visões apresentadas nos textos "Morar Sozinho" e "Solidão pode causar mais males à saúde que obesidade e tabagismo", exponha o seu ponto de vista e aponte as possíveis consequências do fenômeno *morar sozinho* para a vida em sociedade.

Morar Sozinho

Revista IstoÉ.com

Sozinho em um apartamento. Assim vive grande parcela da população mundial nas próximas décadas. De acordo com as análises de tendências. A largada foi dada pelos países mais ricos, em especial os localizados na península escandinava. Ali estão os três com mais moradores arlobos do mundo: Noruega, Finlândia e Dinamarca – em todos, mais de um terço das casas tem um só habitante. No Brasil, o fenômeno ainda desponta, mas com bastante vigor. Entre o censo de 2000 e o de 2010, o número total de moradias com um só habitante subiu 41%. Hoje são cerca de sete milhões de casas de sozinhos. A equação que explica o fenômeno, seja aqui, seja na Noruega, tem em sua base três fatores comuns. São eles a estabilidade econômica, a independência feminina e a revolução da comunicação", disse à IstoÉ Eric Klinenberg, pesquisador da Universidade de Nova York e autor do livro "Vivendo Sozinho" (tradução livre, Editora Penguin, 2012). Nesse somatório, explica Klinenberg, cada elemento influencia a seu modo. O dinheiro é fundamental para pagar os custos, que são maiores. A independência feminina permitiu às mulheres bancar um estilo de vida independente, tornando-se parcela significativa dessa população. E os meios de comunicação facilitam a convivência, evitando que os sozinhos se tornem solitários.

Solidão, inclusive, não é boa palavra para definir essa nova geração de quem mora só. O que se vê nesses lares nem de

longe lembra a imagem estigmatizada da excêntrica tia solteirona sem amigos e cheia de gatos. "Moro só por opção e nem bicho de estimação tenho, pois passo quase o dia todo fora", diz o engenheiro Frederico Lainer, 30 anos, que vive em um espaçoso quarto e sala na Cidade Baixa, bairro tradicional de Porto Alegre. Lainer resolveu assumir a casa por sua conta e risco após um período de vida em comum com uma namorada. Há seis anos a lista dos sozinhos de Porto Alegre, capital brasileira líder no ranking de gente que vive só. Por lá, 21,4% das residências abriga um único morador, índice bem acima da média do País, que é de 12,2%. Mas quem imagina que morar só é coisa de jovem, se engana. Os maiores de 60 anos representam 42% das casas de um único habitante no Brasil. "O fenômeno ocorre atrelado ao envelhecimento da população", afirma a pesquisadora do Instituto Brasileiro de Geografia e Estatística (IBGE) Ana Lúcia Sobhia. Mais velhos, com boa saúde e estabilidade financeira, os idosos têm optado por seguir a vida em suas próprias casas, bancando as despesas.

Enquanto cada vez mais gente opta por um estilo de vida sozinho (apenas nos Estados Unidos, 36% de sua população está morando só até 2020), teóricos começam a discutir outra face desse fenômeno: sua sustentabilidade. "A quantidade de alimento que se compra é grande, então se perde muito, sem contar os gastos fixos, como eletricidade, que não são divididos", afirma o professor Sany Dana, da Escola de Economia da Fundação Getúlio Vargas de São Paulo. Todo esse consumo concentrado

em uma só pessoa tem feito vários pesquisadores rotularem esse estilo de vida como insustentável em larga escala. O próprio Dana, porém, faz a defesa de quem mora só e diz que há necessidade de se estudar melhor o tema. "Se por um lado se gasta mais eletricidade, de outro se economiza combustível, porque a maior parte das residências dos sozinhos está nas regiões centrais, então se gasta menos com deslocamento".

Disponível em: www.istoé.com.br. Acesso em: 8 jul. 2013.

Jornal O Globo

Solidão pode causar mais males à saúde que obesidade e tabagismo

Roberta Jansen
Publicado: 26/01/11 - 09:00

RIO – A solidão, aquela sensação ruim de ser incompreendido, de não poder contar com ninguém, de estar sozinho no mundo, pode causar mais males à saúde do que a obesidade e o tabagismo, tradicionalmente ligados a problemas cardíacos e cânceres, entre diversos outros problemas. Mas, enquanto os dois últimos são fatores de risco muito bem estabelecidos do ponto de vista médico e aceitos pela sociedade, o isolamento social raramente é analisado, num contexto mais amplo, como potencial deturador de doenças.

Disponível em: www.oglobo.globo.com. Acesso em: 8 jul. 2013.

Fonte: Acervo Celpe-Bras.

A tarefa 3 (à esquerda) faz uso de apenas uma reportagem da revista IstoÉ. A tarefa 4 (à direita), por sua vez, se vale de dois textos distintos, sendo o primeiro uma reportagem da revista IstoÉ e o segundo um trecho de uma reportagem do Jornal O Globo.

O próximo *header*, <texto não verbal: >, aborda o uso de textos não verbais na composição das tarefas 3 e 4, visto que as especificações preveem a avaliação de leitura multimodal. Portanto, espera-se que, em cada edição, pelo menos um elemento não verbal seja incorporado às tarefas de leitura e escrita. Essa variável assume dois valores: *sim* e *não*. Nas tarefas ilustradas acima, ambas fizeram uso de textos multimodais, portanto, seus respectivos *headers* são <texto não verbal: sim>.

A *header* seguinte, <cor: >, também oferece apenas duas possibilidades: *sim* e *não*. Esse *header* diz respeito ao uso de impressão colorida nas tarefas 3 e 4. Apesar das capas dos cadernos de questões serem sempre coloridas, o conteúdo interno nem sempre segue o mesmo padrão, como ilustrado no exemplo da edição 2014-2.

Figura 20 – Edição 2014-2 com a capa do caderno de questões impressa colorida

Parte Escrita Caderno de Questões

Celpe-Bras

Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 3 | CONHEÇA O MUNDO E AJUDE A MUDÁ-LO **Página 6**

Tarefa 4 | LEITURA E TECNOLOGIA **Página 8**

2014 / 2

INEP

Fonte: Acervo Celpe-Bras.

Os quatro próximos *headers*, que inserimos manualmente, são relativos ao texto de insumo. A título de exemplo, os *headers* dos dois textos que constituem o texto de insumo da tarefa 4, ilustrada acima, são: *<texto1:entrevista, revista, fragmento, Época>* e *<texto2:entrevista, site, fragmento, Interrogação>*.

Nesse *header*, inicialmente, informamos o gênero discursivo do texto, abrangendo 13 elementos: anúncio de campanha, artigo de opinião, carta do leitor, crônica, editorial, entrevista, FAQ, guia de orientações, notícia, previsão do tempo, relato, reportagem. Após essa informação, abordamos a categoria do suporte em que o texto original foi publicado, contendo 5 variantes: jornal, revista, nd, site e site de governo. O termo “nd” denota “não disponível”, indicando a impossibilidade de resgatar o texto original via busca na internet, impossibilitando a confirmação do suporte. A distinção entre “site” e “site de governo” foi adotada para considerar separadamente os sites sob domínio gov.br. Os sites não governamentais ou de mídia estão categorizados sob o termo “site”. Por exemplo, a segunda entrevista da tarefa 4, ilustrada acima, é proveniente do site “Interrogação”, hospedado em domínio WordPress, e se designa como um site de conteúdo cultural. As variantes dessas duas variáveis foram estabelecidas a partir de dados extraídos do *corpus* de estudo, qualitativamente categorizadas mediante análise e discussão com as duas orientadoras e uma colega pesquisadora do grupo Avalia.

A informação subsequente diz respeito à forma como o texto original foi transposto para a tarefa do exame, com três variantes: completo, fragmento e nd. É importante destacar que o termo “fragmento” no contexto do Celpe-Bras não denota frases ou trechos isolados e

desconexos, mas sim uma adaptação para acomodar o texto original em uma página. Por exemplo, o texto original da tarefa 3, mencionada anteriormente, “Conheça o mundo e ajude a mudá-lo”, contém vinte e dois parágrafos, dos quais apenas sete foram usados na prova.

A última informação se refere ao nome da fonte do texto original, identificando 36 fontes distintas, totalizando 37 variáveis, incluindo a categoria “nd”. As fontes são: A escotilha, Almanaque de Cultura Popular, Araraquara, Arte na rede, Bons Fluidos, ComCiência, Correio Braziliense, Correio Cidadania, Correio Web, CPRM (Serviço Geológico do Brasil), Época, Estadão, Estado de Minas, Folha de São Paulo, Free the essence, Gol, Interrogação, IstoÉ, Jornal do Comércio, Jornal Em Dia, National Geographic, Nova Escola, O Globo, Pampulha, Pensamento Verde, Pequenas Empresas & Grandes Negócios, Planeta, Revista Galileu, Revista Superinteressante, Secretaria Municipal de Saúde, UFPA, Veja, Viagem e Turismo, Você S.A. e Zero Hora.

Em resumo, nessa etapa de descrição, observamos as seguintes características das tarefas.

Quadro 14 – Headers do corpus Celpe

Variável	Variantes
Número de textos	1, 2
Textos não verbal	sim, não
Cor	sim, não
Gênero de leitura	Anúncio de campanha, artigo de opinião, carta do leitor, crônica, edital, editorial, entrevista, FAQ, guia de orientações, notícia, previsão do tempo, relato, reportagem
Suporte do texto original	Jornal, revista, nd, site, site de governo
Texto original completo ou fragmento	Completo, fragmento, nd
Fonte do texto original	A escotilha, Almanaque de Cultura Popular, Araraquara, Arte na rede, Bons Fluidos, ComCiência, Correio Braziliense, Correio Cidadania, Correio Web, CPRM (Serviço Geológico do Brasil), Época, Estadão, Estado de Minas, Folha de São Paulo, Free the essence, Gol, Interrogação, IstoÉ, Jornal do Comércio, Jornal Em Dia, National Geographic, nd, Nova Escola, O Globo, Pampulha, Pensamento Verde, Pequenas Empresas & Grandes Negócios, Planeta, Revista Galileu, Revista Superinteressante, Secretaria Municipal de Saúde, UFPA, Veja, Viagem e Turismo, Você S.A., Zero Hora

Fonte: Elaborado pela autora.

Para concluir o estudo descritivo, conduzimos análises detalhadas dos *headers* anotados relacionados às categorizações propostas pelo grupo Avalia, previamente apresentadas na seção 4.1 (SCHOFFEN *et al.*, 2018). Para cada variável, conduzimos análises de frequência de ocorrência das variantes, *bootstrapping*, contagem de palavras por variantes e frequência de ocorrência de variantes por ano de aplicação do Celpe-Bras.

O grupo Avalia, no livro “Estudo descritivo das tarefas da Parte Escrita do exame Celpe-Bras”, publicado em 2018, apresentou uma proposta de sete categorias para a análise e descrição geral das tarefas aplicadas entre os anos de 1998 e 2017. Para nossa pesquisa, os dados posteriores a 2017 foram obtidos no site do grupo (<https://www.ufrgs.br/grupoavalia/>), que mantém um banco de dados atualizado com as edições das provas. As sete categorias são:

Quadro 15 – Categorias do grupo Avalia.

Categoria	Definição	Variantes
Esfera de atuação	Contexto social e discursivo no qual a interlocução projetada acontece	Artístico-cultural, Científico-educacional, Comercial, Comunitária, Jornalística, Pessoal, Político-cidadã, Profissional
Temática	Assunto principal da interlocução projetada	Alimentação, Ambiente, Ciência e tecnologia, Consumo, Educação, Esporte, Estilos de vida, Lazer, Mídia, Mundo do trabalho, Patrimônio cultural, Políticas e cidadania, Saúde, Transporte, Turismo
Propósito	Objetivo principal da interlocução projetada	Aconselhar, Apresentar, Divulgar, Incentivar, Indicar, Orientar, Posicionar-se, Relatar experiência, Solicitar, Sugerir
Relação de interlocução	Interlocutores envolvidos na interlocução projetada. De quem para quem?	Amigo para amigo, Colaborador de veículo de comunicação para leitor, Empresa para candidato a emprego, Empresa para cliente, Funcionário para gestor, Funcionário para outro funcionário, Gestor para funcionário, Instituição para a população, Indivíduo para instituição, Instituição para outra instituição, Indivíduo para profissional, Integrante de comunidade educacional para comunidade educacional, Integrante de grupo social para seus pares, Leitor/ouvinte para veículo de comunicação, Organização para associados
Natureza da interlocução	Tipo de interlocução, pessoal ou social; tipo de circulação da interlocução, restrita ou pública; e hierarquia social dos interlocutores envolvidos na interlocução projetada	Pessoal; Pública, assimétrica ascendente; Pública, assimétrica descendente; Pública, simétrica; Restrita, assimétrica ascendente; Restrita, assimétrica descendente; Restrita, simétrica
Gênero do discurso	Gênero discursivo da interlocução projetada	Abaixo-assinado, Anúncio de emprego, Apresentação, Artigo, Artigo de opinião, Capítulo de livro, Carta aberta, Carta do leitor, Carta/e-mail, Convite, Crônica, Depoimento, Diário de viagem, Editorial, Guia de orientações, Item de catálogo, Notícia, Panfleto, Pergunta, Pôster científico, Projeto, Propaganda, Redação para concurso, Relatório, Resposta à pesquisa de opinião, Resumo, Seção de guia

Suporte	Local de publicação ou método de envio da interlocução projetada	Abaixo-assinado, Blog, Cardápio, Carta/e-mail, Catálogo, Encarte de CD, Formulário, Guia, Jornal, Jornal escolar, Livro, Mural, Panfleto, Pôster, Processo, Programa, Projeto, Relatório, Revista, Site
---------	--	---

Fonte: Elaborado pela autora.

6.5 MEDIDAS DE COMPLEXIDADE TEXTUAL

O trabalho com as métricas do NILC-Matrix teve início na fase de análises preliminares, por meio da extração dos índices utilizando a ferramenta de interface gráfica disponível online. A interface web dessa ferramenta é capaz de processar um único texto, contendo, no máximo, duas mil palavras por vez. Assim, cada texto passou por processamento individual por meio dessa interface, e os resultados foram baixados em formato .csv, posteriormente revisados e agregados aos seus respectivos textos para a realização da análise no RStudio. Com a disponibilização do código fonte da ferramenta no segundo semestre de 2022, em resposta à nossa solicitação, o código e o *container* contendo os recursos necessários para o funcionamento adequado da ferramenta em nosso próprio servidor foram fornecidos. Todavia, mesmo após diversas tentativas e um diálogo direto com Sidney Evaldo Leal, um dos autores do código, não conseguimos resolver algumas falhas que se apresentaram durante esse processo. Nesse contexto, as métricas finais relativas aos nossos três *corpora* foram extraídas por ele, em seu próprio servidor, sendo gentilmente compartilhadas conosco¹³².

Como mencionado anteriormente, nem todas as 200 métricas estabelecem uma relação clara com a complexidade textual. Além disso, mediante pesquisa, constatamos que alguns dos índices carecem de evidências suficientes para validação, seja devido à falta de informações ou à fragilidade teórica e metodológica. Ademais, consultando a documentação da ferramenta, pudemos constatar que algumas métricas não apresentaram resultados satisfatórios nos testes conduzidos pelos desenvolvedores ou não apresentaram dados completos sobre os testes. Considerando esse cenário, concluímos que tais métricas apenas introduziriam ruído em nossas análises. Por essa razão, estabelecemos os seguintes critérios de exclusão das métricas: a) ausência de relação direta com a complexidade textual, conforme apontado pelos próprios criadores da ferramenta; b) fragilidade na validação; c) baixa acurácia nos resultados dos testes realizados pelos desenvolvedores; e d) resultados incompletos sobre os testes realizados pelos

¹³² Registramos novamente nossos agradecimentos a Sidney Evaldo Leal.

desenvolvedores. A tabela a seguir elenca as 70 métricas que foram excluídas de nossas análises (quadro completo com as 200 métricas no anexo E).

Quadro 16 – Métricas do NILC-Metrix excluídas das análises

GRUPOS	DESCRIÇÃO	MÉTRICAS	CRITÉRIO DE EXCLUSÃO
1. Medidas Descritivas	10 métricas de estatísticas básicas que descrevem o <i>corpus</i>	<ol style="list-style-type: none"> 1. Quantidade média de sentenças por parágrafo no texto (id: 155) <i>sentences_per_paragraph</i> 2. Proporção de subtítulos em relação à quantidade de sentenças do texto (id: 156) <i>subtitles</i> 	As 2 métricas apresentam ausência de relação direta com a complexidade textual, de acordo com os autores.
2. Simplicidade Textual	8 métricas que indicam a extensão das sentenças, uso de palavras de conteúdo e de pronomes	<ol style="list-style-type: none"> 1. Proporção de conjunções fáceis em relação à quantidade de palavras do texto (id: 187) <i>easy_conjunctions_ratio</i> 2. Proporção de conjunções difíceis em relação à quantidade de palavras do texto (id: 188) <i>hard_conjunctions_ratio</i> 3. Proporção de palavras de conteúdo simples em relação a todas as palavras de conteúdo do texto (id: 193) <i>simple_word_ratio</i> 	As 3 métricas apresentam fragilidade na validação.
3. Coesão Referencial	9 métricas que indicam cadeias de correferência intraparágrafo e interparágrafo	<ol style="list-style-type: none"> 1. Quantidade média de referentes que se repetem nos pares de sentenças do texto (id: 4) <i>arg_ovl</i> 2. Quantidade média de radicais de palavras de conteúdo que se repetem nos pares de sentenças do texto (id: 5) <i>stem_ovl</i> 	As 2 métricas apresentam baixa acurácia nos resultados.
5. Medidas Psicolinguísticas	24 métricas provenientes de experimentos psicolinguísticos com base em escala Likert de 7 níveis	<ol style="list-style-type: none"> 1. Proporção de palavras com valor de concretude entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 160) <i>concretude_1_25_ratio</i> 2. Proporção de palavras com valor de concretude entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 161) <i>concretude_25_4_ratio</i> 3. Proporção de palavras com valor de concretude entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto (id: 162) <i>concretude_4_55_ratio</i> 4. Proporção de palavras com valor de concretude entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 163) <i>concretude_55_7_ratio</i> 5. Média dos valores de concretude das palavras de conteúdo do texto (id: 164) <i>concretude_mean</i> 6. Desvio padrão do valor de concretude das palavras de conteúdo do texto (id: 165) <i>concretude_std</i> 7. Proporção de palavras com valor de familiaridade entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 166) <i>familiaridade_1_25_ratio</i> 	As 24 métricas apresentam fragilidade na validação.

		<p>8. Proporção de palavras com valor de familiaridade entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 167) familiaridade_25_4_ratio</p> <p>9. Proporção de palavras com valor de familiaridade entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto (id: 168) familiaridade_4_55_ratio</p> <p>10. Proporção de palavras com valor de familiaridade entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 169) familiaridade_55_7_ratio</p> <p>11. Média dos valores de familiaridade das palavras de conteúdo do texto (id: 170) familiaridade_mean</p> <p>12. Desvio padrão dos valores de familiaridade das palavras de conteúdo do texto (id: 171) familiaridade_std</p> <p>13. Proporção de palavras com valor de idade de aquisição entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 172) idade_aquisicao_1_25_ratio</p> <p>14. Proporção de palavras com valor de idade de aquisição entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 173) idade_aquisicao_25_4_ratio</p> <p>15. Proporção de palavras com valor de idade de aquisição entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto (id: 174) idade_aquisicao_4_55_ratio</p> <p>16. Proporção de palavras com valor de idade de aquisição entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 175) idade_aquisicao_55_7_ratio</p> <p>17. Média dos valores de idade de aquisição das palavras de conteúdo do texto (id: 176) idade_aquisicao_mean</p> <p>18. Desvio padrão dos valores de idade de aquisição das palavras de conteúdo do texto (id: 177) idade_aquisicao_std</p> <p>19. Proporção de palavras com valor de imageabilidade entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 178) imageabilidade_1_25_ratio</p> <p>20. Proporção de palavras com valor de imageabilidade entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 179) imageabilidade_25_4_ratio</p> <p>21. Proporção de palavras com valor de imageabilidade entre 4 e 5,5 em</p>	
--	--	--	--

		<p>relação a todas as palavras de conteúdo do texto (id: 180) <code>imageabilidade_4_55_ratio</code></p> <p>22. Proporção de palavras com valor de imageabilidade entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 181) <code>imageabilidade_55_7_ratio</code></p> <p>23. Média dos valores de imageabilidade das palavras de conteúdo do texto (id: 182) <code>imageabilidade_mean</code></p> <p>24. Desvio padrão dos valores de imageabilidade das palavras de conteúdo do texto (id: 183) <code>imageabilidade_std</code></p>	
6. Diversidade Lexical	15 métricas que descrevem o tipo de léxico utilizado no <i>corpus</i>	<ol style="list-style-type: none"> 1. Proporção de types de adjetivos em relação à quantidade de tokens de adjetivos no texto (id: 62) <code>adjective_diversity_ratio</code> 2. Proporção de types de palavras de conteúdo em relação à quantidade de tokens de palavras de conteúdo no texto (id: 64) <code>content_word_diversity</code> 3. Proporção máxima de palavras de conteúdo em relação à quantidade de palavras das sentenças (id: 65) <code>content_word_max</code> 4. Proporção Mínima de palavras de conteúdo por quantidade de palavras nas sentenças (id: 66) <code>content_word_min</code> 5. Desvio padrão das proporções entre as palavras de conteúdo e a quantidade de palavras das sentenças (id: 67) <code>content_word_standard_deviation</code> 6. Proporção de types de palavras funcionais em relação à quantidade de tokens de palavras funcionais no texto (id: 68) <code>function_word_diversity</code> 7. Proporção de types de pronomes em relação à quantidade de tokens de pronomes no texto (id: 72) <code>pronoun_diversity</code> 8. Proporção de types de verbos em relação à quantidade de tokens de verbos no texto (id: 76) <code>verb_diversity</code> 	<p>As métricas 1, 2, 6, 7 e 8 apresentam ausência de relação direta com a complexidade textual, de acordo com os autores.</p> <p>As métricas 3, 4 e 5 apresentam baixa acurácia nos resultados.</p>
8. Léxico Temporal	12 métricas que avaliam o uso de tempos e modos verbais e de conectivos temporais	<ol style="list-style-type: none"> 1. Proporção de Verbos no Pretérito Mais que Perfeito do Indicativo em relação à quantidade de verbos flexionados no texto (id: 143) <code>indicative_pluperfect_ratio</code> 	A métrica apresenta baixa acurácia nos resultados.
9. Complexidade Sintática	27 métricas que descrevem as estruturas sintáticas	<ol style="list-style-type: none"> 1. Quantidade média de apostos por oração do texto (id: 21) <code>apposition_per_clause</code> 2. Distância na árvore de dependências (id: 24) <code>dep_distance</code> 	<p>As métricas 1 e 4 apresentam baixa acurácia nos resultados.</p> <p>As métricas 2, 3 e 5 apresentam</p>

		<ol style="list-style-type: none"> 3. Fórmula de Complexidade Sintática de Frazier (id: 25) frazier 4. Desvio-padrão do tamanho dos sintagmas nominais do text (id: 41) std_noun_phrase 5. Fórmula de Complexidade Sintática de Yngve (id: 45) yngve 	resultados incompletos sobre os testes realizados pelos desenvolvedores.
10. Densidade de Padrões Sintáticos	4 métricas sobre sintagmas nominais	<ol style="list-style-type: none"> 1. Média dos tamanhos médios dos sintagmas nominais nas sentenças (id: 60) mean_noun_phrase 	A métrica apresenta baixa acurácia nos resultados dos testes realizados pelos desenvolvedores.
11. Informações Morfossintáticas de Palavras	42 métricas relativas às classes gramaticais da língua	<ol style="list-style-type: none"> 1. Proporção de Advérbios em relação à quantidade de palavras do texto (id: 91) adverbs 2. Proporção de types de advérbios em relação à quantidade de tokens de advérbios no texto (id: 92) adverbs_diversity_ratio 3. Proporção máxima de advérbios em relação à quantidade de palavras das sentenças (id: 93) adverbs_max 4. Proporção mínima de advérbios em relação à quantidade de palavras das sentenças (id: 94) adverbs_min 5. Proporção de verbos flexionados em relação a todos os verbos do texto (id: 104) inflected_verbs 6. Proporção de verbos no gerúndio, particípio ou infinitivo em relação a todos os verbos do texto (id: 105) non-inflected_verbs 7. Proporção máxima de substantivos em relação à quantidade de palavras das sentenças (id: 107) nouns_max 8. Proporção mínima de pronomes em relação à quantidade de palavras das sentenças (id: 116) pronouns_min 9. Proporção de Pronomes Relativos em relação à quantidade de pronomes do texto (id: 120) relative_pronouns_ratio 10. Proporção de pronomes possessivos nas terceiras pessoas em relação à quantidade de pronomes possessivos do texto (id: 123) third_person_possessive_pronouns 11. Proporção de pronomes pessoais nas terceiras pessoas em relação à quantidade de pronomes pessoais do texto (id: 124) third_person_pronouns 12. Proporção máxima de verbos por palavras em relação à quantidade de palavras das sentenças (id: 126) verbs_max 	As métricas de 1 a 8 e 11 a 13 apresentam ausência de relação direta com a complexidade textual, de acordo com os autores. As métricas 9 e 10 apresentam baixa acurácia nos resultados dos testes realizados pelos desenvolvedores.

		13. Proporção mínima de verbos em relação à quantidade de palavras das sentenças (id: 127) verbs_min	
12. Informações Semânticas de Palavras	11 métricas referentes aos significados das palavras usadas no <i>corpus</i>	<ol style="list-style-type: none"> 1. Proporção de sentidos dos adjetivos do texto em relação à quantidade de adjetivos do texto (id: 130) adjectives_ambiguity 2. Proporção de sentidos dos advérbios do texto em relação à quantidade de advérbios do texto (id: 131) adverbs_ambiguity 3. Proporção de sentidos dos substantivos do texto em relação à quantidade de substantivos do texto (id: 137) nouns_ambiguity 4. Proporção de sentidos dos verbos do texto em relação à quantidade de verbos do texto (id: 139) verbs_ambiguity 5. Quantidade Média de Hiperônimos por verbo nas sentenças (id: 133) hypernoms_verbs 6. Proporção de substantivos abstratos em relação à quantidade de palavras do texto (id: 129) abstract_nouns_ratio 7. Média de sentidos por palavra de conteúdo do texto (id: 132) content_words_ambiguity 8. Proporção de Nomes Próprios em relação à quantidade de palavras do Texto (id: 135) named_entity_ratio_text 9. Proporção de palavras de polaridade negativa em relação a todas palavras do texto (id: 136) negative_words 10. Proporção de palavras de polaridade positiva em relação a todas palavras do texto (id: 138) positive_words 	As métricas de 1 a 7 apresentam fragilidade na validação. As métricas 8, 9 e 10 apresentam ausência de relação direta com a complexidade textual, de acordo com os autores.
14. Índices de Leiturabilidade	5 métricas superficiais de complexidade textual (apenas duas adaptadas ao português: Flesch e Dale Chall)	<ol style="list-style-type: none"> 1. Índice Gunning Fog (id: 199) gunning_fox 	A métrica apresenta fragilidade na validação.

Fonte: Adaptado pela autora de <http://fw.nilc.icmc.usp.br:23380/metrixdoc>. Acesso em: 15 mar. 2023.

Uma vez em posse dos quantitativos das métricas do NILC-Metrix referentes aos nossos três *corpora*, procedemos à seleção das métricas produtivas que demonstrassem capacidade preditiva em relação à complexidade textual de textos da esfera jornalística. Para isso, replicamos o procedimento realizado por Leal *et al.* (2021), utilizando o algoritmo Boruta como método de seleção de variáveis. Através desse modelo, os autores conseguiram eliminar 147 das 200 métricas, resultando em um subconjunto de 53 métricas que foram capazes de distinguir os textos originais complexos dos textos manualmente simplificados.

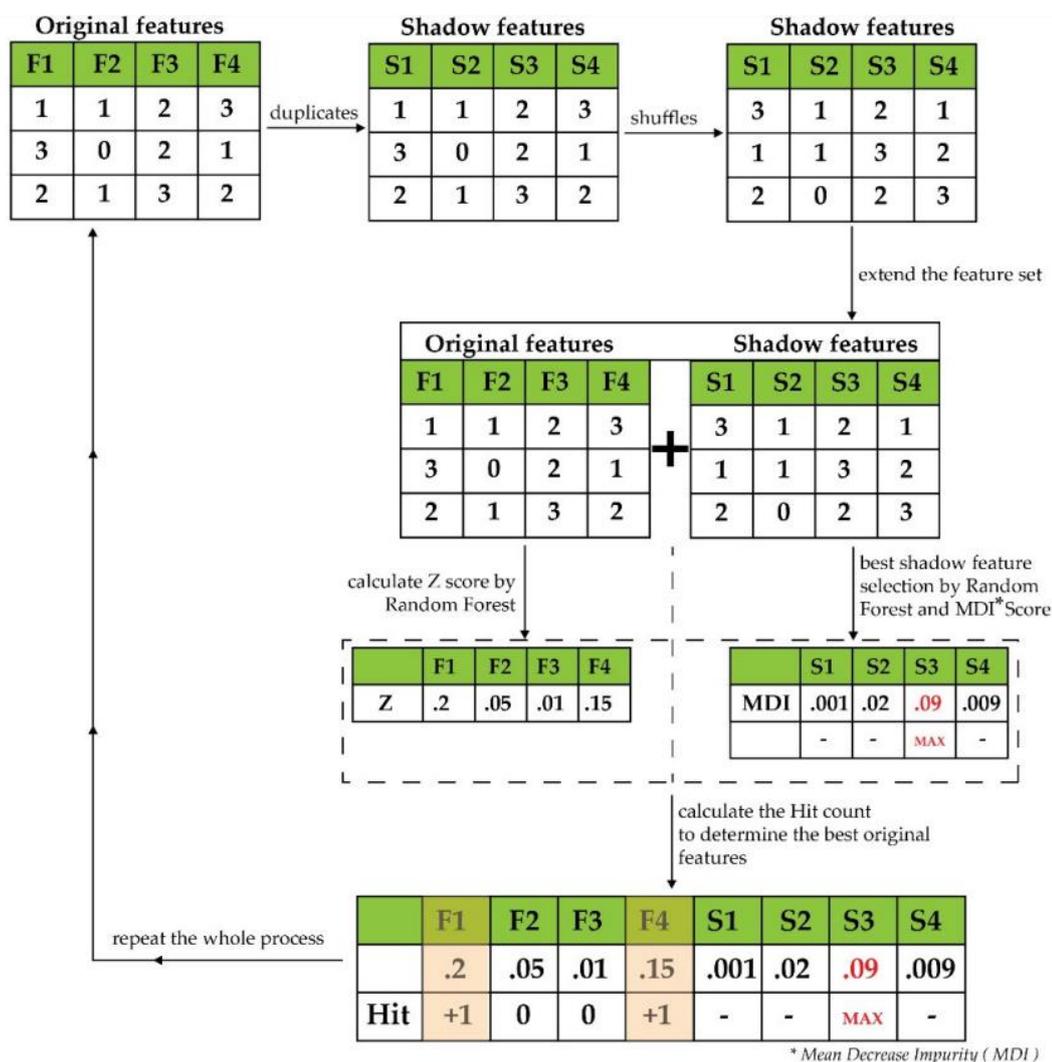
Com as nossas 130 métricas incluídas para as análises, avançamos para a etapa de redução de variáveis múltiplas, trabalhando com os dois *corpora* de referência, os quais se contrastam em termos de complexidade textual. O algoritmo Boruta funciona do seguinte modo:

1. Em primeiro lugar, adiciona aleatoriedade ao conjunto de dados, criando cópias aleatórias de todas as variáveis (variáveis sombra).
2. Em seguida, treina um classificador de random forest no conjunto de dados ampliado e aplica uma medida de importância das variáveis (Mean Decrease Accuracy) para avaliar a importância de cada variável, sendo que uma variável com valor mais alto significa maior importância.
3. Em cada iteração, o algoritmo verifica se uma variável real tem uma importância mais elevada do que a melhor das suas variáveis sombra (ou seja, se a variável tem um escore Z mais elevado do que o escore Z máximo das suas variáveis sombra) e remove constantemente as variáveis consideradas sem importância.
4. Finalmente, o algoritmo termina a análise quando todas as variáveis são aceitas ou rejeitadas ou quando atinge um determinado limite de execuções da random forest. (DUTTA, 2016, sem paginação)¹³³.

Uma representação das etapas do algoritmo é ilustrada na figura a seguir, adaptada de Hasan *et al.* (2020), para mostrar o processo iterativo que se repete até que todas as variáveis sejam categorizadas.

¹³³ No original, lê-se: 1. Firstly, it adds randomness to the given data set by creating shuffled copies of all features (which are called shadow features). 2. Then, it trains a random forest classifier on the extended data set and applies a feature importance measure (the default is Mean Decrease Accuracy) to evaluate the importance of each feature where higher means more important. 3. At every iteration, it checks whether a real feature has a higher importance than the best of its shadow features (i.e. whether the feature has a higher Z score than the maximum Z score of its shadow features) and constantly removes features which are deemed highly unimportant. 4. Finally, the algorithm stops either when all features gets confirmed or rejected or it reaches a specified limit of random forest runs.

Figura 21 – Etapas do algoritmo Boruta de seleção de variáveis

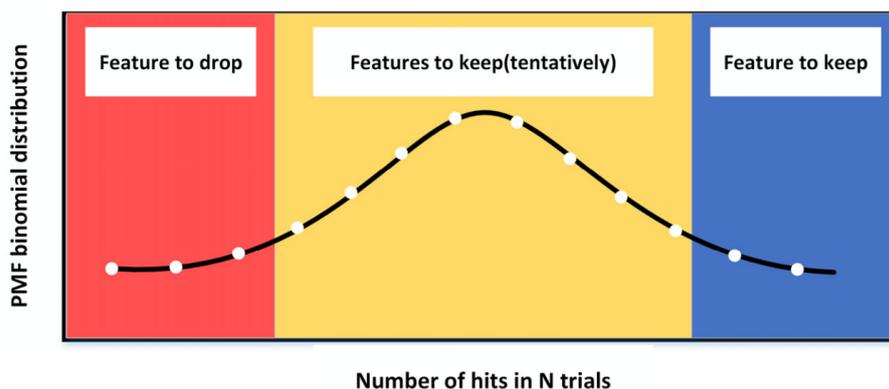


Fonte: Hasan *et al.*, 2020, p.5¹³⁴.

As variáveis são classificadas em três níveis: *recusa*, correspondente às variáveis de ruído; *aceite*, representando as variáveis preditivas; e *indecisão*, onde o pesquisador deve investigar qualitativamente as variáveis a fim de decidir se serão recusadas ou aceitas como relevantes, como ilustra o gráfico a seguir.

¹³⁴ Tradução nossa: Variáveis originais. Duplica. Variáveis sombra. Embaralha. Variáveis sombra. Coleção de dados ampliada. Variáveis originais + Variáveis sombra. Calcula escore Z por meio de Random Forest. Escolha da melhor variável sombra por meio de Random Forest e escore Mean Decrease Impurity. Calcula o número de acertos para escolher as variáveis originais. Repete o processo completo.

Gráfico 1 – Distribuição binomial de experimentos repetidos



Fonte: Lin *et al.*, 2022, p. 23¹³⁵.

As múltiplas iterações expandem o intervalo de confiança. É possível notar que as variáveis estão distribuídas em três faixas: a faixa vermelha denota as variáveis classificadas como ruído, a faixa azul engloba as variáveis consideradas preditivas e a faixa amarela, que segue uma distribuição normal, abrange as variáveis que podem ser submetidas a análises pelo pesquisador como possivelmente preditivas. Essa abordagem considera que correlações aleatórias entre variáveis podem levar à criação de dependências falsas, que podem passar nos testes estatísticos de validade, o que é particularmente possível de acontecer em análises de amostras pequenas. Dessa forma, destaca-se a importância de que a decisão final sobre a relevância das variáveis seja tomada por especialistas nos dados. Segundo Kursu, Jankowski e Rudnicki (2010, p. 283), “a importância do fator no classificador de aprendizagem de máquina pode ser usada como uma sugestão para a existência de uma relação entre as variáveis do sistema de informação e o fator de decisão, e não como uma prova decisiva”¹³⁶.

Após a seleção das variáveis preditivas de complexidade textual indicadas pelo algoritmo Boruta, procedemos à análise dos textos em nosso *corpus* de estudo com o objetivo de identificar o nível de complexidade textual desses textos. Essas análises foram complementadas por uma exploração qualitativa dos índices e de trechos de textos.

Ao concluir essa fase, realizamos uma análise abrangente dos dados, levando em consideração os resultados obtidos em todas as etapas da pesquisa, buscando uma visão holística a partir de uma perspectiva interpretativa.

¹³⁵ PMF = Probability mass function (função massa de probabilidade), $f(x)=P[X=x]=\binom{n}{x}p^x(1-p)^{n-x}$

¹³⁶ No original, lê-se: the importance of the attribute in the machine learning classifier may be used as a hint for existence of a relationship between variables of the information system and decision attribute, not as a decisive proof.

7 RESULTADOS

Neste capítulo, nossos resultados estão organizados em duas seções. Na primeira seção, exploramos os textos dos *corpora*, reportando informações sobre a distribuição de palavras nos textos e sobre os *headers* anotados no *corpus* de estudo com o propósito de descrever o conteúdo das tarefas. Na segunda seção, examinamos os índices obtidos por meio da ferramenta NILC-Metrix, a fim de compreender a complexidade textual dos textos. Antes de iniciar as descrições, recapitulamos a composição inicial dos três *corpora*, anteriormente apresentada na seção 6.1.

Tabela 5 – Composição inicial dos corpora

	<i>Corpus</i>	n	Total de <i>tokens</i>	Média de <i>tokens</i> (sd)	Total de <i>types</i>	Média de <i>types</i> (sd)
<i>Corpus</i> de estudo	Celpe	90	52.668	585 (268)	27.220	302 (85)
<i>Corpus</i> de referência (<i>proxy</i> complexidade textual alta)	PorSimples	148	61.525	415 (144)	33.183	224 (65)
<i>Corpus</i> de referência (<i>proxy</i> complexidade textual baixa)	PorPopular	148	68.334	461 (137)	35.644	240 (59)

Fonte: Elaborado pela autora.

Esses dados iniciais mostram que as médias de *tokens* e *types* são mais elevadas no *corpus* de estudo. Além disso, o desvio padrão (*standard deviation - sd*) no *corpus* Celpe-Bras é maior, o que indica uma variação mais ampla na dispersão dos valores em torno da média. Esse resultado sugere que os textos do *corpus* Celpe-Bras são mais longos e menos uniformes em comparação com os textos dos *corpora* de referência. Isso mostra que os textos de insumo das tarefas que integram leitura e escrita não apresentam a constância esperada entre as diferentes edições, no que se refere ao número de palavras por texto. Cabe observar que em jornais, cada texto é submetido às restrições que lhe são impostas pelo projeto gráfico (largura de coluna, mancha de texto e de imagem, fontes e corpo da fonte, etc.) e pela otimização do espaço disponível para a matéria (coluna, reportagem, notícia, etc.). No caso do Celpe-Bras, conforme dito anteriormente, a única restrição explicitada nas orientações para a elaboração de tarefas é o texto ocupar uma única página, e, como mostram os dados, isso não é suficiente para garantir uma uniformidade de extensão de textos.

Discrepâncias como essa podem comprometer a validade de construto de testes padronizados de larga escala e forte impacto social. Isso porque é pressuposto que as edições

sejam equivalentes, com conteúdo e dificuldade similares (SCARAMUCCI, 2016; GREEN; FULCHER, 2020; JIN, 2021). Além disso, é interessante observar que a média de frequência de *tokens* mais próxima à do *corpus* de estudo é encontrada nos dados dos jornais populares. Esse dado entra em contradição com o que a literatura sugere: os textos dos jornais populares analisados nesta tese não são textos verbais curtos, conforme descrevem Finatto *et al.* (2011), mas apresentam um número mais elevado de palavras por texto do que os textos dos jornais de referência.

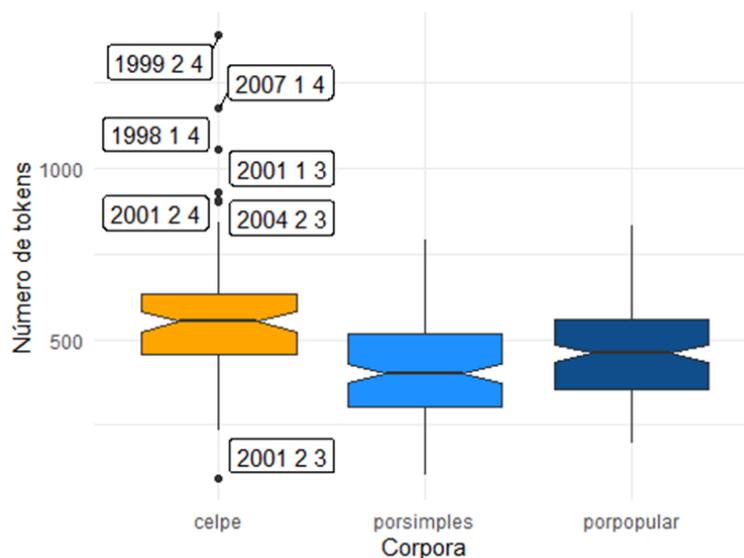
7.1 DESCRIÇÃO DAS TAREFAS

Como já mencionado, algumas edições das provas do Celpe-Bras foram excluídas das análises. Na primeira edição de 2001, na tarefa 4, os candidatos foram solicitados a preencher um quadro de respostas, portanto, esse quadro não foi considerado como texto de insumo para leitura, conforme caracterizado na concepção teórica de tarefa adotada neste estudo. Além disso, devido à utilização das sete categorias propostas por Schoffen *et al.* (2018) como variáveis nos *headers* dos textos, as tarefas 3 de 1998 e 4 de 1999-1 foram excluídas, uma vez que as autoras as consideraram como perguntas de compreensão que não envolvem habilidades integradas. Por fim, a tarefa 3 de 1999-1 também foi excluída do nosso conjunto de dados, pois as autoras a trataram como duas tarefas em uma só. Consequentemente, essa tarefa possui dois grupos de categorizações, o que nos impediria de manter a padronização dos nossos dados. Como resultado, do total inicial de 90 tarefas transcritas, demos início ao processo de investigação com o *corpus* de estudo composto por 86 tarefas.

7.1.1 Medidas de tendência central e dispersão

Iniciamos a exploração dos dados identificando a distribuição dos textos de acordo com o número de *tokens*, com o objetivo de detectar *outliers* nos três *corpora*, ou seja, textos com número de *tokens* fora do padrão, que poderiam enviesar os resultados das análises que contabilizam frequência de palavras.

Gráfico 2 – Distribuição de textos por tokens



Fonte: Elaborado pela autora.

Vemos claramente que o *corpus* de estudo exhibe uma dispersão de dados mais ampla e assimétrica, incluindo valores discrepantes. Essa visualização possibilitou a detecção de tarefas que apresentam um número de palavras divergente em relação aos demais dados que constituem o conjunto de observações do *corpus* Celpe (Primeiro Quartil = 458, Mediana = 551.5, Terceiro Quartil = 633, Amplitude Interquartil = 175). Somente o *corpus* de estudo revelou a presença de *outliers*. Esses valores atípicos são identificados pelos pontos que extrapolam os limites dos fios de bigode superior e inferior, que se estendem desde os entalhes da caixa até o valor máximo e mínimo, que não ultrapassam 1,5 vezes o valor do intervalo interquartil de 175.

Adicionalmente, os *boxplots* permitiram visualizar a não sobreposição dos entalhes das medianas, com destaque para a distância maior no *corpus* de estudo. Isso pode indicar uma diferença estatisticamente significativa entre o número de *tokens* das amostras. Contudo, para confirmar essas diferenças visuais identificadas pelo *boxplot*, são necessários testes estatísticos.

Os *outliers* detectados no *corpus* Celpe são referentes às seguintes tarefas:

- 1998-1, T4, 1057 *tokens*;
- 1999-2, T4, 1391 *tokens*;
- 2001-1, T3, 930 *tokens*;
- 2001-2, T4, 906 *tokens*;
- 2004-2, T3, 904 *tokens*;
- 2007-1, T4, 1175 *tokens*;

- 2001-2, T3, 93 *tokens*.

A tarefa 4 de 1998 utiliza dois textos de leitura, uma crônica e uma reportagem, e, por isso, o número total de palavras difere dos demais. A tarefa 4 de 1999-2 apresenta o número máximo de *tokens* no conjunto de dados, com 1391 palavras, e também utiliza dois textos de insumo, uma entrevista e uma reportagem. A tarefa 3 de 2001-1 é uma reportagem completa da *Folha de São Paulo*, e a tarefa 4 de 2001-2 utiliza dois textos, um editorial e uma reportagem. Em 2004-2, a tarefa 3 apresenta uma entrevista da *Folha de São Paulo*. Em 2007-1, a tarefa 4 também utilizou uma entrevista da revista *IstoÉ*, que ocupa duas páginas da prova. Por fim, abaixo do limite inferior temos a tarefa 3 de 2001-2, que utilizou um anúncio de promoção de aniversário do *Jornal do Brasil* com 93 palavras, representando o valor mínimo de nossos dados.

A variação identificada no *corpus* de estudo, com observações que variam de 93 até 1391 *tokens*, enfatiza a necessidade de orientações claras, objetivas e minuciosas para guiar os elaboradores na escolha do material de leitura para as tarefas (GREEN; FULCHER, 2020; JIN, 2021). Por se tratar de um teste padronizado, essa alta variação no número de palavras dos textos de insumo utilizados nas tarefas é uma variável que deve ser controlada, uma vez que a extensão do texto pode influenciar a dificuldade textual e o desempenho na leitura (MESMER; HIEBERT, 2015). As especificações contidas no Documento Base não indicam informações sobre a extensão dos textos utilizados como insumo para a leitura. Como mencionado anteriormente, o documento “Especificações para a elaboração de tarefas do exame Celpe-Bras - Parte Escrita” apenas sugere que o texto não exceda uma página; essa não é uma informação que atende o detalhamento técnico recomendado pela literatura. Isso porque essa orientação não esclarece de maneira eficaz qual é a extensão adequada do texto para compor a tarefa, uma vez que a formatação e a diagramação podem permitir a inclusão de textos com números de palavras significativamente distintos em uma página do tamanho A4, como ilustrado nos exemplos a seguir.

Figura 22 – Comparação de número de palavras em duas tarefas

2014/2 Celpe Bras
Certificado de Proficiência em Língua Portuguesa para Estrangeiros
Tarefa 4 | LEITURA E TECNOLOGIA | Página 8

Você é escritor e foi convidado a redigir um artigo de opinião sobre o tema "leitura e tecnologia" para uma revista de circulação nacional. Em seu texto, discuta as perspectivas de Philip Roth e de Robert Darnton e posicione-se a respeito do tema.

Philip Roth: "A cultura literária vai acabar em 20 anos"

O escritor americano afirma que a tecnologia deve acabar com o livro em papel e que a literatura tende a perder a influência na formação dos jovens.

Por Luis Antônio Gilson, de Nova York

Os 78 anos, 52 de carreira, ele é tido por críticos respeitados como o maior escritor vivo e figura há décadas na lista de possíveis ganhadores do Prêmio Nobel. É o único autor vivo a merecer a edição de suas obras na editora The Library of America, dedicada a escritores consagrados. Sob sua supervisão, o nono e último volume com os romances curtos deverá sair em 2013. Dele fará parte *Nemesis*, seu 22º romance, recém-lançado no Brasil.

ÉPOCA – Não há nenhum computador nesta sala. O que o senhor pensa sobre os avanços tecnológicos como tablets e e-readers? Eles melhoram a compreensão do mundo?

Philip Roth – Não sou fanático por tecnologia. Tenho o mesmo telefone celular há anos e não pretendo trocá-lo. Escrevo em computador, como fiz antes com a máquina de escrever. É óbvio que as máquinas facilitam a finalização de um texto. Só que as coisas estão se transformando muito rapidamente para meu gosto. Não consigo achar graça em ler livros em formato eletrônico em e-reader. Outro dia passei em uma loja Apple com a forte disposição de comprar um iPad. Cheguei lá, vi tanta gente se acotovelando para ver como funcionava o aparelho e cheguei a testá-lo. Acabei desistindo. Não sei por que, mas o iPad não me convenceu, talvez porque pareça chato escrever nele, e ler nele é desperdício. Quem vai conseguir ler um livro inteiro mesa naquela *tablet*? É mais um totem do culto à tecnologia. Hoje, toda a cultura se encontra a nossa disposição. E isso me preocupa. A cultura literária como conhecemos vai acabar em 20 anos. Ela já está agonizando. Obras de ficção não despertam mais interesse dos jovens, e tenho a impressão de que não são mais lidas. Hoje, a atenção é voltada para o mais novo celular, o mais novo tablet. Daqui a poucas décadas, a relação do público e do escritor com a cultura será muito diferente. Não sei como será, mas os livros em papel vão acabar. Surgirá outro tipo de literatura, com recursos audiovisuais e o que mais inventarem.

Disponível em: <http://weblogepoca.globo.com>. Acesso em: 10 set. 2011.

O Futuro do Livro: Robert Darnton

Por Luis Antônio Gilson, de Nova York

"Mas as pessoas dizem que o futuro é digital. Claro que é digital. O presente também é digital."

Quando o assunto é a especulação a respeito do futuro do livro, é difícil não ouvir aquela já batida pergunta: "o livro digital vai matar o livro físico?". Mas por que um deve necessariamente aniquilar o outro? Não seria possível haver uma coexistência? Robert Darnton, historiador e diretor da biblioteca de Harvard, a maior biblioteca universitária do mundo, não só defende a ideia de que os dois podem coexistir, mas que são complementares, que um reforça o outro.

Darnton não é só conhecido como um grande historiador norte-americano, mas também como um dos pioneiros na área da história do livro. Um de seus livros, *A Questão dos Livros – passado, presente e futuro* (The Case for Books, 2010), traduzido por Daniel Pellizzari e lançado pela Companhia das Letras, aborda justamente este questionamento a respeito do futuro dos livros.

Mesmo com o barulho em torno dos livros digitais, sabemos que eles vão conviver um bom tempo com os livros em papel. Quais as principais questões que ainda não mereceram a devida atenção quando se fala nisso?

Darnton – Tenho sido convidado para tantas conferências sobre a morte do livro que acredito que o livro está muito mais vivo. Tenho algumas estatísticas sobre a produção de livros: a produção de livros impressos aumenta a cada ano e, em breve, teremos 1 milhão de títulos novos a cada ano. Sei que no Brasil o mercado editorial está florescendo. Portanto, a ideia de que o livro morreu me parece absurda. Contudo, levo a sério o fato de que livros eletrônicos são a onda do futuro. E como conduzir essa transição de um mundo onde o livro impresso é dominante para um futuro onde os livros digitais dominarão? Estamos passando por um tipo de fase de transição onde ambos vão conviver. E isso, para mim, é o momento mais excitante e interessante, porque podemos trabalhar em novos tipos de livros, depois de repensar a noção do que é um livro. Como escritor, acho que eles são complementares, e não contradições.

Se você analisar a História da Comunicação, descobrirá que uma mídia não toma o lugar de outra, elas vivem num tipo de co-habitação, que é mutuamente benéfica. Aprendemos que, no tempo de Gutenberg, apesar de sabermos que a impressão foi uma tremenda nova força, os livros manuscritos, que tinham uma tradição de 100, 200 cópias, permaneceram. Ambos reforçaram um ao outro, por serem incomparáveis. E agora temos isso novamente, com o livro impresso e eletrônico a se reforçarem.

Disponível em: <http://historiagac.com.br>. Acesso em: 7 jul. 2011.

Celpe Bras PARTE COLETIVA

Tarefa III

SOLUÇÕES

Sem medo de avião

- Aceite o medo e tente enfrentá-lo. Saiba que o avião é o mais seguro meio de transporte. Nos últimos dez anos, a média de acidentes foi de um para cada 1 milhão de decolagens.
- Informe-se sobre o funcionamento da nave e a preparação dos profissionais que a comandam.
- Não cultive idéias catastróficas apenas porque o avião balançou numa turbulência. Caso se sinta inseguro, converse com um comissário.
- Respire devagar e corretamente. Ajuda a relaxar o corpo e acalma a mente. Para ficar mais tranquilo, vale rezar, falar mantra ou tentar se distrair.
- Procure uma distração: ler um livro, fazer palavras cruzadas, ouvir música, assistir a um filme ou simplesmente conversar com o passageiro ao lado.
- Use roupas leves. Viaje sempre com meias para ficar descalço no avião. Ficar confortável ajuda.
- Não tome bebidas alcoólicas. O efeito do álcool em voo é potencializado, aumentando a chance de ficar embriagado ou ter enjojo.
- Hidrate-se sempre, mas evite bebidas gasificadas. Um copo de água ou suco a cada hora é suficiente.
- Prefira alimentos leves, como sopas e saladas. Pratos gordurosos e pesados dificultam a digestão e aumentam o desconforto natural do voo.
- Evite ingerir alimentos e bebidas com cafeína. Eles prejudicam o sono e aumentam a sensação de preocupação.
- Jamais tome medicamentos por conta própria. Caso não consiga controlar o medo, consulte um especialista.

Época, 25 de julho de 2005

Página 5

Fonte: Acervo Celpe-Bras.

O texto de insumo da tarefa 4 de 2014-1 (imagem à esquerda) contém 840 palavras, que é o novo valor máximo dos dados após a exclusão de *outliers*. Essa quantidade excede em três vezes o número de palavras do texto de insumo utilizado na tarefa 3 de 2005-2 (imagem à direita), que possui 234 palavras, representando o novo valor mínimo do *corpus* de estudo. As imagens das tarefas evidenciam a falta de padronização no *layout*, o que dificulta o controle apropriado do número de palavras nos textos de leitura. Independentemente das sub-habilidades de leitura avaliadas ou dos propósitos de escrita solicitados em cada tarefa, pressupõe-se a leitura completa do texto de insumo pelo candidato. Considerando que o tempo estimado de leitura para adultos proficientes na língua, fora de um contexto de avaliação, é de 220 a 250 palavras por minuto (ROSA, *et al.*, 2016; BRYLSBAERT, 2019), e que o tempo de realização da PE é fixo em 3 horas, entendemos como necessário para a validade do conteúdo do exame

que a extensão do texto de insumo seja levada em consideração no momento da escolha do material¹³⁷.

Além disso, ao observar as duas tarefas mencionadas, podemos notar amplas diferenças nos aspectos tipográficos, como o tipo e tamanho da fonte, margens, espaçamentos entre letras, palavras e linhas, e o contraste entre a fonte e a cor do fundo da página. Esses aspectos são considerados como características que podem favorecer ou desfavorecer a compreensão leitora, até mesmo para leitores proficientes (LOURENÇO, 2016). Além disso, a apresentação do texto pode potencialmente influenciar a tensão que o teste gera no candidato e a motivação para se envolver na tarefa (SOUSA; HÜBNER, 2015). Em nossos dados, identificamos algumas tarefas em que o *layout* exibe o texto de leitura no formato de orientação paisagem, como no exemplo da tarefa 3 da edição 2012-1. Isso pode influenciar no desempenho da tarefa, uma vez que requer que o candidato manipule o material que exibe textos em ambas as orientações, paisagem e retrato.

¹³⁷ Conforme já mencionado, as orientações de elaboração do exame não são públicas. No entanto, durante a defesa desta tese, foi relatada uma praxe dos elaboradores de buscar um equilíbrio entre as extensões dos textos nas tarefas 3 e 4 (um mais e outro menos extenso) em cada edição. Uma pesquisa futura poderia analisar se essa praxe tem de fato sido observada ao longo dos anos.

Figura 23 – Tarefa 3 da edição 2012-1 em orientação retrato

Celpe
Bras
2012/1

Certificado de Proficiência
em Língua Portuguesa para Estrangeiros

Tarefa 3 | Faxina no Céu
Página 6

Preocupado com a situação do lixo abandonado na órbita da Terra, você decidiu escrever para o Greenpeace, solicitando que a instituição inclua essa discussão em sua agenda de reivindicações. No seu texto, você deverá apresentar a origem do problema e seus possíveis desdobramentos, bem como sugerir algumas soluções.

Faxina no céu

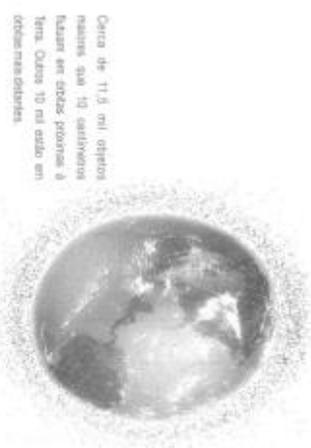
O espaço está cheio de veículos descartados e fragmentos metálicos girando em alta velocidade. Como eliminar esses detritos?

OUTROS CIENTISTAS DA NASA estão empenhados em levar astronautas ao espaço ou enviar sondas interplanetárias a Júpiter, mas o que preocupa Nicholas Johnson é um cenário de pesadelo chamado "síndrome de Kessler", em homenagem a seu colega Donald Kessler, o cientista que a descreveu pela primeira vez, na década de 1970. O cenário básico é uma órbita saturada de objetos. Duas peças de equipamentos enormes — como satélites ou foguetes de lançamento — colidem a uma velocidade superior a 32 mil quilômetros por hora, e ambas se transformam em centenas de fragmentos. Um desses fragmentos, em seguida

choca-se com outro objeto volumoso, que por sua vez se desfaz em centenas de pedaços — e assim por diante, em uma reação em cadeia que culmina na formação de um amêlo de detritos espaciais densos demais para ser atravessado em segurança.

Ale o ano passado, diz Johnson, o responsável pelo Departamento do Programa de Detritos Orbitais da NASA, "tal perigo era apenas uma questão acadêmica". Mas, em 10 de fevereiro de 2009, foi registrado o primeiro choque direto entre objetos em hipervelocidade, acima de 10 mil quilômetros por hora. Um satélite da Índia colidiu com um satélite russo obsoleto à altitude de 800 quilômetros sobre a Sibéria. Esse único acidente acrescentou cerca de 2 mil grandes fragmentos à nuvem de detritos que gira em torno da Terra.





Cerca de 11,5 mil objetos maiores que 10 centímetros flutuam em órbitas próximas à Terra. Outros 10 mil estão em órbitas mais distantes.

National Geographic, jul. 2010.

Agências espaciais de todo o mundo monitoram os fragmentos maiores, evitando que haja colisões entre eles e os veículos espaciais, sobretudo os tripulados. E, em 2007, a ONU pediu a adoção de medidas preventivas, como o esgotamento do combustível dos foguetes utilizados para que não explodam, e a recomendação de que não se usem satélites desativados como alvos para testes de mísseis — o que a China havia feito no ano anterior.

Tais regras básicas, contudo, não vão impedir colisões acidentais. "Nos próximos 50 anos", diz Johnson, "é razoável que haja uma colisão de dois grandes objetos a cada cinco anos." É provável que isso não seja suficiente para desencadear o pesadelo de Kessler. Por outro lado, tampouco se vê no horizonte algum esquema de faxina vácuo. Está sendo discutidas várias maneiras de lidar com o lixo espacial. Um cabo comido e capaz de conduzir eletricidade poderia ser fixado aos satélites obsoletos, colocando-os sob a influência do campo magnético do planeta, que acabaria por arrastá-los para a atmosfera, onde seriam consumidos.

Outra solução: um satélite coletor, uma espécie de cambião de lixo espacial, recolhida os detritos para jogá-los fora perto da atmosfera para que se desintegrassem ao cair.

No caso de fragmentos menores, um satélite dotado de um laser poderoso daria conta de destruí-los. Outra abordagem seria menos violenta: uma gigantesca bola de espuma móvel no espaço. Os detritos não seriam capturados na espuma, que apenas absorveria deles quantidade de energia suficiente para que caíssem na atmosfera. Mas é evidente que, como reconhece Johnson, "talvez seja difícil lançar ao espaço uma bola com 1,6 mil metros de diâmetro". — Por Michael D. Lemonick

ILUSTRAÇÃO DE BEAN WILMINGTON
FONTE: DEPARTAMENTO DO PROGRAMA DE DETRITOS ORBITAIS DA NASA

De acordo com Lourenço (2016), o ato de ler exige esforço do sujeito leitor, e características tipográficas que causem desconforto podem não apenas dificultar a compreensão, mas também levar ao abandono da leitura. Portanto, uma uniformização dos elementos tipográficos pode ajudar a controlar variáveis externas indesejáveis à avaliação pretendida. As características tipográficas e de edição de *layout* do material são detalhes técnicos que também devem ser fornecidos nas especificações para que a reprodução de edições equivalentes seja possível (WEIGLE, 2004; SCARAMUCCI, 2016; GREEN; FULCHER, 2020; SHIN, 2021).

Os valores centrais, assim como a distribuição e a dispersão de palavras nos três *corpora*, estão apresentados na tabela a seguir.

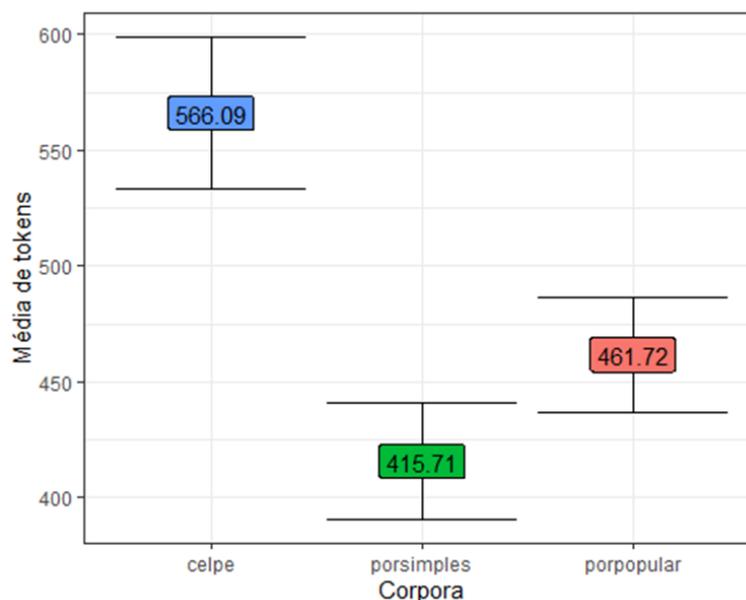
Tabela 6 – Medidas de tendência central e de dispersão de palavras dos três corpora

	<i>Corpus</i>	n	Valor mínimo	Primeiro quartil	Mediana	Média	Terceiro quartil	Valor máximo
<i>Corpus</i> de estudo	Celpe	86	93.0	458.0	551.5	566.1	633.0	1391.0
<i>Corpus</i> de referência	PorSimples	148	101.0	301.2	401.5	415.7	517.2	791.0
<i>Corpus</i> de referência	PorPopular	148	194.0	355.8	458.5	461.7	560.5	833.0

Fonte: Elaborado pela autora.

Assim como a mediana do *corpus* de estudo revelou-se mais elevada do que as medianas dos *corpora* de referência, a tabela acima também indica que a média de palavras do *corpus* Celpe é superior às médias do PorSimples e do PorPopular. A fim de aprofundar a análise dessa diferença, optamos por representar graficamente as médias de *tokens* por *corpora*.

Gráfico 3 – Média de tokens por corpora



Fonte: Elaborado pela autora.

É evidente que o *corpus* de estudo apresenta uma média de palavras por texto superior àquelas dos dois *corpora* de referência. Além disso, o *corpus* PorPopular é constituído por textos mais extensos quando comparado ao *corpus* PorSimples. A representação gráfica do *boxplot* reforça a suposição de que os nossos dados contradizem o pressuposto teórico de que os textos do jornalismo popular se caracterizam por serem textos verbais curtos (FINATTO *et al.*, 2011). De fato, sua distribuição e média de *tokens* revelam-se mais elevadas em comparação com o outro *corpus* de referência, constituído por textos oriundos de jornais de maior prestígio. A fim de obter uma avaliação mais precisa das diferenças no número de *tokens* entre os *corpora*, realizamos um modelo estatístico de regressão.

Tabela 7 – Resultados da regressão linear de número de tokens e corpora

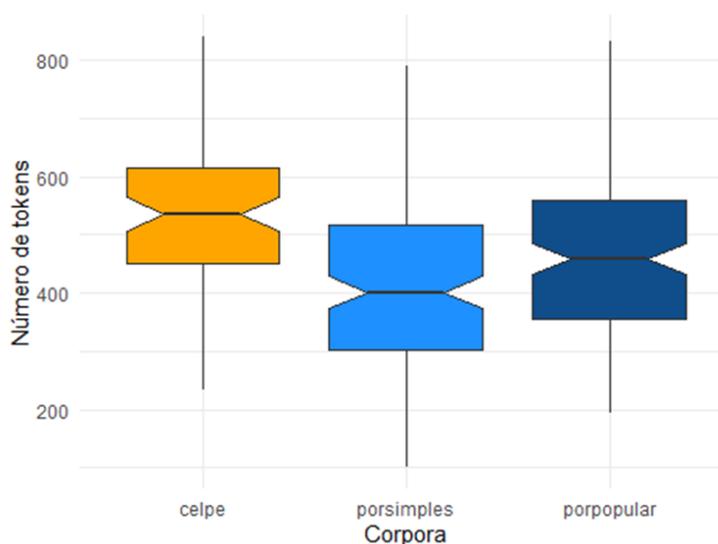
	<i>Corpus</i>	n	Estimativa	Erro padrão	Valor t	Valor p
<i>Corpus</i> de estudo	Celpe (Intercept)	86	566.09	16.72	33.867	<0.001***
<i>Corpus</i> de referência	PorSimples	148	-104.38	21.02	-4.966	<0.001***
<i>Corpus</i> de referência	PorPopular	148	-150.38	21.02	-7.155	<0.001***
Resíduo	155			$F_{2,381} = 25.76$		$p < 0.001$
R^2	R^2 Múltiplo = 0.1197 R^2 Ajustado = 0.115					

Fonte: Elaborado pela autora.

O modelo confirma que existe uma diferença estatisticamente significativa no número de palavras por texto entre os três *corpora* analisados. A análise mostra que a média de palavras por texto no *corpus* Celpe é de 566; em comparação, um texto do PorSimples apresentará uma redução de 150 palavras, enquanto um texto do PorPopular terá uma diminuição de 104 palavras. Apesar das estimativas do teste t terem um valor de significância $p = <0.001$, o valor do coeficiente de determinação ajustado (R^2 ajustado) indica que os *corpora* explicam apenas 11% da variação no número de *tokens* nos textos. Isso sugere que, embora o modelo estatístico sinalize diferenças no número de palavras por texto entre os três *corpora*, e que essas diferenças não ocorrem ao acaso, o tamanho do efeito do *corpus* na variação do número de *tokens* não é substancialmente forte para explicar as diferenças.

Após a detecção e exclusão de *outliers* no *corpus* Celpe, observamos uma maior aproximação na distribuição de textos em relação ao número de *tokens* entre os três *corpora*, conforme evidenciado no *boxplot* a seguir.

Gráfico 4 – Distribuição de textos por tokens após exclusão de outliers



Fonte: Elaborado pela autora.

O gráfico proporciona uma visualização de *corpora* mais equilibrados após a remoção dos valores discrepantes, embora os entalhes das medianas ainda não se sobreponham. Mesmo que os dados continuem a indicar uma diferença estatisticamente significativa no número de *tokens* entre os *corpora* ($p = <0,001$), o modelo estatístico de regressão linear reafirma que a diferença no número de *tokens* é minimamente determinada pelo *corpus* (Erro padrão dos resíduos = 139, R^2 ajustado = 0,08). Esse dado mostra que foi possível trabalhar com *corpora*

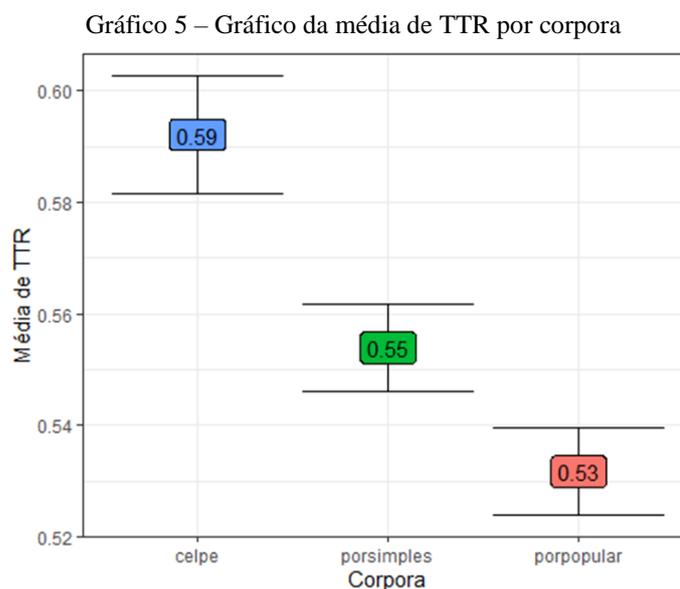
balanceados e comparáveis, à exceção da complexidade textual direcionada a públicos-alvo distintos, que é precisamente o efeito que desejamos observar. Isso nos permite obter resultados significativos para responder às nossas perguntas de pesquisa relacionadas à complexidade textual de maneira ética e válida, seguindo o rigor preconizado pela LC (EGBERT, 2017; XI, 2017; GABLASOVA, 2020). Após a exclusão dos dados discrepantes, a composição dos três *corpora* é a seguinte:

Tabela 8 – Composição final dos corpora

	Corpus de estudo		Corpus de referência	
	Celpe	PorSimples	PorPopular	
n	79	148	148	
Total de <i>tokens</i>	42.228	61.525	68.334	
Valor mínimo	234.0	101.0	194.0	
Primeiro quartil	451.0	301.2	355.8	
Mediana	535.0	401.5	458.5	
Terceiro quartil	615.5	517.2	560.5	
Valor máximo	840.0	791.0	833.0	
Média de <i>tokens</i> (sd)	534.53 (129.18)	415.71 (144.87)	461.72 (137.1)	
Total de <i>types</i>	24.698	33.183	35.644	
Média de <i>types</i> (sd)	312.63 (64)	224.21 (65)	240.84 (59)	
<i>Type-Token Ratio</i>	0.59	0.55	0.53	

Fonte: Elaborado pela autora.

O conjunto de textos do *corpus* PorSimples é caracterizado pelo menor número total de palavras, bem como pela menor média de *tokens* e *types* por texto. Já o *corpus* PorPopular apresenta o maior número total de palavras entre os três *corpora*. No entanto, a proporção de TTR é a mais baixa nesse conjunto, pois quanto mais extenso o texto, menor tende a ser a diversidade lexical (LEAL *et al.*, 2021). Por sua vez, devido ao seu tamanho reduzido em termos de número de textos, o *corpus* de estudo exhibe o menor número total de palavras, mas a maior média de *tokens*, *types* e TTR, revelando textos mais extensos e com uma maior variedade lexical. No gráfico a seguir, são apresentados os resultados do modelo de regressão linear, destacando o efeito de cada *corpus* na variável TTR. As barras representam os intervalos de confiança (95%) em torno da média para cada *corpus*.



Fonte: Elaborado pela autora.

O *boxplot* revela diferenças estatisticamente significativas entre os valores de TTR nos três *corpora*, uma vez que os intervalos de confiança não se sobrepõem. Isso indica que os conjuntos de textos possuem distintas densidades lexicais. Os dados apresentados na tabela anterior indicaram que, em termos de média de *tokens* e de *types*, o *corpus* Celpe (média de *tokens* = 535 e média de *types* = 313) se assemelha mais ao PorPopular (média de *tokens* = 462 e média de *types* = 241). Entretanto, apesar de serem textos mais longos, quando consideramos o TTR, o *corpus* Celpe (0.59) se aproxima mais do PorSimples (0.55). Uma maior taxa de TTR indica uma densidade lexical mais elevada nos textos, que, segundo Leal *et al.* (2020), indica textos mais complexos, visto que quanto maior a riqueza lexical, maior será a complexidade textual.

Os textos do *corpus* de estudo se aproximam em extensão dos textos presentes nos jornais populares. No entanto, apresentam uma complexidade lexical mais próxima dos textos provenientes dos jornais de referência, que tendem a ser mais breves, porém, com uma maior diversidade de palavras. Esses dados indicam que os textos do Celpe-Bras são os mais longos e densos lexicalmente em comparação aos *corpora* de referência, podendo sugerir, teoricamente, que são os mais complexos.

Para uma análise mais detalhada da riqueza lexical, analisamos a recorrência de palavras de conteúdo na tarefa 4 da edição 2016-2 do Celpe-Bras, anteriormente discutida no capítulo 4, que apresenta uma média de TTR de 0.55, exatamente a média geral do *corpus* de estudo. Na

figura a seguir, as palavras destacadas em amarelo representam as palavras que se repetem pelo menos uma vez, considerando suas formas lematizadas. Por exemplo, “perdida” e “perdeu” são variações do *lemma* “perder”.

Figura 24 – Recorrência de palavras na tarefa 4 da edição do Celpe-Bras 2016-2

Depois de ler o texto “Shopping em casa”, você decidiu escrever para a seção Cartas do Leitor do jornal O Estado de São Paulo, defendendo seu ponto de vista em relação às questões levantadas por Denise Fraga.

Shopping em casa

— Que bonita esta **saia!**
 — Pois é, estava **perdida** — ela me respondeu.
 — Você **perdeu** uma **saia**?
 — Sim, dentro do meu **armário**.
 Arrumando as **coisas** pra **mudança**, eu **achei**.

Esta conversinha banal com minha amiga aumentou a minha angústia. Será que **preciso** de uma **mudança** pra conseguir **achar** o que anda **perdido** **diante** dos meus **olhos**? Muitas **vezes**, tenho a sensação de não ter o que vestir **diante** do meu guarda-roupa **lotado** e fico doida cobiçando um **novo** **vestido** na **vitrine** do **shopping**. No outro dia, **comprei** uma **blusa** que tinha amado e, quando **fui** **pendurá-la** no **armário**, vi que sua **vizinha** de cabide era sua prima. O **curioso** é que, se fosse a **blusa** antiga que estivesse na loja, eu a **compraria** de **novo**. **Mais curioso**: ela estava lá, eu sabia que tinha e **pouco** a **usava**. A **blusa** **nova** me devolvera a **velha**, que passei a **usar** com prazer. Mas por que **comprei** uma **nova** se havia uma quase igual **pendurada**? Que fenômeno é esse de **precisar** do que já se tem? Que cegueira é esta para o que está ao seu lado? Será a galinha do **vizinho** sempre **mais** gorda? **Adquirir** é melhor do que ter?

Não sou exatamente **consumista**, mas **olho** a minha **casa** e, apesar de morrer de amores por meus objetos, sinto que há uma infinidade de **coisas** que eu **realmente** poderia **viver** sem. O **curioso** é que, se descuidar, se me **forem** oferecidas, estou arriscada a **comprá-las** de **novo**.

É **realmente** muito violento o impulso **consumista** a que **fomos** doutrinados. Nem **precisamos** **ir** ao **shopping**. **Vivemos** em uma feira constante. Meu **e-mail** está **lotado** de propostas de vendas que nem desconfio como chegaram ali. A internet virou, como previsto, um mar de publicidade que, a um clique, desfalca nossa conta bancária. Tudo pela estrela maior: o dinheiro, o circular da moeda. Cada **vez** **mais** produtos, cada **vez** **mais** necessidades desnecessárias. Não entendo nada de economia, mas **gostaria** de **viver** num mundo com **menos** coisas.

Gostaria de morar numa **casa** com **menos** coisas, ter um **armário** com **menos** roupas.

Resolvi fazer um exercício: promover as **vitrines** a uma exposição de **arte**. A **arte** de um artista poderoso e midiático que se chama Mercado e de quem estou longe de **poder** **adquirir** uma obra. Mesmo sem ter que **mudar** de apartamento, **vou** tirar as coisas de meu **armário**, colocar sobre a cama para **comprar** com meu **novo** **olhar** os meus **velhos** **vestidos**. Liberdade é **viver** com **pouco**.



Ilustração de Zé Vitorino

Disponível em: www1.folha.uol.com.br/colunas/denisefraga/2016/02/1739053-shopping-em-casa.shtml. Acesso em: 5 ago. 2016 (adaptado).

Fonte: Acervo Celpe-Bras.

Ao visualizar a imagem com as palavras destacadas em amarelo, torna-se evidente que na crônica de Denise Fraga existe uma manutenção consistente dos tópicos entre os parágrafos. Pode-se notar o uso frequente de palavras recorrentes que pertencem aos mesmos campos

semânticos, tais como: *saia, blusa, vestido e roupa; casa e armário; shopping e vitrine; comprar, adquirir e consumismo*. Nesse texto, a colunista Denise Fraga compartilha um diálogo com uma amiga que a levou a refletir sobre como a grande quantidade de roupas que ela possui não a impede de desejar adquirir mais. A partir dessa revelação, a autora contempla o impacto da publicidade, especialmente a virtual, nesse impulso consumista. A cronista conclui seu texto relatando que planeja retirar as roupas do armário para apreciar o que já possui em casa, na tentativa de saciar a necessidade de adquirir novas peças, assemelhando esse processo a um “shopping em casa”.

Ao compor seu texto, como evidenciado na imagem, Denise Fraga utiliza palavras que podem ser diretamente relacionadas ao título “Shopping em casa”, o que pode favorecer a compreensão global do texto e, conseqüentemente, a compreensão do ponto de vista da autora, aspecto crucial para o cumprimento da tarefa. No entanto, a formulação do enunciado da tarefa não é precisa ao solicitar que o candidato tome uma posição em relação ao ponto de vista da autora acerca do consumismo. Ao convidar o examinando a defender um ponto de vista “em relação às questões levantadas por Denise”, o enunciado pode levar a interpretações ambíguas sobre o que deve ser abordado na resposta. Essa interpretação ambígua é possível porque o termo “questão” em português brasileiro, especialmente em contextos educacionais, é frequentemente empregado com o significado de “pergunta para obtenção de uma informação ou de um esclarecimento”¹³⁸. Nesse sentido, essa ambigüidade que o enunciado pode suscitar é reforçada pelas perguntas literais feitas ao longo do texto: *Será que preciso de uma mudança pra conseguir achar o que anda perdido diante dos meus olhos?; Mas por que comprei uma nova se havia uma quase igual pendurada?; Que fenômeno é esse de precisar do que já se tem?; Que cegueira é esta para o que está ao seu lado?; Será a galinha do vizinho sempre mais gorda?; Adquirir é melhor do que ter?*

Mendel (2017) ressalta que a leitura dessa crônica exige uma leitura reflexiva, compreensão das diferentes estratégias discursivas características do gênero crônica e a habilidade de realizar inferências. A autora também aponta para incongruências na tarefa que podem prejudicar o desempenho dos candidatos. Por um lado, há a indicação da fonte *O Estado de São Paulo* e da autora Denise Fraga no enunciado. Por outro lado, há a omissão da autoria no texto de insumo e a citação de um jornal diferente como fonte, ao pé da página. Examinandos com níveis mais baixos de proficiência, vocabulário limitado ou experiência restrita com tarefas

¹³⁸ <https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=quest%C3%A3o> Acesso em: 13 ago. 2023.

integradas podem se confundir ao não entender por que devem escrever para o jornal *O Estado de São Paulo*, a que “questões” devem se posicionar e quem é Denise Fraga.

Apesar disso, Sirianni (2020) observa que Denise Fraga aborda diversos aspectos relacionados à temática do consumismo em seu texto. Portanto, a autora considera que o enunciado da tarefa permite que o candidato se posicione a respeito do assunto de modo mais amplo. Desse modo, poderíamos inferir que, embora a formulação e a orientação da tarefa possam ser consideradas problemáticas, a recorrência das palavras relacionadas aos campos semânticos que aludem ao consumismo de roupas facilita a leitura e a liberdade de resposta simplifica a produção textual. No entanto, a pesquisa de Schlatter *et al.* (2021), apresentada anteriormente, indica que essa tarefa foi a mais desafiadora na edição de 2016-2, apresentando uma alta incidência de notas que não certificam e a média mais baixa entre as quatro tarefas, conforme ilustra a tabela abaixo.

Tabela 9 – Médias das notas finais das quatro tarefas das edições 2015-2, 2016-1, 2016-2 e 2017-1

	N	Média (DP) T1	Média (DP) T2	Média (DP) T3	Média (DP) T4
2015-2	4471	3,01 (0,95)	2,77 (1,22)	2,79 (1,03)	2,93 (1,11)
2016-1	4603	2,91 (1,20)	2,92 (1,07)	2,85 (1,18)	2,56 (1,04)
2016-2	4729	2,77 (0,99)	2,88 (0,87)	3,06 (1,05)	2,56 (1,05)
2017-1	3968	3,61 (0,96)	3,05 (1,02)	3,02 (0,90)	3,01 (0,90)

Fonte: Adaptado de Schlatter *et al.* (2021).

Isso mostra que a análise de características específicas dos textos, como o TTR, não é suficiente para formular conclusões precisas sobre a dificuldade do texto. Esse cenário é particularmente válido quando se trata de investigações que envolvem a avaliação de leitura integrada à produção textual. Conforme já dissemos, isso ocorre porque não basta apenas ser capaz de ler o texto; é imprescindível realizar a leitura prevista pelo enunciado e alcançar a interpretação preferível, não apenas outras interpretações possíveis, para reagir ao texto de acordo com as expectativas estabelecidas e cumprir plenamente os objetivos da tarefa.

Além disso, essa análise também corrobora alguns pressupostos teóricos discutidos na revisão bibliográfica: um enunciado claro é fundamental para o sucesso da tarefa integrada (BROWN, 2004; NAGASAWA, 2018; VICENTINI, 2019; DELL’ISOLA; PORDEUS, 2021; LOBO, 2021; NAGASAWA; SCHLATTER, 2021); o gênero discursivo influencia na

complexidade textual (EGBERT, 2017; GAZZOLA *et al.*, 2019; BIBER *et al.*, 2021c; BIBER *et al.*, 2021); a dificuldade do texto está diretamente relacionada tanto à tarefa quanto ao leitor (GOMES, 2009; SPENCER *et al.*, 2019); medidas linguísticas devem ser analisadas de acordo com o construto teórico do exame (EGBERT, 2017; XI, 2017); os textos não verbais devem colaborar na construção de sentidos (LINDE, 2011 *apud* LACERDA; FARBIARZ, 2018; SERAFINI, 2012; SIRIANNI, 2020); e a predição da complexidade da tarefa só é possível de ser realizada perante testagem (GOMES, 2009; KREMMEL *et al.*, 2021; PLAKANS, 2021; SCHLATTER *et al.*, 2021).

Por fim, cabe dizer que ambiguidades e incongruências como as observadas na tarefa ilustrada acima podem ser identificadas antes da aplicação, se pré-testadas, de modo a evitar ao máximo a presença de variáveis externas à avaliação pretendida. Com base no exposto, ressaltamos a importância de uma abordagem cuidadosa e padronizada na elaboração e revisão das tarefas integradas, assegurando que o conteúdo e a complexidade das provas estejam de acordo com o planejado em todas as edições.

7.1.2 Composição das tarefas

Até este ponto, delineamos uma visão geral dos *corpora*, analisando algumas medidas de tendência central e dispersão de palavras para descrever a composição dos conjuntos de textos e características específicas das tarefas do Celpe-Bras. Agora, voltamos nossa atenção para as características das edições do exame Celpe-Bras, buscando um panorama das tarefas já aplicadas por meio da observação das variáveis referentes aos *headers* do *corpus* de estudo. Vamos começar examinando os dados relativos às aplicações do exame Celpe-Bras.

Conforme mencionado anteriormente, a aplicação do Celpe-Bras tradicionalmente ocorre bianualmente. Em 2008 e 2016, houve a realização de reaplicação das provas, resultando em três edições por ano. Além disso, existem dois anos recentes em que houve somente uma aplicação do exame. Esse dado sugere a possibilidade de que alguma mudança organizacional interna tenha ocorrido em 2018, levando à interrupção do padrão histórico de aplicação semestral do exame Celpe-Bras.

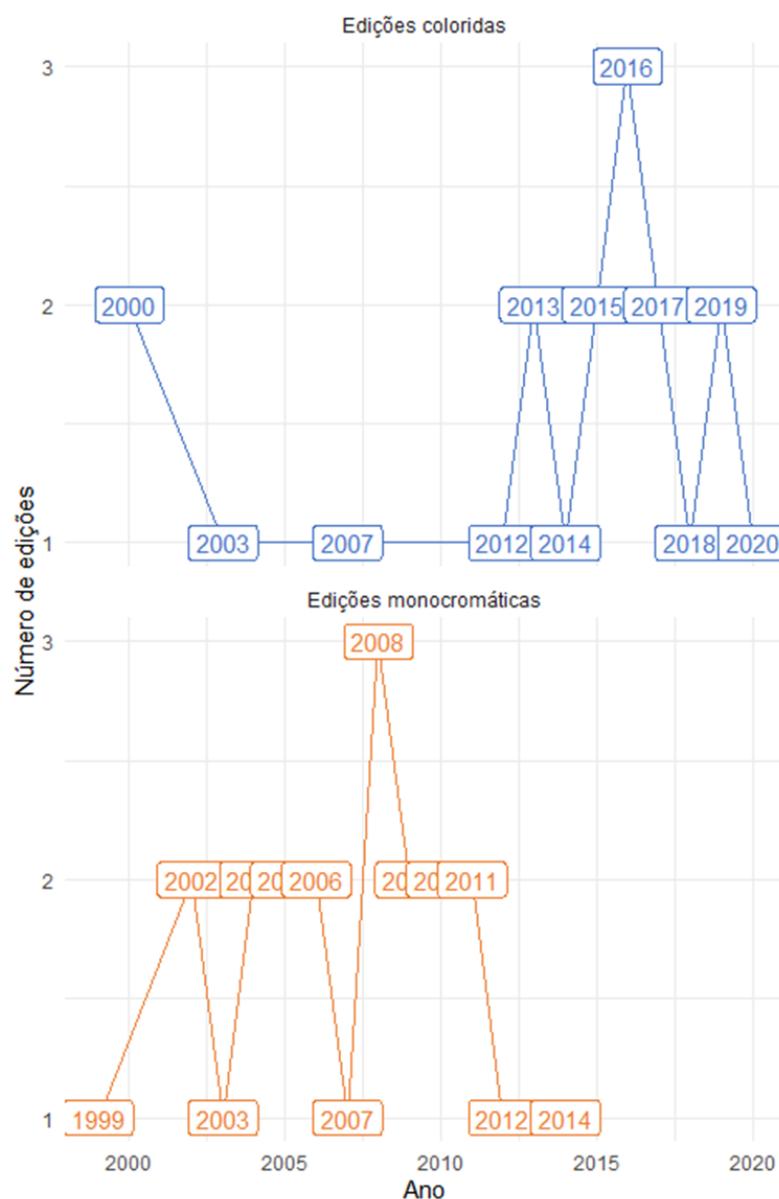
No que diz respeito à composição das 79 tarefas que analisamos, 71 delas utilizaram apenas um texto de insumo (90%), enquanto 8 tarefas fizeram uso de dois textos como material de leitura para a mesma tarefa (10%), sendo que a última vez que houve dois textos foi em 2014.

O uso de textos não verbais foi observado em 61 tarefas (77%), o que significa que 18 tarefas foram compostas exclusivamente por texto de insumo escrito (23%). Uma vez que as especificações afirmam que nas tarefas 3 e 4 são avaliadas a “leitura multimodal e produção escrita” (BRASIL, 2020, p. 58), se espera que ao menos uma delas seja composta por texto verbal e texto não verbal. Essa característica, portanto, é uma variável a ser controlada em cada edição, buscando uma padronização no uso de textos não verbais, que, como discutido anteriormente, são textos que influenciam na construção de sentidos (LINDE, 2011 *apud* LACERDA; FARBIARZ, 2018; SERAFINI, 2012; SIRIANNI, 2020).

Além disso, 41 tarefas utilizaram textos originais completos (52%), 30 tarefas utilizaram fragmento dos textos, ou seja, adaptaram de alguma forma o texto original (38%), e 8 tarefas usaram texto de insumo que não foi possível ser resgatado na internet, por serem textos de tarefas mais antigas, impossibilitando a verificação de adaptações no material (10%).

Em relação ao uso de cores na impressão das provas, além da capa do caderno de questões, das 41 edições avaliadas, 22 delas foram impressas em escala de cinza, ou seja, monocromáticas (54%), e 19 edições utilizaram impressão colorida em pelo menos uma das tarefas que integram leitura e escrita (46%).

Gráfico 6 – Número de edições em impressão colorida e monocromática



Fonte: Elaborado pela autora.

Apesar do número total de observações indicar que historicamente mais edições foram impressas em escala de cinza, o gráfico destaca que as tarefas mais recentes, desde o ano 2014, foram impressas em multicores. Esse dado é positivo, considerando que, como discutido previamente, aspectos tipográficos do texto estão ligados à legibilidade, um fator que pode influenciar o desempenho na leitura (LOURENÇO, 2016; SOUSA; HÜBNER, 2015).

Deslocamos, a partir daqui, o nosso foco para os textos selecionados como texto de insumo para compor as 41 edições de nosso *corpus* de estudo. Vamos analisar a tabela de frequência dos gêneros discursivos dos textos de insumo presentes no *corpus* Celpe, lembrando

que nesse caso o *n* aumenta para 87, já que estamos considerando separadamente os dois gêneros usados nas 8 tarefas que empregam dois textos diferentes para leitura.

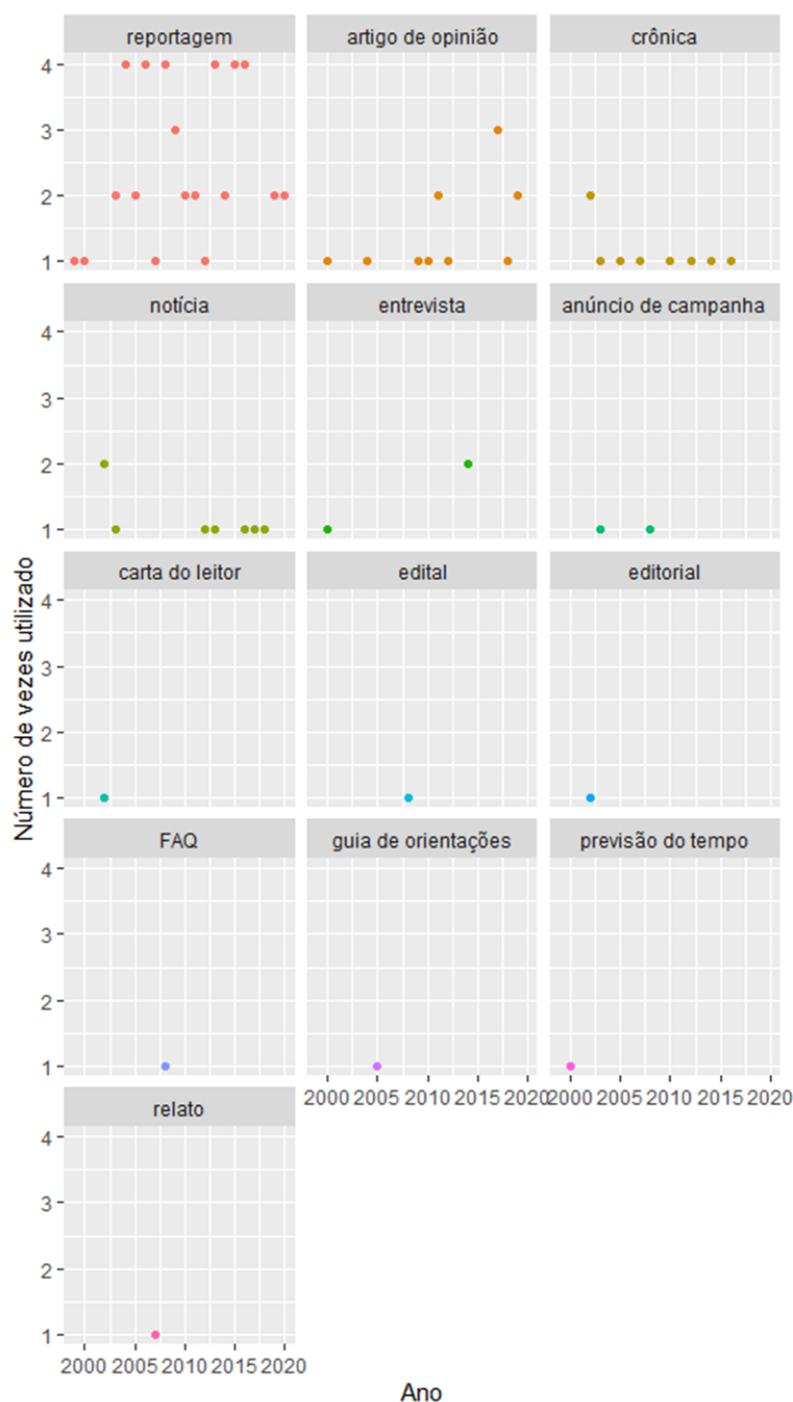
Tabela 10 – Total de gêneros de leitura do corpus Celpe

Gênero de Leitura	Total	n	Porcentagem
1 reportagem	45	87	52%
2 artigo de opinião	13	87	15%
3 crônica	9	87	10%
4 notícia	8	87	9%
5 entrevista	3	87	3%
6 anúncio de campanha	2	87	2%
7 FAQ	1	87	1%
8 carta do leitor	1	87	1%
9 edital	1	87	1%
10 editorial	1	87	1%
11 guia de orientações	1	87	1%
12 previsão do tempo	1	87	1%
13 relato	1	87	1%

Fonte: Elaborado pela autora.

Os dados evidenciam que o *corpus* Celpe abrange 13 distintos gêneros discursivos, sendo que 6 deles ocorreram mais de uma vez, enquanto outros 7 foram usados apenas em uma tarefa. Além disso, como previamente argumentado, os dados confirmam a tendência de que o *corpus* Celpe se aproxima dos *corpora* jornalísticos, pois sua composição é majoritariamente constituída de gêneros encontrados em veículos de comunicação. Os gêneros mais frequentes incluem a reportagem, o artigo de opinião, a crônica e a notícia. A maioria dos gêneros de leitura utilizados apenas uma vez aparecem em tarefas mais antigas, o que sugere uma mudança nas orientações de seleção de textos de insumo em torno do ano 2010, como é possível verificar na figura a seguir.

Figura 25 – Gêneros de leitura por ano do corpus Celpe



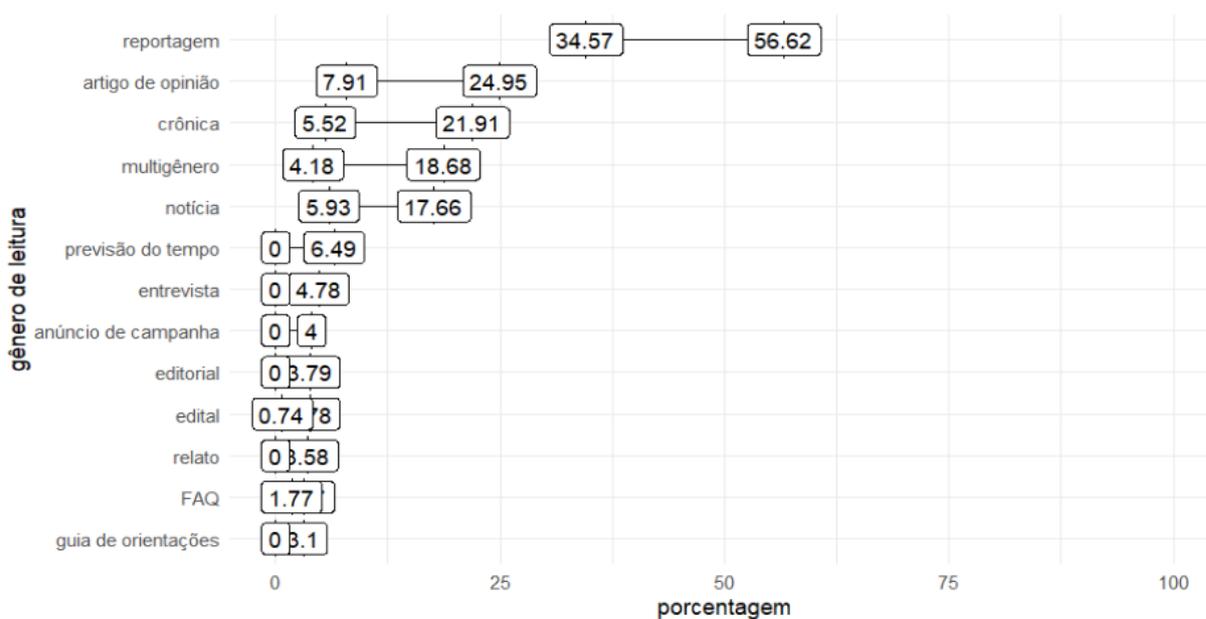
Fonte: Elaborado pela autora.

A figura mostra que a variação nos gêneros de leitura é mais proeminente nas edições anteriores a 2010. A partir desse ponto temporal, com uma exceção em 2014-2 (entrevistas), todas as tarefas empregaram apenas quatro tipos de gêneros discursivos como textos de insumo: “reportagem”, “artigo de opinião”, “crônica” ou “notícia”. Além disso, vemos claramente a

predominância do uso de reportagens, que representam mais da metade do material de leitura utilizado nas tarefas (52%).

A fim de verificar a significância estatística das diferenças observadas, foi conduzido o teste *Bootstrapping*. O resultado desse modelo é ilustrado no gráfico a seguir.

Gráfico 7 – Bootstrapping da distribuição de gêneros de leitura no corpus Celpe



Fonte: Elaborado pela autora.

Verificamos forte indicação de significância estatística na diferença entre o gênero de leitura “reportagem” e os demais gêneros. O gráfico mostra a não sobreposição dos intervalos de “reportagem” com os outros gêneros, indicando, com 95% de confiança, que a proporção de “reportagem” é o gênero selecionado como texto de insumo entre 35% e 57% das vezes. Essa porcentagem é estatisticamente superior e se distancia de todas as outras.

Os resultados de nossas análises apontam, assim, que os gêneros discursivos dos textos de insumo utilizados nas tarefas 3 e 4 já aplicadas são bastante reduzidos se comparados aos explicitados nas especificações para a elaboração de tarefas e no Documento Base, conforme apresentado no quadro a seguir.

Quadro 17 – Gêneros de leitura nas especificações e no corpus Celpe

Especificações para a elaboração de tarefas do exame Celpe-Bras - Parte Escrita	Especificações no Documento Base	Corpus Celpe
Editorial, notícia, entrevista, reportagem, anúncio classificado, publicidade, cartas de leitores, horóscopo, cartuns, quadrinhos, etc.; crônica, conto, poema, texto didático, receita etc.; textos de panfletos, cartazes, avisos, placas de trânsito; textos de telegramas, cartas, bilhetes, e-mails, cartões-postais; textos de diários, agendas, notas, listagens, resenhas, relatórios, currículos, biografias; textos de documentos, formulários, questionários, instruções; de mapas, roteiros, quadro de horários, calendários, programas, cardápios, recibos; textos de dicionários, catálogos, listas telefônicas, letras de música, legendas de filme etc.	Reportagens, entrevistas, artigos de opinião, editoriais, crônicas, colunas assinadas, notícias, informativos, panfletos, propagandas, cartuns, quadrinhos, anúncios, poesias, resenhas, relatórios, currículos, biografias, editais, formulários, questionários, instruções, mapas, roteiros, cardápios, letras de canções, manifestos, cartas do leitor, entre outros.	Predominantemente reportagens. Artigos de opinião, crônicas, notícias, entrevistas, anúncios, FAQs, cartas do leitor, editais, editoriais, guias, previsões do tempo, relatos.

Fonte: Elaborado pela autora.

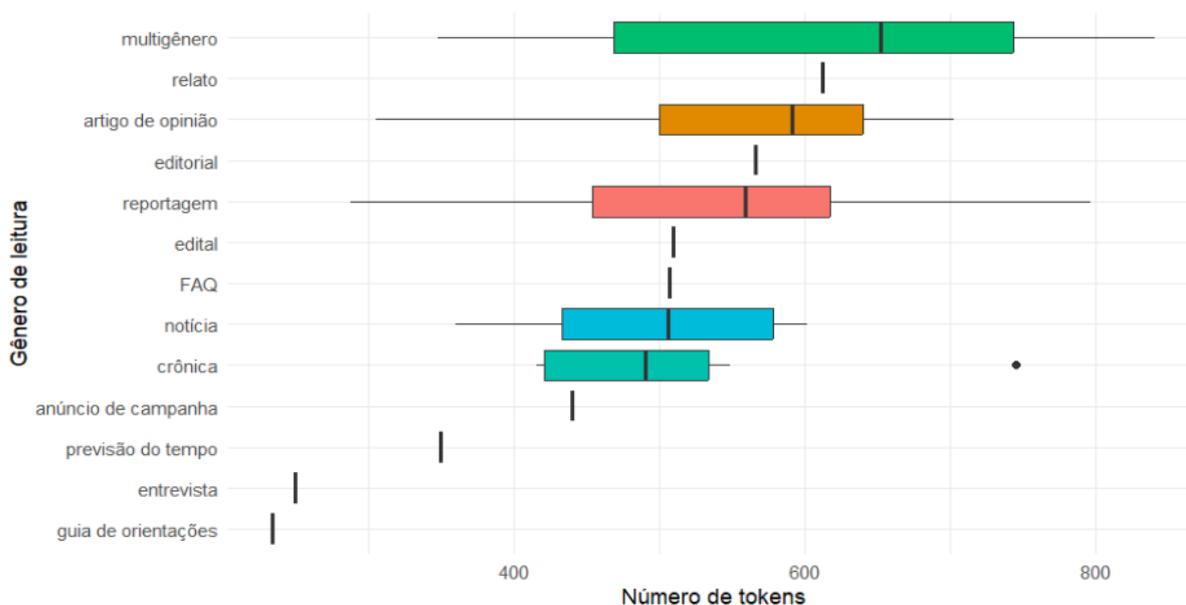
As evidências sugerem que as especificações para a elaboração de tarefas do exame não têm sido seguidas em relação à variedade de gêneros discursivos dos textos de insumo para a leitura, especialmente desde o ano de 2010. A primeira edição do manual do candidato, publicado em 2002, já incluía uma ampla lista de gêneros do discurso. É importante lembrar que, uma vez estabelecido o domínio de uso da linguagem que o exame visa avaliar, a representatividade do conteúdo do teste é uma das evidências que respaldam o argumento de validade (FULCHER; OWEN, 2016). Mesmo que a literatura reconheça a dificuldade de elaborar tarefas que abranjam todos os domínios de uso da língua-alvo (BROWN, 2004; HUGHES; HUGHES, 2020; KANE, 2021), cabe aos responsáveis pelo exame tomar decisões sobre o que fará parte dos instrumentos de avaliação e descrevê-los claramente nas especificações (DOUGLAS, 2000; WEIGLE, 2004; GREEN; FULCHER, 2020; JIN, 2021). Conforme destacado por Shin (2021), não cabe à equipe de elaboração das tarefas selecionar e decidir o que e como os examinandos serão avaliados; ela deve seguir rigorosamente as especificações do exame.

Cabe observar ainda que a decisão por abranger um conjunto menor de gêneros poderia contribuir para a validade do exame se esse repertório representar de modo mais coerente o

domínio de uso da linguagem que o exame busca avaliar. No caso do Celpe-Bras, isso implicaria uma diminuição do repertório de gêneros do discurso listados nas especificações. Da mesma forma, um repertório menor poderia ser positivo tanto para o processo de elaboração quanto para a preparação para o exame. De todo modo, quaisquer que sejam as decisões quanto à abrangência de gêneros representativos do conteúdo que se quer testar, é fundamental que elas sejam visibilizadas aos usuários do exame nas especificações e que sejam efetivamente colocadas em prática pela equipe de elaboração, buscando, assim, contribuir para o argumento de validade do exame.

Passamos para a análise de variância no número de palavras por gênero de leitura do *corpus* Celpe. Para a análise, as edições que fazem uso de dois textos foram categorizadas como “multigênero”, já que é prevista a leitura integral dos dois textos para o cumprimento da tarefa. Abaixo, apresentamos a distribuição do número de palavras por gênero de leitura.

Gráfico 8 – Número de palavras por gênero de leitura do corpus Celpe



Fonte: Elaborado pela autora.

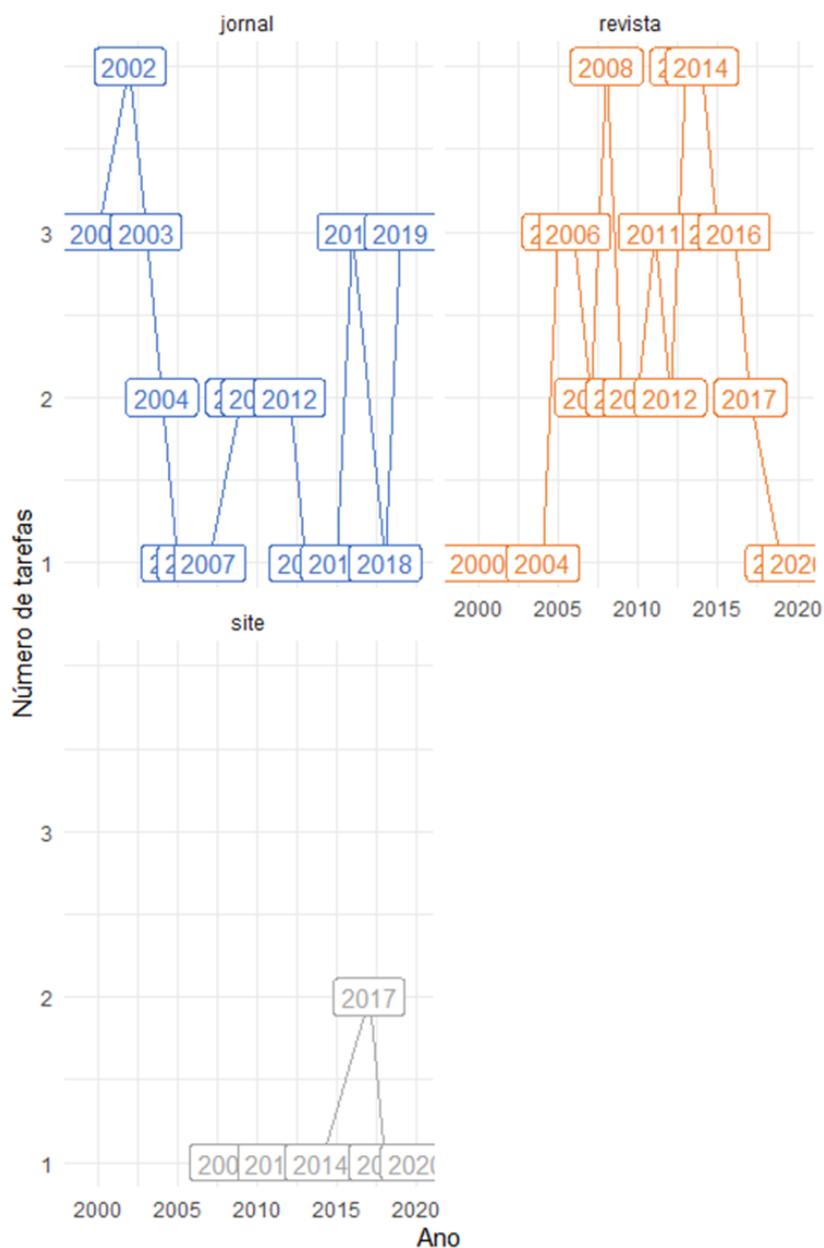
O *boxplot* mostra claramente que há variação no número de palavras por gênero, indicando que a extensão do texto pode divergir conforme o gênero discursivo selecionado como texto de insumo. Notadamente, a categoria “multigênero”, referente às oito tarefas que utilizam dois textos distintos, apresenta a mediana mais alta (652), a distribuição mais assimétrica e a maior amplitude interquartil, revelando observações com alta variação no número de palavras por texto. Retomando as observações detectadas nos textos considerados

como *outliers*, das seis tarefas excluídas por apresentar valor maior ao limite do fio de bigode superior, três delas usam dois textos distintos como texto de insumo.

O mesmo comportamento de ampla dispersão de dados é constatado nos gêneros “artigo de opinião” e “reportagem”, sendo que esta última exhibe a maior variação absoluta. Já o gênero “notícia” apresenta menor variabilidade e uma distribuição mais simétrica dos dados. Por fim, vemos que o gênero “crônica” apresenta a menor variância entre os gêneros de leitura: das 9 tarefas que utilizaram esse gênero, 8 tiveram um número relativamente semelhante de palavras, com apenas uma delas apresentando valor discrepante. Ao desconsiderarmos esse valor atípico, a variação do número de palavras em torno da mediana é a menor, o que nos permite concluir que o gênero “crônica” é o mais padronizado quanto ao número de palavras nos gêneros dos textos selecionados. Embora visualmente seja possível observar diferença no número de palavras por gênero, o teste de regressão revela que a variável “gênero de leitura” também não tem força para explicar o número de palavras por texto (R^2 Ajustado = 0.09). No entanto, “entrevista” e “guia de orientação” apresentaram diferença estatisticamente significativa no número de palavras quando comparados com os demais gêneros ($p = <0.01$).

Quanto à origem dos textos de insumo do *corpus* Celpe, 44 deles são provenientes de revistas, representando 51% do total, enquanto 33 têm origem em jornais, compreendendo 38%. Além desses dois suportes predominantes, há três textos cuja origem não foi identificada (3%), um texto de um site do governo (1%) e 6 textos de outros tipos de sites (7%).

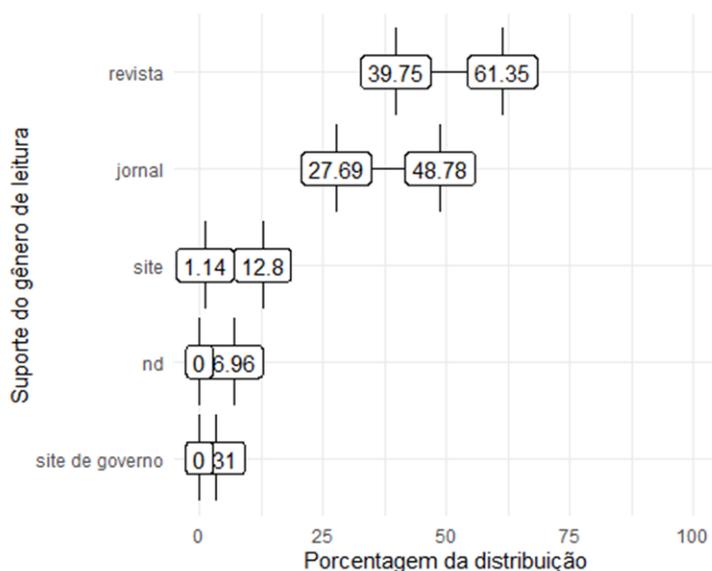
Gráfico 9 – Número de suporte dos textos de leitura por ano do corpus Celpe



Fonte: Elaborado pela autora.

O gráfico acima destaca que a utilização de textos provenientes de sites começou a ser adotada a partir de 2008; no entanto, sua incidência ainda é insignificante em comparação com as outras duas fontes. As proporções já deixam claro que jornais e revistas são as principais fontes escolhidas dos textos de insumo. Entretanto, para verificar a significância dessa diferença observada, realizamos novamente o teste não paramétrico de 1.000 reamostragens com reposição, *Bootstrapping*.

Gráfico 10 – Bootstrapping de distribuição de suportes de gênero de leitura do corpus Celpe



Fonte: Elaborado pela autora.

O modelo mostra uma diferença estatisticamente significativa entre os suportes “revista” e “jornal” e os demais suportes. As tarefas 3 e 4 são constituídas por textos provenientes de revistas em frequências variando de 40% a 61%, enquanto os textos oriundos de jornais correspondem a uma proporção entre 28% e 49% das tarefas aplicadas. Esse resultado nos permite concluir que, para a construção das tarefas de leitura integrada a produção textual do exame, revistas e jornais são as principais fontes escolhidas para seleção dos materiais de leitura. Novamente, podemos constatar que a abrangência de fontes não está alinhada com as informações contidas nas especificações do exame, conforme mostra o quadro a seguir. Uma hipótese plausível para explicar esse resultado pode ser a facilidade de acesso às fontes por parte da equipe de elaboração e a familiaridade da equipe com essas mídias específicas.

Quadro 18 – Fontes dos textos de insumo nas especificações e no corpus Celpe

Especificações para a elaboração de tarefas do exame Celpe-Bras - Parte Escrita	Especificações no Documento Base	Corpus Celpe
Textos de periódicos (jornais e revistas), Textos de livros, Textos de panfletos, cartazes, avisos, placas de trânsito; Textos de telegramas, cartas, bilhetes, e-mails, cartões-postais; Textos de diários, agendas, notas, listagens, resenhas, relatórios, currículos, biografias;	Jornais, revistas, panfletos, informativos, livros, redes sociais, mídia eletrônica em geral, entre outras.	Revistas, jornais e websites.

Textos de documentos, formulários, questionários, instruções; de mapas, roteiros, quadro de horários, calendários, programas, cardápios, recibos; Textos de dicionários, catálogos, listas telefônicas, letras de música, legendas de filme etc.		
---	--	--

Fonte: Elaborado pela autora.

Mais uma vez, levanta-se a questão crucial sobre a necessidade de as especificações do exame refletirem o que está sendo praticado ou que as especificações sejam seguidas de modo mais fidedigno no processo de elaboração das tarefas. As especificações, que são disponibilizadas publicamente aos usuários do exame (Documento Base) e destinadas a guiar a criação de edições do exame que sejam equivalentes e representativas do seu construto, parecem não estar sendo rigorosamente seguidas durante o processo de elaboração das tarefas.

Quanto às diferentes fontes dos textos de leitura, foi possível identificar 35 fontes distintas, sendo que em quatro tarefas não foi possível identificar a fonte (nd).

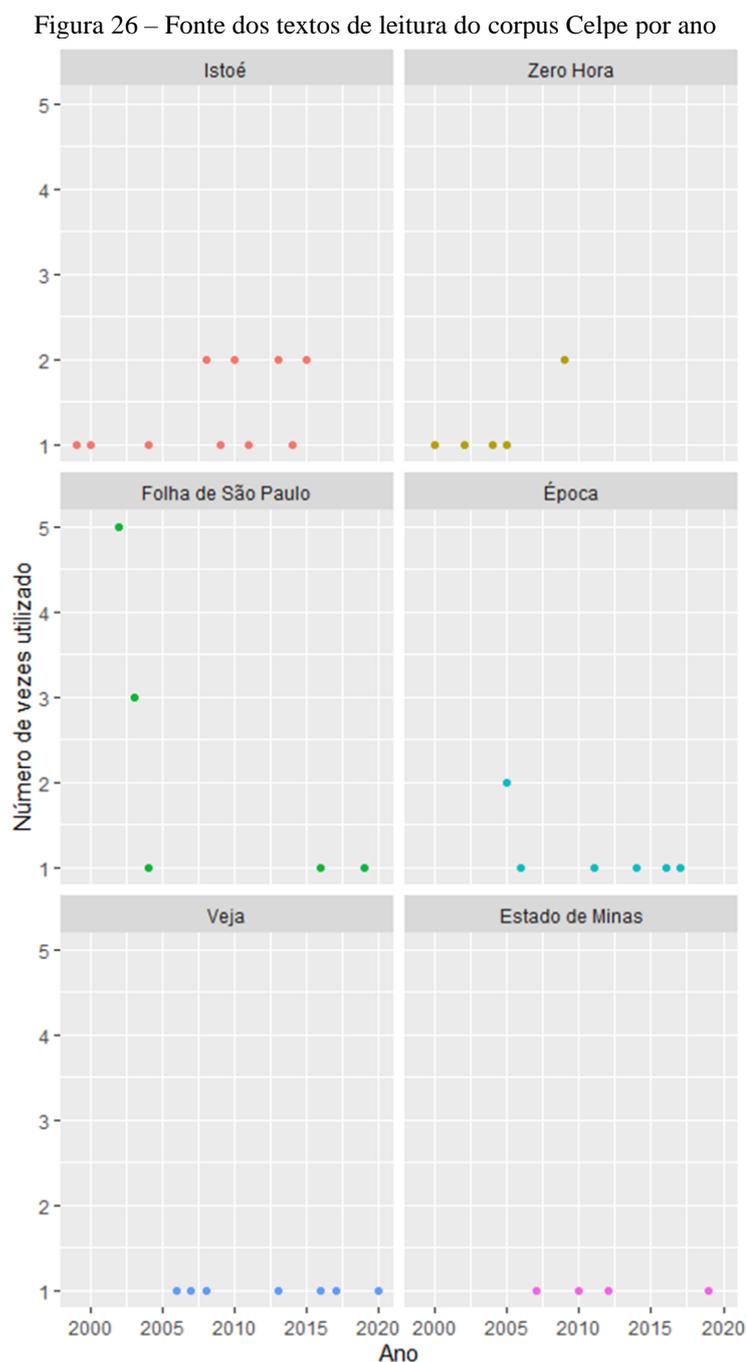
Tabela 11 – Fonte dos textos de leitura do corpus Celpe

Fonte	n	total	porcentagem
1 Istoé	14	87	16.09
2 Folha de São Paulo	11	87	12.64
3 Época	7	87	8.05
4 Veja	7	87	8.05
5 Zero Hora	6	87	6.90
6 Estado de Minas	4	87	4.60
7 nd	4	87	4.60
8 Gol	3	87	3.45
9 Almanaque de Cultura Popular	2	87	2.30
10 Correio Braziliense	2	87	2.30
11 Estadão	2	87	2.30
12 A escotilha	1	87	1.15
13 Araraquara	1	87	1.15
14 Arte na rede	1	87	1.15
15 Bons Fluidos	1	87	1.15
16 ComCiência	1	87	1.15
17 Correio Cidadania	1	87	1.15
18 Correio Web	1	87	1.15
19 CPRM	1	87	1.15
20 Free the essence	1	87	1.15
21 Interrogação	1	87	1.15
22 Jornal do Comércio	1	87	1.15
23 Jornal Em Dia	1	87	1.15
24 National Geographic	1	87	1.15
25 Nova Escola	1	87	1.15
26 O Globo	1	87	1.15
27 Pampulha	1	87	1.15
28 Pensamento Verde	1	87	1.15
29 Pequenas Empresas & Grandes Negócios	1	87	1.15
30 Planeta	1	87	1.15
31 Revista Galileu	1	87	1.15
32 Revista Superinteressante	1	87	1.15
33 Secretaria Municipal de Saúde	1	87	1.15
34 UFPA	1	87	1.15
35 Viagem e Turismo	1	87	1.15
36 Você S.A.	1	87	1.15

Fonte: Elaborado pela autora.

Os dados confirmam jornais e revistas como as fontes primárias dos textos de insumo. Os jornais *Folha de São Paulo* e *Zero Hora*, que integram o *corpus* PorSimples, estão entre as seis fontes mais utilizadas, juntamente das revistas *IstoÉ*, *Época*, *Veja* e o jornal *Estado de Minas*. Como mencionado anteriormente, de acordo com o relatório do Grupo Mídia, esses jornais e revistas representam meios de comunicação de grande circulação em todo o país, tanto em suas versões impressas quanto digitais, sendo predominantemente consumidos por leitores das classes A e B. Ressalta-se ainda que o acesso a parte do conteúdo desses veículos de comunicação não é gratuito, e seus custos são altos quando comparados aos valores do

jornalismo popular, acessíveis para leitores pertencentes aos estratos socioeconômicos C, D e E. É importante também enfatizar que as fontes *Diário Gaúcho* e *Massa!*, que integram o *corpus* PorPopular, não aparecem como fontes de texto de insumo no Celpe-Bras. Vejamos como se dá a distribuição das fontes mais utilizadas para os textos de insumo durante os anos de aplicação do Celpe-Bras.



Fonte: Elaborado pela autora.

Analisando a distribuição das fontes dos textos de insumo ao longo dos anos de aplicação do Celpe-Bras, observa-se que, embora tenham sido frequentemente usados e estejam entre as seis fontes mais utilizadas, os jornais *Zero Hora* e *Folha de São Paulo* deixaram de ser fontes de texto de insumo nos últimos anos. Os dados mostram que o jornal *Zero Hora* foi utilizado pela última vez em 2009, nas duas edições daquele ano. Quanto à *Folha de São Paulo*, foi fonte de texto de insumo apenas em duas ocasiões desde 2004, apesar de ter sido recorrentemente utilizada em anos anteriores.

A análise da distribuição das outras fontes de insumo por ano de aplicação mostra um comportamento mais variável na última década. A mudança observada nos dados, que ocorreu aproximadamente a partir do ano de 2010, pode possivelmente refletir alterações na composição da equipe de elaboradores ou nas orientações relacionadas à elaboração das tarefas da PE. Essa possível mudança no processo de elaboração pode se relacionar a alteração do órgão do governo responsável pelo exame, que em 2009 passou a ser de responsabilidade do Inep. Além disso, em torno da mesma época, em 2012, foi publicado o Chamamento Público para a formação do Banco de Colaboradores, além de uma Portaria que instituiu um novo grupo de especialistas para a CTC¹³⁹.

Com base nesses dados, podemos inferir que, mais recentemente, a elaboração das tarefas do Celpe-Bras tem recorrido a textos de insumo de fontes mais variadas, não se limitando à utilização exclusiva dos “jornais de referência”, considerados prestigiosos e de credibilidade entre formadores de opinião. Ainda assim, ao analisar a lista de fontes utilizadas, é possível observar que elas também são majoritariamente veículos de comunicação não gratuitos voltados para leitores de maior escolaridade e poder aquisitivo.

Em síntese, a análise dos gêneros discursivos e das fontes das quais os textos são selecionados para compor as tarefas que integram leitura e escrita no Celpe-Bras, além de mostrar que o conteúdo de insumo das tarefas 3 e 4 aplicadas nas diferentes edições do exame é mais restrito do que descrito nas especificações, suscita indagações acerca de uma possível alteração no construto do exame e, conseqüentemente, no modo como é operacionalizado nas tarefas de leitura e escrita em torno do ano 2010.

¹³⁹Portaria de 27 de março de 2012. Disponível em: <https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2021/11/27-03-2012Portaria-de-nomeacao-de-Comissao-2012-compactado.pdf> Acesso em: 30 ago. 2023.

7.1.3 Categorias do Grupo Avalia

Considerando que o objetivo deste estudo é analisar o conteúdo das tarefas que integram leitura e escrita focalizando especialmente os textos de insumo e com o objetivo de verificar a consistência das tarefas 3 e 4 ao longo das edições, julgamos pertinente dar seguimento ao estudo de Schoffen *et al.* (2018). As autoras descreveram as características das tarefas, fornecendo a frequência de ocorrência das variáveis em números absolutos do total das tarefas por intervalos de cinco anos, bem como por tarefa de modo individualizado. Em nosso estudo, trazemos a frequência de ocorrências com intervalos de confiança estatística, contudo relativas ao conjunto das tarefas 3 e 4 (e não à totalidade das quatro tarefas). Além disso, a linguagem de programação R nos permitiu a manipulação das variáveis e a observação da relação das categorias com os textos de insumo. Todas as variáveis relacionadas às categorias usadas pelo grupo Avalia para descrever as tarefas foram investigadas em relação à sua variação de porcentagem nas tarefas, à sua distribuição por ano e em relação à frequência de *tokens* nos textos de insumos. Os resultados do *bootstrapping* não indicaram diferenças estatisticamente significativas em nenhuma das testagens realizadas. Apresentamos os resultados que, do nosso ponto de vista, trazem contribuições relevantes para descrever o que foi colocado em prática na elaboração de tarefas nas edições analisadas.

Iniciamos com as análises relativas às esferas de atuação das tarefas. As esferas de atuação predominantes são a *Jornalística* (35%) e a *Político-cidadã* (23%). Além disso, é interessante observar o comportamento das esferas *Pessoal* (5%) e *Comunitária* (6%), que apresentam uma mudança em torno do ano 2010, momento que coincide com o término da *Pessoal* e a introdução da *Comunitária*.

Figura 27 – Esfera de atuação da produção textual do corpus Celpe por ano

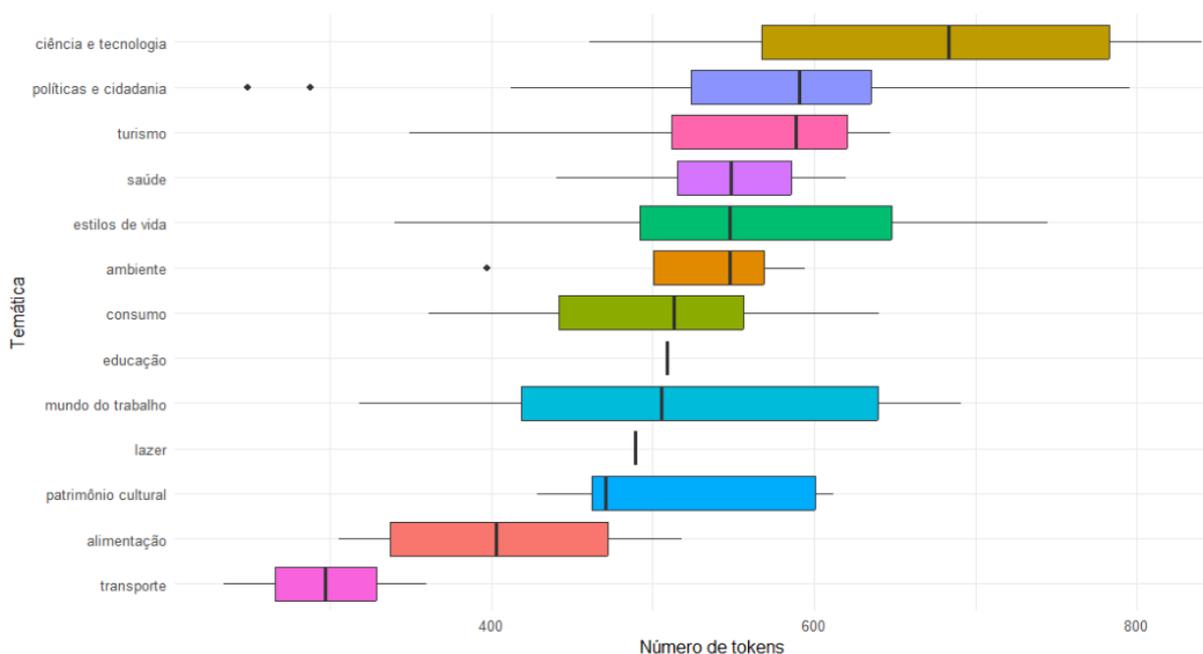


Fonte: Elaborado pela autora.

Nessa figura é possível observar que à medida que a esfera de atuação *Pessoal* foi deixando de ser utilizada, as esferas *Científico-educacional*, *Comunitária* e *Artístico-cultural* foram introduzidas. Essa transição é observada aproximadamente no mesmo período histórico das mudanças mencionadas anteriormente, em torno do ano 2010.

Em relação à temática das tarefas, se destacam *Estilos de vida* (27%) e *Políticas e cidadania* (16%). Vamos examinar a distribuição do número de palavras nos gêneros de leitura de acordo com as temáticas das tarefas.

Gráfico 11 – Número de palavras por temática do corpus Celpe

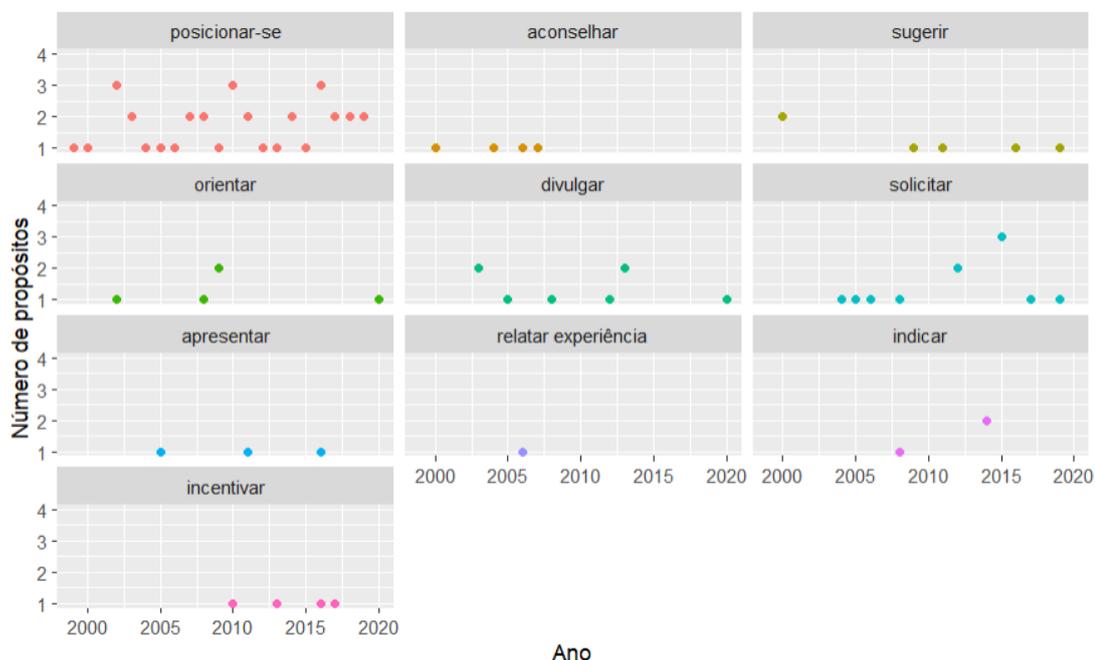


Fonte: Elaborado pela autora.

No gráfico de caixas, observamos uma dispersão ampla das medianas por temática, sugerindo que as temáticas parecem ter alguma influência no número de palavras por texto. Nesse contexto, a temática *Ciência e tecnologia* se destaca ao apresentar a mediana mais alta (683) e uma distribuição próxima à normal, apesar da variabilidade significativa. Vale ressaltar que o texto do *corpus* Celpe com o maior número de palavras (840) está classificado nessa temática.

Quanto aos propósitos da produção textual solicitados nas tarefas, as maiores porcentagens de ocorrências se concentram em *Posicionar-se* (43%) e *Solicitar* (14%). Ao analisarmos essa variável, observamos novamente uma mudança em torno do ano de 2010: o propósito *Aconselhar* (5%) deixou de ser utilizado, enquanto os propósitos *Incentivar* (5%) e *Indicar* (4%) foram introduzidos.

Figura 28 – Propósito da produção textual do corpus Celpe por ano

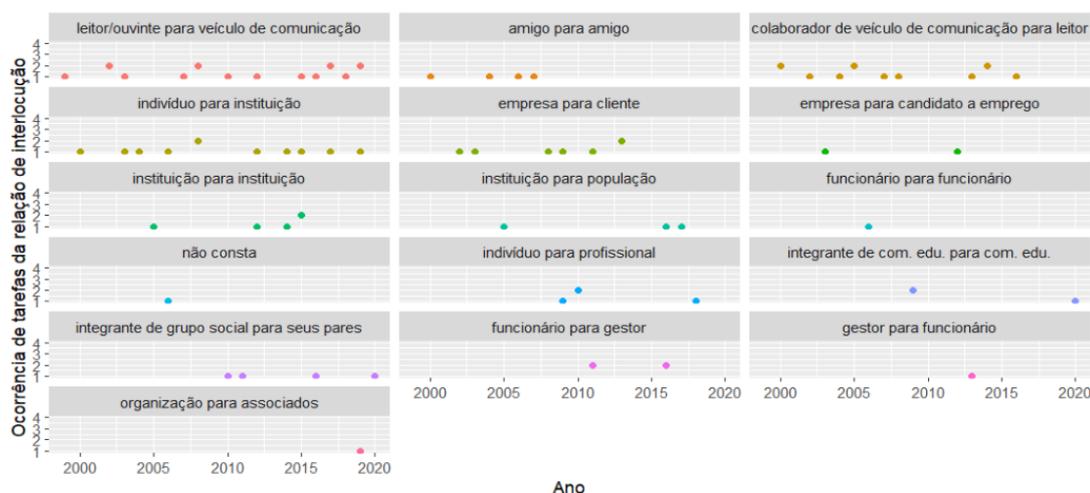


Fonte: Elaborado pela autora.

A imagem mostra a predominância do propósito *Posicionar-se*, cuja distribuição se mantém uniforme ao longo dos anos de aplicação do exame. No entanto, é importante lembrar que o modelo *bootstrapping* não revelou diferenças estatisticamente significativas entre *Posicionar-se* e os demais propósitos de escrita solicitados nas tarefas do exame.

No que diz respeito às relações de interlocução predominantes, destacam-se duas categorias relacionadas aos veículos de comunicação: *Leitor/ouvinte para veículo de comunicação* (20%) e *Colaborador de veículo de comunicação para leitor* (15%). Considerando que os gêneros discursivos se inter-relacionam com outros gêneros dentro de esferas de atuação específicas, é notável que essas interlocuções estejam em sintonia com o uso predominante de gêneros jornalísticos como texto de insumo. Mais uma vez, observamos uma mudança nessa variável em consonância com a alteração na esfera de atuação *Pessoal*. Nesse contexto, a relação *Amigo para amigo* (5%) deixou de existir em 2007, quando a quarta tarefa com essa relação foi utilizada, e novas relações foram incluídas, como pode ser observado na figura a seguir.

Figura 29 – Relação de interlocução da produção textual do corpus Celpe por ano



Fonte: Elaborado pela autora.

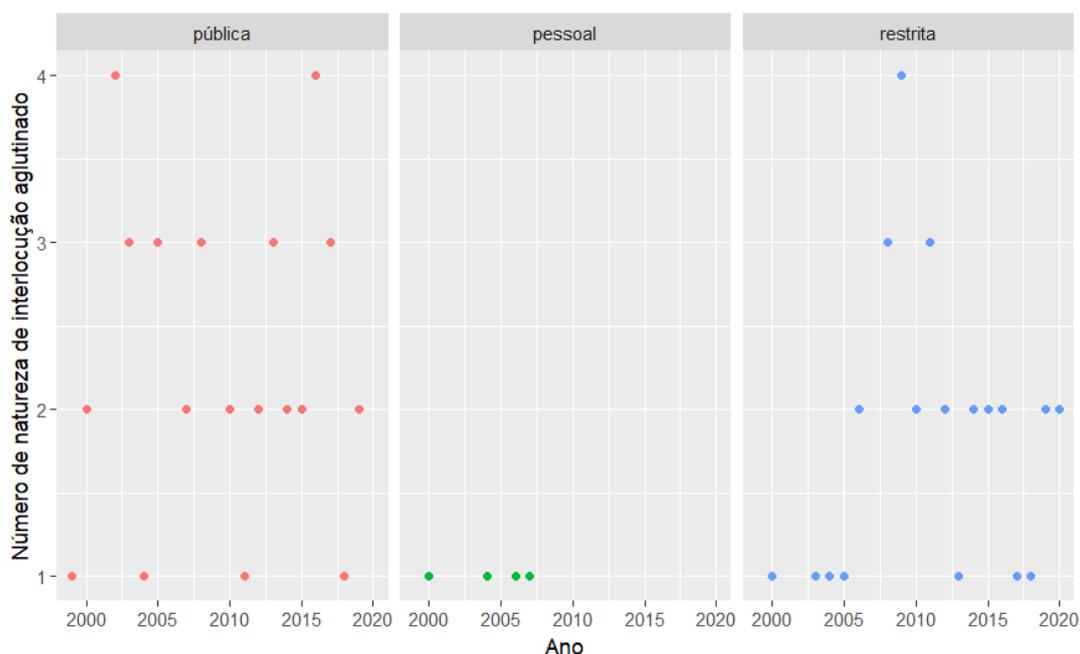
A representação visual mostra que, apesar da predominância das relações de interlocução de natureza jornalística, a partir do período próximo ao ano de 2010, emergiram novas relações, tais como *Indivíduo para profissional*, *Integrante de comunidade educacional para comunidade educacional*, *Integrante de grupo social para seus pares*, *Funcionário para gestor*, *Gestor para funcionário* e *Organização para associados*. Essas novas relações parecem estar vinculadas às esferas de atuação *Comunitária* e *Profissional*.

Vale ressaltar que as três tarefas categorizadas como *Integrante de comunidade educacional para comunidade educacional* tratam de interlocuções projetadas em contexto escolar de Ensino Básico e não de Ensino Superior. Existe uma tarefa 3, utilizada na reaplicação da edição 2008-2, na qual é utilizado um edital de mobilidade acadêmica de uma universidade pública brasileira como texto de leitura. O candidato deve reagir a esse texto por meio de uma carta de apresentação pessoal. Essa tarefa está classificada na categoria de relação de interlocução *Indivíduo para instituição*. É relevante notar que, dentre as 79 tarefas incluídas em nossa análise, apenas esta simula uma situação de comunicação em um contexto universitário brasileiro.

No que tange à variável *Natureza da interlocução*, as ocorrências mais proeminentes correspondem a *Restrita, assimétrica ascendente* (33%) e *Pública, assimétrica descendente* (25%). A primeira categoria engloba tarefas como, por exemplo, aquelas relacionadas à esfera *Político-cidadã*, em que um indivíduo escreve uma carta a uma instituição solicitando apoio financeiro para um projeto. A segunda categoria reúne tarefas tal como as da esfera *Jornalística*, em que um colaborador de um veículo de comunicação é convidado a redigir um texto para os leitores, expressando sua posição sobre uma temática específica. Nesses contextos, a instituição

é considerada um interlocutor de natureza restrita, enquanto o jornal, por ser de acesso público, caracteriza-se como um interlocutor de natureza pública. Para proporcionar uma perspectiva mais abrangente dessa categoria, agrupamos as tarefas de natureza pública e as de natureza restrita.

Figura 30 – Natureza da interlocução da produção textual do corpus Celpe por ano

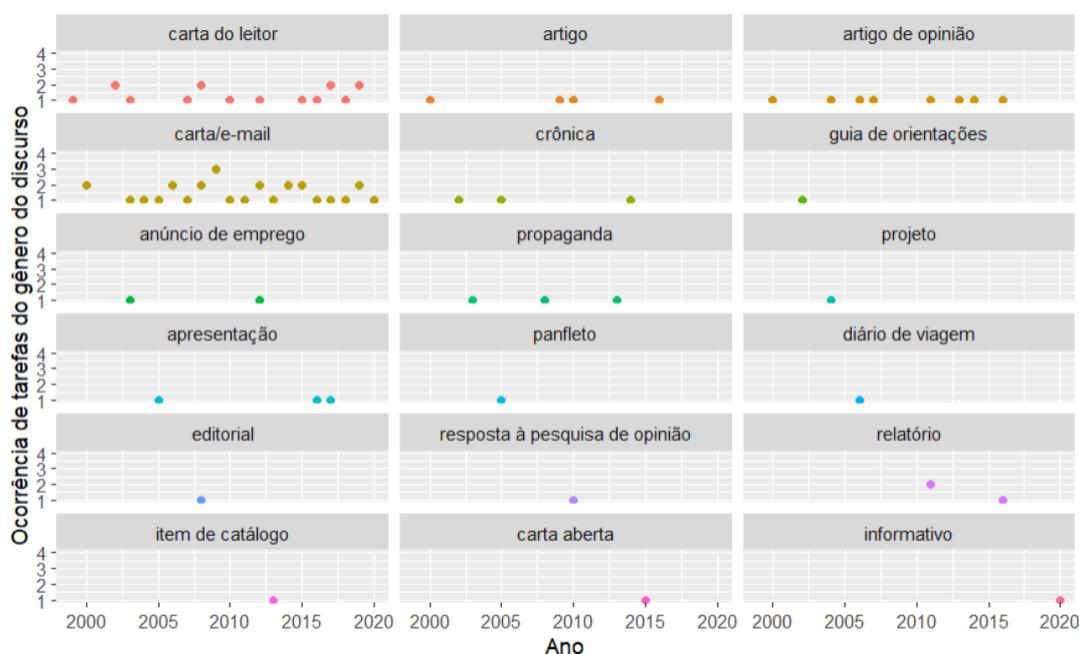


Fonte: Elaborado pela autora.

A figura mostra que as tarefas com interlocução de natureza pessoal deixaram de existir em 2007. As tarefas de interlocução de natureza pública são as mais frequentes (52%) e exibem um padrão com dois picos de alta frequência seguidos por decréscimos. O primeiro pico ocorreu entre os anos 2000 e 2005, enquanto o segundo pico ocorreu entre os anos 2010 e 2020, aproximadamente. Em contrapartida, as tarefas com interlocução de natureza restrita (41%) apresentam um padrão de alta frequência entre os anos 2005 e 2015, aproximadamente.

A variável *Gênero do discurso* de produção possui *carta/e-mail* (35%) e *carta do leitor* (20%) como os gêneros mais recorrentes. A próxima figura ilustra a distribuição dos gêneros de produção por ano de aplicação do Celpe-Bras.

Figura 31 – Gênero do discurso da produção textual do corpus Celpe por ano

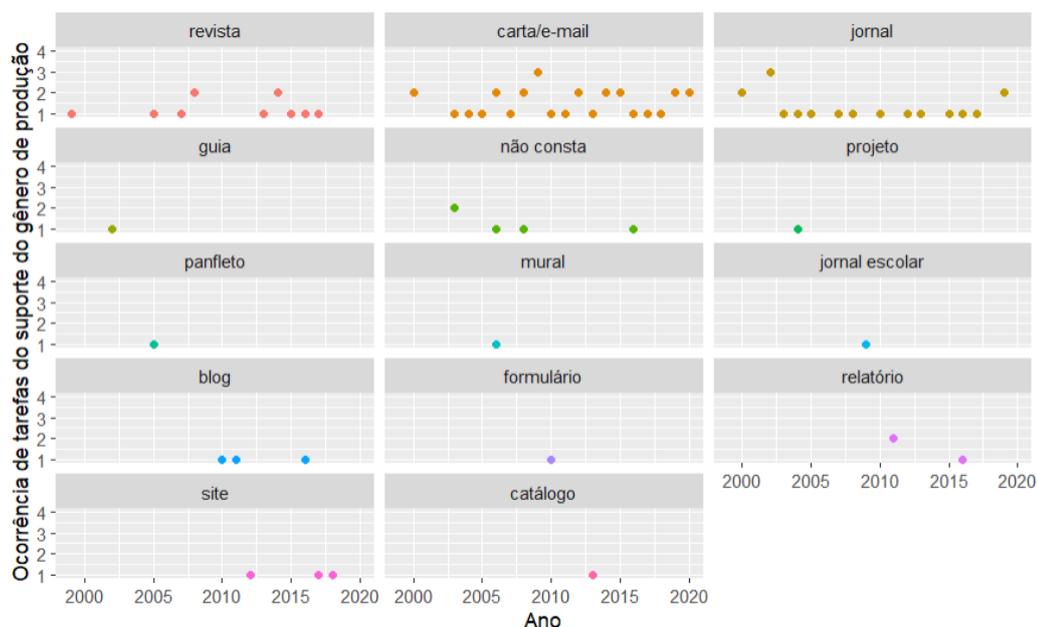


Fonte: Elaborado pela autora.

Novamente, a clara predominância dos dois gêneros – *carta do leitor* e *carta/e-mail* – em comparação aos demais é evidente, e as distribuições se mantêm uniformes ao longo dos anos de aplicação do exame. Embora se distingam em relação a propósitos, natureza da interlocução e suporte, os dois gêneros de produção mais frequentes, que juntos representam 55% das tarefas, compartilham uma formatação textual semelhante, ambos sendo cartas.

Por fim, na variável *Suporte*, temos a predominância de *carta/e-mail* (37%), seguido de *jornal* (23%) e *revista* (14%), o que reforça os resultados anteriores que indicam uma tendência em direção a interlocuções da esfera jornalística. É importante notar que nos dados do grupo Avalia, as tarefas com o gênero de produção *carta/e-mail* só aparecem no suporte correspondente, enquanto as *cartas do leitor* variam entre os suportes *jornal* e *revista*. Além disso, nesta variável, conforme pode-se observar na figura a seguir, também é perceptível o surgimento de novos suportes nas aplicações mais recentes, tais como *blog*, *formulário*, *relatório*, *site* e *catálogo*, embora a predominância dos três suportes mais comuns permaneça ao longo de toda a história do Celpe-Bras.

Figura 32 – Suporte de produção textual do corpus Celpe por ano



Fonte: Elaborado pela autora.

Os resultados de nossas análises das variáveis do grupo Avalia permitiram observar comportamentos das variáveis que corroboram o estudo de Schoffen *et al.* (2018). Devido à possibilidade de manipular as variáveis juntamente ao *corpus* de estudo no R, foi possível evidenciar resultados adicionais que demonstraram o comportamento das variáveis ao longo de todas as aplicações do exame relativas às tarefas 3 e 4 combinadas. Identificamos períodos de mudanças que coincidiram em diferentes variáveis. Além disso, foi possível observar a distribuição do número de *tokens* por temática das tarefas e agrupar a categoria natureza da interlocução, possibilitando analisar o comportamento da variável desde uma perspectiva mais abrangente.

7.2.2 Complexidade textual no corpus de estudo

A etapa seguinte da análise de complexidade textual foi analisar o *corpus* de estudo utilizando o modelo de regressão logística criado com o conjunto de 33 métricas indicadas pelo modelo Boruta e os dados dos *corpora* PorSimples e PorPopular. Esse modelo foi então usado para classificar as tarefas do *corpus* de estudo automaticamente, produzindo uma probabilidade para cada tarefa de se aproximar mais de uma complexidade alta, PorSimples, ou baixa, PorPopular. Durante esta fase do estudo, optou-se por não eliminar as tarefas caracterizadas como *outliers* nas análises descritivas, dado que a análise da complexidade textual dos textos

de insumo não está sujeita a distorções significativas geradas pela presença de um número de palavras fora da curva de normalidade. Além disso, ressaltamos que o modelo de classificação foi criado apenas com os dados do PorSimples e do PorPopular. Os dados do *corpus* de estudo foram somente classificados por esse modelo, então os *outliers* aqui não afetam a qualidade da classificação. É pertinente destacar que a única exceção foi a tarefa 4 da edição 2001-1, que, por não disponibilizar texto de insumo, não foi incluída nesta análise, conforme explicitado na tabela abaixo, que indica que o *corpus* de estudo compreende 89 tarefas.

Tabela 12 – Tarefas do corpus de estudo classificadas por nível de complexidade

Classificação	n	Total	Porcentagem
PorSimples	59	89	67%
PorPopular	27	89	30%
Indeciso	3	89	3%

Fonte: Elaborado pela autora.

Os dados mostram que o *corpus* de estudo se aproxima mais ao PorSimples, textos de alta complexidade. O número de tarefas classificadas como PorPopular representa menos da metade do número de tarefas classificadas como PorSimples. O resultado indicou que apenas 3 tarefas não puderam ser classificadas com 100% de certeza em nenhuma das categorias. Esse resultado indica que, para a maioria das tarefas, os textos selecionados apresentam complexidade textual alta, considerada adequada a um público-alvo em nível “Proficiente” do Inaf, correspondente ao grau mais alto de letramento brasileiro. Conforme foi apresentado, na faixa “Proficiente” encontram-se os leitores das estratificações socioeconômicas A e B, com Ensino Médio a Superior completo, que representam aproximadamente 3% da população. Esse resultado não surpreende, uma vez que os dados descritivos já haviam indicado a predominância de fontes dos textos de insumo como veículos de comunicação voltados para esse público.

Retomando Spencer *et al.* (2019), caso o objetivo do exame não seja avaliar a capacidade de decodificação, a complexidade textual não deve exigir um esforço excessivo, o que poderia prejudicar o desempenho do candidato nas demais habilidades de leitura que necessitam ser avaliadas. Xi (2017) adiciona que a complexidade textual deve estar no nível adequado, conforme o uso da língua-alvo que o exame almeja avaliar. Para realizar essa avaliação, no entanto, seria imprescindível ter uma clara compreensão do público-alvo do Celpe-Bras e do objetivo de uso de seus resultados, o que não se apresenta como tarefa simples.

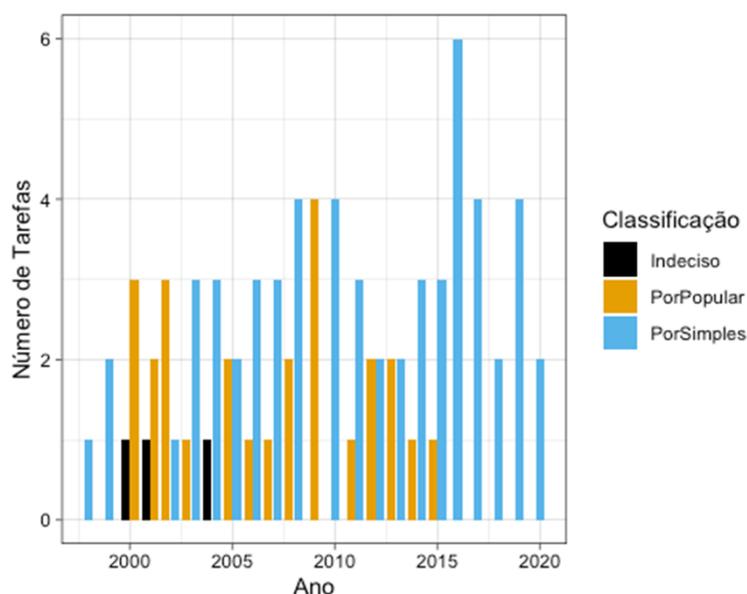
O Documento Base não é explícito ao tratar desse tema, exibindo um público bastante diversificado. As especificações mencionam os seguintes públicos-alvo: indivíduos maiores de

16 anos que tenham o ensino fundamental completo, indivíduos em empresas e instituições de ensino em nível global, ingressantes em cursos de graduação e pós-graduação no Brasil, profissionais estrangeiros com diplomas ou registro em entidades de classe a serem validados no Brasil e usuários de português como primeira língua de socialização de países africanos e do Timor-Leste ingressantes em IES brasileiras (BRASIL, 2020, p. 18), além de pessoas que tenham “familiaridade com diferentes práticas de letramento de que participam cidadãos escolarizados” (BRASIL, 2020, p. 30). É viável contemplar, em um único exame, as necessidades de uso da língua portuguesa por todos esses usuários?

Após a categorização dos textos de acordo com os níveis de complexidade textual, conduzimos testes de regressão linear para avaliar o efeito das variáveis *gênero de leitura* e *fonte do gênero de leitura* na definição da complexidade textual. Contudo, ambos os testes não revelaram efeito ou significância estatística na determinação da complexidade textual.

Por fim, analisamos a distribuição do número de tarefas por ano de aplicação do exame Celpe-Bras, conforme o nível de complexidade textual do texto de insumo.

Gráfico 12 – Número de tarefas do corpus de estudo por ano e nível de complexidade



Fonte: Elaborado pela autora.

Analisando a distribuição de número de tarefas por ano, de acordo com a classificação de complexidade textual, é possível observar o gradual abandono de uso de textos menos complexos no decorrer das aplicações. A predominância de textos mais complexos nos últimos anos de aplicação é claramente evidenciada. É possível cogitar que essa mudança gradual

observada na complexidade textual dos textos de insumo acompanhe as mudanças percebidas anteriormente nos resultados do estudo descritivo.

Parece relevante retomar um questionamento proposto por Shin (2021): quem são os elaboradores atuais das tarefas? Além de os resultados descritivos terem indicado a predominância de conteúdo que circula na camada privilegiada da sociedade e a ausência de textos do jornalismo popular, vemos aqui que tais textos são altamente complexos até mesmo para os parâmetros de leitura da população brasileira. Conforme propõem Randall *et al.* (2022, p. 6), entre as perguntas para um modelo de validade antirracista e orientado para a justiça social, entre outras, cabe a pergunta: “Quais grupos serão privilegiados por essa avaliação? A curto prazo? A longo prazo?”¹⁴⁰.

O comportamento evidenciado pelos resultados de nossas análises mostra que, especialmente nos últimos anos, as tarefas de leitura e escrita são compostas por materiais que contribuem para a manutenção de uma língua dominante, representativa de um conjunto muito específico de práticas letradas privilegiadas. Por qual motivo os textos selecionados promovem o uso desses usos da linguagem hegemônico brancos, que marginaliza outros repertórios comunicativos, potencialmente prejudicando o indivíduo que não tem acesso a tais práticas linguísticas, mesmo sendo “capaz de engajar-se em diferentes situações de uso da língua portuguesa no mundo, mostrando adequação às demandas dos vários contextos” (BRASIL, 2020, p. 28)?

Frequentemente, um argumento utilizado para a escolha de textos complexos é que se o texto não for suficientemente desafiador, a tarefa não será capaz de distinguir os candidatos conforme o grau de proficiência, especialmente nos níveis mais avançados. Em relação a esse tópico, torna-se relevante revisitar a concepção de dificuldade textual, entendida como o resultado da inter-relação da tríade texto-tarefa-leitor. Até aqui, os resultados sugerem que a complexidade textual pode potencialmente excluir candidatos de nível intermediário do exame por tornar o conteúdo inacessível para a conclusão da tarefa.

Contudo, como amplamente discutido, a complexidade textual do texto de insumo, por si só, não define a dificuldade textual, nem a complexidade da tarefa, que pode, por exemplo, exigir uma leitura superficial do texto, aspecto que não foi analisado aqui. No entanto, para evitar o risco de excluir candidatos devido à elevada complexidade textual da leitura, é importante que os textos de insumo apresentem um nível de complexidade adequado e equivalente entre as edições, que deve ser cuidadosamente monitorado para que o texto não se

¹⁴⁰ No original, lê-se: What groups will be privileged by this assessment? In the short term? The long term?

constitua como uma barreira intransponível para os candidatos (THE COLLEGE BOARD, 2015). A nosso ver, utilizar textos de leitura menos complexos possibilita ao mesmo tempo calibrar o texto para o nível intermediário previsto e monitorar o tipo de tarefa, controlando, desse modo, a dificuldade textual e, por consequência, a complexidade da tarefa.

No contexto do Celpe-Bras, isso implicaria a seleção de textos com complexidade textual intermediária, ajustando a elaboração do enunciado para o grau de complexidade da tarefa desejado, uma vez que as sub-habilidades de leitura são avaliadas por meio do cumprimento dos propósitos comunicativos evidenciados por meio da produção textual. A seguir, apresentamos dois exemplos de tarefas categorizadas como de complexidade textual PorPopular, com o objetivo de analisar essa possibilidade de elaborar tarefas que utilizem textos de leitura menos complexos, mas com potencial de distinguir os níveis de proficiência.

O primeiro exemplo é a tarefa 4 de 2013-1, “Viagem a dois”, na qual o candidato é convidado a assumir a posição social de gerente de uma empresa. No papel de gerente, o candidato deve escrever um e-mail aos demais funcionários com o objetivo de incentivá-los a se cadastrar em um programa de caronas. O enunciado, ilustrado a seguir, aponta três propósitos a serem atendidos, os quais destacamos em diferentes cores para facilitar a compreensão do quadro seguinte.

Figura 33 – Tarefa 4 de 2013-1

The image shows a screenshot of the Celpe-Bras 2013/1 exam interface. At the top left is the logo 'Celpe-Bras 2013/1'. To its right, the text reads 'Certificado de Proficiência em Língua Portuguesa para Estrangeiros'. Below this, a dark bar contains 'Tarefa 4 | Viagem a Dois' on the left and 'Página 8' on the right. The main content area is a light gray box with the following text: 'Você é gerente de Recursos Humanos de uma empresa que acabou de se cadastrar no site Caronetas. Escreva um e-mail para os funcionários da empresa, incentivando-os a participar do programa de caronas. Seu texto deverá' followed by a bulleted list:

- apresentar o funcionamento da plataforma;
- explicitar os benefícios do sistema; e
- incentivar a abertura de contas no site.

Fonte: Acervo Celpe-Bras.

Embora não possamos afirmar, é possível supor que, para cumprir adequadamente os propósitos dessa tarefa, o candidato necessite mobilizar, especialmente, as sub-habilidades de “localizar informações explícitas no texto” e “selecionar informações do texto necessárias para cumprir o propósito da tarefa”. Evidentemente, considerando que as outras sub-habilidades

relacionadas à situação de comunicação e aos gêneros do discurso são inevitavelmente mobilizadas, de certo modo, em todas as tarefas que integram leitura e produção textual, elas também estarão sendo testadas. O texto de insumo é apresentado no quadro a seguir no qual destacamos as informações explícitas que podem ser utilizadas pelo candidato em sua produção escrita para cumprir os três propósitos solicitados no enunciado. As cores utilizadas no quadro destacam informações que podem ser usadas pelo candidato para responder a cada um dos três propósitos destacados no enunciado acima.

Quadro 19 – Texto de insumo da tarefa 4 de 2013-1

EDIÇÃO Priscilla Santos

Ur

Urbanidade

VIAGEM A DOIS

SITE PREMIADO AJUDA A ORGANIZAR CARONAS ENTRE COLEGAS NO TRAJETO CASA-TRABALHO

Boa parte dos automóveis que circulam em grandes centros urbanos brasileiros carrega poucos passageiros. Em São Paulo, por exemplo, 64% dos carros só levam o próprio motorista. Além de congestionar o trânsito, ainda polui o ar. Foi diante disso que o engenheiro especialista em logística Márcio Nigro, de São Paulo, criou o site Caronetas, eleito em junho pela Universidade de Michigan e pela Rockefeller Foundation como a melhor solução mundial de mobilidade urbana. A plataforma, voltada para empresas, ajuda a organizar caronas entre funcionários. Lançada em 2011 e atualmente com 293 mil usuários, é responsável por mais de 5 mil caronas diárias em 19 estados brasileiros.

A grande sacada é solucionar a questão da segurança, um dos maiores temores das pessoas ao oferecer ou aceitar caronas. Como o círculo é restrito à empresa, essa insegurança diminui. Hoje já são quase 2 mil corporações usando o sistema, que é gratuito. Primeiro, a companhia se cadastra e, então, seus funcionários fazem suas contas individuais no site. Ao preencher seu trajeto diário, a plataforma indica pessoas que fazem caminhos parecidos e as coloca em contato.

Outro atrativo é financeiro. A carona é cobrada, levando em conta gasto de combustível, pedágios e estacionamento. "O ideal é que motorista e passageiro definam um valor por viagem", afirma Márcio. O pagamento pode ser feito diretamente a quem oferece a carona. Mas uma das novidades da plataforma é um sistema de créditos pré-pagos, que facilita a falta de troco, por exemplo. Quem pega carona pode transferir parte de seus créditos virtualmente para quem dirige. Ao compartilhar o custo dos trajetos, o motorista pode economizar entre R\$ 3.500 e R\$ 7 mil ao ano.

Além de se economizar dinheiro e a natureza, as caronas podem ser uma maneira de voltarmos a respirar aliviados, como já fizemos um dia. "Se os proprietários dos 5 milhões de carros de São Paulo transportarem mais uma pessoa, tiraríamos 2,5 milhões de veículos das ruas, um cenário parecido ao de 20 anos atrás".

Vanessa Vieira

20 AGOSTO 2012

Revista Galileu, 20 ago. 2012.

Fonte: Acervo Celpe-Bras.

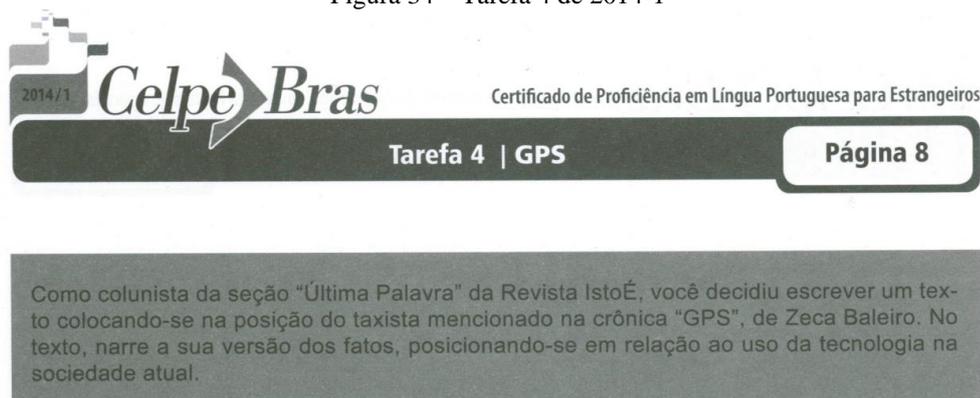
No quadro acima, vemos que as informações para cumprir os dois primeiros propósitos são facilmente localizadas no texto, visto que são fornecidas de maneira explícita. Por outro lado, o terceiro propósito de incentivar a abertura de contas no site se coloca como mais desafiador, pois envolve a recontextualização do texto lido em um novo gênero - um e-mail do gerente de Recursos Humanos para os funcionários para incentivá-los a se cadastrar -, e que

envolve explicitar o convite justificando-o a partir da seleção das informações do texto que dizem respeito aos dois primeiros propósitos.

As informações explícitas podem ser facilmente parafraseadas por candidatos com níveis de proficiência mais baixos ou podem ser utilizadas como base para a elaboração de textos mais autorais, no qual o candidato articula recursos informacionais e linguísticos de seu próprio repertório, conforme já mostrou Mendel (2019), Kunrath (2019) e Sirianni (2020) em relação a outras tarefas (ver seção 4.2). Tendo em vista as sub-habilidades mobilizadas para a compreensão do texto e os propósitos comunicativos, podemos considerar que este é um exemplo de tarefa de baixa complexidade. No entanto, a tarefa é capaz de distinguir níveis de proficiência, visto que, como as pesquisas citadas apontam, a quantidade e a qualidade das informações selecionadas, bem como a forma de recontextualização dessas informações são determinantes para distinguir o nível de proficiência dos candidatos.

O segundo exemplo refere-se à tarefa 4 de 2014-1, “GPS”, na qual o enunciado solicita que o candidato assuma a posição social de um colunista de revista que decidiu escrever uma crônica. Esse texto de insumo também foi classificado pelo algoritmo Boruta como de complexidade PorPopular e trata-se de uma crônica na qual dois personagens dialogam dentro de um táxi. O enunciado, ilustrado abaixo, solicita que o candidato reescreva a crônica, narrando os mesmos fatos, porém, a partir do ponto de vista do personagem oposto, mostrando seu posicionamento em relação ao uso da tecnologia, temática da tarefa.

Figura 34 – Tarefa 4 de 2014-1



2014/1 **Celpe-Bras** Certificado de Proficiência em Língua Portuguesa para Estrangeiros

Tarefa 4 | GPS **Página 8**

Como colunista da seção “Última Palavra” da Revista IstoÉ, você decidiu escrever um texto colocando-se na posição do taxista mencionado na crônica “GPS”, de Zeca Baleiro. No texto, narre a sua versão dos fatos, posicionando-se em relação ao uso da tecnologia na sociedade atual.

Fonte: Acervo Celpe-Bras.

Do mesmo modo, ainda que não possamos afirmar, supomos que, para cumprir adequadamente os propósitos dessa tarefa, o candidato necessite mobilizar mais sub-habilidades, tais como: identificar a ideia principal do texto, identificar detalhes e ideias secundárias, inferir palavras desconhecidas e implícitos no texto e reconhecer atitudes, emoções

e pontos de vista do autor quando expressos explícita ou implicitamente no texto. Conforme se pode verificar no texto de insumo apresentado a seguir, não há informações isoladas destacadas por nós que possam embasar uma produção textual que responde ao enunciado. Nesta tarefa, a compreensão integral da crônica é crucial, exigindo a compreensão dos implícitos e a capacidade de posicionar-se em relação à temática. Além disso, a tarefa exige a habilidade de demonstrar a compreensão ao narrar os mesmos fatos do enredo do texto de leitura.

Quadro 20 – Texto de insumo da tarefa 4 de 2014-1

ZecaBaleiro
cantor e compositor
Última palavra
Conheço pessoas que não se deslocam mais à esquina para comprar pão sem que façam uso do GPS, Google Maps e o escambau
GPS
Entrei no táxi e falei o meu destino.
— Rua Araribóia, por favor.
— Araribóia? Espera um minuto!... — rebateu o homem.
Programou então o GPS e arrancou.
— Não precisa de GPS, amigo. Sei mais ou menos onde fica. Posso lhe orientar.
— Ah, não. Não saio mais de casa sem isto — declarou.
Resmunguei em silêncio. E lá se foi o taxista seguindo seu brinquedinho falante — “vire à esquerda”; “a 50 metros você vai virar à direita”; “daqui a 300 metros faça o retorno à esquerda”...
De repente, entre uma e outra prosa, vi ele se afastando da direção que eu julgava ser a correta.
— Amigo, acho que você está na direção contrária.
Tinha que ter entrado naquela rua à direita, melhor fazer o retorno na frente.
— Não, não, olha aqui — apontou pra geringonça, orgulhoso como ele só. É esse mesmo o caminho.
Cocei a cabeça irritado. Embora eu não soubesse exatamente qual trajeto a seguir, sabia que aquele caminho que ele fazia era estupidamente mais longo e complexo.
Argumentei mais uma vez, já na iminência de explodir.
— Moço, desculpe, mas tenho quase certeza que você está fazendo um caminho muito mais longo do que devia.
— Não esquentar a cabeça não, companheiro. Tá aqui no GPS, ó. Não vou discutir com a tecnologia.
Sim, eu havia ouvido aquilo. E mais que uma frase de efeito de um chofer de praça, aquilo era uma senha que explicava muita coisa, talvez explicasse até toda uma época. O sujeito deixava de lado sua inteligência (se é que a tinha), a experiência de anos perambulando a bordo do seu táxi pelas quebradas da cidade e o próprio poder de dedução para seguir uma engenhoca surda e cega — mas “tecnológica” — sem questioná-la, e sem que eu também pudesse fazê-lo.
Não quero parecer um dinossauro (embora por vezes eu inevitavelmente pareça), mas sempre defendi um uso inteligente, comedido e crítico dos apetrechos eletrônicos. Conheço pessoas que, por comodidade, condicionamento ou deslumbramento com o novo mundo cibernético, não se deslocam mais à esquina para comprar pão sem que façam uso de GPS, Google Maps e o escambau.
Tenho um sobrinho, um pensador irreverente de botequim, que gosta de dizer o seguinte:
— As rodas de bar ficaram muito chatas depois do iPhone. Ninguém mais pode ter dúvida alguma. Se alguém perguntar: “como é o nome daquele cantor que cantava aquela música? “; ou então: “quem era o centroavante da seleção de 86?”, logo algum bobo alegre vai acessar a internet e buscar a resposta. E aí acaba com a graça, a mágica e o mistério... Não sobra assunto pro próximo encontro.
Outro amigo filósofo de padaria, tem uma tese/profecia tenebrosa sobre o uso sem critério dos tecnobreguetes: Diz ele:
— Num futuro próximo, as pessoas deixarão de ter memória. Para que lembrar, se tudo cabe num HD externo?
É. Faz bastante sentido a tese do meu amigo. Aliás, há tempos não o vejo, o... o... Como é mesmo o nome dele, gente? Aníbal, não. Átila, não... É um nome assim meio histórico... Desculpem aí, vou ter que espiar na

agenda do meu celular.
Disponível em: www.istoe.com.br. Acesso em: 24 fev. 2014.

Fonte: Acervo Celpe-Bras.

É possível perceber, no texto de insumo, que o autor se posiciona como passageiro de um táxi e, durante sua viagem, seus comentários indicam um posicionamento contrário ao uso generalizado da tecnologia. Em contrapartida, o taxista, mesmo que não diga explicitamente, manifesta um posicionamento mais favorável ao uso da tecnologia, já que utiliza o GPS para chegar ao destino final. A leitura dessa crônica demanda a compreensão de sentidos mais implícitos, uma vez que as opiniões dos personagens não são diretamente explicitadas. Além de a compreensão ser mais desafiadora do que no primeiro exemplo, a produção textual também é mais elaborada. O gênero crônica exige o uso de recursos linguísticos e estilísticos mais específicos e o propósito de posicionar-se exige habilidades argumentativas.

Assim, diferentemente do exemplo anterior, esta tarefa apresenta dificuldade textual mais elevada, por consequência, o nível de complexidade da tarefa também é mais alto, considerando as sub-habilidades mobilizadas, os gêneros discursivos de leitura e escrita e os propósitos comunicativos necessários para o cumprimento da tarefa. Embora o texto de insumo seja de complexidade textual baixa, os objetivos de leitura e escrita são desafiadores, possibilitando que a tarefa seja capaz de distinguir níveis de proficiência dos candidatos.

É interessante destacar que, nos dois exemplos apresentados, os textos de insumo utilizados pertencem à esfera jornalística, ambos os textos são provenientes de revistas. Parece-nos que, com a devida atenção, é possível selecionar textos com uma complexidade textual mais acessível sem colocar em risco a capacidade da tarefa em distinguir os níveis de proficiência.

Conforme descrito no Documento Base, o examinando do Celpe-Bras certificado no nível intermediário “é capaz de selecionar algumas informações a partir da interpretação de textos de diferentes gêneros orais e escritos, evidenciando problemas de compreensão e dificuldades no trabalho de recontextualização” (BRASIL, 2020, p.68). Sirianni (2020) reforça que “aqueles que recebem certificado de nível Intermediário demonstram domínio parcial da língua portuguesa, no entanto são capazes de compreender e produzir em português, possibilitando a comunicação, mesmo que limitada, na língua.” (SIRIANNI, 2020, p. 66).

Entretanto, o que os descritores não explicitam é que o candidato em nível intermediário do Celpe-Bras demonstra esse domínio parcial da língua portuguesa ao compreender textos de alta complexidade. Essa tarefa demanda a leitura de textos escritos direcionados à elite brasileira, ou seja, direcionados a indivíduos com elevado grau de letramento e poder aquisitivo.

Como amplamente discutido, no exame Celpe-Bras, não basta ao examinando simplesmente compreender o texto; é necessário que ele evidencie essa compreensão por meio da produção textual em resposta à tarefa. O texto deve recontextualizar informações relevantes, selecionadas do texto de insumo, de modo consistente e coeso, mobilizando, para tanto, recursos linguísticos que incluam repertório pessoal, que transcenda o conteúdo informacional do material de leitura. Ademais, além de realizar todas essas ações, é essencial cumprir os propósitos comunicativos solicitados no enunciado da tarefa.

Isso posto, os resultados referentes à complexidade textual do *corpus* de estudo evidenciaram a predominância do uso de textos de alta complexidade textual direcionados para a faixa Proficiente do Inaf.

7.2 PREDIÇÃO DE COMPLEXIDADE TEXTUAL

Conforme explicado no capítulo de metodologia, esta fase da pesquisa começou com a geração das 200 métricas de complexidade textual por meio da ferramenta NILC-Matrix aplicadas a todos os textos presentes nos três *corpora*. A extração dos valores numéricos dessas métricas foi realizada pelo pesquisador Sidney Evaldo Leal, um dos criadores da ferramenta, que nos forneceu os arquivos com os respectivos resultados.

Dentre os dados recebidos, excluímos os valores correspondentes aos 70 índices que foram considerados como ruídos, normalizamos os valores das 130 métricas para a mesma escala numérica e demos início ao processo de seleção de variáveis preditivas utilizando o algoritmo Boruta.

Utilizamos os dois *corpora* de referência para que o algoritmo identificasse padrões capazes de distinguir os dois conjuntos de textos e o modelo aceitou 31 índices e apontou mais 9 como indecisos para a predição de complexidade textual. Em outras palavras, o modelo rejeitou 90 das 130 variáveis iniciais, categorizando essas variáveis como não relevantes para a discriminação de textos entre os dois conjuntos. A seguir um modelo de regressão foi criado com 80% dos dados (randomicamente selecionados), com 20% reservados para a testagem do modelo. A eficácia do modelo foi verificada e apresentou uma acurácia de 100%. Isso significa que conduzimos testes com amostras aleatórias de textos provenientes dos dois *corpora*, e esses foram corretamente categorizados pelo algoritmo em todas as ocasiões. Assim, o modelo criado é capaz de distinguir entre textos do PorSimples e do PorPopular usando as variáveis selecionadas pelo Boruta.

7.2.1 Métricas preditivas de complexidade textual

Até esta etapa da pesquisa, conduzimos o teste de um classificador automático binário para selecionar índices preditivos capazes de distinguir nossos dois conjuntos de textos de acordo com a complexidade textual. Em geral, os estudos concluem as análises após a redução das multivariáveis, apontando quais variáveis são altamente relacionadas à complexidade textual e a acurácia do modelo em identificá-las. No entanto, alinhando-se aos princípios da Avaliação de línguas informada por corpus, neste trabalho, a redução de variáveis representa o início da análise, que visou permitir a interpretação coletiva desses traços e o modo como se inter-relacionam na materialidade dos textos.

Conforme mencionado, o modelo apontou 9 métricas em nível de indecisão, das quais apenas 2 foram indicadas por pesquisas prévias da língua portuguesa, apresentadas no quadro 8, da seção 5.2 deste trabalho. Desse modo, optamos por incluir essas duas e excluir as outras 7 métricas do modelo. A primeira métrica aceita é “verbs”, que calcula a proporção da frequência de palavras etiquetadas como verbos pelo *tagger* da ferramenta em relação ao total de palavras no texto. A segunda métrica é “words_before_main_verb”, que calcula a média do número de palavras que antecedem os verbos principais das principais orações em cada sentença e depende do tokenizador e do *parser*. Ambos os índices estão vinculados à análise morfológica de verbos, que são interpretados pela ferramenta como representantes de orações. Quanto maior o número de orações nas sentenças, maior é a complexidade, e quanto mais distante do início da oração, maior é a carga de memória requerida para a compreensão textual. Levando em consideração que a literatura da área indica que a extensão e a quantidade de sentenças no texto estão relacionadas à complexidade do texto, logo, consideramos relevante a inclusão dessas variáveis em nosso conjunto.

As 31 métricas apontadas pelo modelo Boruta como indicadoras de complexidade textual, somada às 2 adicionalmente aceitas, estão listadas no quadro abaixo. Destacamos em negrito os índices que foram identificados por estudos anteriores como preditivos de complexidade textual.

Quadro 21 – Métricas do NILC-Metrix aceitas pelo modelo Boruta

GRUPOS	DESCRIÇÃO	MÉTRICAS
1. Medidas Descritivas	10 métricas de estatísticas básicas que descrevem o <i>corpus</i>	<ol style="list-style-type: none"> 1. Quantidade de Parágrafos no texto (id: 152) paragraphs 2. Quantidade de Sentenças no texto (id: 154) sentences 3. Quantidade média de sílabas por palavra no texto (id: 157) syllables_per_content_word 4. Média de Palavras por Sentença (id: 159) words_per_sentence 5. Quantidade Mínima de palavras por sentença (id: 184) sentence_length_min
2. Simplicidade Textual	8 métricas que indicam a extensão das sentenças, uso de palavras de conteúdo e de pronomes	<ol style="list-style-type: none"> 6. Proporção de Sentenças Muito Longas em relação a todas as sentenças do texto (id: 189) long_sentence_ratio 7. Proporção de Sentenças Curtas em relação a todas as sentenças do texto (id: 192) short_sentence_ratio
3. Coesão Referencial	9 métricas que indicam cadeias de correferência intraparágrafo e interparágrafo	<ol style="list-style-type: none"> 8. Média de candidatos a referente, na sentença anterior, por pronome demonstrativo anafórico (id 195) demonstrative_pronoun_ratio
4. Coesão Semântica	11 métricas que capturam relações semânticas e similaridade entre palavras	<ol style="list-style-type: none"> 9. Média de similaridade entre pares de sentenças adjacentes no texto (id: 7) lsa_adj_mean 10. Desvio padrão de similaridade entre pares de sentenças adjacentes no texto (id: 8) lsa_adj_std 11. Média de similaridade entre todos os pares de sentenças no texto (id: 9) lsa_all_mean 12. Média do *givenness* da cada sentença do texto, a partir da segunda (id: 11) lsa_givenness_mean 13. Desvio padrão do *givenness* da cada sentença do texto, a partir da segunda (id: 12) lsa_givenness_std 14. Média de similaridade entre pares de parágrafos adjacentes no texto (id: 13) lsa_paragraph_mean 15. Desvio padrão entre parágrafos adjacentes no texto (id: 14) lsa_paragraph_std 16. Média do *span* da cada sentença do texto, a partir da segunda (id: 15) lsa_span_mean 17. Desvio padrão do span da cada sentença do texto, a partir da segunda (id: 16) lsa_span_std 18. Média da entropia cruzadas das sentenças do texto (id: 6) cross_entropy

8. Léxico Temporal	12 métricas que avaliam o uso de tempos e modos verbais e de conectivos temporais	19. Proporção de verbos no particípio em relação a todos os verbos do texto (id: 146) <i>participle_verbs</i>
9. Complexidade Sintática	27 métricas que descrevem as estruturas sintáticas	20. Proporção de orações relativas em relação à quantidade de orações do texto (id: 32) <i>relative_clauses</i> 21. Proporção de orações subordinadas pela quantidade de orações do texto (id: 42) <i>subordinate_clauses</i> 22. Quantidade Média de palavras antes dos verbos principais das orações principais das sentenças (id: 44) <i>words_before_main_verb</i>
11. Informações Morfossintáticas de Palavras	42 métricas relativas às classes gramaticais da língua	23. Proporção de Adjetivos em relação à quantidade de palavras do texto (id: 87) <i>adjective_ratio</i> 24. Proporção mínima de substantivos em relação à quantidade de palavras das sentenças (id: 108) <i>nouns_min</i> 25. Desvio padrão das proporções entre substantivos e a quantidade de palavras das sentenças (id: 109) <i>nouns_standard_deviation</i> 26. Proporção de preposições em relação à quantidade de orações no texto (id: 112) <i>prepositions_per_clause</i> 27. Quantidade Média de preposições por sentença no texto (id: 113) <i>prepositions_per_sentence</i> 28. Proporção de Verbos em relação à quantidade de palavras do texto (id: 125) <i>verbs</i> 29. Desvio padrão das proporções entre verbos e a quantidade de palavras das sentenças (id: 128) <i>verbs_standard_deviation</i>
13. Frequência de Palavras	10 métricas de frequência normalizada das palavras do <i>corpus</i>	30. Média dos valores das frequências das palavras do texto na escala logarítmica Zipf, via <i>Corpus</i> Brasileiro (id: 80) <i>freq_bra</i> 31. Média das frequências absolutas das palavras de conteúdo mais raras das sentenças do texto, via Banco de Português (BP) (id: 82) <i>min_cw_freq</i>
14. Índices de Leitabilidade	5 métricas superficiais de complexidade textual (apenas duas adaptadas ao português: Flesch e Dale Chall)	32. Índice Flesch (id: 198) <i>flesch</i> 33. Estatística de Honoré (id: 200) <i>honore</i>

Fonte: Adaptado pela autora de <http://fw.nilc.icmc.usp.br:23380/metrixdoc>. Acesso em: 15 mar. 2023.

Como mostra o quadro, o algoritmo Boruta não indicou índices dos grupos “6. Diversidade Lexical”, “7. Conectivos”, “10. Densidade de Padrões Sintáticos” e “12. Informações Semânticas de Palavras”. Além disso, nossa redução de variáveis também incluiu um número significativo dos grupos “1. Medidas Descritivas”, “11. Informações Morfossintáticas de Palavras” e “14. Índices de Leiturabilidade”, do mesmo modo que foi observado nos estudos prévios citados anteriormente. No total, 12 índices indicados por nosso modelo corroboram resultados de pesquisas anteriores.

O grupo com mais índices indicados é o “4. Coesão Semântica”, com 10 das 11 métricas sendo apontadas como preditivas. Se destacam também os grupos “1. Medidas Descritivas” e “14. Índices de Leiturabilidade” que tiveram a metade do total de suas métricas aceitas pelo modelo Boruta. O grupo “11. Informações Morfossintáticas de Palavras” também aparece com um número alto de índices; 7 das 42 métricas foram consideradas preditivas.

Antes de prosseguir para as análises, reforçamos que, nesta etapa da pesquisa, o *corpus* PorSimples, está sendo utilizado como substituto para complexidade textual alta e o *corpus* PorPopular, como baixa. A descrição dos *proxies* é sumarizada no quadro abaixo.

Quadro 22 – Descrição dos proxies de complexidade textual

<i>Corpus</i>	<i>Proxy</i>	Público-alvo brasileiro
PorSimples	Nível de complexidade textual alta	Leitores na faixa Proficiente, estratificações socioeconômicas A e B, Ensino Médio a Superior completo, aproximadamente 3% da população brasileira, de acordo com dados do Inaf.
PorPopular	Nível de complexidade textual baixa	Leitores nas faixas Elementar e Intermediário, estratificações socioeconômicas C, D e E, Anos Finais do Ensino Fundamental a Superior completo, aproximadamente 54% da população brasileira, de acordo com dados do Inaf.

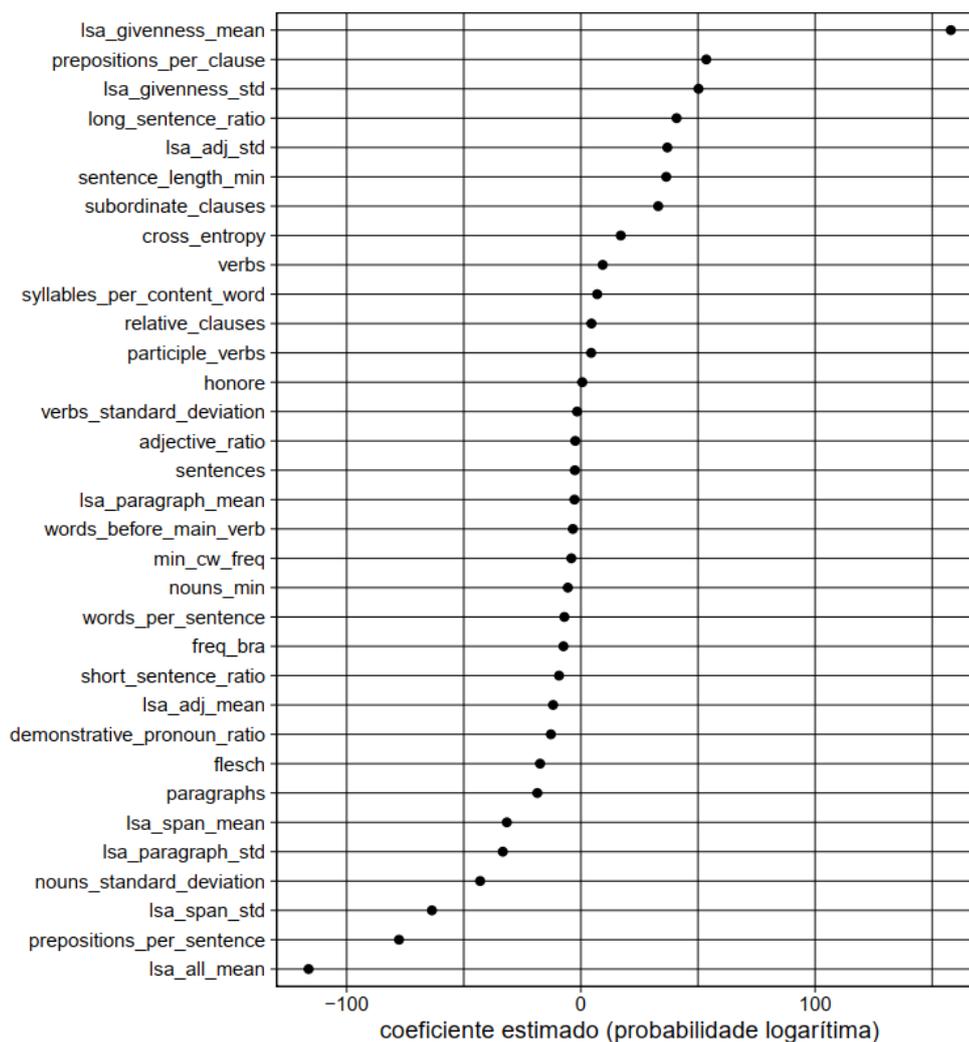
Fonte: Elaborado pela autora.

Após a redução de multivariáveis, exploramos o efeito de cada um dos 33 índices na capacidade de distinção dos textos. Para as análises que seguem, utilizamos como variável dependente o *corpus* de onde cada texto originou. A variável dependente então tem dois níveis: o *corpus* PorSimples e o *corpus* PorPopular. A referência para essa variável é o PorSimples (valor 1 na classificação binária), com os textos do PorPopular sendo colocados em oposição (valor 0 na classificação binária). Assim os resultados indicam a probabilidade de uma observação ser classificada como sendo do *corpus* PorSimples.

É imperativo enfatizar, no entanto, que o conjunto de 33 traços linguísticos apresentado é o resultado de um modelo estatístico capaz de identificar múltiplas variáveis que coocorrem com nível de magnitude do efeito e significância estatística suficientes para apontar um padrão que distingue os textos que compõem nossos *corpora*. No entanto, não estamos afirmando que esse padrão encontrado seja generalizável para outros gêneros discursivos, mas podem servir como indicativos para futuras análises de complexidade textual.

O gráfico abaixo mostra o efeito de cada variável na probabilidade de um texto ser categorizado como PorSimples, ou seja, de complexidade textual alta. Quanto mais alto o coeficiente estimado, pontos para o lado direito do gráfico, maior o tamanho do efeito daquele índice para distinguir o texto como complexidade alta. Para o lado oposto, os índices com valores negativos têm maior força para distinguir textos com complexidade textual baixa, na categoria PorPopular, identificados pelos pontos para o lado esquerdo do gráfico. Os índices que se aproximam do zero, situados no centro do gráfico, embora tenham um efeito menor na predição, integram o modelo multivariável que leva em consideração os 33 atributos sinalizados pelo algoritmo Boruta simultaneamente.

Gráfico 13 – Efeito dos índices NILC-Metrix na probabilidade de ser PorSimples



Fonte: Elaborado pela autora.

O gráfico mostra que a maior parte dos índices possui uma força de predição distinta de zero, exibindo uma curva mais acentuada à esquerda, o que indica índices com maior efeito para identificar características do padrão encontrado no PorPopular. Esses dados, no entanto, apenas revelam a magnitude do efeito de cada índice dentro do padrão de traços linguísticos que ocorrem com frequência significativa nos textos analisados.

Contudo, conforme anteriormente elucidado, cada métrica do NILC-Metrix possui sua própria descrição e interpretação sobre como se relaciona com a complexidade textual, conforme definido pelos desenvolvedores da ferramenta. Certos índices sinalizam maior complexidade textual ao apresentarem um valor numérico elevado, enquanto outros indicam o oposto. Essa informação é crucial para a interpretação dos resultados. Por exemplo, no caso dos índices de leiturabilidade, no Índice Flesch, quanto menor o valor resultante da métrica, maior

a complexidade. Os resultados do modelo acima confirmam essa relação, indicando que quanto maior o Índice Flesch, menor a probabilidade de o texto ser complexo (ou seja, relação inversa). De maneira oposta, na Estatística Honoré, quanto maior o valor, maior a complexidade. De novo no modelo acima a Estatística Honoré é positiva, mas com um efeito pequeno. Ou seja, quanto maior a Estatística Honoré, maior a probabilidade de o texto ser complexo, mas com um efeito bem atenuado. Isso significa que é necessário um aumento muito maior da Estatística Honoré para aumentar a probabilidade de o texto ser complexo.

Além disso, ao apresentar apenas o nome das métricas preditivas de complexidade textual, tira-se do leitor a oportunidade de atuar como coanalista dos dados junto ao pesquisador que as divulga, visto que se torna impossível discernir se a métrica prevê alta complexidade quando a característica é predominante no texto, ou o contrário. Por exemplo, indicar a métrica “verbs” sem esclarecer se uma alta ou baixa frequência nos textos os classifica como complexos é manter o leitor às escuras. Nesse contexto, naturalmente, a interpretação dos dados será orientada por uma abordagem *top-down*, na qual parte-se de conceitos estáticos preconcebidos e aceitos como fatos (BIBER *et al.*, 2021c). Em nosso exemplo anterior, a maior probabilidade é que a métrica “verbs” seja interpretada da seguinte forma: quanto mais verbos, mais orações e, portanto, mais extenso o texto, resultando, assim, em maior complexidade textual.

Portanto, adotando uma abordagem *bottom-up* para viabilizar uma análise empírica das evidências e evitar a concepção apriorística de padrões de complexidade textual, reescrevemos as informações do gráfico de efeito dos índices na classificação binária. O quadro apresenta os mesmos resultados, porém utilizamos o nome das métricas e as organizamos por grupo, para facilitar a leitura. Para indicar a direção do valor que se relaciona com alta complexidade textual, utilizamos o símbolo de uma seta. Quando valores mais altos dos resultados significam maior complexidade, a seta aponta para cima. O oposto será representado por uma seta direcionada para baixo. Ou seja, a seta para cima significa que o valor alto do métrica significa maior complexidade. A seta para baixo significa que o valor baixo da métrica significa maior complexidade, conforme os desenvolvedores da ferramenta. Na última coluna, indicamos o nível de complexidade (PorSimples ou PorPopular) para o qual a variável tem efeito maior na probabilidade de classificação, que no primeiro gráfico foi ilustrado pelos pontos à direita e à esquerda do valor 0.

Quadro 23 – Efeito dos índices NILC-Metrix na probabilidade de classificação

Grupo	Métrica	Direção	Complexidade
1. Medidas Descritivas	1. Quantidade de Parágrafos no texto	↑	PorPopular
	2. Quantidade de Sentenças no texto	↑	PorPopular
	3. Quantidade média de sílabas por palavra no texto	↑	PorSimples
	4. Média de Palavras por Sentença	↑	PorPopular
	5. Quantidade Mínima de palavras por sentença	↑	PorSimples
2. Simplicidade Textual	6. Proporção de Sentenças Muito Longas em relação a todas as sentenças do texto	↑	PorSimples
	7. Proporção de Sentenças Curtas em relação a todas as sentenças do texto	↓	PorPopular
3. Coesão Referencial	8. Média de candidatos a referente, na sentença anterior, por pronome demonstrativo anafórico	↑	PorPopular
4. Coesão Semântica	9. Média de similaridade entre pares de sentenças adjacentes no texto	↓	PorPopular
	10. Desvio padrão de similaridade entre pares de sentenças adjacentes no texto	↑	PorSimples
	11. Média de similaridade entre todos os pares de sentenças no texto	↓	PorPopular
	12. Média do *givenness* da cada sentença do texto, a partir da segunda	↓	PorSimples
	13. Desvio padrão do *givenness* da cada sentença do texto, a partir da segunda	↑	PorSimples
	14. Média de similaridade entre pares de parágrafos adjacentes no texto	↓	PorPopular
	15. Desvio padrão entre parágrafos adjacentes no texto	↑	PorPopular
	16. Média do *span* da cada sentença do texto, a partir da segunda	↓	PorPopular
	17. Desvio padrão do span da cada sentença do texto, a partir da segunda	↑	PorPopular

	18. Média da entropia cruzadas das sentenças do texto	↑	PorSimples
8. Léxico Temporal	19. Proporção de verbos no particípio em relação a todos os verbos do texto	↓	PorSimples
9. Complexidade Sintática	20. Proporção de orações relativas em relação à quantidade de orações do texto	↑	PorSimples
	21. Proporção de orações subordinadas pela quantidade de orações do texto	↑	PorSimples
	22. Quantidade Média de palavras antes dos verbos principais das orações principais das sentenças	↑	PorPopular
11. Informações Morfossintáticas de Palavras	23. Proporção de Adjetivos em relação à quantidade de palavras do texto	↑	PorPopular
	24. Proporção mínima de substantivos em relação à quantidade de palavras das sentenças	↑	PorPopular
	25. Desvio padrão das proporções entre substantivos e a quantidade de palavras das sentenças	↑	PorPopular
	26. Proporção de preposições em relação à quantidade de orações no texto	↑	PorSimples
	27. Quantidade Média de preposições por sentença no texto	↑	PorPopular
	28. Proporção de Verbos em relação à quantidade de palavras do texto	↑	PorSimples
	29. Desvio padrão das proporções entre verbos e a quantidade de palavras das sentenças	↑	PorPopular
13. Frequência de Palavras	30. Média dos valores das frequências das palavras do texto na escala logarítmica Zipf, via <i>Corpus Brasileiro</i>	↓	PorPopular
	31. Média das frequências absolutas das palavras de conteúdo mais raras das sentenças do texto, via Banco de Português	↓	PorPopular
14. Índices de Leiturabilidade	32. Índice Flesch	↓	PorPopular
	33. Estatística de Honoré	↑	PorSimples

Fonte: Elaborado pela autora.

A primeira observação é que, individualmente, os índices não indicam diferenças, e não necessariamente todos apontam para uma alta complexidade no PorSimples. Isso reafirma a relevância de que os traços linguísticos não sejam interpretados de maneira isolada, mas sim a partir de suas inter-relações de forma holística nos textos (BIBER *et al.*, 2021a).

Segundo o modelo estatístico, quanto maior a média do “givenness” da cada sentença do texto (Isa_givenness_mean) maior a probabilidade de o texto ser complexo. O índice de “givenness” calcula a similaridade do trecho analisado com o contexto que o precede. A medida de desvio padrão de similaridade entre pares de sentenças adjacentes no texto (Isa_adj_std), outra medida de coesão semântica, também favorece maior complexidade, como o quadro mostra. Adicionalmente, dentro deste grupo, a métrica de entropia cruzada foi indicada como preditiva de complexidade textual alta. Esta métrica avalia o grau de desvio das combinações das palavras nas sentenças em relação ao modelo de língua estatístico treinado. Um maior grau de “surpresa” do modelo diante da sequência de palavras indica uma complexidade textual mais elevada. Em nosso estudo, a métrica de entropia cruzada foi considerada válida, visto que o gênero jornalístico está incluído entre os gêneros utilizados para o treinamento do modelo.

O desvio padrão mais alto nesses índices indica uma menor uniformidade desses elementos coesivos. O índice de entropia cruzada sugere que os textos do PorSimples causam mais estranhamento ao modelo de língua estatístico treinado pelo algoritmo, apresentando mais sentenças com combinações não usuais de palavras na sequência. Os resultados dessa métrica, contudo, podem ser enviesados pela ausência do uso de jornais populares na constituição de *corpora* da língua portuguesa utilizados para o treinamento do algoritmo da ferramenta NILC-Matrix, interferindo no cálculo de “surpresa” identificada pelo algoritmo. Uma questão já apontada por Pasqualini (2018), que destaca que, em geral, os *corpora* de grande porte da língua portuguesa não abrangem textos representativos da diversidade da população brasileira.

Os dois valores com maior força preditiva, que se situam acima de 100% de probabilidade no gráfico de efeito dos índices apresentado anteriormente, pertencem a esse grupo de métricas: “Coesão Semântica”. A Análise Semântica Latente avalia a presença de palavras semanticamente relacionadas no texto em vários níveis: por sentença, por parágrafo e levando em consideração o texto como um todo. A similaridade entre as partes do texto está ligada à manutenção de tópico, que, teoricamente, de acordo com os criadores da ferramenta, está associada a uma leitura menos complexa. O modelo estatístico indicou o índice de média de similaridade entre todos os pares de sentenças no texto (Isa_all_mean) na extrema esquerda, portanto, como indicativo de maior probabilidade de baixa complexidade.

As métricas do grupo “Medidas Descritivas”, identificadas pelo algoritmo Boruta, indicam que a extensão do texto está relacionada com a complexidade textual. Foram indicadas métricas que calculam o número de sentenças, número de parágrafos, quantidade mínima de palavras por sentença, média de palavras por sentença e média de sílabas por palavra. A relação entre a extensão do texto e a complexidade textual também é sinalizada pelas métricas da categoria “Simplicidade Textual”, que indicam a presença de sentenças muito longas no PorSimples e sentenças curtas no PorPopular. Além disso, o índice do grupo “Frequência de palavras” corrobora essa compreensão, através da métrica de média de frequências das palavras de conteúdo mais raras no texto, que é calculada em comparação aos *corpora Corpus Brasileiro* e *Banco do Português*. O modelo apontou ambas as métricas para o PorSimples.

O conjunto dessas métricas se relaciona diretamente ao grupo “Índices de Leiturabilidade”, validando a eficácia desses índices como preditores de complexidade textual, quando considerados parte das 33 métricas preditivas do modelo multivariável. A Estatística de Honoré avalia a densidade lexical através do uso de fórmulas estatísticas mais válidas do que o TTR e o Índice Flesch contabiliza a média de sílabas por palavras e a média de palavras por sentença. O modelo indicou os dois índices como maior probabilidade de o texto ser PorSimples.

Esses dados sugerem que os textos classificados como PorSimples são compostos por uma maior quantidade de sílabas por palavras, palavras por sentenças e sentenças por parágrafos. Simultaneamente, apresentam uma menor quantidade de sentenças e de parágrafos, visto que as sentenças são mais extensas e que os parágrafos são compostos por um número maior de sentenças. Esses dados se alinham aos resultados do Índice Flesch, que também classifica os dois *proxies* em faixas distintas de nível de escolaridade de acordo com esses cálculos. As categorias do Índice Flesch são descritas no quadro abaixo.

Quadro 24 – Níveis do Índice Flesch

Índice Flesch	Nível de dificuldade	Nível de escolaridade indicado
90-100	Muito fáceis	Até o 5º ano do Ensino Fundamental
80-89	Fáceis	Até o 9º ano do Ensino Fundamental
70-79	Razoavelmente fáceis	Até o 9º ano do Ensino Fundamental
60-69	Padrão	Até o 9º ano do Ensino Fundamental
50-59	Razoavelmente difíceis	Cursando Ensino Médio ou universitário
30-49	Difíceis	Ensino Médio ou universitário completo

0-29	Muito difíceis	Áreas acadêmicas específicas
------	----------------	------------------------------

Fonte: Adaptado de Evers (2013, p. 100)¹⁴¹.

A média do PorSimples é 42, situando-o na categoria “Difíceis”, correspondente aos níveis de escolaridade do Ensino Médio ou Superior completo. Já a média do PorPopular é 55, classificando seus textos como “Razoavelmente difíceis”, adequados para estudantes cursando o Ensino Médio ou a universidade.

A categoria “Informações Morfossintáticas de Palavras” sinaliza que a quantidade de palavras etiquetadas como adjetivos e substantivos é mais alta no PorPopular, apontando menor complexidade textual. No entanto, o PorSimples exibe um valor mais alto na proporção de posições, verbos gerais e de verbos no particípio, indicando maior complexidade.

Dentro do grupo “Coesão Referencial”, a média de candidatos a referentes na sentença anterior por pronome demonstrativo anafórico indica que um maior número de candidatos resulta em maior complexidade textual, pois requer que o leitor identifique qual dos substantivos anteriores ao pronome demonstrativo anafórico está sendo referenciado. O modelo apontou que a métrica sugere maior probabilidade de classificar o texto como PorPopular, indicando uma observação mais alta de candidatos a referente na sentença que antecede o uso do pronome demonstrativo anafórico.

Na categoria “Complexidade Sintática”, o modelo apontou três métricas. A primeira está relacionada à proporção de orações subordinadas em geral. A segunda indica orações subordinadas relativas iniciadas por “cujo”, “que”, “o qual”, etc. Ambas exigem mais esforço cognitivo para seu processamento. E a última é referente à quantidade média de palavras antes do verbo principal da oração, que requer um maior esforço de memória durante a leitura. Os primeiros índices apresentam um valor mais alto no PorSimples, corroborando a característica de que esses textos, apesar de exibirem uma quantidade mais baixa de sentenças, essas sentenças são mais extensas. Entretanto, a quantidade média de palavras antes dos verbos principais das orações principais das sentenças foi indicada pelo modelo como demonstrando uma probabilidade levemente maior de o texto ser classificado como PorPopular.

O resultado do modelo Boruta apresentou um conjunto de 33 métricas preditivas que abrangem aspectos relacionados à semântica, morfossintaxe de palavras e orações, coesão referencial e extensão do texto, mostrando a inter-relação multifacetada do processo de

¹⁴¹ O texto foi atualizado para corresponder ao novo sistema escolar brasileiro de nove anos de Ensino Fundamental.

avaliação da complexidade textual. Nossa análise corrobora resultados de pesquisas prévias discutidas anteriormente, como quanto mais alto o número de sílabas por palavras, palavras por sentença e sentenças por parágrafos, mais complexo o texto. Assim como a presença de sentenças subordinadas, palavras de uso raro e densidade lexical.

Embora a interação dinâmica entre esses diversos traços linguísticos que ocorrem no texto não seja facilmente observável pelo olhar humano, é importante buscar compreender de que modo eles se inter-relacionam de forma holística nos textos analisados. Para tanto, é crucial realizar análises de trechos dos textos a fim de embasar as análises e asserções, uma vez que entende-se que os dados quantitativos devem ser observados dentro de seus contextos discursivos (BIBER *et al.*, 2021a). Com o objetivo de realizar uma análise qualitativa, retomamos o exemplo de texto do *corpus* PorSimples apresentado anteriormente na seção 6.1.

Quadro 25 – Exemplo de análise de texto do corpus PorSimples

O trabalho de pesquisadores da USP está revelando uma série de novas espécies de um tipo todo especial de fungo: Pequenos cogumelos que emitem uma misteriosa luminosidade verde no escuro. As criaturas, antes desconhecidas no Brasil, podem ajudar a elucidar o mecanismo bioquímico que leva à produção de luz em fungos. Além disso, com um pouco mais de estudo, poderiam servir como sensores vivos de poluição ou mesmo fontes de moléculas úteis para a biotecnologia.

Segundo Cassius Vinicius Stevani, químico da USP que coordena os estudos, é possível que o material recolhido abranja pelo menos dez espécies novas. Não é pouca coisa, já que no mundo todo se conhecem só 42 espécies do fungo, quase todas restritas ao Sudeste Asiático. Já temos uma descrita, a *Gerronema viridilucens*, e seis já estão em fase de descrição, contou Stevani à Folha.

Mistério antigo.

Em fungos, a bioluminescência, trocando em miúdos, a luminosidade produzida por seres vivos, é um fenômeno conhecido há milênios, tendo sido descrito pelo filósofo grego Aristóteles (século 4º a.C.). Os cientistas, no entanto, continuam sem saber como ela funciona. Foi na esperança de entender o fenômeno que Stevani e seus colaboradores se puseram a procurar cogumelos do tipo no Brasil. É provável que nenhum pesquisador tivesse identificado espécies no Brasil porque é preciso procurá-los à noite. Além do mais, a luz é muito fraca, até sob o luar, explica o químico.

Com base em relatos de moradores e muita paciência na coleta, a equipe chegou aos cogumelos em dois locais: O Parque Estadual Turístico do Alto Ribeira, em Iporanga (SP) e uma área de mata ciliar no município de Costa Rica, Mato Grosso do Sul. Outros relatos sugerem que fungos do mesmo tipo podem ser encontrados no Tocantins e no Amazonas. A identificação e descrição está sendo feita com a ajuda de Marina Capelari, do Instituto de Botânica de São Paulo, e Dennis Desjardin, da Universidade Estadual de San Francisco (Unidos). A emissão de luz é efêmera. Os cogumelos, que na verdade são apenas uma espécie de órgão sexual do fungo, duram entre 24 e 48 horas. Eles é que emitem luz, que aparentemente pode vir tanto da parte inferior do chapéu quanto do “cabo” ou do cogumelo inteiro. Algumas espécies surgem da casca de árvores vivas, enquanto outras se alimentam de matéria vegetal em decomposição.

Apesar da rapidez do processo, os pesquisadores descobriram como cultivar o *G. Viridilucens* em laboratório e verificaram que ele pode funcionar como uma espécie de sensor de poluição. É que a luz parece variar de acordo com a concentração de poluentes aplicados sobre o fungo.

Dessa forma, você teria um parâmetro rápido de medir e fácil de avaliar, ao contrário de outros biossensores com fungos, que exigem tempo maior de aplicação e tratamento estatístico mais complicado, explica Stevani. Como os fungos ajudam a reciclar nutrientes, têm um importante papel na manutenção da saúde dos ecossistemas, outra razão para usá-los como indicadores.

Finalmente, com informações mais detalhadas sobre o mecanismo da luminescência, uma série de aplicações úteis pode surgir. É que os estudos de biologia molecular dependem muito de reações que geram luz e de proteínas fluorescentes de diversas cores. Quando um pesquisador quer saber se determinado gene está ativo ou não, por exemplo, gruda nele o trecho de DNA correspondente à GFP (proteína fluorescente verde). Se o

gene for ativado, a proteína “acende”, dedurando-o.
Como as técnicas desse tipo têm limitações, é sempre bom contar com novas opções, avalia Stevani, com a ressalva de que os estudos ainda estão no começo.

Fonte: Elaborado pela autora.

O texto em questão contém 12 parágrafos, 28 sentenças e 588 palavras. Nota-se uma variação nos números de sentenças por parágrafo, variando desde parágrafos que contêm apenas 1 sentença até parágrafos mais longos com 5 sentenças. Apesar de ter uma sentença com apenas duas palavras, percebe-se que a maioria das sentenças são mais extensas, apresentando mais de uma oração, sendo que a mais longa apresenta 44 palavras.

A análise do texto nos permite observar um alto uso de adjetivos, desde o primeiro parágrafo – com termos como “novas”, “especial”, “pequenos”, “misteriosa” e “escuro” – até o último, que utiliza “bom” e “novas”. Adicionalmente, observa-se a utilização de palavras de uso raro ao longo do texto, como os substantivos “bioluminescência” e “biossensores”, os quais também indicam um alto número de sílabas. A presença significativa de preposições é igualmente perceptível durante a leitura.

Da mesma forma, é possível identificar que, em sentenças mais extensas, a quantidade de palavras antecedendo o verbo principal é alta. Por exemplo, a seguinte sentença é constituída por 13 palavras antes do verbo principal e evidencia o uso de orações subordinadas: “Em fungos, a bioluminescência, trocando em miúdos, a luminosidade produzida por seres vivos, é um fenômeno conhecido há milênios, tendo sido descrito pelo filósofo grego Aristóteles (século 4º a.C.)”. Além disso, leitores proficientes podem ser capazes de perceber sentenças com combinações não usuais de palavras (entropia cruzada), como em “Foi na esperança de entender o fenômeno que Stevani e seus colaboradores se puseram a procurar cogumelos do tipo no Brasil”, na qual a típica combinação de sujeito-verbo-objeto não é utilizada. As características associadas às relações semânticas e similaridade entre palavras podem ser as mais desafiadoras de identificar durante a leitura, porém, no texto que analisamos, percebe-se a manutenção da temática, que ao longo do texto vai gradualmente sendo aprofundada com a introdução de novo conteúdo informacional.

Esta breve análise do texto não apenas evidencia a acurácia do algoritmo Boruta em automaticamente identificar características linguísticas que efetivamente coocorrem nele, mas também mostra que algumas características podem ser qualitativamente examinadas durante a leitura cuidadosa na seleção de textos para compor textos de insumo de tarefas que integram leitura e escrita. Além disso, reiteramos que, embora tenhamos explicado e exemplificado

alguns dos 33 índices de maneira individualizada, eles não são preditivos de complexidade textual quando analisados isoladamente e sem uma análise qualitativa complementar.

7.3 CONCLUSÕES ACERCA DO CONTEÚDO DAS TAREFAS E DA DIFICULDADE TEXTUAL

O estudo descritivo apontou que os textos utilizados como texto de insumo de leitura nas tarefas integradas da PE do Celpe-Bras apresentam variação ampla em relação ao número de palavras por texto. Isso indica uma falta de padronização da extensão dos textos selecionados para compor as tarefas. Mesmo após a exclusão de *outliers*, o *corpus* de estudo evidenciou média de palavras por texto mais alta do que os *corpora* de referência, o que sugere que os textos utilizados nas tarefas do Celpe-Bras tendem a ser mais longos do que os textos dos jornais populares e dos jornais de referência que compõem esses dois *corpora*. Além disso, em comparação, o *corpus* Celpe é constituído de textos mais longos, aproximando-se do *corpus* PorPopular, porém, mais densos lexicalmente, assim como os textos que compõem o PorSimples.

No que tange à densidade lexical, as análises qualitativas da tarefa 4 de 2016-2, cujo texto de insumo foi uma crônica, mostraram que somente análises lexicais são insuficientes para informar sobre a complexidade textual de textos de forma isolada. Nesse texto, caracterizado por narrativas reflexivas sobre aspectos do dia a dia, a análise lexical indicou a manutenção de tópico e o uso de vocabulário do cotidiano, mas isso não possibilitou embasar asserções sobre complexidade textual. Mostramos que, para além do texto de insumo, a dificuldade textual se relaciona com o enunciado, considerando o gênero discursivo e os objetivos de leitura e de produção, que, em conjunto, influenciam a complexidade da tarefa e, conseqüentemente, o desempenho dos candidatos. As análises realizadas ressaltaram a importância de testar tarefas de forma a evitar incongruências e ambigüidades, bem como garantir um controle mais eficaz das interpretações não previstas pelos elaboradores. A pilotagem das tarefas pode fornecer indicativos relacionados à complexidade das tarefas, que, como evidenciado, só pode ser mensurada quando as tarefas são efetivamente realizadas.

Tendo em vista a ampla variação do número de palavras por texto evidenciada pelos dados, nossas análises igualmente indicaram a importância de definir critérios mais precisos sobre a seleção de textos de insumo. Isso deve considerar aspectos como o número de textos

por tarefa, uma contagem média de palavras por texto e a combinação de textos verbais e não verbais.

Com base nos resultados da revisão bibliográfica e da análise dos dados, consideramos que o uso de apenas um texto por tarefa seria o mais indicado, visto que o uso de dois textos na mesma tarefa tende a aumentar consideravelmente o número de palavras. Ademais, recomendamos a padronização de uso de textos compostos por linguagem verbal e não verbal em todas as tarefas, visto que as especificações do exame indicam a avaliação da leitura multimodal e as diferenças entre edições deve ser evitada. É fundamental que as especificações também forneçam detalhes sobre as características desses textos não verbais, a fim de orientar a seleção de textos que contribuam para a construção dos sentidos pretendidos na leitura e não o contrário. Além disso, com base em nossos dados, a média de palavras em nossos três *corpora* de estudo oscila aproximadamente entre 400 e 550 palavras por texto, uma medida que poderia ser adotada como parâmetro referencial de extensão dos textos a serem selecionados.

Quanto à diagramação das tarefas, nossos resultados indicam a necessidade de especificações que englobem detalhes técnicos relacionados aos aspectos tipográficos, a fim de reduzir a ampla variação observada entre uma edição e outra. Embora o ideal fosse contar com a consultoria de um especialista da área de diagramação para definir parâmetros e criar um projeto gráfico adequado para a leitura em um teste como o Celpe-Bras, algumas características se destacaram em nossas análises. Sugerimos atenção à orientação do texto, uso de cores, fontes, margens e espaçamentos, que são algumas características que podem ser padronizadas para garantir que a diagramação não afete a avaliação da leitura e o desempenho do candidato, mantendo equivalência entre as edições.

Os enunciados também podem ser elaborados para evitar ambiguidades e incongruências que prejudiquem o desempenho do candidato. Dado que as tarefas da PE são compostas por textos de insumo de modalidades distintas e, portanto, envolvem diferentes compreensões, seria interessante que as especificações do exame explicitassem conjuntos de propósitos de escrita que demandassem diferentes sub-habilidades de compreensão para cada uma das tarefas. Além disso, considerando que os gêneros discursivos influenciam na complexidade textual, sugere-se combinar diferentes gêneros na composição das edições. Por exemplo, as tarefas 3 e 4 apresentarem textos de insumo de diferentes gêneros discursivos na mesma edição; que os gêneros discursivos selecionados não pertençam à mesma esfera de comunicação; ou ainda, que os dois textos não apresentem o mesmo tipo textual predominante,

como dois textos narrativos¹⁴². Dessa forma, seria possível garantir uma maior harmonia na distribuição das sub-habilidades de compreensão, bem como a equivalência na complexidade textual de gêneros discursivos de leitura em cada edição do exame.

Com base nos dados analisados, percebemos que as aplicações das provas do Celpe-Bras eram historicamente bianuais; no entanto, a partir de 2018, observamos uma transição para a possibilidade de aplicações anuais¹⁴³. Vemos que as tarefas são predominantemente compostas por um único texto de insumo, que é reproduzido integralmente, semelhante ao original. Além disso, na maioria das vezes, o texto escrito é acompanhado por textos não verbais e as provas têm sido produzidas sistematicamente em impressão colorida desde 2014.

Os dados permitiram observar algumas mudanças nas características das tarefas a partir de 2010, aproximadamente. Até esse período, os gêneros dos textos de insumo eram mais diversificados na composição das tarefas 3 e 4. Após esse período, as tarefas passaram a utilizar “reportagem”, “artigo de opinião”, “crônica” ou “notícia”. No entanto, a análise dos dados do grupo Avalia mostrou que a partir dessa mesma época, as práticas comunicativas projetadas para a produção textual nos enunciados passaram a ser mais variadas, incluindo novas esferas de atuação, assim como diferentes propósitos comunicativos e relações de interlocução. A esfera de atuação e a natureza da interlocução *pessoal* deixaram de existir, e esferas relacionadas à comunidade, artes e educação foram introduzidas. O propósito de escrita *aconselhar* deixou de ser uma opção, e, apesar das semelhanças, os propósitos *incentivar* e *indicar* passaram a ser usados. A mesma mudança foi notada no suporte de produção textual, que nos anos mais recentes inclui o uso de novos veículos, como *blog* e *site*.

As análises das categorias do grupo Avalia também mostraram uma predominância da esfera de atuação *jornalística*. As relações de interlocução mais frequentes estão vinculadas à mídia, envolvendo leitores e colaboradores de veículos de comunicação. Isso, por sua vez, reflete na predominância de interlocuções de natureza pública. As naturezas da interlocução restrita e pública mostraram variações bem marcadas nas tarefas em diferentes períodos, quando examinada a distribuição dessa categoria ao longo dos anos de aplicação. Além disso, notamos a predominância da temática *estilo de vida*, e a análise da distribuição de palavras nos textos de insumo por tema pareceu indicar que a temática pode ter alguma influência na extensão do texto

¹⁴² Durante a defesa desta tese, membros da banca examinadora indicaram que tais procedimentos seriam tradicionalmente executados pela equipe responsável pela elaboração, apesar de não constarem publicamente nas especificações do exame. Até o momento da escrita desta tese, não temos conhecimento de pesquisas que tenham verificado os resultados dessa prática.

¹⁴³ Durante a defesa desta tese, membros da banca examinadora explicaram que a pandemia de Covid-19 teria motivado a realização de apenas uma aplicação do exame nos anos de 2020, 2021 e 2022, informação não divulgada publicamente.

selecionado. O propósito comunicativo *posicionar-se* se mostrou dominante na história do exame. Embora muitas das variáveis analisadas indiquem a tendência de interlocuções da esfera jornalística, o suporte de produção textual mais utilizado é *carta/e-mail*, seguido de *carta do leitor*. Dado que esses suportes juntos representam mais da metade da porcentagem dos suportes analisados, esse dado aponta para a necessidade de os professores de cursos preparatórios e candidatos estudarem o gênero discursivo *carta* e suas diversas possibilidades de propósitos, interlocutores e locais de publicação, conforme já apontado por Schoffen e Segat (2020).

As análises das categorias do grupo Avalia se mostraram produtivas para compreender as características das tarefas, especialmente quando observamos as distribuições por ano de aplicação na história do exame. Com base nessas análises, é possível considerar que essas características poderiam ser incluídas nas especificações para a elaboração das tarefas, a fim de que possam ser controladas. Isso permitiria a elaboração de tarefas que, ao longo do tempo, reflitam a variação das práticas discursivas que o exame visa contemplar, sem criar períodos de predominância de apenas algumas delas.

O Documento Base, em nota de rodapé, adiciona ao quadro de especificações da Parte Escrita do Exame Celpe-Bras que

[a]s especificações da parte escrita que constam deste documento, a serem utilizadas a partir de 2020, juntamente com as da Parte Oral, explicitadas mais adiante, foram mais bem detalhadas, buscando refletir o que temos efetivamente realizado ao longo dos anos de implementação do Exame (BRASIL, 2020, p. 58).

Entretanto, as evidências obtidas por nossas análises não corroboram essa afirmação. Os dados revelaram que os gêneros do discurso mais selecionados para a leitura pertencem à esfera jornalística, indicando uma preferência estatisticamente significativa pelo uso de reportagens como texto de insumo durante a história de aplicação do Celpe-Bras. Do mesmo modo, as fontes dos textos de insumo que compõem as tarefas que integram leitura e escrita apontaram para a predominância de meios de comunicação, com porcentagens estatisticamente significativas de uso de jornais e revistas, muitos dos quais são considerados de prestígio nacional, são dirigidos a leitores das classes mais altas e não estão integralmente disponíveis gratuitamente. Observou-se também a ausência de uso de veículos de comunicação do jornalismo popular, bem como dos outros tipos de fontes indicadas nas especificações.

As análises de complexidade textual dos textos de insumo do Celpe-Bras evidenciaram o uso de textos majoritariamente de alta complexidade linguística, equiparando-se aos textos

do *corpus* de referência PorSimples. Ademais, observamos que a seleção de textos para compor as tarefas de leitura e escrita dos últimos cinco anos utilizaram exclusivamente textos com essa característica. Conforme a descrição e a análise do *corpus* PorSimples, a complexidade textual nesse nível não é direcionada ao grau de letramento brasileiro Intermediário, ela é relacionada a leitores na faixa Proficiente do Inaf, das classes A e B, com Ensino Médio a Superior completo. Com base nesse resultado, concluímos que as tarefas que fazem uso de textos de insumo de alta complexidade textual podem dificultar ou até mesmo inviabilizar que candidatos, potencialmente proficientes em nível intermediário do Celpe-Bras, acessem o conteúdo das tarefas e possam respondê-las dentro do tempo previsto. Tal conclusão tem como base a complexidade textual que, como vimos, é um dos fatores que, na inter-relação com a tarefa (propósitos de leitura e de escrita) e o leitor, pode ter um potencial preditivo da dificuldade textual. Nesse sentido, cabe sugerir, a partir dos resultados apresentados aqui, o desenho de estudos que verifiquem, a partir do mesmo enunciado, se e até que ponto textos de insumo de complexidade textual variada podem ter um impacto no desempenho dos candidatos no exame.

A literatura na área de avaliação ressalta a importância da transparência das especificações dos testes, garantindo que estas descrevam precisamente o que e como será avaliado, assegurando a equivalência entre as edições, a fim de evitar prejuízos para os candidatos e garantir a validade do exame. Com base na análise de nossos dados, constituídos pelas tarefas já aplicadas e pelos documentos oficiais, foi possível observar que as especificações carecem de detalhes sobre o processo de seleção de texto de insumo, bem como de informações sobre o processo de elaboração e o perfil dos profissionais envolvidos. Além disso, o conteúdo e as características observados nas tarefas que integram leitura e escrita não correspondem plenamente às especificações, tendo em vista a predominância na leitura de textos da esfera jornalística.

Nossas análises forneceram evidências que não respaldam o argumento de validade de que o conteúdo das tarefas que integram leitura e escrita no Celpe-Bras é representativo das diferentes demandas de leitura que os candidatos podem encontrar nos contextos de suas necessidades de uso da língua-alvo, de acordo com as especificações do exame. Contudo, uma vez que as especificações não são transparentes, pormenorizadas e bem definidas, e que esta investigação observou apenas a parte dos instrumentos de avaliação que integram leitura e escrita, esta validação é inevitavelmente incompleta e fragmentada. Embora as tarefas 1 e 2 não envolvam especificamente a leitura de textos escritos (mas compreensão oral e multimodal),

essa análise pode acrescentar dados importantes em relação à variedade de gêneros discursivos, fontes e também de níveis de complexidade textual dos textos de insumo.

Pelo fato de as especificações do exame e as características dos textos de insumo evidenciadas em nossas análises não se alinharem, torna-se desafiador respaldar o argumento de validade da concepção de leitura avaliada nas tarefas que integram leitura e escrita no Celpe-Bras. Os resultados não parecem refletir a noção de leitura desde uma perspectiva de uso da linguagem em práticas sociais que abrange também a democratização da cultura, crenças, valores, sentimentos, identidades, e a diversidade de interações e ações significativas e socialmente situadas que acontecem em língua portuguesa. Devido à falta de transparência nas especificações do exame, muitas perguntas permanecem sem respostas. O conteúdo é representativo da língua de qual público-alvo? Quando os dados indicam a predominância da temática *estilos de vida*, a qual comunidade eles se referem? Qual o público do exame que constitui o grupo de leitores das revistas *IstoÉ*, *Época* e *Veja* e é esse público específico que é o público-alvo projetado pelo exame?

Conforme amplamente abordado neste trabalho, testes de línguas de alto impacto são instrumentos de políticas linguísticas capazes de instituir determinados usos da linguagem e padrões culturais. O Documento Base enfatiza a relação indissociável entre língua e cultura, um dos princípios fundamentais subjacentes ao construto teórico do exame. O documento diz operacioná-lo “ao incluir, em sua estrutura, gêneros discursivos, verbais e multissemióticos, que circulam em diferentes domínios da vida social brasileira, os quais estão plenos de cultura” (BRASIL, 2020, p. 29).

No entanto, nossas análises das tarefas 3 e 4 revelaram que os textos utilizados como material de leitura não representam diferentes domínios da vida social brasileira e não parecem promover as “culturas brasileiras” plurais, como declarado (BRASIL, 2020, p. 19). Pelo contrário, os dados sugerem que o conteúdo das tarefas que integram leitura e escrita na PE do Celpe-Bras está diretamente relacionado a textos altamente complexos que circulam predominantemente em contextos de prestígio, representativos das classes altas do país e, conseqüentemente, da norma privilegiada dominante, potencialmente excluindo candidatos que são proficientes em outros contextos de letramento. Consoante à perspectiva crítica de testes de línguas, os resultados da análise dos textos de leitura apontam para a imposição de uma língua nomeada dominante e a promoção da supremacia branca e de uma ideologia monolíngue, que nega a diversidade de repertórios comunicativos usados nas práticas letradas em diferentes esferas sociais das quais os candidatos ao exame precisam ou desejam participar.

Levando isso em consideração e visando a construção de argumentos de validade mais consistentes, vemos a necessidade de uma revisão das especificações do exame, que busque definições mais claras quanto aos objetivos e públicos-alvo do exame e que levem em conta, com base em Randall *et al.* (2022), uma abordagem antirracista e orientada para a justiça de testes de línguas de modo a combater a supremacia branca na área de avaliação e a exclusão de grupos minoritarizados e, acrescentamos, o uso exclusivo de textos de alta complexidade textual. Some-se a isso, a composição de uma equipe de elaboradores diversa e representativa da população brasileira, incluindo especialistas no ensino e na avaliação de PLA com experiências diversas e amplas de repertórios comunicativos.

Com base nos resultados desta pesquisa, podemos sugerir as seguintes orientações para as especificações de elaboração de tarefas que integram leitura e escrita no Celpe-Bras:

- estabelecer o perfil dos elaboradores de tarefas para compor uma equipe qualificada, diversificada e coesa;
- capacitar a equipe de elaboradores e monitorar seu desempenho;
- seguir rigorosamente as especificações de elaboração de tarefas;
- selecionar apenas um texto para cada tarefa;
- selecionar textos constituídos de texto verbal e não verbal que se relacionem de maneira articulada e cooperativa para construir um sentido em comum;
- estabelecer o número de palavras por texto, aproximadamente entre 400 e 550 palavras;
- estabelecer um projeto gráfico com critérios fixos relacionados aos aspectos tipográficos;
- detalhar nas especificações quais sub-habilidades são avaliadas nas tarefas que são compostas por textos de insumo de diferentes modalidades (vídeo, áudio e texto escrito);
- não utilizar na mesma edição tarefas que fazem uso de textos de gêneros discursivos da mesma esfera de comunicação;
- não utilizar na mesma edição tarefas que fazem uso de textos com predominância do mesmo tipo textual;
- estabelecer agrupamentos de gêneros discursivos a serem utilizados na mesma edição;
- categorizar as tarefas elaboradas com as características discursivas propostas pelo grupo Avalia de modo a diversificar a composição final das edições;
- utilizar fontes que forneçam textos de insumo que representem a variabilidade dos repertórios comunicativos brasileiros, além de veículos tradicionais da mídia;

- avaliar o nível de complexidade dos textos selecionados por meio de análises quantitativas e qualitativas minuciosas, realizadas por especialistas em tais análises;
- pilotar as tarefas elaboradas para identificar a dificuldade textual e a complexidade da tarefa;
- dar acesso público às especificações sobre as elaborações das tarefas.

8 CONSIDERAÇÕES FINAIS

Esta pesquisa teve como objetivo descrever o conteúdo das tarefas que integram leitura e escrita da Parte Escrita do Celpe-Bras e a complexidade textual de seus textos de insumo. Assim como verificar a capacidade preditiva das métricas automáticas de complexidade textual do NILC-Metrix. A descrição das tarefas e o levantamento de indicadores de complexidade textual nos permitiram obter evidências para verificar o argumento de validade relacionado ao domínio de uso da linguagem e a capacidade preditiva das métricas do NILC-Metrix para distinguir textos jornalísticos por nível de complexidade. O tópico da leitura em testes de alto impacto permeia o objetivo central desta tese, representando a motivação inicial, bem como o incentivo para persistir com nossos propósitos.

Desde a concepção inicial deste projeto de tese, submetido durante o processo seletivo de doutorado, muitos aspectos sofreram modificações, exceto o desejo de fomentar a reflexão e o diálogo acerca da leitura, desde a perspectiva de práticas sócio-histórico-culturalmente situadas, e o poder social dos testes que avaliam essa habilidade.

Associar essa visão de linguagem dialógica com procedimentos metodológicos que envolvem análises linguísticas usando ferramentas automáticas foi um dos nossos primeiros desafios. Na primeira etapa desta pesquisa, conduzimos um levantamento bibliográfico na área de Avaliação de línguas, nos centrando em conceitos-chave fundamentais para o estudo: texto, leitura, letramento, habilidades de leitura, proficiência, avaliação de leitura, argumento de validade, tarefa integrada e complexidade textual. Além disso, procedemos com uma revisão bibliográfica na área de Linguística de *Corpus*, delineando abordagens teóricas e metodologias de pesquisa implementadas tanto internacionalmente quanto no Brasil, com especial ênfase na análise da complexidade textual e dificuldade textual. O cruzamento de fronteiras entre essas duas áreas da Linguística Aplicada implicou que a etapa de revisão de literatura se tornasse uma investigação ampla e pormenorizada, resultando na identificação da esfera de estudos resultante da fusão desses dois campos teóricos: a Avaliação de línguas informada por corpus.

Ao identificarmos uma significativa lacuna em estudos brasileiros nessa zona de interseção teórica, nos deparamos com o seguinte questionamento: quem são nossos interlocutores? Seguras de que esta é uma área de estudos promissora, que vem demonstrando significativos avanços em pesquisas internacionais, optamos por projetar um público-alvo possível, composto tanto por especialistas em avaliação interessados nas amplas possibilidades de investigações proporcionadas pela Linguística de *Corpus*, quanto por linguistas de *corpus*

interessados em investigar *corpora* de testes, especialmente o exame Celpe-Bras. Assim, buscamos aprofundar ambos os campos de conhecimento, resultando em uma revisão bibliográfica mais abrangente do que inicialmente havíamos planejado, contudo, focada nas temáticas da nossa investigação.

Durante a fase de compilação e seleção dos *corpora* para o estudo, enfrentamos diversas indecisões e ajustes; diferentes *corpora* foram inicialmente adotados e, mediante os resultados de avaliações preliminares, foram posteriormente desconsiderados, ao percebermos que tais *corpora* não satisfaziam as exigências de nossa pesquisa. Entre as decisões relacionadas às exclusões, a mais difícil foi desconsiderar o *corpus* de referência constituído pelas tarefas de redação do Enem. Enquanto docente, parecia vital fomentar a discussão sobre a leitura no maior exame do Brasil, um teste de larga escala e grande impacto social, que também funciona como mecanismo de acesso à educação superior. Esse interesse era estimulado sobretudo pelo desejo de compreender a concepção de leitura subjacente à prática de ler trechos fragmentados de textos motivadores para escrever uma redação, bem como a relação dessa prática avaliativa com os parâmetros curriculares que orientam o ensino da educação básica, o qual o Enem visa avaliar. Por outro lado, enquanto pesquisadora, entendemos a necessidade de limitar o escopo desta pesquisa que, devido à revisão bibliográfica, já se mostrava mais extensa do que havíamos pretendido.

Do mesmo modo, a etapa de análise dos dados se revelou extremamente desafiadora, uma vez que, conforme salientado previamente, a nossa formação em Linguística Aplicada não nos proporciona o letramento necessário para o uso de linguagens de programação para a condução de análises estatísticas de dados empíricos. A quantidade de conhecimento que adquirimos durante esta trajetória é imensa, conhecimentos que não só expandem nossas capacidades investigativas e ampliam nossas percepções sobre a linguagem, mas também nos aprimoram como professora formadora de professores e especialista em avaliação.

Nossos resultados nos permitiram responder às nossas perguntas de pesquisa que retomamos aqui:

1. Em que medida os padrões de características observados nas tarefas que integram leitura e escrita na PE do Celpe-Bras estão em consonância com as especificações publicadas no Documento Base?

Os nossos resultados indicaram uma falta de diversidade na seleção de gêneros discursivos e fontes para a composição das tarefas 3 e 4. Como mencionado previamente, a não ser que exista alguma orientação específica (que desconhecemos) em relação a essa escolha, é possível presumir que a seleção de gêneros e fontes possa estar relacionada à facilidade de acesso e à familiaridade da equipe responsável pela elaboração com esses materiais. Sobretudo ao longo da última década, as tarefas têm sido compostas por textos provenientes exclusivamente de revistas e jornais nacionais de prestígio, apresentando textos dos gêneros discursivos “artigo de opinião”, “crônica”, “notícia” e, sobretudo, “reportagem” para a avaliação integrada de leitura e escrita da Parte Escrita. A menos que se prove o contrário, as evidências resultantes das análises dos textos de insumo das tarefas que integram leitura e escrita, junto aos documentos referentes ao exame, não respaldam o argumento de validade de que o conteúdo das tarefas está em consonância com as especificações do próprio exame que constam no Documento Base.

2. Quais métricas do NILC-Matrix apresentam capacidade preditiva de complexidade textual em textos da esfera jornalística?

Os resultados das nossas análises indicaram que 33 das 200 métricas do NILC-Matrix são preditivas de complexidade textual dos textos jornalísticos dos *corpora* de referência que, nesta etapa da pesquisa, foram utilizados como *proxy* de complexidade textual alta e baixa. A acurácia da capacidade de distinção dessas métricas foi testada e o modelo de classificação binária demonstrou uma precisão de 100% na classificação automática de textos por nível de complexidade textual. Essas 33 métricas foram identificadas pelo algoritmo Boruta como um conjunto de traços linguísticos que coocorrem e se inter-relacionam nos textos que constituem nossos *corpora* de textos da esfera jornalística, permitindo a distinção de graus de complexidade textual. As características linguísticas significativas identificadas por essas métricas se relacionam a aspectos semânticos, morfossintáticos de palavras e orações, coesão referencial e extensão textual. Algumas dessas características foram identificadas mediante análise qualitativa de um exemplar de texto. No entanto, o conjunto de 33 métricas preditivas não é generalizável para textos de outros gêneros discursivos ou esferas de comunicação, nem podem ser consideradas individualmente como indicadoras de complexidade textual.

3. De acordo com as métricas do NILC-Metrix e os *corpora* de referência, qual é o nível de complexidade textual observado nos textos de insumo das tarefas que integram leitura e escrita na PE do Celpe-Bras, em relação ao nível de leitura brasileiro, de acordo com o Indicador de Alfabetismo Funcional?

A maioria dos textos do *corpus* de estudo (67% do total) foram classificados em nível de complexidade textual alta, especialmente nos últimos anos de aplicação, nos quais os textos selecionados foram exclusivamente categorizados como altamente complexos. A complexidade textual observada na maior parte dos textos de insumo das tarefas de leitura e escrita no exame Celpe-Bras é correspondente ao grau mais alto de letramento brasileiro do Inaf; leitores na faixa Proficiente, das estratificações socioeconômicas A e B, com Ensino Médio a Superior completo, que representam aproximadamente 3% da população brasileira.

Como contribuições desta pesquisa apontamos as sugestões de orientação para o processo de elaboração das tarefas e sugestão de nível de complexidade textual que permita o acesso de candidatos em nível intermediário. Além disso, os resultados desta pesquisa fornecem indicadores que permitem a condução de pesquisas experimentais que busquem avaliar o impacto da complexidade textual dos textos de insumo e das características da tarefa na complexidade da tarefa e no desempenho de leitura dos examinandos. Os resultados de uma pesquisa desta natureza podem informar tanto o processo de elaboração das tarefas do exame, como os professores de PLA sobre a seleção de textos e a elaboração de tarefas para compor materiais didáticos. Tais resultados têm o potencial de possibilitar a criação de edições do Celpe-Bras estáveis e equivalentes, contribuindo para conferir validade às inferências feitas a partir das notas das provas. Além disso, com devidos ajustes, a metodologia proposta nesta tese pode ser replicada para descrever as outras duas tarefas da PE, possibilitando uma maior compreensão das composições das edições do Celpe-Bras. Com o apoio da abordagem da Avaliação de línguas informada por corpus, a nossa metodologia de análise multivariável pode ser utilizada para investigar padrões que distinguem respostas de candidatos certificados em diferentes níveis do Celpe-Bras, dando continuidade às discussões iniciadas por pesquisas anteriores. Da mesma forma, a metodologia pode ser ajustada para analisar outros testes de línguas, como foi atestado por nosso processo inicial de análise de textos de insumo da tarefa de redação do Enem.

Por se tratar de uma área relativamente nova no Brasil, consideramos como uma contribuição para a área da Linguística Aplicada a revisão bibliográfica realizada do campo de

estudos de Avaliação de línguas informada por corpus, uma vez que ela fornece subsídios para futuros pesquisadores interessados em realizar investigações a partir desta perspectiva interdisciplinar. Os resultados deste estudo informam ainda os pesquisadores interessados em utilizar as métricas do NILC-Metrix com o objetivo de analisar a complexidade textual de textos em língua portuguesa. Nossas análises evidenciaram que a complexidade textual é multifacetada e multidimensional. Assim, é possível que nem todas as 200 métricas se revelem produtivas na predição da complexidade de diferentes gêneros discursivos e, por isso, consideramos essencial que análises estatísticas de seleção de métricas sejam conduzidas e complementadas por análises qualitativas de trechos dos textos.

Como produto complementar desta tese, disponibilizamos o *corpus* de estudo na página <https://picoral.shinyapps.io/insumo-celpe/> para acesso aberto para que pesquisadores, professores e estudantes interessados possam acessar os textos de insumo do Celpe-Bras em formato .txt, assim como realizar buscas por palavras-chave.

Uma limitação desta pesquisa está relacionada ao uso de *corpora* restritos a poucos veículos de comunicação que não podem ser considerados representativos da esfera jornalística brasileira. Além disso, o fato de nossos *corpora* de referência terem sido compilados por outros pesquisadores limitou a nossa compreensão mais aprofundada sobre suas constituições e os procedimentos adotados para suas compilações. Similarmente, outra limitação do estudo é o uso do indicador de alfabetismo Inaf, que é uma pesquisa de baixa representatividade, apenas 2.002 participantes, e baixa transparência em relação ao processo de coleta de dados. Em relação ao *corpus* de estudo, apesar de o foco de nossa investigação ter sido a leitura nas tarefas da PE que integram leitura e escrita, a investigação de apenas essa parte individualizada dos instrumentos de avaliação não fornece resultados generalizáveis para a avaliação do argumento de validade do exame na íntegra. Além disso, houve a tentativa de poder utilizar o código aberto da ferramenta NILC-Metrix em nosso próprio servidor que não foi concretizada.

Entretanto, todo o processo de investigação desenvolvido para a realização desta tese nos leva a concluir que seu principal mérito é fornecer evidências para os agentes envolvidos com o Celpe-Bras preocupados em fornecer maior transparência dos objetivos e procedimentos avaliativos do exame. Ademais, é importante reiterar que o órgão responsável pelo Celpe-Bras negou a existência de especificações relativas à elaboração das tarefas da Parte Escrita. Ao considerarmos, de maneira hipotética, a inexistência de especificações detalhadas que orientam o processo de elaboração de tarefas e a seleção de material de leitura, conforme declarado pelo Inep, a validade do exame estaria irremediavelmente comprometida. Em contrapartida, o fato

de que esse documento existe, mas sua existência foi omitida, não apenas coloca em xeque a transparência do exame, mas também suscita preocupações éticas e morais referentes à integridade dessa instituição e à violação do direito constitucional dos cidadãos de acessar informações de caráter público¹⁴⁴.

O Documento Base garante que:

[...] as especificações, elaboradas segundo a análise de necessidades do público-alvo, são componente fundamental de um exame. Uma vez definido o conteúdo do exame, com base nessas necessidades, são as especificações que relacionam as referidas necessidades aos itens do exame Celpe-Bras (BRASIL, 2020, p.57).

Essa afirmação está alinhada com as recomendações da literatura na área, que preconiza que as avaliações devem apresentar evidências empíricas que sustentem a validade do teste em simular o uso da língua-alvo de seu público. Entretanto, nossa pesquisa revelou que a análise das necessidades de uso da língua pelo público-alvo foi realizada na primeira metade da década de 1990, quando o público-alvo do exame era composto por estudantes não-lusófonos candidatos a ingressar em IES brasileiras.

Constatamos não apenas que o Celpe-Bras, ao longo do tempo, passou a ser utilizado para finalidades distintas das inicialmente pretendidas pelos criadores, mas também que o público-alvo foi sendo ampliado. Isso não seria uma ameaça para a validade do teste caso análises tivessem sido conduzidas sistematicamente de forma a gerar novas evidências de que os instrumentos de avaliação estão adequados aos usos pretendidos de seus resultados, fundamentam as suas interpretações e operacionalizam adequadamente seu construto teórico. Por outro lado, se os usos dos resultados e públicos-alvo incluídos posteriormente não foram planejados e não estão de acordo com os propósitos do exame e seus instrumentos de avaliação, entendemos que os documentos oficiais não devem respaldá-los, especialmente não devem incluí-los nas especificações.

Nosso estudo destaca a necessidade de que os responsáveis pelo Celpe-Bras conduzam novas pesquisas de validação do exame e sejam transparentes e objetivos em relação ao conteúdo das tarefas de leitura e escrita. Para atingir esse objetivo, inicialmente é fundamental estabelecer precisamente o público-alvo e a finalidade do exame, aspectos que não estão claramente explícitos nas especificações mais atuais.

¹⁴⁴ Lei nº 12.527, de 18 de novembro de 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm Acesso em: 23 set. 2023.

É fundamental replicar todos os procedimentos adotados pela primeira comissão técnica, que garantiram a validade e a confiabilidade do teste em sua criação. Não apenas devido às ampliações dos usos dos resultados e do público-alvo, mas também porque os usos da linguagem estão em constante mudança, alterações significativas foram impulsionadas pelo avanço da tecnologia. Assim, ajustes e refinamentos sistemáticos do construto teórico e sua operacionalização devem ser realizados de modo a assegurar a constante validade do teste. Conforme abordado anteriormente, os hábitos e as necessidades de leitura contemporâneos, indiscutivelmente, se diferem daqueles do final do século passado. Embora estejamos cientes de que tais procedimentos são custosos e demorados, e demandem o engajamento de uma ampla equipe de especialistas, trata-se de uma imensa responsabilidade social garantir que sejam realizados.

Uma vez definidos com precisão o público-alvo e o propósito do exame, análises minuciosas e rigorosas da língua-alvo podem ser realizadas por meio de uma abordagem da Avaliação de línguas informada por corpus. Isso permitirá identificar características dos repertórios linguísticos relativos às necessidades de uso da língua do público-alvo. Análises qualitativas podem ser realizadas para identificar quais leituras esse público realiza, quais suas motivações, de que modo se dá essa leitura e de que maneiras reagem a ela, levando em consideração o atual período de inevitabilidade da tecnologia. As evidências empíricas e atualizadas, junto da experiência dos especialistas, devem embasar as tomadas de decisões para a atualização e o detalhamento das especificações relativas ao processo de elaboração das tarefas.

Conforme a revisão bibliográfica discutida neste trabalho, a criação do exame Celpe-Bras teve duas motivações principais: a necessidade de avaliar o potencial de estudantes estrangeiros para participar das práticas universitárias que se dão por meio da língua portuguesa nas IES brasileiras e o redirecionamento do ensino de PLA.

Os especialistas envolvidos no desenho do exame previam uma mudança no ensino, passando de um foco em regras gramaticais e vocabulário para um ensino mais orientado ao uso da linguagem, incentivando os estudantes a interagir em uma variedade maior de contextos sociais por meio de diferentes gêneros discursivos. O exame Celpe-Bras, enquanto um mecanismo de política linguística, tem desempenhado esse papel estratégico com sucesso na qualificação de educadores, pesquisadores e estudantes, servindo como referência para o ensino e a elaboração de currículos desde a sua implementação. Além de ter desempenhado um papel

importante na promoção do letramento em avaliação dos profissionais envolvidos diretamente com o exame.

Tendo em vista o efeito retroativo positivo desejado, além das necessidades de uso da língua pelo público-alvo inicial, uma ampla lista de gêneros discursivos está presente nas especificações do exame desde a primeira publicação do manual do candidato, em 2002, conforme abordado previamente. No entanto, nossos dados indicam que, ao longo dos vinte e cinco anos de aplicação do exame, os gêneros discursivos efetivamente utilizados para a leitura tendem a ser gêneros jornalísticos. Ademais, a partir de 2010, limitam-se a artigos de opinião, crônicas, notícias e, especialmente, reportagens¹⁴⁵; e desde 2015, esses textos são altamente complexos, produzidos e direcionados, em sua maioria, para leitores brasileiros no nível mais alto de letramento do Inaf.

Conforme discutido anteriormente, a certificação do Celpe-Bras tem poder decisório na vida de muitos examinandos, cujas vidas e direitos são controlados pelos resultados desse exame. Portanto, os candidatos que necessitam da certificação no Celpe-Bras têm o direito de estar cientes de que uma preparação eficaz para a leitura deve incluir uma atenção especial à leitura de jornais e revistas de prestígio nacional. Isso não descarta a importância de explorar a leitura de quadrinhos, poemas e letras de canções durante seus estudos da língua portuguesa. Cientes de possíveis implicações aos direcionamentos didático-pedagógicos para o ensino e a avaliação de PLA nas salas de aulas, no entanto, consideramos ser fundamental que candidatos e professores de cursos preparatórios compreendam que a maioria dos gêneros discursivos listados nas especificações do exame não fazem parte dos requisitos de leitura da PE do Celpe-Bras.

Após uma cuidadosa consideração das discussões teóricas realizadas, depreendemos que o propósito principal de um teste de língua deve ser avaliar a proficiência no uso da língua-alvo que seja representativa das necessidades linguísticas do público-alvo, uma vez que o agente mais importante nesse contexto é o candidato. Com isso queremos dizer que cabe aos legisladores definir parâmetros curriculares que orientem o ensino de PLA, assegurando a presença de gêneros discursivos nos materiais didáticos e garantindo que o uso significativo e socialmente situado da linguagem seja o objetivo central da aprendizagem¹⁴⁶. Não cabe mais

¹⁴⁵ Com exceção da tarefa 4 de 2014-2, que utilizou duas entrevistas como texto de insumo.

¹⁴⁶ A exemplo da recente coleção “Propostas curriculares para o ensino de português no exterior”, publicada em 2020, de iniciativa do Ministério das Relações Exteriores brasileiro. Disponível em: https://funag.gov.br/biblioteca-nova/produto/21-21-propostas_curriculares_para_ensino_de_portugues_no_exterior_sete_volumes_ Acesso em: 10 out. 2023.

ao Celpe-Bras manter em suas especificações gêneros discursivos que, embora sirvam como direcionadores didáticos, comprometem a transparência e a validade do teste.

Embora enquanto docentes desejemos que o ensino e aprendizagem de PLA inclua as diversas linguagens, especialmente as linguagens artísticas, e nossa ampla diversidade cultural brasileira, como pesquisadoras, nos sentimos na obrigação de assumir uma consciência moral e uma responsabilidade social e política, apontando a falta de transparência e nos opondo aos usos indevidos do exame Celpe-Bras que prejudicam os examinandos.

Considerando as observações de especialistas em avaliação de que muitos testes acabam sendo utilizados como instrumentos de políticas sociais, a proficiência no uso das línguas que esses testes avaliam frequentemente não é essencial para uma participação social funcional e não deriva de resultados de pesquisas, mas sim é determinada pela facilidade em serem testadas. Devido à falta de transparência das especificações do exame em relação aos textos de insumo e às habilidades de leitura focalizadas, torna-se desafiador compreender, a partir das descrições existentes do conteúdo das tarefas do exame, de que maneira a leitura de artigos de opinião, crônicas, notícias e reportagens são essenciais ou suficientes para que o Celpe-Bras avalie “o desempenho em relação ao uso da língua portuguesa em diferentes práticas de linguagem de que seu público-alvo necessita participar, no Brasil ou no exterior” (BRASIL, 2020, p. 26).

Por fim, outra questão importante a ser ressaltada em relação a essa funcionalidade social das línguas é o fato de que, durante o processo desta pesquisa, identificamos que as quatro famílias de habilidades de leitura subjacentes às sub-habilidades avaliadas na PE do exame Celpe-Bras são obrigatórias para a participação no exame. Apesar da complexidade em separar tais sub-habilidades, como discutido anteriormente, decodificar o código, participar da compreensão, usar funcionalmente e analisar criticamente o texto são práticas dinâmicas, independentes entre si, mas não autossuficientes individualmente para assegurar o acesso às outras famílias ou ao letramento do sujeito. Portanto, mesmo que pessoas incapazes de decodificar o código dominem as outras três habilidades envolvidas na compreensão de textos e sejam capazes de participar ativamente de interações sociais que se dão em torno da língua escrita, elas são excluídas do exame. Diferente de outros testes brasileiros, no Celpe-Bras, embora seja possível solicitar atendimento especializado para pessoas com baixa visão, cegueira, visão monocular, deficiência física, deficiência auditiva, surdez, deficiência intelectual (mental), surdocegueira, dislexia, déficit de atenção, transtorno do espectro autista, gestante, lactante, idoso e/ou pessoa com outra condição específica, o acesso a leitor e transcritor ainda não é uma opção. Além disso, o exame não oferece correção diferenciada para

as pessoas que a necessite. O edital do Celpe-Bras apresenta recursos que podem ser solicitados para atendimento especializado, mas não inclui o auxílio de leitor e transcritor como opções disponíveis, o que se configura como um obstáculo e, potencialmente, torna o exame inacessível para alguns indivíduos com essas condições¹⁴⁷.

Concluimos estas considerações finais nos questionando, “o quê?”, “por quê”, “de que modo?”, “para quê?” e “para quem?” estamos avaliando o que avaliamos, ao afirmarmos estar avaliando o uso adequado da língua para agir no mundo?

Conforme anunciamos na introdução, considerando as consequências de testes de alto impacto na vida das pessoas, é fundamental que eles apresentem argumentos de validade, isto é, que possam explicar mediante evidências e base teórica se e até que ponto as inferências sobre o significado dos resultados são verdadeiras (FULCHER; OWEN, 2016). São vários os tipos de evidência que têm sido focalizados para construir a validação de testes, que, conforme vimos, incluem fundamentação e análises de especialistas sobre o conteúdo, estudos sobre o processo de construção de respostas pelos examinandos, análise estatística dos dados das respostas dos examinandos, relação do construto com outras variáveis, consequências dos testes (CHAPELLE; LEE, 2021).

É interessante observar que, embora a relação entre o domínio de uso da linguagem testada e seu uso em contextos externos ao exame seja um dos tipos fundamentais para a validade, conforme discutimos aqui, nos parece que, em geral, as pesquisas não questionam a seleção dos usos da linguagem fonte para os testes. Com base nos resultados apresentados e discutidos nesta tese, nos alinhamos com Cushman (2016), Randall (2021) e Randall *et al.* (2022), que afirmam que os dados que tradicionalmente têm sido usados para a validação de exames de larga escala naturalizam iniquidades históricas, pois estão fundamentados em domínios de uso da linguagem, valores e identidades hegemônicas brancas. Na proposta de um modelo de validade antirracista orientado para a justiça social, Randall *et al.* (2022) consideram fundamental mudar as perguntas sobre validade para incluir uma perspectiva crítica que confronte sistemas de opressão e de exclusão, o que deve incluir uma autorreflexão e uma análise sobre as ideologias de linguagem e relações de poder, começando-se com as dos próprios agentes responsáveis pela construção do teste. Esse deslocamento inclui mudar as perguntas de validade relacionadas à articulação do construto, ao conteúdo, às consequências,

¹⁴⁷ Conferimos o edital mais recente de abertura de inscrições, edital nº 8, de 1º de fevereiro de 2023. Disponível em: <https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2023/02/EDITAL-No-8-DE-1o-DE-FEVEREIRO-DE-2023-EDITAL-No-8-DE-1o-DE-FEVEREIRO-DE-2023-DOU-Imprensa-Nacional.pdf> Acesso em: 27 set. 2023.

aos processos de respostas dos examinandos, à estrutura interna do exame e a outras variáveis. Para todas as perguntas tradicionalmente utilizadas, tais como se o conteúdo do teste representa o domínio de uso da linguagem pretendido, se as tarefas funcionam de modo diferente para públicos-alvo diversos ou se as interpretações das tarefas pelos examinandos respondem às expectativas, interessa saber quais parâmetros de uso da linguagem estão sendo utilizados, de quais segmentos sociais, quem é privilegiado e quem é excluído e marginalizado a partir das escolhas feitas. E para além de alterar as perguntas, é fundamental explicitar, de modo detalhado e transparente, nas especificações do exame, de que modo essas perguntas foram respondidas no processo de validação.

Posfácio

Depois de sessenta anos desde sua chegada ao Brasil, Shigeru Nagasawa, agora com oitenta anos, percebe, dia após dia, que as palavras, outrora rabiscadas em seu caderninho, estão, aos poucos, sendo apagadas de sua memória. A frase que muitas vezes lhe proporcionou proteção, “não sei português”, hoje lhe provoca aflição em momentos em que deseja se expressar nessa língua, mas as palavras teimam em não se materializar.

Já não é no convés de navios que estrangeiros chegam ao Brasil e talvez já não sejam em manuscritos que as novas palavras sejam registradas, mas o que realmente mudou desde então? O que é “saber português” para que os falantes de outras línguas possam ser considerados proficientes nas práticas sociais de nosso país? Que práticas sociais são essas que projetamos como indispensáveis para que um cidadão seja considerado funcionalmente social? O poder social do monolinguismo da língua portuguesa no Brasil fragmenta e apaga muitos caderninhos com nossas tantas variações linguísticas e outras línguas minoritárias e minoritarizadas que são efetivamente utilizadas neste vasto território.

Se é no e pelo uso da linguagem que nos tornamos sujeitos ativos e pertencentes da sociedade, quem é e a qual lugar pertence Shigeru Nagasawa? Ele é brasileiro, não somente no e pelo uso da língua portuguesa, mas através de uma vida dedicada e compartilhada com a comunidade, por meio da língua portuguesa ou apesar dela. Uma vida que lhe foi possibilitada por uma certidão de naturalização. Sua profissão era ler o mundo e reescrevê-lo por meio de fotografias. Em seu quarto escuro, revelava suas reescritas de mundo para que seus interlocutores do jornal Zero Hora pudessem ler o mundo transformado, conscientemente, através das lentes de sua câmera – sua leitura preferida, permitida e aceita pela comunidade dessa esfera de atuação humana.

É sobre isso que trata esta tese, sobre revelar o poder social das línguas dominantes e a noção de leitura proficiente imposta por aparatos sociais de poder que abrem e fecham portas: os testes de larga escala. Esta tese também é sobre a responsabilidade de ser avaliadora e formadora de professores e de avaliadores de Português como Língua Adicional no Brasil.

Após a condução de três trabalhos de pesquisa, Ellen segue se deparando com novos questionamentos. Esta tese é uma homenagem a Shigeru Nagasawa e a todos os leitores e não-leitores que, como ele, seguem estoicamente lutando por suas palavras, suas leituras, suas reescritas de mundo e, acima de tudo, por novos recomeços no Brasil.

REFERÊNCIAS

ACERVO CELPE-BRAS. **Acervo de provas e documentos públicos do Exame Celpe-Bras**. Disponível em: www.ufrgs.br/acervocelpebras. Acesso em: 22 jan. 2023.

AÇÃO EDUCATIVA; INSTITUTO PAULO MONTENEGRO. **Indicador de alfabetismo funcional: resultados preliminares**. 2018. Disponível em: https://alfabetismofuncional.org.br/wp-content/uploads/2020/03/Inaf2018_Relato%CC%81rio-Resultados-Preliminares_v08Ago2018.pdf. Acesso em: 15 jul. 2022.

ALEXANDER, P. A.; FOX, E. Reading research and practice over the decades: a historical analysis. In: ALVERMANN, D. E.; UNRAU, N. J.; SAILORS, M.; RUDELL, R. B. (Eds.). **Theoretical models and processes of literacy**. 7. ed. New York: Routledge, 2018.

ALUÍSIO, S. M.; PINHEIRO, G. M.; MANFRIN, A. M. P.; OLIVEIRA, L. H. M.; GENOVES JR., L. C.; TAGNIN, S. E. O. The Lácio-Web: Corpora and Tools to Advance Brazilian Portuguese Language Investigations and Computational Linguistic Tools. **Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)**. Lisboa, Portugal, 2004.

ALUÍSIO, S. M.; SPECIA, L.; GASPERIN, C.; SCARTON, C. Readability assessment for text simplification. **Proceedings of the NAACL HLT 2010 fifth workshop on innovative use of NLP for building educational applications**. Los Angeles, California, 2010.

ALUÍSIO, S. M.; CUNHA, A.; SCARTON, C. Evaluating progression of Alzheimer's disease by regression and classification methods in a narrative language test in portuguese. In: SILVA, J.; RIBEIRO, R.; QUARESMAS, P.; ADAMI, A., BRANCO, A. (Eds.). **Computational Processing of the Portuguese Language, PROPOR**, 2016.

AMARAL, M. F. **Jornalismo Popular**. São Paulo: Contexto, 2006.

ANDERSEN, G. Corpus linguistics and sociolinguistics. In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge Handbook of Corpus Linguistics**. Abingdon; New York: Routledge, 2010.

ANSARIFAR, A.; SHAHRIARI, H.; PISHGHADAM, R. Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. **Journal of English for Academic Purposes**, v. 31, 2018.

ANTUNES, I. **Lutar com palavras: coesão e coerência**. São Paulo: Parábola Editorial, 2005.

ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA. **Critério de classificação econômica Brasil**. Associação brasileira de empresas de pesquisa, 2022.

BACHMAN, L. F. **Statistical analyses for language testing**. Cambridge: Cambridge University Press, 2004.

BALAGE FILHO, P. P.; ALUÍSIO, S. M.; PARDO, T. A. S. An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology – STIL**. Fortaleza, 2013.

BALYAN, R.; MCCARTHY, K. S.; MCNAMARA, D. S. Combining machine learning and natural language processing to assess literary text comprehension. In: A. HERSHKOVITZ, A.; PAQUETTE, L. (Orgs.). **Proceedings of the 10th International Conference on Educational Data Mining**. Wuhan, 2017.

BALYAN, R. MCCARTHY, K. S.; MCNAMARA, D. S. Comparing machine learning classification approaches for predicting expository text difficulty. In: BRAWNER, K.; RUS, V. (Orgs.). **Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference**. Florida, 2018.

BALYAN, R.; MCCARTHY, K. S.; MCNAMARA, D. S. Applying natural language processing and hierarchical machine learning approaches to text difficulty classification. **International Journal of Artificial Intelligence in Education**, v. 30, 2020.

BARKAOUI, K. Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study. **TESOL QUARTERLY**, v. 44, n. 1, 2010.

BARKAOUI, K. Effects of marking method and rater experience on ESL essay scores and rater performance. **Assessment in Education: principles, policy and practice**, v. 18, n. 3, p. 279-193, 2011.

BARKER, F. Corpora and language assessment: trends and prospects. **Research Notes**, v. 26, 2006.

BARKER, F. How can corpora be used in language testing? In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge Handbook of Corpus Linguistics**. Abingdon; New York: Routledge, 2010.

BARKER, F.; SALAMOURA, A.; SAVILLE, N. Learner corpora and language testing. In: GRANGER, S.; GILQUIN, G.; MEUNIER, F. (Eds.). **The Cambridge Handbook of Learner Corpus Research**. Cambridge: Cambridge University Press, 2015.

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

BERBER SARDINHA, T.; KAUFFMANN, C.H.; ACUNZO, C. A multidimensional analysis of register variation in Brazilian Portuguese. **Corpora**, v. 9, n. 2, 2014.

BERBER SARDINHA, T.; PINTO, M. V.; ACUNZO, C. M.; ZUPPARDI, M. C.; KAUFFMANN, C. H. Adding registers to a previous multi-dimensional analysis. In:

BERBER SARDINHA, T.; PINTO, M. V. (Eds.). **Multi-Dimensional Analysis: Research Methods and Current Issues**. Bloomsbury Publishing, 2019.

BERBER SARDINHA, T.; MOREIRA FILHO, J. L.; ALAMBERT, E. **Manual Corpus Brasileiro**. Pontifícia Universidade Católica de São Paulo - CEPRIL, LAEL. São Paulo, [2010?]. Disponível em: https://www.linguateca.pt/Repositorio/manual_cb.pdf Acesso em: 4 ago. 2023.

BIBER, D. **Variation across Speech and Writing**. New York: Cambridge University Press, 1988.

BIBER, D. Corpus-based and corpus-driven analyses of language. In: HEINE, B.; NARROG, H. (Eds.). **The Oxford handbook of linguistic analysis**. Oxford University Press, 2012.

BIBER, D.; GRAY, B. Discourse Characteristics of Writing and Speaking Task Types on the TOEFL iBT Test: A Lexico-Grammatical Analysis. **TOEFL iBT Research Report**, TOEFL iBT-19, 2013.

BIBER, D.; GRAY, B.; POONPON, K. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? In: BIBER, D.; GRAY, B.; STAPLES, S.; EGBERT, J. (Eds.). **The register-functional approach to grammatical complexity: theoretical foundation, descriptive research findings, application**. Routledge, 2021.

BIBER, D.; GRAY, B.; STAPLES, S.; EGBERT, J. Theoretical and Descriptive Linguistic Foundation of the Register-Functional Approach to Grammatical Complexity. In: BIBER, D.; GRAY, B.; STAPLES, S.; EGBERT, J. (Eds.). **The register-functional approach to grammatical complexity: theoretical foundation, descriptive research findings, application**. Routledge, 2021a.

BIBER, D.; GRAY, B.; STAPLES, S.; EGBERT, J. Overview of the Analytical Methods Used in Register-Functional Complexity Research. In: BIBER, D.; GRAY, B.; STAPLES, S.; EGBERT, J. (Eds.). **The register-functional approach to grammatical complexity: theoretical foundation, descriptive research findings, application**. Routledge, 2021b.

BIBER, D.; GRAY, B.; STAPLES, S.; EGBERT, J. Reflecting on the Register-Functional approach to Grammatical Complexity: What Do We Know and Where Do We Go from Here? In: BIBER, D.; GRAY, B.; STAPLES, S.; EGBERT, J. (Eds.). **The register-functional approach to grammatical complexity: theoretical foundation, descriptive research findings, application**. Routledge, 2021c.

BIDERMAN, M. T. C. **Dicionário Didático de Português**. São Paulo: Editora Ática, 1998.

BIJANKHAN, M.; SHAYESTEFAR, P. Assessing the content typicality and construct of Persian language proficiency test (PLPT) for non-Persian speakers: a corpus-informed study. **Language Testing in Asia**, v. 13, n. 11, 2023.

- BOTARLEANU, R.; DASCALU, M.; WATANABE, M.; CROSSLEY, S. A.; MCNAMARA, D. S. Age of exposure 2.0: estimating word complexity using iterative models of word embeddings. **Behavior Research Methods**, v. 54, p. 3015-3042, 2022.
- BRASIL. **Manual do candidato do exame Celpe-Bras**. Brasília: Secretaria de Educação Superior (SESu), Ministério da Educação, 2002.
- BRASIL. **Manual do examinando do exame Celpe-Bras**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Ministério da Educação, 2011.
- BRASIL. **Documento base do exame Celpe-Bras**. Brasília: Instituto Nacional de Estudos e Pesquisas Anísio Teixeira, Ministério da Educação, 2020.
- BROWN, H. D. **Language assessment: principles and classroom practices**. Longman, 2004.
- BRUNFAUT, T. Assessing reading. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge handbook of language testing**. 2. ed. London; New York: Routledge, 2021.
- BRYSSBAERT, M. How many words do we read per minute? A review and meta-analysis of reading rate. **Journal of memory and language**, v. 109, 2019.
- BULLA, G. S.; KUHN, T. Z. ReVEL na escola: português como língua adicional no Brasil - perfis e contextos implicados. **ReVEL**, v. 18, n. 35, 2020.
- CAMARGO, D. C. Pesquisas em tradução e linguística de corpus. **Anais do Encontros Nacionais da ANPOLL**. Goiás, 2008.
- CAMEIRÃO, M. L.; VICENTE, S. Age-of-acquisition norms for a set of 1,749 Portuguese words. **Behavior Research Methods**, v. 42, n. 2, 2010.
- CANDIDO, M. D. **Avaliação da interação face a face no exame Celpe-Bras: as características dos elementos provocadores e a atuação dos examinadores-interlocutores**. 2015. Dissertação (Mestrado em Linguística Aplicada) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2015.
- CARRELL, P. The effect of writers' personalities and raters' personalities on the holistic evaluation of writing. **Assessing Writing**, v. 2, n. 2, p. 153-190, 1995.
- CARRO, R. B. In: NEWMAN, N.; FLETCHER, R.; ROBERTSON, C. T.; EDDY, K.; NIELSEN, R. K. **Reuters Institute Digital News Report 2022**. Reuters Institute for the Study of Journalism, 2022. Disponível em: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf Acesso em: 8 ago. 2023.
- CARVALHO, S. C.; SCHLATTER, M. Repertórios indígenas, voz e agência na escrita de relatórios de pesquisa de mestrado. **Trabalhos em Linguística Aplicada**, Campinas, v. 61, n. 3, 2022.

- CHAN, S.; INOUE, C.; TAYLOR, L. Developing rubrics to assess the reading-into-writing skills: a case study. **Assessing Writing**, v. 26, p. 20-37, 2015.
- CHAPELLE, C. A. Validity in Language Assessment. In: WINKE, P.; BRUNFAUT, T. (Eds.). **The Routledge handbook of second language acquisition and language testing**. New York: Routledge, 2020.
- CHAPELLE, C. A.; VOSS, E. Utilizing technology in language assessment. In: SHOHAMY, E.; OR, I. G.; MAY, S. (Eds.). **Language Testing and Assessment**. 3. ed. Auckland: Springer, 2017.
- CHAPELLE, C. A.; LEE, H. Conceptions of validity. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.
- CROSSLEY, S.; SALSBUURY, T.; MCNAMARA, D. S. Assessing lexical proficiency using analytic ratings: a case for collocation accuracy. **Applied Linguistics Advance**, v. 22, n. 1, 2014.
- CROSSLEY, S. A.; SKALICKY, S.; DASCALU, M.; MCNAMARA, D. S.; KYLE, K. Predicting text comprehension, processing, and familiarity in adult readers: new approaches to readability formulas. **Discourse Processes**, 2017.
- CROSSLEY, S. A.; SKALICKY, S.; DASCALU, M. Moving beyond classic readability formulas: new methods and new models. **Journal of Research in Reading**, 2019.
- CUMMING, A. Expertise in evaluating second language compositions. **Language Testing**, v. 7, 1990.
- CUMMING, A. ESF/EFL instructors' practices for writing assessment: specific purposes or general purposes? **Language Testing**, v. 18, 2001.
- CUNNINGHAM, J. W.; MESMER, H. A. Quantitative measurement of text difficulty: what's the use? **The Elementary School Journal**, v. 115, n. 2, 2014.
- CUNNINGHAM, J. W.; HIEBERT, E. H.; MESMER, H. A. Investigating the validity of two widely used quantitative text tools. **Reading and Writing**, v. 31, n. 4, 2018.
- CUSHING, S. T. Corpus linguistics in language testing research. **Language Testing**. v. 34, n. 4, 2017.
- CUSHING, S. T. Corpus linguistics and language testing. **The Routledge Handbook of Language Testing**. 2 ed. London; New York: Routledge, 2021.
- CUSHMAN, E. Decolonizing Validity. **The Journal of Writing Assessment**, v. 9, n. 1, p. 1-6, 2016.

- DASCALU, M.; CROSSLEY, S.; MCNAMARA, D.; DESSUS, P.; TRAUSAN-MATU, S. "Please ReaderBench this text": A multi-dimensional textual complexity assessment framework. In: CRAIG, S. (Org.). **Tutoring and intelligent tutoring systems**. Nova Science Publishers, 2018.
- DAVIS, L. Rater and interlocutor training. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.
- DELL'ISOLA, R. L. P.; PORDEUS, I. R. Os enunciados de tarefas integradas de leitura e escrita do exame Celpe-Bras. **Inventário**, n. especial, 2021.
- DINIZ, L. R. A.; BIZON, A. C. C. Discursos sobre a Relação Brasil/ África "Lusófona" em Políticas Linguísticas e de Cooperação Educacional. **Línguas e Instrumentos Linguísticos**, v. 36, 2015.
- DIONÍSIO, C. I. B. O exame Celpe-Bras como mecanismo de política linguística para candidatos ao PEC-G. **Fórum linguística**, v. 19, n. 1, p. 7608-7620, 2022.
- DIVINO, L. S. **Índices lexicais de análise para a caracterização dos níveis intermediário e avançado superior no exame Celpe-Bras**: uma pesquisa guiada por corpus. Trabalho de Conclusão de Curso (Licenciatura em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.
- DOUGLAS, D. **Assessing languages for specific purposes**. The Cambridge Language Assessment Series, 2000.
- DUTTA, D. How to perform feature selection (i.e. pick important variables) using Boruta Package in R? **Analytics Vidhya**, 2016.
- EDUCATIONAL TESTING SERVICE. Validity evidence supporting the interpretation and use of TOEFL iBT. **TOEFL Research Insight Series**, v. 4. Princeton, NJ: ETS, 2011.
- EGBERT, J. Corpus linguistics and language testing: navigating uncharted waters. **Language Testing**, v. 34, n. 4, 2017.
- EVERS, A. **Processamento de língua natural e níveis de proficiência do português**: um estudo de produções textuais do exame Celpe-Bras. 2013. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.
- EVISON, J. What are the basics of analysing a corpus? In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge Handbook of Corpus Linguistics**. Abingdon; New York: Routledge, 2010.
- FARIA, E. R. **Avaliação de proficiência oral em língua estrangeira**: foco nos elementos provocadores na interação face a face do exame Celpe-Bras. Tese (Doutorado em Linguística) – Universidade Federal de São Carlos, São Carlos, 2017.

FINATTO, M. J. B. Projeto PorPopular, frequência de verbos em português e no jornal popular brasileiro. In: ISQUERDO, A. N.; SEABRA, M. C. T. C. (Orgs.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. Campo Grande: Editora da UFMS/Laboratório de Edição da FALE-UFMG, 2012.

FINATTO, M. J. B.; SCARTON, C. E.; ROCHA, A.; ALUISIO, S. M. Características do jornalismo popular: avaliação da inteligibilidade e auxílio à descrição do gênero. **Anais do VIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana**, Cuiabá, v. 1. p. 30-39, 2011.

FINATTO, M. J. B.; STEFANI, M.; PASQUALINI, B. F.; CIULLA, A.; EVERS, A.; SORTICA, M. **Leitura**: um guia sobre teoria(s) e prática(s). [recurso eletrônico]. Porto Alegre: Instituto de Letras – UFRGS, 2015. Disponível em: [https://www.ufrgs.br/textecc/traducao/teorias/e-book/Leitura%20-%20um%20guia%20sobre%20teoria\(s\)%20e%20pr%C3%A1tica\(s\).pdf](https://www.ufrgs.br/textecc/traducao/teorias/e-book/Leitura%20-%20um%20guia%20sobre%20teoria(s)%20e%20pr%C3%A1tica(s).pdf) Acesso em: 8 jan. 2023.

FINATTO, M. J.; MOTTA, E. Terminologia e acessibilidade: novas demandas e frentes de pesquisa. **Revista GTLex.**, v. 2, n. 2, 2017.

FINATTO, M. J. B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (Orgs.). **Linguística de corpus**: perspectivas [recurso eletrônico]. Porto Alegre: Instituto de Letras – UFRGS, 2018. Disponível em: https://www.ufrgs.br/ppgletras/ebooks/2018Linguisticadecorpus_NOVO_FINAL.pdf. Acesso em: 13 mar. 2023.

FISCHER, L. A.; GARCEZ, P. M. Leitura para o neoleitor: conversa com animadores de leitura. GT Transculturalidade, linguagem e educação. **ANPOLL**, 2012.

FONSECA, L. C.; REBECHI, R.; TOSQUI-LUCKS, P. Festschrift Para Stella Esther Ortweiler Tagnin. **TradTerm**, São Paulo, v. 37, n. 1, 2021.

FRANÇA, A. I. Relatório do Workshop de Complexidade Textual - Hotel Transamérica Prime, 19 de janeiro de 2018. In: GUTEN. **Complexidade textual**: contribuição do tema para a progressão dos estudantes brasileiros em leitura, 2018.

FULCHER, G.; OWEN, N. Dealing with the demands of language testing and assessment. In: HALL, G. (Ed.). **The Routledge Handbook of English Language Teaching**. Routledge, 2016.

FULCHER, G.; HARDING, L. (Orgs.). **The Routledge handbook of language testing**. 2. ed. London; New York: Routledge, 2021.

GABLASOVA, D. Corpora for second language assessments. In: WINKE, P.; BRUNFAUT, T. (Eds.) **The Routledge handbook of second language acquisition and language testing**. New York: Routledge, 2020.

GABRIELATOS, C.; MARCHI, A. Keyness: appropriate metrics and practical issues. **Corpus-assisted Discourse Studies International Conference**, Itália, 2012.

GAZZOLA, M. G.; LEAL, S. E.; ALUÍSIO, S. M. Predição da complexidade textual de recursos educacionais abertos em português. **Anais do XII Symposium in Information and Human Language Technology**. Salvador, 2019.

GEBRIL, A.; PLAKANS, L. Assembling validity evidence for assessing academic writing: rater reactions to integrated tasks. **Assessing Writing**, v. 21, p. 56-73, 2014.

GEE, J. P. Reading as situated language: A sociocognitive perspective. In: ALVERMANN, D. E.; UNRAU, N. J.; SAILORS, M.; RUDELL, R. B. (Eds.) **Theoretical models and processes of literacy**. 7. ed. New York: Routledge, 2019.

GOMES, M. S. **A complexidade de tarefas de leitura e produção escrita no exame Celp-Bras**. Dissertação (Mestrado em Linguística Aplicada) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

GOTZ, S. Learner corpora to inform testing and assessment. In: JABLONKAI, R. R.; CSOMAY, E. (Eds.). **Routledge Handbook of Corpora and Language Teaching and Learning**. Londres: Routledge, 2022.

GRAESSER, A.C.; MCNAMARA, D.S.; LOUWERSE, M. M. What do readers need to learn in order to process coherence relations in narrative and expository text. In: SWEET, A. P.; SNOW, C. E. (Eds.). **Rethinking reading comprehension**. New York: Guilford Publications, 2003.

GRAESSER, A. C.; MCNAMARA, D. S.; LOUWERSE, M. M; CAI, Z. Coh-Metrix: analysis of text on cohesion and language. **Behavior Research Methods, Instruments, & Computers**, v. 36, n. 2, 2004.

GRAESSER, A. C.; MCNAMARA, D. S.; CAI, Z.; CONLEY, M.; LI, H.; PENNEBAKER, J. Coh-Metrix measures text characteristics at multiple levels of language and discourse. **The Elementary School Journal**, v. 115, n. 2, 2014.

GREEN, A.; FULCHER, G. Teste Design Cycle. In: WINKE, P.; BRUNFAUT, T. (Eds.) **The Routledge handbook of second language acquisition and language testing**. New York: Routledge, 2020.

GRUPO DE MÍDIA SÃO PAULO. **Midiadados Brasil 2021 para todxs**. Grupo de Mídia São Paulo, 2021. Disponível em: <https://midiadadosgmsp.com.br/2021/midia-dados-2021.pdf> Acesso em: 8 ago. 2023.

GU, P. Y. An argument-based framework for validating formative assessment in the classroom. **Frontiers in Education**, v. 6, 2021.

GUTEN. **Complexidade textual**: contribuição do tema para a progressão dos estudantes brasileiros em leitura. 2018.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. (Eds.). **Multivariate data analysis**. 7. ed. Pearson Prentice Hall, 2009.

HALL, G. (Ed.). **The Routledge Handbook of English Language Teaching**. Routledge, 2016.

HAMP-LYONS, L. Fairnesses in language testing. In: KUNNAN, A. J. (Ed.). **Fairness and validation in language assessment**, Studies in Language Testing. Cambridge: Cambridge University Press, 2000.

HARTMANN, N.; CUCATTO, L.; BRANTS, D.; ALUÍSIO, S. M. Automatic classification of the complexity of nonfiction texts in Portuguese for early school years. **International Conference on Computational Processing of the Portuguese Language**. Springer, 2016.

HASAN, J.; KIM, J.; KIM, C. H.; KIM, J. Health State Classification of a Spherical Tank Using a Hybrid Bag of Features and K-Nearest Neighbor. **Applied Sciences**, v. 10, 2020.

HEATH, S. B. What no bedtime story means: narrative skills at home and school. In: DURANTI, A. (Org.). **Linguistic Anthropology**: a reader. Oxford: Blackwell, (2001) [1983].

HIEBERT, E. H.; PEARSON, P. D. Understanding text complexity: introduction to the Special Issue. **The Elementary School Journal**, v. 115, n. 2, 2014.

HILL, Y. Z.; LIU, O. L. Is there any interaction between background knowledge and language proficiency that affects TOEFL iBT reading performance? **TOEFL iBT Report**, v. 18. Princeton: ETS, 2012.

HUGUES, A.; HUGHES, J. **Testing for language teachers**. Cambridge University Press, 2020.

HUNSTON, S. **Corpora in applied linguistics**. 2. ed. Cambridge; New York: Cambridge University Press, 2022.

INAF. **Alfabetismo no Brasil**, 2018?. Disponível em: <https://alfabetismofuncional.org.br/alfabetismo-no-brasil/>. Acesso em: 7 ago. 2023.

JABLONKAI, R. R.; CSOMAY, E. (Eds.). **Routledge Handbook of Corpora and Language Teaching and Learning**. Londres: Routledge, 2022.

JANCZURA, G. A.; CASTILHO, G. M.; ROCHA, N. O.; VAN ERVEN, T. J. C. HUANG, T. P. **Normas de concretude para 909 palavras da língua portuguesa**. *Psic.: Teor. e Pesq.* [online], v. 23, n. 2, p.195-204, 2007. Disponível em: <<http://www.scielo.br/pdf/ptp/v23n2/a10v23n2.pdf>>. Acesso em: 28 jun. 2023.

JI, M.; LIU, U.; ZHAO, M.; LYU, Z.; ZHANG, B.; LUO, X.; LI, Y.; ZHONG, Y. Use of machine learning algorithms to predict the understandability of health education materials: development and evaluation study. **JMIR Medical Informatics**, v. 9, n. 5, 2021.

JIN, Y. Test specifications. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.

JOUVE, V. **A leitura**. São Paulo: Editora da UNESP, 2002.

KAMAKURA, W.; MAZZON, J. A. Critérios de estratificação e comparação de classificadores socioeconômicos no Brasil. **Revista de Administração de Empresas**. São Paulo, v. 56, n. 1, 2016.

KANE, M. T. Articulating a validity argument. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.

KAUFFMANN, C. H. **O corpus do jornal: variação linguística, gênero e dimensões da imprensa diária escrita**. 2005. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2005.

KAUFFMANN, C. H. Cognição e variação linguística de gêneros/registros jornalísticos: um estudo baseado em corpus. **Domínios de Lingu@gem**, Uberlândia, v. 16, n. 4, 2022.

KILGARRIFF, A.; GREFFENSTETTE, G. Web as corpus. **Computational Linguistics**, v. 29, n. 3, p. 333-347, 2003.

KILLNER, M. **A presença de textos literários no exame Celpe-Bras**. Tese (Doutorado em Letras) – Universidade Estadual Paulista, Araraquara, 2021.

KIM, J.; CHI, Y.; HUENSCH, A.; JUN, H.; LI, H.; ROULLION, V. A case study on an item writing process use of test specifications, nature of group dynamics, and individual item writer's characteristics. **Language Assessment Quarterly**, v. 7, 2010.

KLEIMAN, A. (Org.). **Os significados do letramento: uma nova perspectiva sobre a prática social da escrita**. Campinas: Mercado de Letras, 1995.

KLEIMAN, A. B. Avaliando a compreensão: letramento e discursividade nos testes de leitura. In: RIBEIRO, V. M. (Org.). **Letramento no Brasil: reflexões a partir do INAF**. São Paulo: Global Editora, 2003.

KNOCH, U.; SITAJALABHORN, W. A closer look at integrated writing tasks: towards a more focused definition for assessment purposes. **Assessing Writing**, v. 18, p. 300-308, 2013.

KOCH, I. G. V. **Desvendando os segredos do texto**. 8. ed. São Paulo: Cortez, 2018.

KREMMELE, B.; EBERHARTER, K.; HOLZKNECHT, F. Pre-operational testing. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.

KÜBLER, N.; ASTON, G. Using corpora in translation. In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge Handbook of Corpus Linguistics**. Abingdon; New York: Routledge, 2010.

KUNRATH, S. P. **Os descritores gerais e a progressão dos níveis de proficiência do exame Celpe-Bras**. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.

KURSA, M. B. JANKOWSKI, A.; RUDNICKI, W. R. Boruta – A system for feature selection. **Fundamenta Informaticae**, v. 10, 2010.

KURSA, M. B.; RUDNICKI, W. R. Feature selection with the Boruta package. **Journal of statistical software**, v. 36, n. 11, 2010.

LACERDA, M. G.; FARBIARZ, J. L. A formação visual do leitor por meio do Design na Leitura: livros de literatura para Educação Infantil e Ensino Médio. **13º Congresso Brasileiro de Pesquisa e Desenvolvimento em Design**. Joinville, 2018.

LEAL, S. E. **Predição da complexidade sentencial do português brasileiro escrito, usando métricas linguísticas, psicolinguísticas e de rastreamento ocular**. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2021.

LEAL, S. E.; DURAN, M. S.; ALUÍSIO, S. M. A nontrivial sentence corpus for the task of sentence readability assessment in Portuguese. **Proceedings of International Conference on Computational Linguistic**. Santa Fé, Novo México, 2018.

LEAL, S. E.; DURAN, M. S.; SCARTON, C. E.; HARTMANN, N. S.; ALUÍSIO, S. M. NILC-Metrix: assessing the complexity of written and spoken language in Brazilian Portuguese. **arXiv:2201.03445 [cs.CL]**, 2021.

LIN, L.; ZHANG, J.; ZHANG, N.; SHI, J.; CHEN, C. Optimized LightGBM Power Fingerprint Identification Based on Entropy Features. **Entropy**, v. 24, 2022.

LOBO, T. R. M. **As consignas do exame Celpe-Bras - Parte escrita**. Dissertação (Mestrado em Linguística) – Universidade Federal do Ceará, Fortaleza, 2021.

LOURENÇO, D. A. **Desenvolvimento de um método para avaliação do rendimento de leitura com crianças brasileiras envolvendo legibilidade, leiturabilidade e fundamentos da lectoescrita Método Lêcom**. Tese (Doutorado em Design) – Universidade Federal de Pernambuco, Recife, 2016.

LUKE, A.; FREEBODY, P. Literacies programs: Debates and demands in cultural context. **Prospect: Australian Journal of TESOL**, v. 5, n. 7, p. 7-16, 1990.

LUKE, A.; FREEBODY, P. **Further Notes on the Four Resources Model**, 1999. Disponível em:

<http://kingstonnetworknumandlitteam.wikispaces.com/file/view/Further+Notes+on+the+Four+Resour+ces+Model-Allan+Luke.pdf> Acesso em: 10 out. 2023.

MAIA, M. Relatório do Workshop de Complexidade Textual - Hotel Transamérica Prime, 19 de janeiro de 2018. In: GUTEN. **Complexidade textual**: contribuição do tema para a progressão dos estudantes brasileiros em leitura. 2018.

MARQUES, J. F.; FONSECA, F. L.; MORAIS, A. S.; PINTO, I. A. Estimated age of acquisition norms for 834 Portuguese nouns and their relation with other psycholinguistic variables. **Behavior Research Methods**, v. 39, n. 3, 2007.

MARTINS, M. H. **O que é leitura?** São Paulo: Brasiliense, 1982.

MARTINS, T. B. F.; GHIRALDELO, C. M.; NUNES, M. G. V.; OLIVEIRA JUNIOR, O. N. Readability formulas applied to textbooks in Brazilian portuguese. **Notas do ICMC**, n. 28, 1996.

MAZIERO, E.G.; PARDO, T.A.S.; DI FELIPPO, A.; DIAS-DA-SILVA, B.C. A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. **VI Workshop em tecnologia da informação e da linguagem humana (TIL)**, p. 390-392, 2008.

MCNAMARA, T.; ROEVER, C. **Language Testing: The Social Dimension**. Malden: Blackwell Publishing Limited, 2006.

MCNAMARA, T.; SHOHAMY, E. Language tests and human rights. **International Journal of Applied Linguistics**, v. 18, n. 1, 2008.

MENDEL, K. **A avaliação integrada de leitura e escrita no Exame Celpe-Bras**. Trabalho de Conclusão de Curso (Licenciatura em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.

MENDEL, K. **Proficiência e autoria na avaliação integrada de leitura e escrita do exame Celpe-Bras**. Dissertação (Mestrado em Linguística Aplicada) – Instituto de Letras, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2019.

MESMER, H. A.; HIEBERT, E. H. Third graders' reading proficiency reading text varying in complexity and length: responses of students in an urban, high-needs school. **Journal of Literacy Research**, v. 47, n. 4, p. 473-504, 2015.

MOTA, M. Entrevista por Skype sobre complexidade textual em 17 de janeiro de 2018. In: GUTEN. **Complexidade textual**: contribuição do tema para a progressão dos estudantes brasileiros em leitura. 2018.

MUNIZ, M. C. M. **A construção de recursos linguístico-computacionais para o português do Brasil: o projeto de Unitex-PB**. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2004.

MURAKAWA, C. de A. A. Maria Tereza Camargo Biderman, Dicionário didático do português. **Todas As Letras - Revista De Língua E Literatura**, v. 3, n. 1, 2009.

NAGASAWA, E. Y. **Português como língua adicional para fins específicos**: preparação ao exame Celpe-Bras. Dissertação (Mestrado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

NAGASAWA, E. Y.; SCHLATTER, M. A complexidade de enunciados de tarefas de produção escrita no Celpe-Bras e no Enem. **Tradterm**, v. 37, n. 1, p. 330-363, 2021.

NAGASAWA, E. Y.; PICORAL, A. Observações preliminares acerca da complexidade textual de textos de leitura da Parte Escrita do Celpe-Bras. **Encontro Mundial sobre o Ensino de Português**. Apresentação individual de artigo, 2022. Disponível em: <https://sites.google.com/view/xiemep/programaprogram> Acesso em: 25 maio 2023.

NATER, C. S.; MARQUES, J. Elementos provocadores do exame Celpe-Bras: recorrência de gêneros discursivos e de temas nas aplicações do triênio 2017-2019. **ReVEL**. v. 18, n. 35, 2020.

NATIONAL GOVERNORS ASSOCIATION CENTER FOR BEST PRACTICES, COUNCIL OF CHIEF STATE SCHOOL OFFICERS. **Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects**: Appendix A. Washington D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers Publisher, 2010.

NEGRI, M. F. **Representações sociais e interculturalidade**: um estudo documental sobre elementos provocadores e roteiros de interação do exame Celpe-Bras. Tese (Doutorado em Linguística) – Universidade Federal de São Carlos, São Carlos, 2019.

NESI, H.; GARDNER, S. Complex, but in what way? A step towards greater understanding of academic writing proficiency. In: DANJO, C.; MEDDEGAMA, I.; O'BRIEN, D.; PRUDHOE, J.; WALZ, L.; WICAKSONO, R. (Eds.). **Online Proceedings of the 51st Annual Meeting of the British Association for Applied Linguistics**: Taking Risks in Applied Linguistics. British Association for Applied Linguistics, 2019.

OHTA, R.; PLAKANS, L. M.; GEBRIL, A. Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis. **Assessing Writing**, v. 38, p. 21-36, 2018.

O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge Handbook of Corpus Linguistics**. Abingdon; New York: Routledge, 2010.

OUSHIRO, L. **Introdução à estatística para linguistas**. Editora Abralín, 2022.

PARAGUASSU, L. B. **Tradução especializada acessível (TEA): revisão do tema e proposta de disciplina para cursos de graduação em tradução**. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

PASQUALINI, B. F. **Corpop: um corpus de referência do português popular escrito do Brasil**. Tese (Doutorado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, UFRGS, 2018.

PEARSON, P. D.; HIEBERT, E. H. The state of the field: qualitative analyses of text complexity. **The Elementary School Journal**, v. 115, n. 2, 2014.

PHILIP, G. Arriving at equivalence: making a case for comparable general reference corpora in translation studies. In: BEEBY, A.; INÉS, P. R.; SÁNCHEZ-GIJÓN, P. (Eds.). **Corpus use and translating**. Amsterdam; Philadelphia: John Benjamins, 2008.

PICORAL, A.; STAPLES, S.; REPPEN, R. Automated annotation of learner English. *International Journal of Learner*. **Corpus Research**, v. 7, n. 1, 2021.

PLAKANS, L. Writing integrated tasks. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.

PLAKANS, L.; GEBRIL, A. Exploring the relationship of organization and connection with scores in integrated writing assessment. **Assessing Writing**, v. 31, p. 98-112, 2017.

PROVA IELTS: qual das versões devo fazer? Publicado pelo canal British Council. Disponível em: <https://www.britishcouncil.org.br/exame/ielts/que-teste> Acesso em: 22 jun. 2023.

RAND Reading Study Group. **Reading for understanding: toward an R&D program in reading comprehension**. Santa Monica: RAND, 2002.

RANDALL, J. Color-neutral is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*, v. 40, n. 4, 82–90, 2021.

RANDALL, J.; SLOMP, D.; POE, M.; OLIVERI, M. E. Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. **Educational Assessment**, v. 27, n. 2, p. 1-9, 2022.

REBECHI, R. R. O discurso dos deputados na votação do impeachment: a LC combinada à ACD. In: FINATTO, M. J. B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (Orgs.). **Linguística de corpus: perspectivas [recurso eletrônico]**. Porto Alegre: Instituto de

Letras – UFRGS, 2018. Disponível em:

https://www.ufrgs.br/ppglettras/ebooks/2018Linguisticadecorpus_NOVO_FINAL.pdf. Acesso em: 13 jun. 2023.

RÉVÉSZ, A.; BRUNFAUT, T. Validating assessments for research purposes. In: WINKE, P.; BRUNFAUT, T. (Eds.) **The Routledge handbook of second language acquisition and language testing**. New York: Routledge, 2020.

RIBEIRO, M. S. L. **Jornal Massa!**: uma estratégia de ampliação de público. Trabalho de Conclusão de Curso (Bacharelado em Comunicação) – Faculdade de Comunicação, Universidade Federal da Bahia, Salvador, 2013.

RIGATTI, P. C.; FONTES, A. B. A. L. Relações entre experiência de leitura, habilidades linguísticas e acesso lexical na L2. **Caderno de Letras**, n. 35, 2019.

RIO GRANDE DO SUL. **Referenciais Curriculares do Estado do Rio Grande do Sul: Linguagens, Códigos e suas Tecnologias**. Secretaria de Estado da Educação. Porto Alegre: SE/DP, 2009.

RODRIGUES, E. Entrevista por Skype sobre complexidade textual em 23 de janeiro de 2018. In: GUTEN. **Complexidade textual**: contribuição do tema para a progressão dos estudantes brasileiros em leitura. 2018.

ROSA, A. M.; FARINHA, C. L.; RADNER, W.; DIENDORFER, G.; LOUREIRO, M. F.; NURTA, J. N. Development of Portuguese version of a standardized reading test: the Radner-Coimbra Charts. **Arquivos Brasileiros de Oftalmologia**, v. 79, n. 4, 2016.

SÁ, T. M. M.; SANTOS, A. A.; SOUZA, R. A.; AMARAL, L.; ALMEIDA, V. N. Um teste de verificação lexical de português brasileiro (TVLPB) como língua adicional: criação e validação. **Caligrama**, Belo Horizonte, v. 26, n. 2, p. 125-153, 2021.

SANTOS, J. **Avaliação de habilidades de leitura no ensino superior a partir da análise de provas de nivelamento**. Dissertação (Mestrado em Estudos da Linguagem) – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

SANTOS, L. B.; DURAN, M. S.; HARTMANN, N. S. CANDIDO JUNIOR, A.; PAETZOLD, G. H.; ALUÍSIO, S. M. A lightweight regression method to infer psycholinguistic properties for Brazilian Portuguese. **International Conference on Text, Speech, and Dialogue**, Proceedings, Springer, Lecture Notes in Artificial Intelligence, v. 10415, p. 281-289, 2017.

SCARAMUCCI, M. V. R. Validade e consequências sociais das avaliações em contextos de ensino de línguas. **Linguarum Arena**, Porto, v. 2, p. 103-120, 2011.

SCARAMUCCI, M. V. R. A avaliação de habilidades integradas na Parte Escrita do Exame Celpe-Bras. In: ALVAREZ, M. L. O.; GONÇALVES, L. (Orgs.). **O mundo do português e**

o português no mundo afora: especificidades, implicações e ações. Campinas: Pontes Editores, 2016.

SCARAMUCCI, M. V. R.; DINIZ, L. R.A. Avaliação de proficiência em português em processos de naturalização no Brasil. **D.E.L.T.A.**, v. 38. n. 4, 2022.

SCARAMUCCI, M. V. R.; SCHLATTER, M. A formação de avaliadores da Parte Oral do Celpe-Bras. In: SCHOFFEN, J. R.; SCHLATTER, M. (Orgs.). **Impactos pedagógicos, políticos e sociais de exames de alta relevância: o Celpe-Bras em pauta.** Porto Alegre, no prelo.

SCARTON, C.; ALUÍSIO, S. Herança automática das relações de hiperonímia para a WordNet.Br. **Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional.** NILC - ICMC - USP, São Carlos, 2009.

SCARTON, C.; ALUÍSIO, S. 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, 2010.

SCHLATTER, M. CELPE-BRAS: Certificado de Língua Portuguesa para Estrangeiros - Breve Histórico. In: CUNHA, M. J.; SANTOS, P. (Orgs.). **Ensino e pesquisa em português para estrangeiros.** Brasília: Editora da UnB, 1999.

SCHLATTER, M. O ensino de leitura em língua estrangeira na escola: uma proposta de letramento. **Calidoscópico**, v. 7, n. 1, p. 11-23, 2009.

SCHLATTER, M. Celpe-Bras: avaliação, ensino e formação de professores de português como língua adicional. In: **Um pouco de história**, Acervo Celpe-Bras, 2014. Disponível em: <<http://www.ufrgs.br/acervocelpebras/um-pouco-de-historia>>. Acesso em: 26 fev. 2019.

SCHLATTER, M. Ler, escrever e compartilhar crônicas para construir-se como autor. **Revista Na Ponta do Lápis**, v. 32, p. 13-23, 2018.

SCHLATTER, M.; SCARAMUCCI, M. V. R., PRATI, S., ACUÑA, L. Celpe-Bras e Celu: impactos da construção de parâmetros comuns de avaliação de proficiência em português e em espanhol. In: ZOPPI-FONTANA, M. (Org.) **O português do Brasil como língua transnacional.** Campinas: Editora RG, 2009.

SCHLATTER, M.; NUNES, L. N.; KUNRATH, S. P. Análise descritiva da Parte Escrita do exame Celpe-Bras. **Examen: Política, Gestão e Avaliação da Educação**, v. 4, n. 4, 2021.

SCHOFFEN, J. R. **Gêneros do discurso e parâmetros de avaliação de proficiência em português como língua estrangeira no exame Celpe-Bras.** Tese (Doutorado em Linguística Aplicada) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

SCHOFFEN, J. R.; COURA SOBRINHO, J.; DELL'ISOLA, R. P.; SCARAMUCCI, M.; DINIZ, R. R. A.; FURTOSO, V. B.; MOUTINHO, R.; LIMA, R. A.; SANTOS, L. G.; TOSATTI, N.; KUNRATH, S. P. **Especificações para a elaboração de tarefas do exame Celpe-Bras – Parte Escrita**, 2012.

SCHOFFEN, J. R.; MENDEL, K. As especificações do exame Celpe-Bras e a descrição das tarefas da Parte Escrita: convergências e divergências. **Domínios de Lingu@gem**, v. 12, n. 2, p. 1091-1122, 2018.

SCHOFFEN, J. R.; SCHLATTER, M.; KUNRATH, S. P.; NAGASAWA, E. Y.; SIRIANNI, G. R.; MENDEL, K.; TRUYLLIO, L. R.; DIVINO, L. S. **Estudo descritivo das tarefas da Parte Escrita do exame Celpe-Bras: edições de 1998 a 2017**. [recurso eletrônico]. Porto Alegre: Instituto de Letras - UFRGS, 2018. Disponível em: <https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2021/12/Schoffen-et-al-2018.pdf>. Acesso em: 15 mar. 2023.

SCHOFFEN, J. R.; SEGAT, G. L. O gênero carta/e-mail no exame Celpe-Bras: reflexões para a preparação de examinandos e para o ensino de português como língua adicional. **REVISTA VIRTUAL DE ESTUDOS DA LINGUAGEM**, v. 18, p. 398-427, 2020.

SERAFINI, F. Expanding the four resources model: reading visual and multi-modal texts. **Pedagogies: An International Journal**, v. 7, n. 2, p. 150–164, 2012.

SHEEHAN, K. M.; KOSTIN, I.; NAPOLITANO, D.; FLOR, M. The textevaluator tool: helping teachers and test developers select texts for use in instruction and assessment. **The Elementary School Journal**, v. 115, n. 2, 2014.

SHIN, D. Item writing and item writers. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.

SHIN, S. Y.; EWERT, D. What accounts for integrated reading-to-write task scores. **Language Testing**, v. 32, n. 2, 2015.

SHOHAMY, E. The social responsibility of the language testers. In COOPER, R. L.; SHOHAMY, E.; WALTERS, J. (Eds.). **New Perspectives and issues in educational language policy: a festschrift for Bernard Dov Spolsky**. John Benjamins Publishing Company, 2001.

SHOHAMY, E. Assessing multilingual Competencies: adopting construct valid assessment policies. **The Modern Language Journal**, v. 95, n. 3, 2011.

SHOHAMY, E. Critical Language Testing. In: SHOHAMY, E.; OR, I. G.; MAY, S. (Eds.). **Language Testing and Assessment**. 3. ed. Auckland: Springer, 2017.

SIERSCHYNSKI, J.; LOUIE, B.; PUGHE, B. Complexity in picture books. **Reading Teacher**, v. 68, n. 4, p. 287-295, 2015.

SINCLAIR, J. Corpus and Text — Basic Principles. In: WYNNE, M. (Ed.). **Developing linguistic corpora: a guide to good practice**. ahds literature, languages and linguistics. online, 2004. Disponível em: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora>. Acesso em: 11 mar. 2019.

SIRIANNI, G. R. **Entre a certificação e a não certificação no Celpe-Bras**: um estudo sobre os níveis de proficiência na Parte Escrita do exame. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020.

SOARES, A. P.; COSTA, A. S.; MACHADO, J.; COMSAÑA, M.; OLIVEIRA, H. M. The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. **Behavior Research Methods**, v. 49, 2017.

SOARES, M. **Letramento**: um tema em três gêneros. 3. ed. Belo Horizonte: Autêntica Editora, 2009.

SOUZA, L. B.; HÜBNER, L. C. Desafios na avaliação da compreensão leitora: demanda cognitiva e leitura textual. **Revista Neuropsicologia Latinoamericana**, v. 7, n. 1, p. 34-46, 2015.

SPENCER, M.; GILMOUR, A. F.; MILLER, A. C.; EMERSON, A. M.; SAHA, N. M.; CUTTING, L. E. Understanding the influence of text complexity and question type on reading outcomes. **Reading and Writing**, v. 32, 2019.

SPOLSKY, B. History of Language Testing. In: SHOHAMY, E.; OR, I. G.; MAY, S. (Eds.). **Language Testing and Assessment**. 3. ed. Auckland: Springer, 2017.

STAPLES, S.; EGBERT, J.; BIBER, D.; GRAY, B. Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. **Written Communication**, v. 33, n. 2, 149-183, 2016.

STENNER, A. J.; BURDICK, H.; SANFORD, E.; BURDICK, D. How accurate are Lexile text measure? **Journal of Applied Measurement**, v. 7, n. 3, 2006.

STREET, B. V. **Literacy in Theory and Practice**. Cambridge: Cambridge University Press, 1984.

STRICKER, L.; ATTALI, Y. Test takers' attitudes about the TOEFL iBT. In: **TOEFL iBT Report**, v. 13. Princeton, NJ: ETS, 2010.

TAGNIN, S. E A Linguística de Corpus vai desbravando novos horizontes. In: FINATTO, M. J. B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (Orgs.). **Linguística de corpus: perspectivas** [recurso eletrônico]. Porto Alegre: Instituto de Letras – UFRGS, 2018. pp. 11-15. Disponível em: https://www.ufrgs.br/ppgletras/ebooks/2018Linguisticadecorpus_NOVO_FINAL.pdf. Acesso em: 15 mar. 2023.

TEIXEIRA, E. D. Linguística de Corpus no Brasil, ontem e hoje: entrevista atualizada com Stella Tagnin. **Tradterm**, v. 37, n. 2, 2023.

THE COLLEGE BOARD. **Test specifications for the redesigned SAT**. College Board, 2015. Disponível em: <https://satsuite.collegeboard.org/media/pdf/test-specifications-redesigned-sat-1.pdf> Acesso em: 21 jun. 2023.

THOMAS, C.; KESELJ, V.; CERCONI, N.; ROCKWOOD, K.; ASP, E. Automatic Detection and Rating of Dementia of Alzheimer Type through Lexical Analysis of Spontaneous Speech. **IEEE International Conference Mechatronics and Automation**, 2005.

TOGNINI-BONELLI, E. Theoretical overview of the evolution of corpus linguistics. In: O'KEEFFE, A.; MCCARTHY, M. (Eds.) **The Routledge handbook of Corpus Linguistics**. Abingdon; New York: Routledge, 2010.

TOSATTI, N. M. **O desempenho de estudantes de países africanos de língua oficial portuguesa no Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras)**. Tese (Doutorado em Linguística Aplicada) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

TOSATTI, N. M. O gênero carta do leitor produzido por estudantes de PALOP no exame Celpe-Bras: proficiência e letramento. In: KUHN, T. Z.; SCHOFFEN, J. R.; PERNA, C. B. L.; ANTUNES, A. J.; CARILLO, M. S. (Orgs.). **Português Língua Pluricêntrica: das políticas às práticas**. Campinas, SP: Pontes Editores, 2022.

TOULMIN, S. E. **The uses of argument**. Cambridge University Press: New York, 2003.

VALENCIA, S. W.; WIXSON, K. K.; PEARSON, P. D. Putting text complexity in context: refocusing on comprehension of complex text. **The Elementary School Journal**, v. 115, n. 2, 2014.

VICENTINI, M. P. As tarefas de compreensão oral para produção escrita no Celpe-Bras: construto e operacionalização. **Estudos Linguísticos**, v. 48, n. 1, p. 561–580, 2019.

VICENTINI, M. P. **As dimensões do construto compreensão oral para produção escrita no exame Celpe-Bras: percepções, processos, estratégias e desempenhos de examinandos**. Tese (Doutorado em Linguística Aplicada) – Universidade Estadual de Campinas, Campinas, 2022.

VIGEL, M. Z. **Comportamento de consumo de jornais por jovens: uma análise do meio impresso e online**. Trabalho de Conclusão de Curso (Bacharelado em Administração) – Departamento de Ciências Administrativas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.

WAGNER FILHO, J. A.; WILKENS, R.; IDIART, M.; VILLAVICENCIO, A. The brWaC Corpus: A New Open Resource for Brazilian Portuguese. **11th Language Resources and Evaluation Conference (LREC 2018)**. Miyazaki, Japão, 2018.

WALL, D.; HORÁK, T. The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 3, The role of the coursebook, and Phase 4, Describing change. **TOEFL iBT Report**, v. 17. Princeton, NJ: ETS, 2011.

WALTER, E. Using corpora to write dictionaries. In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge Handbook of Corpus Linguistics**. Abingdon; New York: Routledge, 2010.

WEIGLE, S. C. Integrating Reading and Writing in a Competency Test for Non-Native Speakers of English. **Assessing Writing**, v. 9, 2004.

WINKE, P.; BRUNFAUT, T. (Eds.). **The Routledge handbook of second language acquisition and language testing**. New York: Routledge, 2020.

WINKE, P.; BRUNFAUT, T. Perspectives on 'Knowing' a Second Language: What Are We Seeking to Measure? In: WINKE, P.; BRUNFAUT, T. (Eds.). **The Routledge handbook of second language acquisition and language testing**. New York: Routledge, 2020.

XI, X. What does corpus linguistics have to offer to language assessment? **Language Testing**, v. 34, n. 4, 2017.

XI, X.; HIGGINS, D.; ZECHNER, K.; WILLIAMSON, D. M. Automated scoring of spontaneous speech using SpeechRater v1.0. **TOEFL iBT Report**, v. 18. Princeton, NJ: ETS, 2008.

XI, X.; SAWAKI, Y. Methods of test validation. In: SHOHAMY, E.; OR, I. G.; MAY, S. (Eds.). **Language Testing and Assessment**. 3. ed. Auckland: Springer, 2017.

YOUNG, R. Social dimensions of language testing. In: FULCHER, G.; HARDING, L. (Orgs.). **The Routledge Handbook of Language Testing**. 2. ed. London; New York: Routledge, 2021.

ZILIO, L. **VerbLexPor: um recurso léxico com anotação de papéis semânticos para o português**. Tese (Doutorado em Letras) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.

ANEXO A – Especificações para a elaboração de tarefas do Exame Celpe-Bras - Parte Escrita



ESPECIFICAÇÕES PARA A ELABORAÇÃO DE TAREFAS DO EXAME CELPE-BRAS – PARTE ESCRITA¹

I. Enunciado das tarefas

- O gênero, o propósito, o enunciador e o interlocutor devem ser claramente explicitados no enunciado. Deve ficar claro, ainda, o suporte de publicação do texto, ainda que este nem sempre precise ser explicitado.
- As tarefas devem visibilizar, sempre que possível, uma resposta plausível. Nesse sentido, deve-se evitar que o texto-base (vídeo, áudio ou texto escrito) seja um mero pretexto para a elaboração de uma proposta de produção escrita. Para tanto, durante a elaboração da tarefa, deve-se ter em conta a seguinte questão: “Se, na vida real, esse texto fosse lido/ouvido, que outros textos poderiam, na vida real, ser produzidos a partir dele (ou provocados por ele)?”. Deve-se ter cuidado para que a necessidade de avaliar a compreensão (auditiva ou escrita) não comprometa a plausibilidade da tarefa. Não se deve exigir, por exemplo, que o examinando recupere certa informação do texto-base simplesmente para avaliar se ele compreendeu um determinado ponto, caso isso não tenha relação com o propósito do texto a ser escrito.
- Na elaboração do enunciado, dar preferências a palavras de uso mais frequente em português, a fim de evitar que um examinando seja prejudicado pelo desconhecimento de um item lexical específico. Como o exame avalia níveis distintos de proficiência, é importante que os enunciados sejam compreensíveis mesmo para quem está em níveis iniciais, a fim de que o desempenho na tarefa não seja prejudicado pela falta de compreensão do enunciado.
- Dar preferência a tarefas em que o examinando seja levado a escrever um texto de um gênero diferente daquele que lhe foi apresentado como texto-base, a não ser que, no mundo real, a resposta para aquele dado texto seja no mesmo gênero (exemplo: escrever um artigo de opinião argumentando contra as ideias expressas em um artigo anteriormente publicado).
- Os gêneros solicitados devem ser preferencialmente do âmbito público, ainda que possam requerer registro informal (devemos evitar, por exemplo, pedir que o

¹ Elaborado pela Comissão Técnico-Científica do exame Celpe-Bras em julho de 2012 e revisado em julho de 2013. Membros da Comissão Técnico-Científica: Juliana Roquete Schoffen (coordenadora), Jerônimo Coura Sobrinho, Regina Peretti Dell’Isola, Matilde Scaramucci, Leandro Rodrigues Alves Diniz, Viviane Bagio Furtoso, Ricardo Moutinho, Ronaldo Amorim Lima, Leticia Grubert dos Santos, Natalia Tosatti, Simone Paula Kumrath. Atualizado em junho de 2016 pelos membros da Comissão Técnico-Científica: Edleise Mendes, Jerônimo Coura Sobrinho, Juliana Roquete Schoffen, Leandro Rodrigues Alves Diniz, Matilde Scaramucci, Patricia Maria Campos de Almeida e Regina Peretti Dell’Isola.

examinando escreva um e-mail para um amigo, visto que pode haver muita informação pressuposta nessa interlocução que pode dificultar a avaliação).

- As instruções da tarefa não devem criar contradições em relação ao conteúdo do texto-base, inclusive no que diz respeito ao tempo. Exemplo: se o vídeo trata de um projeto que já foi implementado, a tarefa não pode consistir na escrita de uma carta cujo objetivo seja incentivar a implementação do projeto.

II. Seleção dos textos-base

- Evitar textos que exijam uma grande carga de conhecimento cultural prévio.
- Evitar temas que possam ser tabus para algumas culturas ou sujeitos (exemplo: morte, tragédias, guerras, pena de morte);
- Não utilizar textos que reforcem estereótipos. Por outro lado, textos que os questionem podem ser utilizados de maneira produtiva para a elaboração de tarefas.
- Evitar textos que possam privilegiar certos grupos, de forma que a leitura possa ser dispensada e que a tarefa possa ser realizada com base em conhecimentos prévios (textos sobre tratamentos de doenças ou problemas de saúde, em que os examinandos médicos nem precisem ler o texto para realizar a tarefa, por exemplo). Não se trata apenas do tema, mas de como o tema é tratado. Buscar textos em que os assuntos sejam tratados de maneira original, não previsível.
- Evitar textos que tratem de episódios datados (circunscritos em um momento específico), que ficam desatualizados facilmente, perdendo seu sentido após poucos meses.
- Buscar textos que possibilitem a criação de uma situação-problema para ser resolvida, criticada ou discutida pelo examinando.

III. Características dos textos-base

A Tarefa I deve ter como insumo um segmento de vídeo, a Tarefa II um segmento de áudio e as Tarefas III e IV devem ter como insumo texto(s) escrito(s).

- Extensão do vídeo/áudio: 2' a 2'30" (se o vídeo/áudio original for maior, os cortes devem ser indicados e não devem prejudicar a qualidade e a unidade da imagem e do texto).
- Qualidade do vídeo/áudio: a diction deve ser clara e o volume das vozes adequado. Os ruídos não devem prejudicar a compreensão/audição do texto (se houver ruídos, o material deve passar por filtro ANTES de se fazer a tarefa).

- Se houver mais de uma pessoa falando, como no caso de entrevistas, deve-se cuidar para que não haja muita oscilação de volume entre os falantes ou interposição de vozes.
- Extensão do(s) texto(s) escrito(s): é desejável que o texto utilizado como insumo para cada tarefa não exceda uma página. No caso de dois textos, cuidar para que os dois juntos não excedam uma página.
- É preferível que os textos escritos utilizados como insumo tenham título interessante, fotos ilustrativas, gráficos, mapas ou outros elementos que possam ser explorados na tarefa. É preciso cuidar, porém, para que as imagens e fotos ou gráficos não apresentem toda a informação do texto, para não haver risco de o examinando ler apenas as imagens.

IV. Avaliação de Leitura / Compreensão Auditiva

Considerando que o Celpe-Bras avalia as habilidades de maneira integrada, é preciso que as tarefas permitam avaliar não apenas a produção escrita (capacidade de produzir um texto coeso e coerente em língua portuguesa, respeitando o gênero, o propósito e o interlocutor selecionados no enunciado da tarefa), mas também a compreensão auditiva (Tarefas I e II), imagética (Tarefa I) e escrita (Tarefas III e IV).

Algumas indicações importantes nesse sentido:

- O texto-base (vídeo, áudio ou texto escrito) deve ser indispensável para a realização da tarefa, de tal forma que, sem sua compreensão, seja impossível ao examinando produzir o texto solicitado. Assim, é inadequada uma tarefa em que o texto-base serve de mera “inspiração” para a escrita de um texto.
- Evitar textos cujo conteúdo informacional seja muito específico de modo que possa ser facilmente compreendido por uns e dificilmente percebido por outros, devido ao conhecimento prévio de alguns examinandos (Exemplo: Um examinando pode se sair muito bem na Tarefa I mesmo com grandes dificuldades de compreensão auditiva, se ele conhece o Rio de Janeiro e o vídeo se limita a apresentar os principais pontos turísticos dessa cidade).
- Os textos selecionados devem conter elementos suficientes para possibilitar a escrita de textos autônomos, que tenham relação com as informações contidas no material de insumo, mas que favoreçam um trabalho que vá além da mera recuperação de informações. São inadequadas, por exemplo, tarefas que podem ser respondidas por meio da cópia/transcrição ou paráfrase de diferentes trechos do texto-base.



No Manual do Examinando (2010), encontram-se especificadas diferentes sub-habilidades que devem ser avaliadas no exame (ver quadro abaixo). O elaborador deve procurar elaborar tarefas que possam avaliar essas habilidades (se possível, mais de uma em uma mesma tarefa), para que tenhamos um conjunto de possibilidades, e não apenas tarefas que avaliem, por exemplo, localização de informação específica.

SUB-HABILIDADES AVALIADAS NO EXAME CELPE-BRAS

- Reconhecer a situação de comunicação (quem fala, para quem, em que contexto, em que veículo, com que objetivo, em que registro, etc.).
- Localizar e entender informação específica no texto.
- Identificar a ideia principal do texto.
- Fazer distinção entre pontos principais e detalhes de apoio.
- Identificar a finalidade ou o objetivo do texto.
- Relacionar tipografia, **layout**, imagens para compreender o propósito do texto.
- Decidir se o texto é relevante (no todo ou em parte) para as ações a serem desenvolvidas na execução de uma tarefa.
- Reconhecer atitudes, emoções e ponto de vista do autor quando expressos explicita ou implicitamente no texto.
- Expressar, como escritor, suas atitudes, emoções e ponto de vista.
- Destacar os pontos relevantes para resumir o texto, uma ideia específica ou subjacente.
- Acompanhar e registrar o desenvolvimento de um argumento.
- Decidir se o texto é baseado em fato, opinião, pesquisa, etc.
- Interpretar gráficos e outros materiais apresentados visualmente.
- Reescrever informação no mesmo estilo ou em estilo diferente.
- Transferir informação de/para diagramas, tabelas, gráficos.
- Reconhecer marcas linguísticas características de diferentes gêneros do discurso.

Manual do Candidato (2010)

V. Combinação das quatro tarefas para a elaboração da prova

Os gêneros, enunciadores, interlocutores e propósitos das quatro tarefas devem ser tão diversificados quanto possível. Conseqüentemente, os registros a serem empregados em cada tarefa também deverão ser distintos. A fim de garantir essa diversidade, deve-se cuidar para que com que o grupo de elaboradores produza tarefas com propósitos distintos. Para isso, é importante fazer um levantamento dos propósitos que estão aparecendo durante

a elaboração das tarefas, pois é preciso que haja variedade. Abaixo, o quadro com os propósitos, gêneros dos textos de insumo e interlocutores apresentados no Manual do Examinando (2010):

Propósitos:

Narrar, relatar, argumentar, expor, instruir, agradecer, pedir, opinar, comentar, expressar atitudes, confirmar, desculpar-se, informar, reclamar, justificar, persuadir, aconselhar, avisar (além desses, que estão no manual, incluímos aqui também comparar, descrever, discutir).

Gêneros que podem servir de insumo para as tarefas I e II: entrevistas, depoimentos, noticiários, debates, reportagens, documentários, anúncios de produtos ou endereços úteis, programa musical, programa de auditório ou de variedades, previsão do tempo, receita, palestra, aula, instruções, informes de trânsito e de situação nas estradas, mensagens na secretária eletrônica, conversas ao telefone, filmes, seriados, novelas, peças de teatro etc.

Gêneros que podem servir de insumo para as tarefas III e IV:

- Textos de periódicos (jornais e revistas) – editorial, notícia, entrevista, reportagem, anúncio classificado, publicidade, cartas de leitores, horóscopo, cartuns, quadrinhos, etc.;
- Textos de livros – crônica, conto, poema, texto didático, receita etc.;
- Textos de panfletos, cartazes, avisos, placas de trânsito;
- Textos de telegramas, cartas, bilhetes, e-mails, cartões-postais;
- Textos de diários, agendas, notas, listagens, resenhas, relatórios, currículos, biografias;
- Textos de documentos, formulários, questionários, instruções; de mapas, roteiros, quadro de horários, calendários, programas, cardápios, recibos;
- Textos de dicionários, catálogos, listas telefônicas, letras de música, legendas de filme etc.

Interlocutores:

- Falantes de português em geral, em situações que requerem registro formal e informal.

Manual do Candidato (2010)

VI. Recomendações para elaboração da Tarefa I

VII.

Características das imagens: deve haver imagens significativas para a compreensão ou ilustrativas do que está sendo falado, de modo que a tarefa possa explorar também as



imagens. Se o vídeo apresentar informações sobre a cultura brasileira, deve trazer imagens que colaborem para a compreensão dos aspectos culturais, não exigindo tanto conhecimento prévio do examinando.

Observação: É importante lembrar que o vídeo deve conter informação “falada” em quantidade suficiente para realizar-se a avaliação. Vídeos em que a compreensão pode se dar apenas a partir das imagens devem ser evitados. Também devem ser evitados vídeos em que apenas a informação falada seja explorada (senão vira áudio). Na medida do possível, devem ser explorados aspectos multimodais da composição do vídeo. No enunciado, deve estar clara a necessária compreensão da constituição do vídeo em seus aspectos composicionais que relacionam texto oral e imagem.

7

VIII. Recomendações para elaboração da TAREFA II

Como o áudio é o único recurso que os examinandos possuirão para cumprir a tarefa solicitada, precisa-se ter especial atenção à qualidade do som. Áudios com chiados devem passar por filtro ANTES de ser elaborada a tarefa.

IX. Recomendações para elaboração da TAREFA III

A Tarefa III pode solicitar a produção de textos de registro menos formal ou com complexidade menor. Pode-se solicitar que o examinando, para cumprir a tarefa, estabeleça comparação entre dois textos, ou entre um texto e um gráfico ou imagem, ou que use parte das informações de cada texto para responder ao enunciado da tarefa. É importante que a tarefa não suscite cópia literal do(s) texto(s)-base.

X. Recomendações para elaboração da TAREFA IV

Nesta tarefa, pode-se solicitar a produção de gêneros mais formais, nos quais há maior distância entre enunciador e interlocutor. A Tarefa IV deve solicitar, preferencialmente, que o examinando apresente e sustente um posicionamento. O gênero selecionado deve exigir que o examinando escreva textos bem desenvolvidos (panfletos, listas, dicas e instruções devem ser evitados).

Pode-se solicitar que o examinando, para cumprir a tarefa, complete um quadro ou gráfico, defenda claramente seu posicionamento ou contraponha/concorde com a opinião de um especialista. É importante que a tarefa não suscite cópia literal do(s) texto(s)-base.

Fonte: Schoffen *et al.*, 2012, *apud* Mendel, 2019, p. 181.

ANEXO B – Eixos de sub-habilidades das partes escrita e oral do Exame Celpe-Bras

Quadro 8 Eixos de sub-habilidades das partes escrita e oral do Exame Celpe-Bras

(continua)

	Parte Escrita	Parte Oral
Eixos de avaliação	<ul style="list-style-type: none"> • Configuração da relação de interlocução no gênero discursivo proposto na tarefa e realização da ação solicitada • Recontextualização autoral e apropriada das informações necessárias para cumprir o propósito interlocutivo de forma consistente • Produção de texto autônomo, claro e coeso, revelando acionamento de recursos linguísticos apropriados para configurar a relação de interlocução no gênero solicitado 	<ul style="list-style-type: none"> • Compreensão do fluxo de fala do interlocutor • Competência interacional • Fluência • Adequação lexical • Adequação gramatical • Pronúncia (sons, ritmo e entonação)
Sub-habilidades de compreensão avaliadas	<ul style="list-style-type: none"> • Reconhecer a situação de comunicação (quem fala ou escreve, para quem, em que contexto, em que veículo, com que objetivo, em que registro, em que gênero, entre outros) • Reconhecer marcas linguísticas características de diferentes gêneros do discurso • Identificar a ideia principal do texto • Identificar detalhes e ideias secundárias • Localizar informações explícitas no texto • Inferir palavras desconhecidas e implícitas no texto • Selecionar informações do texto necessárias para o cumprimento do propósito da tarefa • Reconhecer atitudes, emoções e pontos de vista do autor quando expressos explícita ou implicitamente no texto • Interpretar gráficos e outros materiais apresentados visualmente 	<ul style="list-style-type: none"> • Compreender o fluxo natural da fala em situação de interação face a face • Compreender a temática abordada ou problematizada na interação face a face • Compreender intervenções e perguntas do interlocutor em uma interação face a face • Reconhecer atitudes, emoções e pontos de vista expressos explícita ou implicitamente pelo interlocutor • Inferir palavras desconhecidas • Compreender aspectos culturais abordados na interação

(conclusão)

Quadro 8 Eixos de sub-habilidades das partes escrita e oral do Exame Celpé-Bras

	Parte Escrita	Parte Oral
Sub-habilidades de produção avaliadas	<ul style="list-style-type: none">• Interpretar gráficos e outros materiais apresentados visualmente• Expressar, como escritor, suas atitudes, emoções e ponto de vista• Sumarizar parte(s) do texto de insumo• Retextualizar informação contida no texto de insumo• Reescrever informação com outro propósito, em outra relação de interlocução, no mesmo gênero ou em gênero diferente• Adaptar informação de/para diagramas, tabelas, gráficos• Utilizar recursos linguísticos e discursivos característicos de diferentes gêneros do discurso	<ul style="list-style-type: none">• Partilhar, oralmente, conhecimento de diversas ordens• Expressar ideias e opiniões sobre assuntos variados• Expressar-se, com base nas intervenções e perguntas feitas pelo interlocutor• Usar diferentes estratégias comunicativas, quando necessário• Utilizar recursos lexicais e gramaticais para discutir tópicos do cotidiano ou assuntos variados• Utilizar sons, ritmo e entonação da língua portuguesa

Fonte: Elaboração própria.

ANEXO C – Quadro dos Parâmetros de avaliação da Parte Escrita

Quadro 2 Parâmetros de avaliação da Parte Escrita

Parâmetros de Avaliação da parte escrita	
5 –	Configura adequadamente a relação de interlocução no gênero discursivo proposto na tarefa, realizando a ação solicitada. Recontextualiza apropriadamente e de maneira autoral as informações necessárias para cumprir o propósito interlocutivo de forma consistente. Eventuais inadequações ou equívocos não comprometem a configuração da interlocução. Produz um texto autônomo, claro e coeso, em que os recursos linguísticos acionados são apropriados para configurar a relação de interlocução no gênero solicitado, e possíveis inadequações raramente comprometem a fluidez da leitura.
4 –	Configura a relação de interlocução no gênero discursivo proposto na tarefa, realizando a ação solicitada. Recontextualiza apropriadamente as informações necessárias para cumprir o propósito interlocutivo, mas possíveis equívocos ou incompletudes podem fragilizar, em momentos localizados, a consistência da interlocução. Os recursos linguísticos acionados são apropriados para configurar a relação de interlocução no gênero proposto, construindo um texto claro e coeso em que possíveis inadequações podem comprometer, em momentos localizados, a fluidez na leitura.
3 –	Configura a relação de interlocução no gênero discursivo proposto na tarefa, realizando a ação solicitada, ainda que a consistência da relação de interlocução possua algumas falhas. Pode recontextualizar de forma pouco articulada e/ou equivocada ou não recontextualizar informações necessárias para cumprir o propósito dentro do contexto de produção solicitado. Os recursos linguísticos acionados são apropriados, podendo apresentar limitações ou inadequações que podem prejudicar, em alguns momentos, a configuração da interlocução no gênero proposto. Problemas de clareza e coesão podem ocasionar, em alguns momentos, dificuldades na leitura.
2 –	Configura a relação de interlocução de forma pouco consistente, realizando superficialmente a ação solicitada. Pode estabelecer uma relação de interlocução próxima à solicitada, não cumprir propósito(s) menor(es) e/ou apresentar problemas na construção do gênero. Pode apresentar trechos do texto que remetem a um gênero diferente, comprometendo a relação de interlocução. A relação entre o propósito do texto e a interlocução configurada não é clara ou não é totalmente adequada. Pode não recontextualizar informações que seriam necessárias para a configuração adequada da interlocução ou não articular claramente essas informações. Equívocos de compreensão podem comprometer parcialmente o cumprimento do propósito. Os recursos linguísticos acionados são limitados e/ou inadequados, podendo prejudicar parcialmente a configuração da relação de interlocução no gênero solicitado. Problemas de clareza e coesão podem ocasionar, em diferentes momentos, dificuldades na leitura.
1 –	Configura com problemas recorrentes ou não configura a relação de interlocução solicitada, realizando muito superficialmente ou não realizando a ação solicitada. Remete-se ao tema, mas pode não considerar o contexto de produção e não construir o gênero discursivo proposto ou apresentar problemas recorrentes na sua construção. Não recontextualiza informações suficientes para o cumprimento do propósito comunicativo, considerando a relação de interlocução configurada. OU Pode apresentar equívocos graves e/ou frequentes de compreensão que comprometem o cumprimento do propósito. Os recursos linguísticos acionados são muito limitados e/ou inadequados, o que prejudica substancialmente o cumprimento do propósito e a configuração da relação de interlocução, comprometendo a construção do gênero solicitado. Problemas frequentes de clareza e coesão ocasionam, em vários momentos, problemas na leitura.
0 –	Não configura, ou configura de forma equivocada, a relação de interlocução, não realizando a ação solicitada. OU Trata de outro tema. OU Demonstra problemas generalizados de compreensão, impedindo o cumprimento do propósito e a configuração da relação de interlocução E/OU Limita-se a reproduzir o(s) texto(s)-base(s), sem marcas de autoria. OU Ignora completamente os texto(s)-base(s). E/OU Problemas generalizados de clareza e coesão e/ou inadequações linguísticas impedem a configuração da relação de interlocução no gênero solicitado, comprometendo a compreensão geral do texto. OU A produção é insuficiente para a avaliação.

ANEXO D – Quadro de níveis de proficiência certificados pelo Exame Celpe-Bras

Quadro 9 Níveis de proficiência certificados pelo Exame Celpe-Bras

(continua)

Avançado Superior

O examinando que atinge o nível Avançado Superior é capaz de produzir textos escritos claros e coesos de diferentes gêneros discursivos sobre assuntos variados, configurando a interlocução de forma adequada e consistente, utilizando recursos lexicais e gramaticais apropriados aos gêneros produzidos. É capaz de recontextualizar, com propriedade, informações relevantes obtidas a partir da interpretação de textos de diferentes gêneros orais e escritos, demonstrando compreensão eficiente e seletiva. Eventuais inadequações pontuais não comprometem o bom cumprimento dos propósitos dos textos produzidos.

É capaz de interagir oralmente com muita autonomia e desenvoltura, utilizando vocabulário amplo e adequado e variedade também ampla de estruturas para expressar ideias e opiniões sobre assuntos variados, contribuindo muito para o desenvolvimento da interação. Apresenta fluência, sem interrupções do fluxo natural da conversa, e pronúncia adequada. Demonstra compreensão do fluxo natural da fala do interlocutor, com rara necessidade de repetição e/ou reestruturação.

Avançado

O examinando que atinge o nível Avançado é capaz de produzir textos escritos claros e coesos de diferentes gêneros discursivos sobre assuntos variados, configurando a interlocução de forma adequada, utilizando recursos lexicais e gramaticais apropriados aos gêneros produzidos. É capaz de recontextualizar adequadamente informações relevantes obtidas a partir da interpretação de textos de diferentes gêneros orais e escritos, demonstrando compreensão eficiente. Inadequações pontuais podem fragilizar partes do texto, ainda que não comprometam o cumprimento dos propósitos dos textos produzidos.

É capaz de interagir oralmente com autonomia e desenvoltura para a expressão de ideias e opiniões sobre assuntos variados, contribuindo para o desenvolvimento da interação. Demonstra fluência, com poucas interrupções do fluxo natural da conversa. Seu vocabulário é amplo e adequado, com poucas interferências de outras línguas. Utiliza uma variedade ampla e adequada de estruturas, com poucas inadequações no uso de estruturas complexas e raras inadequações no uso de estruturas básicas. Sua pronúncia pode apresentar algumas inadequações e/ou interferências de outras línguas. Demonstra compreensão do fluxo natural da fala do interlocutor, com alguma necessidade de repetição e/ou reestruturação ocasionada por palavras menos frequentes e/ou por aceleração da fala.

Intermediário Superior

O examinando que atinge o nível Intermediário Superior é capaz de produzir textos escritos de diferentes gêneros discursivos sobre assuntos variados, podendo configurar a interlocução de forma nem sempre adequada e mobilizando recursos lexicais e gramaticais nem sempre apropriados aos gêneros produzidos, podendo apresentar problemas de clareza, coesão e/ou inadequações que podem comprometer a fluidez da leitura. É capaz de recontextualizar, ainda que com equívocos, informações a partir da interpretação de textos de diferentes gêneros orais e escritos, podendo demonstrar problemas de compreensão. Inadequações podem dificultar o cumprimento dos propósitos dos textos produzidos.

(conclusão)

Intermediário Superior

É capaz de interagir oralmente para a expressão de ideias e opiniões sobre assuntos variados. Demonstra fluência, com algumas pausas e hesitações que às vezes interrompem o fluxo da conversa. Seu vocabulário é adequado, embora apresente algumas interferências de outras línguas. Apresenta algumas inadequações no uso de estruturas complexas e poucas no uso de estruturas básicas. Sua pronúncia contém inadequações e/ou interferências de outras línguas. Demonstra alguns problemas de compreensão do fluxo natural da fala do interlocutor, com necessidade de repetição e/ou reestruturação ocasionada por palavras de uso frequente, em ritmo normal da fala.

Intermediário

O examinando que atinge o nível Intermediário é capaz de produzir textos escritos sobre assuntos variados que, com dificuldade, podem ser reconhecidos como pertencentes a determinados gêneros discursivos, podendo não configurar adequadamente a interlocução. Os recursos lexicais e gramaticais mobilizados são limitados, podendo apresentar problemas de clareza e coesão e/ou inadequações frequentes que comprometem mais frequentemente a fluidez da leitura. É capaz de selecionar algumas informações a partir da interpretação de textos de diferentes gêneros orais e escritos, evidenciando problemas de compreensão e dificuldades no trabalho de recontextualização que podem levar ao cumprimento parcial dos propósitos dos textos produzidos. É capaz de interagir oralmente para a expressão de ideias e opiniões sobre assuntos variados. Apresenta poucas hesitações, com algumas interrupções no fluxo da conversa. Seu vocabulário pode apresentar limitações que podem comprometer o desenvolvimento da interação. Utiliza variedade limitada de estruturas, com algumas inadequações em estruturas complexas e poucas inadequações em estruturas básicas. Sua pronúncia contém inadequações e/ou interferências frequentes de outras línguas. Demonstra alguns problemas de compreensão do fluxo da fala, com necessidade frequente de repetição e/ou reestruturação ocasionada por palavras de uso frequente em nível normal de fala.

Fonte: Brasil, 2020, p. 67.

ANEXO E – As 200 métricas do NILC-Matrix

GRUPOS	DESCRIÇÃO	MÉTRICAS
1. Medidas Descritivas	10 métricas de estatísticas básicas que descrevem o <i>corpus</i>	Quantidade de Parágrafos no texto (id: 152) paragraphs Quantidade de Sentenças no texto (id: 154) sentences Quantidade média de sentenças por parágrafo no texto (id: 155) sentences_per_paragraph Quantidade média de sílabas por palavra no texto (id: 157) syllables_per_content_word Quantidade de Palavras no texto (id: 158) words Média de Palavras por Sentença (id: 159) words_per_sentence Quantidade Máxima de palavras por sentença (id: 153) sentence_length_max Quantidade Mínima de palavras por sentença (id: 184) sentence_length_min Desvio Padrão da quantidade de palavras por sentença (id: 185) sentence_length_standard_deviation Proporção de subtítulos em relação à quantidade de sentenças do texto (id: 156) subtitles
2. Simplicidade Textual	8 métricas que indicam a extensão das sentenças, uso de palavras de conteúdo e de pronomes	Proporção de pronomes pessoais que indicam uma conversa com o leitor em relação à quantidade de pronomes pessoais do texto (id: 186) dialog_pronoun_ratio Proporção de conjunções fáceis em relação à quantidade de palavras do texto (id: 187) easy_conjunctions_ratio Proporção de conjunções difíceis em relação à quantidade de palavras do texto (id: 188) hard_conjunctions_ratio Proporção de Sentenças Muito Longas em relação a todas as sentenças do texto (id: 189) long_sentence_ratio Proporção de Sentenças Longas em relação a todas as sentenças do texto (id: 190) medium_long_sentence_ratio Proporção de Sentenças Médias em relação a todas as sentenças do texto (id: 191) medium_short_sentence_ratio Proporção de Sentenças Curtas em relação a todas as sentenças do texto (id: 192) short_sentence_ratio Proporção de palavras de conteúdo simples em relação a todas as palavras de conteúdo do texto (id: 193) simple_word_ratio
3. Coesão Referencial	9 métricas que indicam cadeias de correferência intraparágrafo e interparágrafo	Média das proporções de candidatos a referentes na sentença anterior em relação aos pronomes pessoais do caso reto nas sentenças (id: 17) adjacent_refs

		<p>Média das proporções de candidatos a referentes nas 5 sentenças anteriores em relação aos pronomes anafóricos das sentenças (id: 18) <code>anaphoric_refs</code></p> <p>Quantidade média de referentes que se repetem nos pares de sentenças adjacentes do texto (id: 1) <code>adj_arg_ovl</code></p> <p>Quantidade média de palavras de conteúdo que se repetem nos pares de sentenças adjacentes do texto (id: 2) <code>adj_cw_ovl</code></p> <p>Quantidade média de radicais de palavras de conteúdo que se repetem nos pares de sentenças adjacentes do texto. (id: 3) <code>adj_stem_ovl</code></p> <p>Quantidade média de referentes que se repetem nos pares de sentenças do texto (id: 4) <code>arg_ovl</code></p> <p>Quantidade média de radicais de palavras de conteúdo que se repetem nos pares de sentenças do texto (id: 5) <code>stem_ovl</code></p> <p>Média de candidatos a referente, na sentença anterior, por pronome anafórico do caso reto (id: 194) <code>coreference_pronoun_ratio</code></p> <p>Média de candidatos a referente, na sentença anterior, por pronome demonstrativo anafórico (id 195) <code>demonstrative_pronoun_ratio</code></p>
4. Coesão Semântica	11 métricas que capturam relações semânticas e similaridade entre palavras	<p>Média de similaridade entre pares de sentenças adjacentes no texto (id: 7) <code>lsa_adj_mean</code></p> <p>Desvio padrão de similaridade entre pares de sentenças adjacentes no texto (id: 8) <code>lsa_adj_std</code></p> <p>Média de similaridade entre todos os pares de sentenças no texto (id: 9) <code>lsa_all_mean</code></p> <p>Desvio padrão de similaridade entre todos os pares possíveis de sentenças do texto (id: 10) <code>lsa_all_std</code></p> <p>Média do <code>*givenness*</code> da cada sentença do texto, a partir da segunda (id: 11) <code>lsa_givenness_mean</code></p> <p>Desvio padrão do <code>*givenness*</code> da cada sentença do texto, a partir da segunda (id: 12) <code>lsa_givenness_std</code></p> <p>Média de similaridade entre pares de parágrafos adjacentes no texto (id: 13) <code>lsa_paragraph_mean</code></p> <p>Desvio padrão entre parágrafos adjacentes no texto (id: 14) <code>lsa_paragraph_std</code></p> <p>Média do <code>*span*</code> da cada sentença do texto, a partir da segunda (id: 15) <code>lsa_span_mean</code></p> <p>Desvio padrão do <code>span</code> da cada sentença do texto, a partir da segunda (id: 16) <code>lsa_span_std</code></p> <p>Média da entropia cruzadas das sentenças do texto (id: 6) <code>cross_entropy</code></p>
5. Medidas	24 métricas provenientes de	Proporção de palavras com valor de

Psicolinguísticas	experimentos psicolinguísticos com base em escala likert de 7 níveis	<p>concretude entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 160) concretude_1_25_ratio</p> <p>Proporção de palavras com valor de concretude entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 161) concretude_25_4_ratio</p> <p>Proporção de palavras com valor de concretude entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto (id: 162) concretude_4_55_ratio</p> <p>Proporção de palavras com valor de concretude entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 163) concretude_55_7_ratio</p> <p>Média dos valores de concretude das palavras de conteúdo do texto (id: 164) concretude_mean</p> <p>Desvio padrão do valor de concretude das palavras de conteúdo do texto (id: 165) concretude_std</p> <p>Proporção de palavras com valor de familiaridade entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 166) familiaridade_1_25_ratio</p> <p>Proporção de palavras com valor de familiaridade entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 167) familiaridade_25_4_ratio</p> <p>Proporção de palavras com valor de familiaridade entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto (id: 168) familiaridade_4_55_ratio</p> <p>Proporção de palavras com valor de familiaridade entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 169) familiaridade_55_7_ratio</p> <p>Média dos valores de familiaridade das palavras de conteúdo do texto (id: 170) familiaridade_mean</p> <p>Desvio padrão dos valores de familiaridade das palavras de conteúdo do texto (id: 171) familiaridade_std</p> <p>Proporção de palavras com valor de idade de aquisição entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 172) idade_aquisicao_1_25_ratio</p> <p>Proporção de palavras com valor de idade de aquisição entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 173) idade_aquisicao_25_4_ratio</p> <p>Proporção de palavras com valor de idade de aquisição entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto (id: 174) idade_aquisicao_4_55_ratio</p> <p>Proporção de palavras com valor de idade de aquisição entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 175) idade_aquisicao_55_7_ratio</p>
-------------------	--	---

		<p>Média dos valores de idade de aquisição das palavras de conteúdo do texto (id: 176) idade_aquisicao_mean</p> <p>Desvio padrão dos valores de idade de aquisição das palavras de conteúdo do texto (id: 177) idade_aquisicao_std</p> <p>Proporção de palavras com valor de imageabilidade entre 1 e 2,5 em relação a todas as palavras de conteúdo do texto (id: 178) imageabilidade_1_25_ratio</p> <p>Proporção de palavras com valor de imageabilidade entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto (id: 179) imageabilidade_25_4_ratio</p> <p>Proporção de palavras com valor de imageabilidade entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto (id: 180) imageabilidade_4_55_ratio</p> <p>Proporção de palavras com valor de imageabilidade entre 5,5 e 7 em relação a todas as palavras de conteúdo do texto (id: 181) imageabilidade_55_7_ratio</p> <p>Média dos valores de imageabilidade das palavras de conteúdo do texto (id: 182) imageabilidade_mean</p> <p>Desvio padrão dos valores de imageabilidade das palavras de conteúdo do texto (id: 183) imageabilidade_std</p>
6. Diversidade Lexical	15 métricas que descrevem o tipo de léxico utilizado no <i>corpus</i>	<p>Proporção de types (despreza repetições de palavras) em relação à quantidade de tokens (computa repetições de palavras) no texto (id: 75) trr</p> <p>Proporção de types de adjetivos em relação à quantidade de tokens de adjetivos no texto (id: 62) adjective_diversity_ratio</p> <p>Proporção de palavras de conteúdo em relação à quantidade de palavras funcionais do texto (id: 63) content_density</p> <p>Proporção de types de palavras de conteúdo em relação à quantidade de tokens de palavras de conteúdo no texto (id: 64) content_word_diversity</p> <p>Proporção máxima de palavras de conteúdo em relação à quantidade de palavras das sentenças (id: 65) content_word_max</p> <p>Proporção Mínima de palavras de conteúdo por quantidade de palavras nas sentenças (id: 66) content_word_min</p> <p>Desvio padrão das proporções entre as palavras de conteúdo e a quantidade de palavras das sentenças (id: 67) content_word_standard_deviation</p> <p>Proporção de types de palavras funcionais em relação à quantidade de tokens de palavras funcionais no texto (id: 68) function_word_diversity</p> <p>Proporção de types de pronomes indefinidos em relação à quantidade de tokens de</p>

		<p>pronomes indefinidos no texto (id: 69) indefinite_pronouns_diversity</p> <p>Proporção de types de substantivos em relação à quantidade de tokens de substantivos no texto (id: 70) noun_diversity</p> <p>Proporção de types de preposições em relação à quantidade de tokens de preposições no texto (id: 71) preposition_diversity</p> <p>Proporção de types de pronomes em relação à quantidade de tokens de pronomes no texto (id: 72) pronoun_diversity</p> <p>Proporção de types de pontuações em relação à quantidade de tokens de pontuações no texto (id: 73) punctuation_diversity</p> <p>Proporção de types de pronomes relativos em relação à quantidade de tokens de pronomes relativos no texto (id: 74) relative_pronouns_diversity_ratio</p> <p>Proporção de types de verbos em relação à quantidade de tokens de verbos no texto (id: 76) verb_diversity</p>
7. Conectivos	12 métricas referentes ao uso de conectivos aditivos, causais e lógicos	<p>Proporção de conectivos aditivos negativos em relação à quantidade de palavras do texto (id: 46) add_neg_conn_ratio</p> <p>Proporção de conectivos aditivos positivos em relação à quantidade de palavras do texto (id: 47) add_pos_conn_ratio</p> <p>Proporção de conectivos causais negativos em relação à quantidade de palavras do texto (id: 49) cau_neg_conn_ratio</p> <p>Proporção de conectivos causais positivos em relação à quantidade de palavras do texto (id: 50) cau_pos_conn_ratio</p> <p>Proporção de Conectivos em relação à quantidade de palavras do texto (id: 51) conn_ratio</p> <p>Proporção de Conectivos Lógicos Negativos em relação à quantidade de palavras do texto (id: 53) log_neg_conn_ratio</p> <p>Proporção de Conectivos Lógicos Positivos em relação à quantidade de palavras do texto (id: 54) log_pos_conn_ratio</p> <p>Proporção do operador lógico E em relação à quantidade de palavras do texto (id: 48) and_ratio</p> <p>Proporção do operador lógico SE em relação à quantidade de palavras do texto (id: 52) if_ratio</p> <p>Proporção de Operadores Lógicos em relação à quantidade de palavras do texto (id: 55) logic_operators</p> <p>Proporção de palavras que denotam negação em relação à quantidade de palavras do texto (id: 56) negation_ratio</p> <p>Proporção do operador lógico OU em relação à quantidade de palavras do texto (id: 57) or_ratio</p>

8. Léxico Temporal	12 métricas que avaliam o uso de tempos e modos verbais e de conectivos temporais	<p>Proporção de conectivos temporais negativos em relação à quantidade de palavras do texto (id: 150) tmp_neg_conn_ratio</p> <p>Proporção de conectivos temporais positivos em relação à quantidade de palavras do texto (id: 151) tmp_pos_conn_ratio</p> <p>Proporção de verbos auxiliares seguidos de participio em relação à quantidade de sentenças do texto (id: 141) aux_plus_PCP_per_sentence</p> <p>Proporção de Verbos no Pretérito Imperfeito do Indicativo em relação à quantidade de verbos flexionados no texto (id: 142) indicative_imperfect_ratio</p> <p>Proporção de Verbos no Pretérito Mais que Perfeito do Indicativo em relação à quantidade de verbos flexionados no texto (id: 143) indicative_pluperfect_ratio</p> <p>Proporção de Verbos no Presente do Indicativo em relação à quantidade de verbos flexionados no texto (id: 144) indicative_present_ratio</p> <p>Proporção de Verbos no Pretérito Perfeito Simples do Indicativo em relação à quantidade de verbos flexionados no texto (id: 145) indicative_preterite_perfect_ratio</p> <p>Proporção de verbos no participio em relação a todos os verbos do texto (id: 146) participle_verbs</p> <p>Quantidade de diferentes tempos-modos verbais que ocorrem no texto (id: 140) verbal_time_moods_diversity</p> <p>Proporção de Verbos no Futuro do Subjuntivo em relação à quantidade de verbos flexionados no texto (id: 147) subjunctive_future_ratio</p> <p>Proporção de Verbos no Pretérito Imperfeito do Subjuntivo em relação à quantidade de verbos flexionados no texto (id: 148) subjunctive_imperfect_ratio</p> <p>Proporção de Verbos no Presente do Subjuntivo em relação à quantidade de verbos flexionados no texto (id: 149) subjunctive_present_ratio</p>
9. Complexidade Sintática	27 métricas que descrevem as estruturas sintáticas	<p>Quantidade Média de palavras antes dos verbos principais das orações principais das sentenças (id: 44) words_before_main_verb</p> <p>Quantidade média de adjuntos adverbiais por oração do texto (id: 19) adjunct_per_clause</p> <p>Proporção de orações com advérbio antes do verbo principal em relação à quantidade de orações do texto (id: 20) adverbs_before_main_verb_ratio</p> <p>Quantidade média de apostos por oração do texto (id: 21) apposition_per_clause</p> <p>Quantidade média de orações por sentença (id: 22) clauses_per_sentence</p> <p>Proporção de conjunções coordenativas em</p>

		<p>relação a todas as orações do texto (id: 23) coordinate_conjunctions_per_clauses</p> <p>Distância na árvore de dependências (id: 24) dep_distance</p> <p>Fórmula de Complexidade Sintática de Frazier (id: 25) frazier</p> <p>Proporção de orações subordinadas reduzidas pela quantidade de orações do texto (id: 26) infinite_subordinate_clauses</p> <p>Proporção de orações que não estão no formato SVO (sujeito-verbo-objeto) em relação a todas orações do texto (id: 27) non_svo_ratio</p> <p>Proporção de orações na voz passiva analítica em relação à quantidade de orações do texto (id: 28) passive_ratio</p> <p>Proporção de sujeitos pospostos em relação a todos os sujeitos do texto (id: 29) postponed_subject_ratio</p> <p>Proporção de conjunções coordenativas em relação a todas as conjunções do texto (id: 30) ratio_coordinate_conjunctions</p> <p>Proporção de conjunções subordinativas em relação a todas conjunções do texto (id: 31) ratio_subordinate_conjunctions</p> <p>Proporção de orações relativas em relação à quantidade de orações do texto (id: 32) relative_clauses</p> <p>Proporção de sentenças com 5 orações em relação a todas as sentenças do texto (id: 33) sentences_with_five_clauses</p> <p>Proporção de sentenças com 4 orações em relação a todas as sentenças do texto (id: 34) sentences_with_four_clauses</p> <p>Proporção de sentenças com 1 oração em relação a todas as sentenças do texto (id: 35) sentences_with_one_clause</p> <p>Proporção de sentenças com 7 ou mais orações em relação a todas as sentenças do texto (id: 36) sentences_with_seven_more_clauses</p> <p>Proporção de sentenças com 6 orações em relação a todas as sentenças do texto (id: 37) sentences_with_six_clauses</p> <p>Proporção de sentenças com 3 orações em relação a todas as sentenças do texto (id: 38) sentences_with_three_clauses</p> <p>Proporção de sentenças com 2 orações em relação a todas as sentenças do texto (id: 39) sentences_with_two_clauses</p> <p>Proporção de sentenças sem verbos em relação a todas as sentenças do texto (id: 40) sentences_with_zero_clause</p> <p>Desvio-padrão do tamanho dos sintagmas nominais do text (id: 41) std_noun_phrase</p> <p>Proporção de orações subordinadas pela quantidade de orações do texto (id: 42) subordinate_clauses</p> <p>Proporção de adjuntos adverbiais de tempo</p>
--	--	--

		em relação a todos os adjuntos adverbiais do texto (id: 43) <code>temporal_adjunct_ratio</code> Fórmula de Complexidade Sintática de Yngve (id: 45) <code>yngve</code>
10. Densidade de Padrões Sintáticos	4 métricas sobre sintagmas nominais	Média dos tamanhos médios dos sintagmas nominais nas sentenças (id: 60) <code>mean_noun_phrase</code> Proporção de verbos no gerúndio em relação a todos os verbos do texto (id: 58) <code>gerund_verbs</code> Máximo entre os tamanhos de sintagmas nominais do texto (id: 59) <code>max_noun_phrase</code> Mínimo entre os tamanhos de sintagmas nominais do texto (id: 61) <code>min_noun_phrase</code>
11. Informações Morfossintáticas de Palavras	42 métricas relativas às classes gramaticais da língua	Proporção de Adjetivos em relação à quantidade de palavras do texto (id: 87) <code>adjective_ratio</code> Proporção de Advérbios em relação à quantidade de palavras do texto (id: 91) <code>adverbs</code> Proporção de palavras de conteúdo em relação à quantidade de palavras do texto (id: 96) <code>content_words</code> Proporção de Palavras Funcionais em relação à quantidade de palavras do texto (id: 99) <code>function_words</code> Proporção de substantivos em relação à quantidade de palavras do texto (id: 106) <code>noun_ratio</code> Proporção de pronomes em relação à quantidade de palavras do texto (id: 114) <code>pronoun_ratio</code> Proporção de Verbos em relação à quantidade de palavras do texto (id: 125) <code>verbs</code> Proporção de Pronomes Pessoais em relação à quantidade de palavras do texto (id: 111) <code>personal_pronouns</code> Proporção máxima de adjetivos em relação à quantidade de palavras das sentenças (id: 88) <code>adjectives_max</code> Proporção mínima de adjetivos em relação à quantidade de palavras das sentenças (id: 89) <code>adjectives_min</code> Desvio padrão das proporções entre adjetivos e a quantidade de palavras das sentenças (id: 90) <code>adjectives_standard_deviation</code> Proporção de tipos de advérbios em relação à quantidade de tokens de advérbios no texto (id: 92) <code>adverbs_diversity_ratio</code> Proporção máxima de advérbios em relação à quantidade de palavras das sentenças (id: 93) <code>adverbs_max</code> Proporção mínima de advérbios em relação à quantidade de palavras das sentenças (id: 94) <code>adverbs_min</code> Desvio padrão das proporções entre advérbios e a quantidade de palavras das sentenças (id:

		<p>95) adverbs_standard_deviation Proporção de pronomes indefinidos em relação a todos os pronomes do texto (id: 100) indefinite_pronoun_ratio</p> <p>Proporção de Verbos no Futuro do Pretérito do Indicativo em relação à quantidade de verbos flexionados do texto (id: 101) indicative_condition_ratio</p> <p>Proporção de Verbos no Futuro do Presente do Indicativo em relação à quantidade de verbos flexionados do texto (id: 102) indicative_future_ratio</p> <p>Proporção de verbos no infinitivo em relação a todos os verbos do texto (id: 103) infinitive_verbs</p> <p>Proporção de verbos flexionados em relação a todos os verbos do texto (id: 104) inflected_verbs</p> <p>Proporção de verbos no gerúndio, particípio ou infinitivo em relação a todos os verbos do texto (id: 105) non-inflected_verbs</p> <p>Proporção máxima de substantivos em relação à quantidade de palavras das sentenças (id: 107) nouns_max</p> <p>Proporção mínima de substantivos em relação à quantidade de palavras das sentenças (id: 108) nouns_min</p> <p>Desvio padrão das proporções entre substantivos e a quantidade de palavras das sentenças (id: 109) nouns_standard_deviation</p> <p>Proporção de pronomes oblíquos em relação a todos os pronomes do texto (id: 110) oblique_pronouns_ratio</p> <p>Proporção de preposições em relação à quantidade de orações no texto (id: 112) prepositions_per_clause</p> <p>Quantidade Média de preposições por sentença no texto (id: 113) prepositions_per_sentence</p> <p>Proporção máxima de pronomes em relação à quantidade de palavras das sentenças (id: 115) pronouns_max</p> <p>Proporção mínima de pronomes em relação à quantidade de palavras das sentenças (id: 116) pronouns_min</p> <p>Desvio padrão das proporções entre pronomes e a quantidade de palavras das sentenças (id: 117) pronouns_standard_deviation</p> <p>Proporção de sinais de pontuação em relação à quantidade de palavras do texto. (id: 118) punctuation_ratio</p> <p>Proporção de palavras funcionais em relação à quantidade de palavras de conteúdo do texto (id: 119) ratio_function_to_content_words</p> <p>Proporção de Pronomes Relativos em relação à quantidade de pronomes do texto (id: 120) relative_pronouns_ratio</p> <p>Proporção de pronomes possessivos nas</p>
--	--	--

		<p>segundas pessoas em relação à quantidade de pronomes possessivos do texto (id: 121) second_person_possessive_pronouns</p> <p>Proporção de pronomes pessoais nas segundas pessoas em relação à quantidade de pronomes pessoais do texto (id: 122) second_person_pronouns</p> <p>Proporção de pronomes possessivos nas terceiras pessoas em relação à quantidade de pronomes possessivos do texto (id: 123) third_person_possessive_pronouns</p> <p>Proporção de pronomes pessoais nas terceiras pessoas em relação à quantidade de pronomes pessoais do texto (id: 124) third_person_pronouns</p> <p>Proporção máxima de verbos por palavras em relação à quantidade de palavras das sentenças (id: 126) verbs_max</p> <p>Proporção mínima de verbos em relação à quantidade de palavras das sentenças (id: 127) verbs_min</p> <p>Desvio padrão das proporções entre verbos e a quantidade de palavras das sentenças (id: 128) verbs_standard_deviation</p> <p>Proporção de pronomes possessivos nas primeiras pessoas em relação à quantidade de pronomes possessivos do texto (id: 97) first_person_possessive_pronouns</p> <p>Proporção de pronomes pessoais nas primeiras pessoas em relação à quantidade de pronomes pessoais do texto (id: 98) first_person_pronouns</p>
12. Informações Semânticas de Palavras	11 métricas referentes aos significados das palavras usadas no <i>corpus</i>	<p>Proporção de sentidos dos adjetivos do texto em relação à quantidade de adjetivos do texto (id: 130) adjectives_ambiguity</p> <p>Proporção de sentidos dos advérbios do texto em relação à quantidade de advérbios do texto (id: 131) adverbs_ambiguity</p> <p>Proporção de sentidos dos substantivos do texto em relação à quantidade de substantivos do texto (id: 137) nouns_ambiguity</p> <p>Proporção de sentidos dos verbos do texto em relação à quantidade de verbos do texto (id: 139) verbs_ambiguity</p> <p>Quantidade Média de Hiperônimos por verbo nas sentenças (id: 133) hypernyms_verbs</p> <p>Proporção de substantivos abstratos em relação à quantidade de palavras do texto (id: 129) abstract_nouns_ratio</p> <p>Média de sentidos por palavra de conteúdo do texto (id: 132) content_words_ambiguity</p> <p>Média das proporções de Nomes Próprios em relação à quantidade de palavras das Sentenças (id: 134) named_entity_ratio_sentence</p> <p>Proporção de Nomes Próprios em relação à quantidade de palavras do Texto (id: 135) named_entity_ratio_text</p>

		<p>Proporção de palavras de polaridade negativa em relação a todas palavras do texto (id: 136) <i>negative_words</i></p> <p>Proporção de palavras de polaridade positiva em relação a todas palavras do texto (id: 138) <i>positive_words</i></p>
13. Frequência de Palavras	10 métricas de frequência normalizada das palavras do <i>corpus</i>	<p>Média das frequências absolutas das palavras de conteúdo do texto, via Banco de Português (BP) (id: 77) <i>cw_freq</i></p> <p>Média dos valores das frequências das palavras de conteúdo do texto na escala logarítmica Zipf, via Corpus Brasileiro (id: 78) <i>cw_freq_bra</i></p> <p>Média dos valores das frequências das palavras de conteúdo do texto na escala logarítmica Zipf, via BrWac (id: 79) <i>cw_freq_brwac</i></p> <p>Média dos valores das frequências das palavras do texto na escala logarítmica Zipf, via Corpus Brasileiro (id: 80) <i>freq_bra</i></p> <p>Média dos valores das frequências das palavras do texto na escala logarítmica Zipf, via BrWac (id: 81) <i>freq_brwac</i></p> <p>Média das frequências absolutas das palavras de conteúdo mais raras das sentenças do texto, via Banco de Português (BP) (id: 82) <i>min_cw_freq</i></p> <p>Média dos valores das frequências das palavras de conteúdo mais raras das sentenças do texto, na escala logarítmica Zipf, via Corpus Brasileiro (id: 83) <i>min_cw_freq_bra</i></p> <p>Média dos valores das frequências das palavras de conteúdo [mais raras das sentenças] do texto na escala logarítmica Zipf, via BrWac (id: 84) <i>min_cw_freq_brwac</i></p> <p>Média dos valores das frequências das palavras mais raras das sentenças do texto, na escala logarítmica Zipf, via Corpus Brasileiro (id: 85) <i>min_freq_bra</i></p> <p>Média dos valores das frequências das palavras mais raras das sentenças do texto, na escala logarítmica Zipf, via BrWac (id: 86) <i>min_freq_brwac</i></p>
14. Índices de Leiturabilidade	5 métricas superficiais de complexidade textual (apenas duas adaptadas ao português: Flesch e Dale Chall)	<p>Índice de Brunet (id: 196) <i>brunet</i></p> <p>Fórmula Dale Chall adaptada (id: 197) <i>dalechall_adapted</i></p> <p>Índice Flesch (id: 198) <i>flesch</i></p> <p>Índice Gunning Fog (id: 199) <i>gunning_fox</i></p> <p>Estatística de Honoré (id: 200) <i>honore</i></p>

Fonte: Adaptado pela autora de <http://fw.nilc.icmc.usp.br:23380/metrixdoc>. Acesso em: 15 mar. 2023.